Computers and Geotechnics 152 (2022) 105040

Contents lists available at ScienceDirect

Computers and Geotechnics

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

Research paper

Prediction of mechanical behavior of rocks with strong strain-softening effects by a deep-learning approach

Shi L.L., Zhang J. *, Zhu Q.Z., Sun H.H.

Key Laboratory of Ministry of Education for Geomechanics and Embankment Engineering, Hohai University, Nanjing 210024, China
College of Civil and Transportation Engineering, Hohai University, Nanjing 210024, China

ARTICLE INFO

Keywords:
Rocks
Constitutive behavior
Strong strain-softening
LSTM
Finite element method

ABSTRACT

Rock materials exhibit various mechanical characteristics, and it is difficult to describe the strain–stress relation with strong strain-softening behavior by a single constitutive law. In the present study, a long short term memory (LSTM) deep learning method is proposed to predict the material's deformation under different loading conditions. Unlike the traditional analysis method focusing on the determination of effective tangent stiffness tensor, the constructed LSTM-based procedure requires only the strain–stress values in certain cases to predict the future mechanical behaviors, even for rock materials with strong strain-softening, indicating that the loading history can be taken into account as time sequence data. In order to validate the accuracy, two applications are provided with the established LSTM model: predicting the deformation of granite and sandstone in conventional triaxial compression tests without introducing any elastoplastic parameters or constitutive laws, where the dataset for training the LSTM model is collected alternatively from an analytical micromechanical damage model and from laboratory experiments by considering a wide range of confining pressure. Comparisons of accuracy and convergence rate with different neural network structures are also carried out to check the best performance of the procedure. Implementation method of the trained LSTM model in Finite Element program as a constitutive relation is also provided and applied to the simulation of sandstone. Comparisons show that the LSTM-FEM method provides a good capacity to predict the mechanical behavior of rocks.

1. Introduction

Rock materials exhibit complicated mechanical properties such as the material hardening/softening, elastic stiffness degradation, and irreversible deformation under compressive stresses due to the influence of the occurrence environment and its own structure (Walsh, 1980). Considering that rock strength and deformation failure laws are crucial to the safety and stability of geotechnical engineering, a large number of scholars have studied the strain–stress behavior of rock materials, with the main research topics including the establishment of rock constitutive models and numerical simulations. Significant progress has been made in modeling plastic damage in quasi-brittle materials such as rocks. Dragon and Mroz (1979), Hansen and Schreyer (1994), De Sciarra (2012) established theoretical frameworks of plastic damage model. Khan et al. (1991), Shao et al. (2006), Parisio et al. (2015) established isotropic and anisotropic plastic damage models for rock materials. At the same time, to better describe the influence of anisotropic microcracks in brittle materials, Carol and Bazant (1997) created a discrete plastic damage model based on microplane theory, while Zhu et al. (2010a,b) used discrete thermodynamic formulations. In addition, in order to establish the physical relation between the two micro physical processes of microcrack propagation and friction sliding and the macro inelastic behavior, Andrieux and Bamberger (1986) and Prat and Bažant (1997) developed a micromechanical model based on linear fracture mechanics, whereas Horii and Nemat-Nasser (1983) and Zhu and Shao (2015) developed a micromechanical model based on homogenization theory. The usual procedure for establishing constitutive models is to introduce additional elastoplastic parameters or constitutive laws in the derivation and then, from the results of experimental tests, to identify via a regression method, the parameters of the model.

The strain–stress curve of rocks (such as granite, argillite and sandstone) in inelastic stage can be divided into the strain-hardening stage before peak strength and the strain-softening at the post-peak stage. The collapse will not immediately arrive after the peak point and makes an important contribution of the load bearing ability of the material in engineering. Therefore, the accurate understanding of strain-softening

* Correspondence to: Key Laboratory of Ministry of Education for Geomechanics and Embankment Engineering, Hohai University, Nanjing 210024, China.
E-mail address: chelseazhangjin@gmail.com (J. Zhang).

https://doi.org/10.1016/j.compgeo.2022.105040
Received 17 June 2022; Received in revised form 21 August 2022; Accepted 14 September 2022
Available online 27 September 2022

stage is crucial for the description of the mechanical behavior. However, at the post-peak stage, the subsequent yield surface evolves with plastic deformation, and the mechanical behavior is relatively complex, which leads to difficulties with classical strength and constitutive theories. Some analytical constitutive models, considering strain-softening behavior of rocks, are carried out by additional assumptions and softening parameters (Tan et al., 2014; Souley et al., 2018; Jin et al., 2020). The difference is that the softening parameters used of different softening models are inconsistent. Tan et al. (2014) assumed that uniaxial compressive strength of intact rock and parameter m_b of H–B criterion were softening parameters of model. Souley et al. (2018) described the strain-softening through parameter m_b and s of H–B criterion. Jin et al. (2020) compared the rationality of the different parameters for describing the softening characteristics of the strain-softening model based on H–B criterion and found that the parameter GSI was best suited for the description. Moreover, assuming a decrease in cohesion with plastic strain and a constant friction angle, Pourhosseini and Shabanimashcool (2014) proposed a pre-peak elastic and post-peak strain-softening constitutive model. Ma et al. (2014) set up the non-simultaneous mobilization of cohesion and friction to capture the post-peak strain-softening and ductile failure behavior of salt rocks. Consequently, an analytical solution is not applicable for different types of rocks due to the various characteristics.

To validate the rationality of the material constitutive model, it is critical to establish a corresponding numerical approach. Numerical integration algorithms are mainly divided into three categories: explicit algorithm, implicit algorithm and semi-implicit algorithm. Explicit algorithms were proposed by Zienkiewicz et al. (1969) and Nayak and Zienkiewicz (1972), in which all quantities are updated at the start of the time step and no iterative procedure is necessary. The implicit closest point projection method (CPMM) was proposed by Simo and Ortiz (1985) as a regression mapping tool for rate independent elastoplastic models. Later, Simo and Taylor (1985) discovered that when the full Newton–Raphson algorithm is utilized, the CPPM with a constant elastoplastic stiffness tensor offers asymptotic global quadratic convergence rates. To improve the numerical computational efficiency, Peng and Chen (2012) developed an implicit integration algorithm for an isotropic elastoplastic model based on a general yielding criterion. Finally, for a semi-implicit algorithm, Ghaei et al. (2010) implemented a plastic model with two yield surfaces by not updating certain variables during the solution process. For plastic models with various yield surfaces, a new semi-implicit algorithm was proposed by Areias et al. (2012).

In recent years, with the growth of machine learning, artificial intelligence methods with powerful nonlinear characterization capabilities have attracted the attention of scholars worldwide. Various artificial intelligence methods have been widely utilized in various areas. The most popular deep learning application topics in geotechnical engineering, such as tunnel construction, slope displacement, and landslide susceptibility, were summarized by Zhang et al. (2021a), and it was demonstrated that the amount of research in this subject is expanding rapidly. For instance, deep learning was utilized by Huang et al. (2018) to locate the source of microseismic occurrences in underground mines. Wu et al. (2020) used domain expertise to design a deep learning model to track tunnel construction operations. To anticipate the surface settlement during tunnel construction, Hu et al. (2019) used three distinct artificial intelligence algorithms. Zhou et al. (2017) used the RFR method to predict ground settlement caused by shield-driven tunnel construction. Kardani et al. (2021) used a hybrid stacking ensemble approach to forecast slope stability. Mahdevari et al. (2017) created an artificial neural network (ANN) to anticipate and analyze the stability of gate roadways based on roof displacements.

In addition, the mechanical behavior of materials can be well predicted by deep learning methods. Deep learning methods, unlike traditional constitutive methods, can directly obtain the neural network model through the strain–stress data without any mechanical assumptions, so as to determine the constitutive behavior of materials. Ghaboussi et al. (1991) first proposed the application of ANN model to the mechanical behavior modeling of concrete, predicting multiple load paths under biaxial loading. Furukawa and Yagawa (1998) proposed an implicit viscoplastic constitutive model based on neural network to study the inelastic behavior of materials. However, the strain–stress relation is affected by the loading history, and the traditional neural network method cannot reflect the stress history-related properties of materials due to the lack of connection between different time steps, the accumulation of errors, and the high computational cost (Zhang et al., 2021b).

To solve the above problems, recurrent neural network (RNN) with node connections between hidden layers was proposed in the field of artificial intelligence (Rumelhart et al., 1986). However, it has been found that RNN suffers from long-term memory loss during training when dealing with long-term history data. For this reason, Hochreiter and Schmidhuber (1997) proposed long short term memory (LSTM), which is a special type of RNN. Currently, LSTM is adopted in a wide number of fields, such as neural machine translation (Sutskever et al., 2014; Bahdanau et al., 2014), speech recognition (Graves et al., 2013), and also vision-and-language applications like image and video captioning (Vinyals et al., 2015; Xu et al., 2015; Baraldi et al., 2017). In the field of engineering, LSTM has been applied to study soil moisture (Fang et al., 2018), soil temperature (Li et al., 2020) and hydro-mechanical coupling effects in porous media (Wang and Sun, 2018). The thrust and cutterhead torque were predicted by a LSTM network model based on data from the rising phase of the TBM tunneling cycles (Li et al., 2021). Liu et al. (2021) developed a lithology prediction model using improved LSTM neuron network methods. Zhang et al. (2021c) compared the prediction accuracy and generalization ability of five ML-based models based on the same set of sand test results, and found that LSTM neural network and its variants are the most suitable for developing constitutive models.

However, LSTM methods have rarely been used to study rock constitutive behavior. Unlike traditional constitutive methods, the LSTM approach, one of data-driven deep-learning methods, can directly obtain the neural network model through the strain–stress data without any mechanical or parameter assumptions, so as to determine the constitutive behavior of materials and well describe the strain-softening phenomenon. Considering the complex mathematical form and the difficulty of parameter assumption of traditional constitutive methods for rock materials with strong strain-softening characteristics and the inappropriateness of general neural network models to deal with long-term history data problems, this study uses numerical and experimental strain–stress data for training and testing based on LSTM model in deep learning.

The present paper is organized in the following way. In Section 2, the traditional analytical constitutive model for rock materials and the basic principle for LSTM network are briefly recalled, and two methods are compared at the end. Section 3 is devoted to providing two applications of the considered LSTM deep-learning model on the mechanical behaviors of granite and sandstone in triaxial compression tests, as well as the prediction of the mechanical response of granite under loading and unloading conditions. In Section 4, the trained LSTM model is implemented in FEM program as a constitutive law, and applied to sandstone. Some concluding points are also provided in the last section.

2. Brief description of constitutive relationship by analysis and LSTM network methods

2.1. Traditional elastoplastic constitutive model of rock-like materials

This section begins with a brief recall of the constitutive model for quasi-brittle materials. Actually, in brittle materials like concrete and rocks, damage caused by microcrack growth (Shao et al., 2003; Zhu

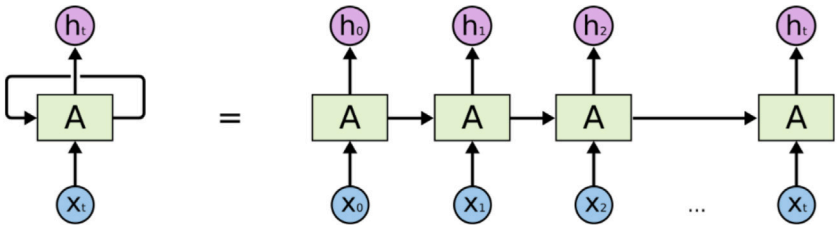


Fig. 1. Expansion of a typical RNN structure (Olah, 2015).

et al., 2008) and voids collapse (Dormieux et al., 2006; Shen and Shao, 2016; Zhang et al., 2017) is usually thought to be the primary cause of inelastic deformation and failure. A damage parameter ω related to material's degradation is consequently introduced to represent the growth and evolution of microscopic cracks, which is usually measured by internal variable such as generalized plastic deformation γ^p.

Through a series of derivation, the final constitutive elastoplastic model considering the damage effect in strain-prescribed loading case is usually taking the following form:

$$\mathrm{d}\sigma_{ij} = C^{tan}_{ijkl}\mathrm{d}\varepsilon_{kl} \tag{1}$$

in which C^{tan}_{ijkl} denotes the effective stiffness tensor to predict the stress increment of damaged material, and $C^{tan}_{ijkl} = C^{0}_{ijkl}$ for the plastic yield stress $f(\sigma_{ij}) < 0$. So that the corresponding stress increment $\mathrm{d}\sigma_{ij}$ is obtained under the specified strain increment $\mathrm{d}\varepsilon_{ij}$.

It can be obviously seen from Eq. (1) that the key point of the analysis constitutive model in terms of elastoplastic assumption is to determine the effective stiffness tensor C^{tan}_{ijkl} beyond the elastic region. However, the strength and deformation rules of different rocks could vary in a wide range, leading to the difficulty of describing the strain–stress relation in a consistent form. In addition, experimental studies (Wawersik and Brace, 1971; Martin and Chandler, 1994) have revealed various modes of microcrack initiation and propagation in rock materials, and the involved damage parameter ω in material property is consequently in different expressions according to the mechanical behavior and failure modes.

As a result, to overcome the above difficulties, a LSTM procedure of deep-leaning network will be used in this work to carry out a widely applicable method to predict the deformation of rocks without introducing the mentioned elastoplastic definitions and the values of material properties.

2.2. Principle technique of LSTM network

Recurrent Neural Network (RNN) is a kind of neural network for processing sequence data. Unlike general neural networks, RNN can take into consideration of the history-related properties. The expansion of a typical RNN structure by time sequence is shown in Fig. 1 (Olah, 2015), where X_i denotes the input layer, A represents the hidden layer, and h_i is the output layer. It can be seen from the network structure that the nodes between the hidden layers of the RNN are connected, i.e., the input of the hidden layer contains not only the input layer at the current moment, but also the output of the hidden layer at the previous moment. Considering that the mechanical property of geomaterials is related to the loading history, the RNN network is chosen in this study due to its main feature of dealing with the time sequence problems.

However, gradient disappearance and explosion during RNN training cannot be ignored concerning the long sequence problems. Compared with RNN, LSTM is able to better solve the long-term dependence problem by a structure of the forget gate, input gate, output gate, and memory cell (The expansion of a LSTM structure is displayed in Fig. 2 (Olah, 2015)). Among them, the input gate is responsible for processing the inputs at the current moment t, while the forget gate decides whether to remember the intermediate state variables C_{t-1}

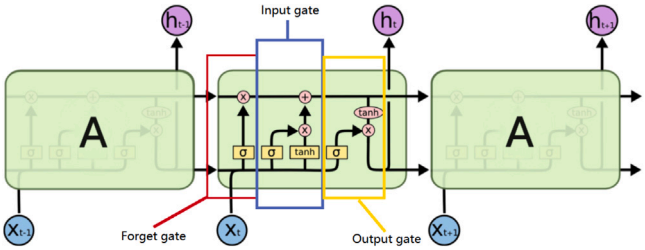


Fig. 2. Expansion of a LSTM unit cell (Olah, 2015).

at the previous moment $t-1$. These gates collaborate to control its ability to learn or forget information from temporal data. The key component of LSTM network is the memory cell, whose function is to make the procedure freely choose what to store at each time step, and the transmission process is similar to a conveyor belt. In addition to the outputs x_t of the input layer at the current moment and the output h_{t-1} of the hidden layer at the previous moment, an obvious difference of LSTM network are the input components, comparing to RNN hidden layer, including also the intermediate state variables C_{t-1} of the previous time step.

In the following part, we will briefly recall the basic principle technique of a typical LSTM network. The forget gate f at t moment is expressed in the form below:

$$f_t = \sigma(W_f h_{t-1} + U_f x_t + b_f) \tag{2}$$

with f_t being the output of the forget gate at t step. U_f and W_f denote respectively the weight of input data and the shared recurrent weights of the recurrent data for the forget gate. And b_f is the corresponding bias for the forget gate. $\sigma(*)$ represents the sigmoid activation function, defined by $\sigma(*) = 1/(1+e^{-*})$.

The input gate i at t step reads:

$$i_t = \sigma(W_i h_{t-1} + U_i x_t + b_i) \tag{3}$$

where i_t is the proportion of the current input that can be saved to the cell memory. Similarly, W_i and U_i are the weights, and b_i represents the corresponding bias for the input gate.

The current updated intermediate state variable C_t is defined by the following equation:

$$C_t = C_{t-1} \odot f_t + i_t \odot \widetilde{C}_t \tag{4}$$

with $\odot$ representing the Hadamard product, and $\widetilde{C}_t$ being the candidate value for cell status

$$\widetilde{C}_t = \tanh(W_C h_{t-1} + U_C x_t + b_C) \tag{5}$$

where W_C and U_C are the weight of cell status, and b_C is the bias for the input unit. $tanh(*)$ represents the activation function, defined as $tanh(*) = (e^{*} + e^{-*})/(e^{*} - e^{-*})$.

The output gate o is defined by the following equation:

$$o_t = \sigma(W_o h_{t-1} + U_o x_t + b_o) \tag{6}$$

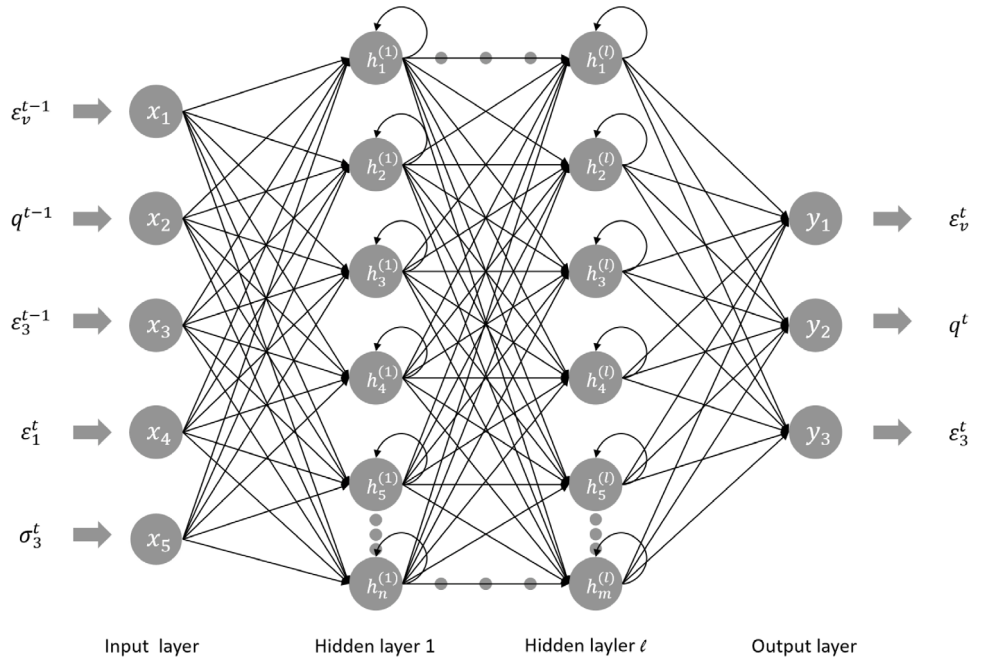


Fig. 3. Multi-layer LSTM neural network topology.

where o_t is the output of the current output gate, and h_t denotes the output data of the LSTM cell

$$h_t = o_t \odot \tanh(C_t) \tag{7}$$

W_o and U_o are the weight of the output gate, and b_f is the bias of the output gate.

In fact, the process of LSTM network training is to determine the model parameters W, U and b, providing a best solution of the final results.

2.3. Application of the LSTM deep-learning method in predicting the deformation of rock in conventional triaxial compression tests

In the conventional triaxial compression experiment, one has the following relations for the stress and strain states of the test sample:

$$\varepsilon_{ij} = \sigma_{ij} = 0 \quad \text{for} \quad i \neq j \tag{8}$$

and the lateral strain and stress components

$$\varepsilon_2 = \varepsilon_3, \quad \sigma_2 = \sigma_3 \tag{9}$$

The deviatoric stress can be calculated by the equation below

$$q = \sigma_1 - \sigma_3 \tag{10}$$

Considering the strain-prescribed loading case (the most common loading mode to obtain the post-peak behavior of rock), the lateral pressure is maintained constant, and the axial strain ε_1 is increased quasi-statically. Consequently, the determination of the constitutive behavior of rock materials by the LSTM deep-learning methods can be transformed into predicting the current deviatoric stress q^t, lateral strain ε_3^t and volumetric strain ε_v^t, with the input data of deviatoric stress q^{t-1}, lateral strain ε_3^{t-1} and volumetric strain ε_v^{t-1} at the previous moment and the axial strain ε_1^t and lateral stress σ_3^t at the current moment. The neural network of deep learning is usually composed of input layer, hidden layer and output layer. Different from the network topology of ANN, LSTM has circular connection in the hidden state (see Fig. 3), which ensures that sequence information is captured in the inputs.

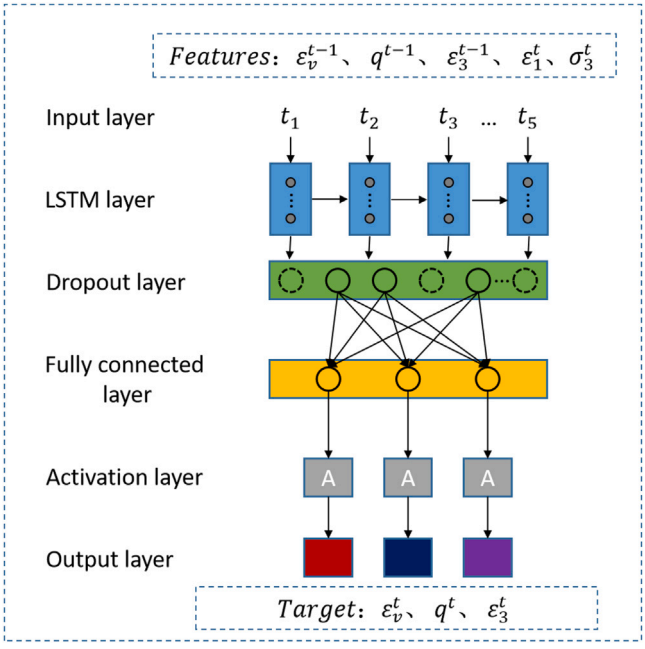


Fig. 4. Architecture of the LSTM deep-learning network.

Actually, we can modify the structure of the network, such as changing the number of layers and neurons of each hidden layer, to obtain the best prediction accuracy, according to the complexity of considered problem. Concerning the conventional triaxial compression tests, LSTM network is adopted as the hidden layer in this paper. As mentioned in the previous part, there are five input parameters in the input layer, including the volumetric strain ε_v^{t-1}, deviatoric stress q^{t-1} and lateral strain ε_3^{t-1} at the previous time and the axial strain ε_1^t and lateral stress σ_3^t at the current time. There are three output parameters in the output layer, involving the current volumetric strain ε_v^t, deviatoric stress q^t and lateral strain ε_3^t. Fig. 4 displays a generally

adopted LSTM network structure, in which the LSTM hidden layer can be reconstructed according to different problems.

The training process of neural network mainly includes forward propagation, back propagation and iterative updating of model parameters. At the beginning of training, the model parameters are generally chosen as random values, so the prediction results obtained by forward propagation are necessarily unreliable. The error between the predicted value and the target value need to be evaluated by a loss function. In this study, the mean square error (MSE) is chosen as the loss function to estimate the accuracy:

$$MSE = \frac{\sum_{i=1}^{n}(f(x) - y_i)^2}{n} \tag{11}$$

where $f(x)$ denotes the predicted value, y_i is the true value and n is the number of samples. A smaller value of MSE indicates the more accurate prediction by the constructed model. Therefore, the process of model training is to determine the model parameters in nature, which minimize the loss function.

To determine the corresponding network parameters, conventional neural networks typically minimize the loss function by using an optimization algorithm. Optimization algorithms commonly include the gradient descent algorithm and the evolutionary one. Considering the computational cost, Adam optimizer (adaptive moment estimation algorithm) based on a gradient descent algorithm, which was firstly proposed by Kingma and Ba (2014), is accepted to train the model in this paper. Unlike the classical gradient descent method, Adam designs independent adaptive learning rates for different parameters by calculating the first-order moment estimation and second-order moment estimation of the gradient, and usually has less parameter adjustment.

In order to evaluate the performance of the proposed LSTM network program, the coefficient of determination (R^2) is used as the evaluation index:

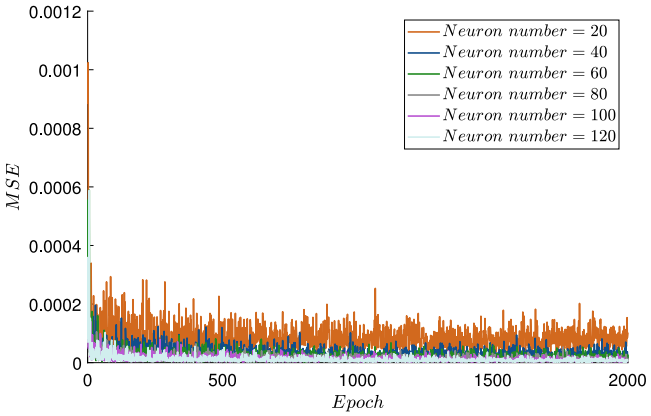
$$R^2 = 1 - \frac{\sum_{i=1}^{n}(f(x) - y_i)^2}{\sum_{i=1}^{n}(\bar{y} - y_i)^2} \tag{12}$$

in which $f(x)$ is the predicted value, y_i is the true value, $\bar{y}$ represents the average of true values, and n is the number of samples.

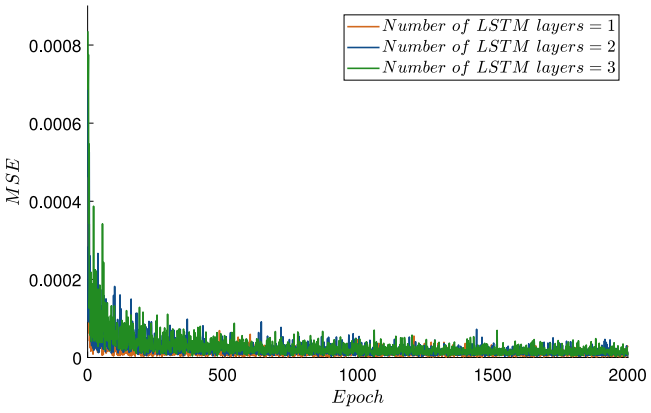
In Section 2.1, the framework of traditional constitutive method is briefly reviewed. The determination of the effective stiffness tensor C_{ijkl}^{tan} is an important part in the derivation, which varies during the load process and leads to the inelastic behavior of rocks. Besides, additional assumption or function to describe the material deformation effect also need to be provided, which will result in burdensome derivation and computation. The implementation of LSTM method is introduced in Sections 2.2 and 2.3. LSTM method, as a special type of RNN, has circular connection in the hidden layers, so it can reflect the stress history-related properties of materials. It can be seen that the LSTM method only needs to be trained by available data, while the traditional constitutive method needs to define the parameters and introduce the constitutive laws, which is consequently more complicated.

3. Predictions of strong strain-softening behaviors of rocks by the LSTM deep-learning procedure

In this section, the constructed LSTM deep-learning procedure will be applied to predict the deformation of granite where the dataset for training the LSTM model is collected from an analytical micromechanical damage law in Section 3.1, and that of sandstone where the dataset is obtained from original laboratory conventional triaxial compression tests in Section 3.2. In order to validate the accuracy, no elastoplastic parameter of the material is introduced for model training, and a wide range of confining pressure is considered.



(a) Evolution of MSE with different number of neurons



(b) Evolution of MSE with different LSTM Layers

Fig. 5. Training process of LSTM model with different structure.

3.1. Mechanical behaviors prediction of granite with training data from analytical model

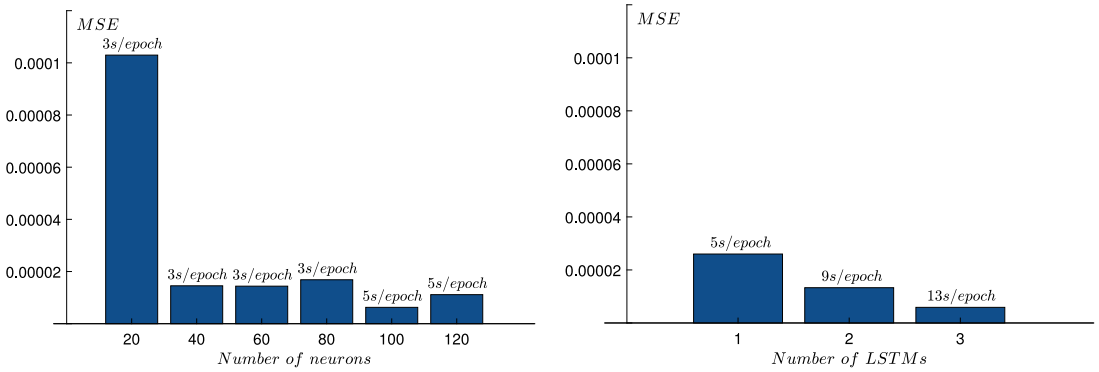
The training data for LSTM network method is crucial to its prediction accuracy. Compared to experimental tests, the analytical constitutive model can provide enough data for better training the proposed procedure. And the validation will take into account a wide range of different loading condition (confining pressure). Consequently, a micromechanical-based constitutive damage model (Zhu et al., 2015) to describe the mechanical properties of Beishan granite is adopted to produce the dataset in this subsection.

Taking into account of the brittle fracture of granite, a damage parameter ω related to the crack density is introduced to describe the material degradation in Zhu et al. (2015). The final friction-damage strength criterion is taking the following form:

$$f_s = q - \sqrt{\frac{3}{2}}\eta p - \sqrt{6R(\omega_c)\chi} \le 0 \tag{13}$$

where p and q are respectively the mean and deviatoric stresses. The derivation and detailed expressions of this model are provided in Appendix. In order to collect the dataset, the above analytical model can be illustrated numerically by an iteration procedure, which contains elastic prediction and inelastic correction phases at each increment. With the imposed strain increment, the stress increment can be correspondingly calculated. It is referred to Yuan et al. (2020) for the iteration program.

The deformation under monotonic load, one of the most important characteristics of rocks in engineering, is investigated by LSTM method in this study. The strategy of the data preparation are arranged



(a) Different number of neurons (b) Different LSTM Layers

Fig. 6. Comparison of the accuracy and computation efficiency with different LSTM structure.

as follows. The model contains five parameters: elastic modulus E^m, Poisson's ratio μ^m, friction coefficient η of crack surface, the maximum value of damage evolution resistance $R(\omega_c)$, and the critical value of damage variable ω_c. The parameters are determined by Zhu et al. (2015), so that $E^m = 70$ GPa, $\mu^m = 0.15$, $\eta = 1.80$, $R(\omega_c) = 9.26 \times 10^{-3}$ MPa, $\omega_c = 7.0$. Through iterative procedure, 1000 strain–stress pairs under each confining pressure are derived as the dataset of LSTM model. Among them, 16000 strain–stress pairs with confining pressures of 0, 2, 4, 6, 8, 10, 12, 14, 16, 18, 20, 22, 24, 26, 28 and 30 MPa are used as the training set of LSTM model, and 2000 strain–stress pairs with confining pressures of 17 and 19 MPa will be predicted by the trained deap-learning procedure and compared to the analytical results, aiming at validating the prediction accuracy.

In addition, considering that the input and output data of the model have different units and might vary greatly in the value, direct application of original data will lead to inaccuracy problem caused by the dominance of large dimension values. Since the gradient descent optimization is used in the model training, the dimensional difference will lead to the decrease of the speed of updating model parameters in each iteration. In order to avoid the above problems, we normalize the input and output data in the preprocessing stage and scale them linearly to the interval of [0,1] by

$$x^* = \frac{x - x_{min}}{x_{max} - x_{min}} \tag{14}$$

where x represents the original values of the input and output data, x_{min} and x_{max} are the minimum and maximum values, and x^* is the objective normalized value which can be used for the training process.

Following the above steps, the proposed LSTM network model has been trained by considering different number of hidden layers and neurons. By analyzing the prediction accuracy and computational cost with different structures, the best solution can be determined. The evaluation of error parameter MSE during the training process is provided in Fig. 5. 1 hidden layer with 20, 40, 60, 80, 100 and 120 neurons have been tried as shown in Fig. 5(a). It can be seen the values of MSE decrease quickly and are stable after 400 epoch of training. The comparisons of the MSE variation during train process between different LSTM layers with 100 neurons are also carried out in Fig. 5(b).

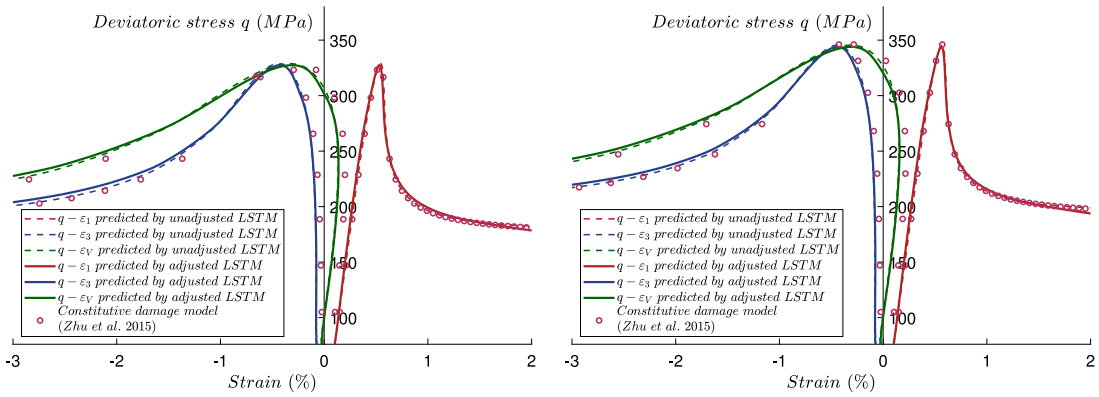
For the sake of clarity, the final value of MSE at the last epoch and calculation efficiency of different LSTM structure are illustrated in Fig. 6. It can be concluded in Fig. 6(a) that except for a layer with 20 neurons, the final value of MSE do not have remarkable changes with the increase of the number of neurons, where it takes the minimum value with 100 neurons. And the computation time of each epoch has very few difference. Besides, as shown in Fig. 6(b), the errors are all acceptable for 1, 2 and 3 layers. Consequently, the LSTM model with 2 hidden layers of 100 neurons is adopted for the training and prediction steps. After 2000 epoch, the loss and accuracy of the test data are 2.79×10^{-5} and 0.99.

Fig. 7 shows the comparison between the predicted mechanical behavior by the proposed LSTM procedure and the data obtained by the analysis model in triaxial compression tests with confining pressure of 17 and 19 MPa. The cycle points represent the strain–stress data from the micromechanical damage constitutive model proposed in Zhu et al. (2015). The red, blue and gray lines denote the variation of deviatoric stress corresponding to the axial strain ($q - \varepsilon_1$), lateral strain ($q - \varepsilon_3$) and volumetric strain ($q - \varepsilon_V$), respectively.

In general, it can be obviously seen that the predicted values have a good agreement to the true values, and the trained model well reflects the strong strain-softening characteristics of granite. In Fig. 7, prediction results are obtained by using the existing data as input (dash line), the accuracy of model prediction cannot be verified for unknown deformation. Therefore, the test procedure should be adjusted. Except for the first set of sequence inputs where analytical results are still used, the axial strain ε_1^t and lateral stress σ_3^t at the current moment determined are retained for the other sequence inputs, and the volumetric strain ε_v^{t-1}, deviatoric stress q^{t-1} and lateral strain ε_3^{t-1} at the previous moment are replaced with the predicted results at previous moment. Comparisons between the predicted values of the strain–stress curve after being adjusted (solid line) and the analytical results are displayed in Fig. 7. It can be seen that the trained model after adjusted can also provide accurate predictions of mechanical behaviors for granite.

In addition, in the analytical constitutive law, the damage variable ω related to material degradation is generally introduced to describe the evolution of microstructure during the loading process, for which the decrease of the effective Young's modulus E is considered as the macroscopic consequence that can be observed in experiments (Liu and Dai, 2018). In order to verify whether the results by the LSTM method can reflect the variation of internal properties of rocks, that is, to explore whether the results of the LSTM model are in accordance with those of the analytical model, Fig. 8 illustrates the evolution of predicted tangent modulus E (blue line) and the damage variable ω (red dash line) by the analytical constitutive model with the confining pressure of 17 MPa. Although no mechanical property parameters are introduced in training the LSTM deep-learning model, the predicted zero point of E is close to the abrupt position of ω, and to the peak point of the strain–stress curve, indicating that the obtained results by LSTM method can capture the main features of mechanical properties and material degradation for rock materials.

To verify the accuracy of the proposed LSTM model in predicting rock deformation under different loading paths, the application of proposed LSTM approach in simulating the behaviors under loading and unloading conditions is also carried out. Similarly, through the iterative process, the loading and unloading process of granite is simulated once before and after the peak strength. The same set of database is accepted for training the model with 3 hidden layer of 140 neurons and predicting the strain–stress values with confining pressure of 17 MPa.



(a) Confining pressure $\sigma_3 = 17MPa$ (b) Confining pressure $\sigma_3 = 19MPa$

Fig. 7. Comparisons between the predicted values by unadjusted and adjusted LSTM model and data obtained by the analysis model under different confining pressures.

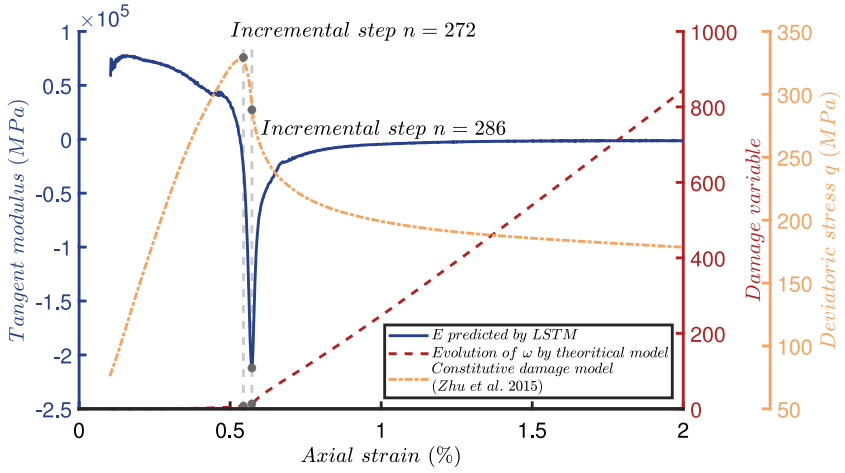


Fig. 8. Comparisons between the evolution of damage parameter and degradation of material by analytical results and LSTM predictions.

After 1000 epoch, the R^2 of the test set reaches 0.9997. Fig. 9 provides the comparisons between the LSTM predictions (solid line) and the data (cycle points) obtained by the analysis mode with σ_3 = 17 MPa under loading and unloading conditions (gray arrows). It can be seen that LSTM method can predict not only the mechanical behavior of rock under monotonic loading path, but also the cyclic response of rock under complex loading history.

3.2. Mechanical behaviors prediction of red sandstone with training data from laboratory experimental tests

In the previous subsection, the proposed LSTM procedure has been tested with dataset from analytical solution, and exhibits a good prediction ability. In order to verify the accuracy on the constitutive behavior of unknown materials, the proposed deep-learning network will be trained with experimental data of red sandstone in conventional triaxial compression tests in this part. The laboratory experiments are carried out on Rock multi-field test system, using standard dry red sandstone with the diameter of 50 mm and height of 100 mm (see Fig. 10).

The triaxial compression test is performed with 7 confining pressure levels of 0, 5, 10, 15, 20, 25 and 30 MPa. The experimental results are displayed in Fig. 11. The red line represents the curve of $q-\varepsilon_1$, and the blue line stands for that of $q-\varepsilon_3$. It can be seen that the reduction of the curve is relatively slow compared to that of granite. Besides, in Fig. 11, the strain-softening phenomenon phase is obvious for 0 MPa confining pressure. But with the increase of confining pressure, the softening of the strain–stress curve is not clear, but still can be observed, which

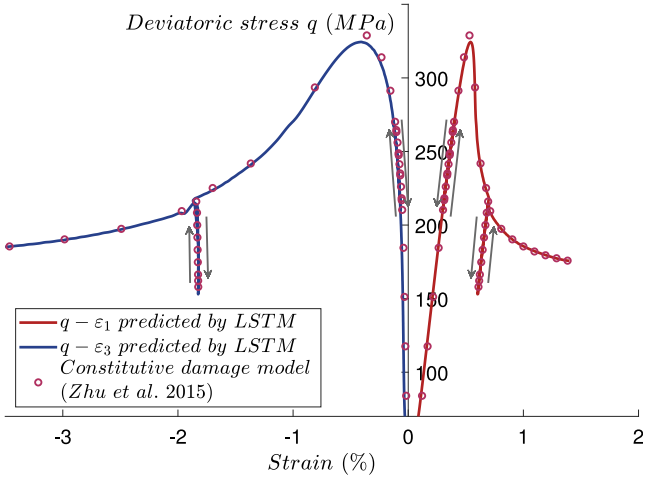


Fig. 9. Comparisons between the numerical results and the LSTM predictions under loading and unloading conditions.

is due to the brittle–ductile transition of sandstone under confining pressure.

A total of 30874 strain–stress pairs are obtained from the triaxial compression test. 25739 experimental results of 0, 5, 10, 15, 25 and 30 MPa confining pressure are accepted as training dataset for the

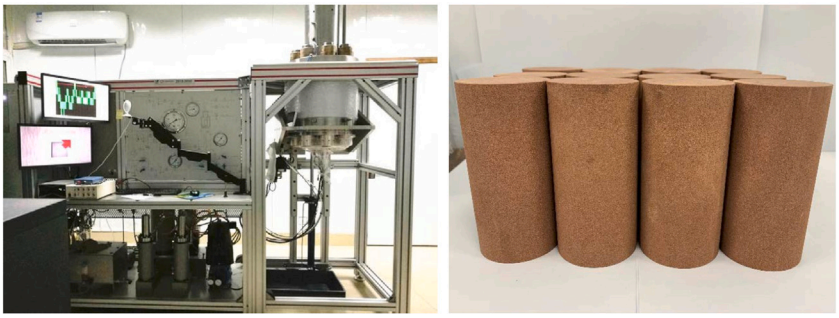


Fig. 10. Rock multi-field test system (left) and red sandstone specimens (right).

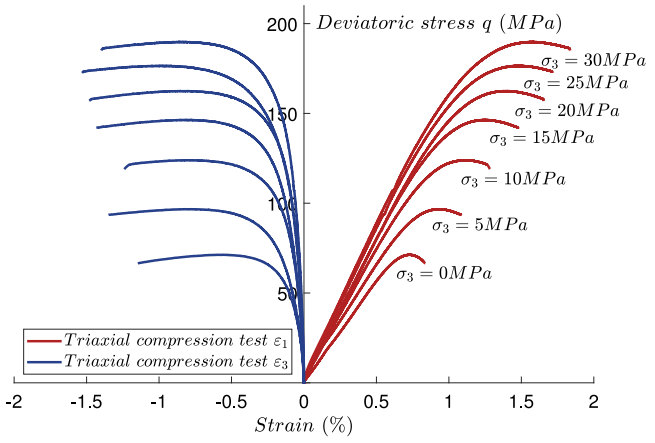


Fig. 11. Strain–stress curves of red sandstone by triaxial compression tests with respect to different confining pressures.

LSTM network with 2 hidden layers of 140 neurons. And 5135 strain–stress data with 20 MPa confining pressure are used as test set to verify the prediction ability of the proposed LSTM-based model. After 100 epoch, the R^2 of the test set reaches 0.9205. Fig. 12 shows the comparisons between the LSTM predictions (solid line) and experimental results (cycle points) of red sandstone with $\sigma_3 = 20$ MPa. It can be seen that, even the material's properties are unknown and the strain-softening phase is not that remarked compared with that of Beishan granite in Section 3.1, the LSTM-based model can also predict the deformation of rock materials with different natures and the brittle–ductile transition can also be predicted through limited experimental data.

4. Method of implementation of LSTM deep-learning model in finite element analysis

4.1. Main steps to implement the trained LSTM model in FEM program

In this section, the proposed LSTM deep-learning network will be implemented in 3 dimensional FEM framework written in Julia (Bezanson et al., 2017) programming language, which could be applied in structure computations. In the FEM program, quadrilateral isoparametric elements are used for spatial discretization, which makes the weak form of governing equations of nonlinear solid mechanics be written as a sum over the element domains (Zienkiewicz and Taylor, 2005). Applying Galerkin method (Galerkin, 1915; Hughes, 2012), displacement $\boldsymbol{u}$ and its virtual form $\delta\boldsymbol{u}$ in weak form are replaced by their polynomial approximations. Then the derived system of nonlinear equations is solved by Newton's method (Burden et al., 2015). The inserted trained model plays a similar role as the constitutive law in FEM program. The key point of the program is to update the new stress tensor $\boldsymbol{\sigma}^{t+\Delta t}$ and Jacobian matrix $\partial\Delta\boldsymbol{\sigma}/\partial\Delta\boldsymbol{\varepsilon}$ with the known current stress $\boldsymbol{\sigma}^t$, strain $\boldsymbol{\varepsilon}^t$, and the prescribed strain increment $\mathrm{d}\boldsymbol{\varepsilon}$.

In LSTM neural network, the input file is sequence data. For example, the input is composed of the strain–stress results of previous and current time steps together. But when solving the stress and Jacobian matrix at a certain time step, the finite element program only provides the stress $\boldsymbol{\sigma}$ and strain $\boldsymbol{\varepsilon}$ of the previous moment and the strain increment $\mathrm{d}\boldsymbol{\varepsilon}$ of the current moment to the constitutive program, which do not constitute the required sequence data for the model. Therefore, in order to obtain the initial sequence data required for the LSTM neural network model, the solution process of the constitutive program is divided into two parts: the imposed elastic stage at beginning and the trained LSTM neural network model obtained in this study. In the initial stage, the constitutive program firstly executes the isotropic linear elastic constitutive relation and saves the results until the sequence data reaches the required length, and then enters into the LSTM neural network model for computation. In addition, aiming at storing the strain–stress data at the previous moment and forming the sequence data input required for the LSTM neural network model, the state variable ($\boldsymbol{STATEV}$) is used for data storage. containing the volumetric strain ε_v^{t-1}, deviatoric stress q^{t-1} and lateral strain ε_3^{t-1} at the previous moment and the axial strain ε_1^t and lateral stress σ_3^t at the current moment in chronological order.

The dimension of the time-sequence input data is imposed as 50, so that $\boldsymbol{STATEV}(1)$ to $\boldsymbol{STATEV}(250)$ will be stored in the computation. In order to ensure that the sequence data is stored, $\boldsymbol{STATEV}(251)$ is set to count the execution times of the constitutive program. The main process of LSTM network implemented in FEM program and the specific storage and update process of $\boldsymbol{STATEV}$ are shown in Fig. 13.

The artificial neural network model trained by experimental results is implemented in FEM consisting of two LSTM layers, a fully connected layer and a Relu activation layer. If the number of model features and target values are noted as j and k, the first and second LSTM layers have m and n neurons, and the number of neurons in the fully connected layer is k, which is the same as the target values, so that there are $(4(m^2 + jm + m) + 4(n^2 + mn + n) + (nk + k))$ internal parameters in the model. In this part, the number of feature and target values of the artificial neural network model trained are imposed as 5 and 3. And there are 140 neurons in the two LSTM layers and 3 neurons in the fully connected layer. Therefore, the model has 239543 internal parameters in total. Inspired from the basic principle technique of LSTM, the training process of each layer for the present application is provided in the following equations.

The first LSTM layer:

$$\begin{cases} f_t^{(1)} = \sigma(W_f^{(1)} h_{t-1}^{(1)} + U_f^{(1)} x_t + b_f^{(1)}) \\ i_t^{(1)} = \sigma(W_i^{(1)} h_{t-1}^{(1)} + U_i^{(1)} x_t + b_i^{(1)}) \\ \widetilde{C}_t^{(1)} = \tanh(W_C^{(1)} h_{t-1}^{(1)} + U_C^{(1)} x_t + b_C^{(1)}) \\ C_t^{(1)} = C_{t-1}^{(1)} \odot f_t^{(1)} + i_t^{(1)} \odot \widetilde{C}_t^{(1)} \\ o_t^{(1)} = \sigma(W_o^{(1)} h_{t-1}^{(1)} + U_o^{(1)} x_t + b_o^{(1)}) \\ h_t^{(1)} = o_t^{(1)} \odot \tanh(C_t^{(1)}) \end{cases} \tag{15}$$

where x_t is the input at the current moment.

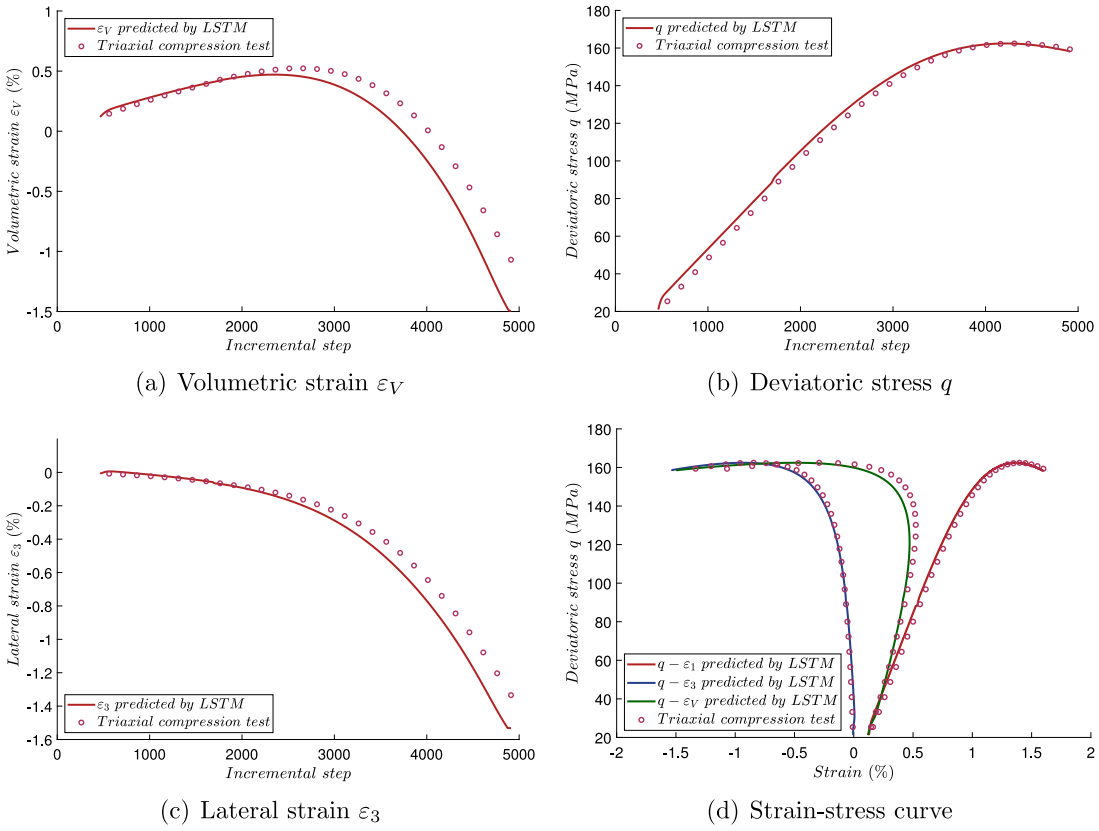


Fig. 12. Comparisons between the LSTM predictions and experimental results of red sandstone in triaxial compression test with $\sigma_3 = 20$ MPa.

The second LSTM layer:

$$\begin{cases} f_t^{(2)} = \sigma(W_f^{(2)} h_{t-1}^{(2)} + U_f^{(2)} h_t^{(1)} + b_f^{(2)}) \\ i_t^{(2)} = \sigma(W_i^{(2)} h_{t-1}^{(2)} + U_i^{(2)} h_t^{(1)} + b_i^{(2)}) \\ \widetilde{C}_t^{(2)} = \tanh(W_C^{(2)} h_{t-1}^{(2)} + U_C^{(2)} h_t^{(1)} + b_C^{(2)}) \\ C_t^{(2)} = C_{t-1}^{(2)} \odot f_t^{(2)} + i_t^{(2)} \odot \widetilde{C}_t^{(2)} \\ o_t^{(2)} = \sigma(W_o^{(2)} h_{t-1}^{(2)} + U_o^{(2)} h_t^{(1)} + b_o^{(2)}) \\ h_t^{(2)} = o_t^{(2)} \odot \tanh(C_t^{(2)}) \end{cases} \tag{16}$$

where $h_t^{(1)}$ is the output of first LSTM layer.

The fully connected layer is defined by:

$$s = W h_t^{(2)} + b \tag{17}$$

where $h_t^{(2)}$ is the output of second LSTM layer.

The Relu activation layer reads:

$$f(s) = max(0, s) \tag{18}$$

where s is the output of the fully connected layer.

To implement the trained LSTM model in the constitutive program, the internal parameters of the trained network are stored and extracted by a Python program. Since the model contains a large number of parameters, the calculation of each network layer is implemented in the constitutive program by means of loop for the sake of simplicity. The implementation process of the constitutive program is shown in Algorithm 1.

4.2. Application to red sandstone

The above constitutive program is embedded into the FEM program to simulate the mechanical behaviors of red sandstone in the conventional triaxial compression tests with the same experimental data.

Algorithm 1: Rock material constitutive program based on LSTM deep-learning network

Input: $\boldsymbol{\varepsilon}^t, \boldsymbol{\sigma}^t, \mathrm{d}\boldsymbol{\varepsilon}^{t+1}, \boldsymbol{STATEV}^t$
Output: $\boldsymbol{\sigma}^{t+1}, \mathbb{C}^{t+1}, \mathrm{d}\boldsymbol{\varepsilon}^{t+1}, \boldsymbol{STATEV}^{t+1}$

$\boldsymbol{STATEV}^{t+1}(249) = \varepsilon_1^t + \mathrm{d}\varepsilon_1^{t+1}$;
$\boldsymbol{STATEV}^{t+1}(250) = \sigma_3^t$;
if $\boldsymbol{STATEV}^{t+1}(251) < m$ **then**
  Isotropic linear elastic constitutive model;
  $\mathbb{C}^{t+1} = \mathbb{C}_0^{t+1}$;
  $\boldsymbol{\sigma}^{t+1} = \boldsymbol{\sigma}^t + \mathbb{C}^{t+1} : \mathrm{d}\boldsymbol{\varepsilon}^{t+1}$;
else
  LSTM neural network model;
  for $j = 1, 2, 3, ..., 50$ **do**
    The output of the first LSTM layer: h_1;
    The output of the second LSTM layer: h_2;
  for $k = 1, 2, 3$ **do**
    The fully connected layer and the Relu activation layer: s;
  $\sigma_1^{t+1} = \boldsymbol{s}(2) \times \boldsymbol{rangeEntries}(2) + \boldsymbol{minEntries}(2)$;
  $\mathrm{d}\varepsilon_2^{t+1} = \boldsymbol{s}(3) \times \boldsymbol{rangeEntries}(3) + \boldsymbol{minEntries}(3)$;
  $\mathrm{d}\varepsilon_2^{t+1} = \boldsymbol{s}(3) \times \boldsymbol{rangeEntries}(3) + \boldsymbol{minEntries}(3)$;
  $\mathbb{C}^{t+1} = \mathbb{C}_0^{t+1}$;
for $i = 1, 2, 3, ..., 245$ **do**
  $\boldsymbol{STATEV}$ update;
  $\boldsymbol{STATEV}^{t+1}(i) = \boldsymbol{STATEV}^t(i+5)$;
$\boldsymbol{STATEV}^{t+1}(246) = \varepsilon_1^t + \mathrm{d}\varepsilon_1^{t+1} + \varepsilon_2^t + \mathrm{d}\varepsilon_2^{t+1} + \varepsilon_3^t + \mathrm{d}\varepsilon_3^{t+1}$;
$\boldsymbol{STATEV}^{t+1}(247) = \sigma_1^{t+1} - \sigma_3^{t+1}$;
$\boldsymbol{STATEV}^{t+1}(248) = \varepsilon_3^t + \mathrm{d}\varepsilon_3^{t+1}$;
$\boldsymbol{STATEV}^{t+1}(251) = \boldsymbol{STATEV}^t(251) + 1$;

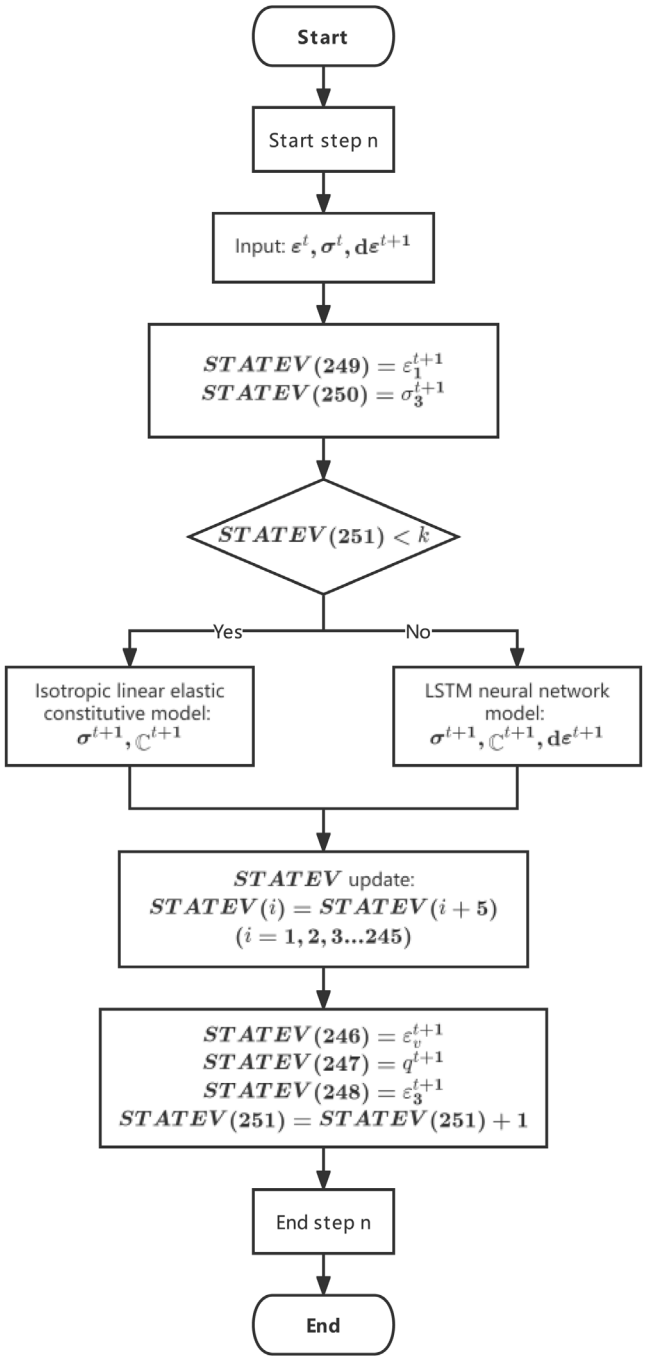


Fig. 13. Flowchart of implementation of LSTM network in FEM program.

Consequently, the parameters of trained LSTM model in Section 3.2 are adopted to predict the deformation of a single Gaussian point under the confining pressure of 20 MPa. The total axial strain is set as 0.02, which is equally divided into $\frac{0.02}{\mathrm{d}\varepsilon_1}$ pieces referring to the increment size $\mathrm{d}\varepsilon_1$.

Computations with respect to different incremental size of the axial strain are carried out by considering 500, 1000, 2500, and 5000 steps (see Fig. 14). The experimental data are represented by cycle points, and the solid lines stand for the numerical simulations by LSTM-FEM program. It can be seen that for all the computations, the obtained results by LSTM-FEM program are acceptable before the peak stress. But the strain–stress curve cannot be accurately predicted if the incremental size is too large. As shown in Fig. 14(d), when the magnitude of strain increment is set as 4×10^{-6} with 5000 incremental steps, the obtained strain–stress curve is in good agreement with the experimental results. And magnitude of strain increment in each time step of the artificial neural network model training set used in the constitutive program is imposed as 10^{-8}, which shows that when the strain increment is closer to the order of magnitude of the model training set, the model can better simulate the main mechanical behavior of rock.

Table 1
Comparisons of the predicted results by LSTM-FEM program to experimental results with respect to different incremental steps.

No.	Increment steps	$\mathrm{d}\varepsilon_1$	MSE	R^2
1	500	4×10^{-5}	94.5586	0.9538
2	1000	2×10^{-5}	30.0291	0.9853
3	2500	8×10^{-6}	8.9733	0.9956
4	5000	4×10^{-6}	9.4021	0.9954

Table 2
Comparisons of the predicted results by LSTM-FEM program to that of LSTM model with respect to different incremental steps.

No.	Increment steps	$\mathrm{d}\varepsilon_1$	MSE	R^2
1	500	4×10^{-5}	156.7304	0.9169
2	1000	2×10^{-5}	54.6694	0.9710
3	2500	8×10^{-6}	8.7939	0.9953
4	5000	4×10^{-6}	0.5817	0.9997

The loss and accuracy of the predicted values by LSTM-FEM numerical simulations compared to experimental results with respect to different incremental sizes are shown in Table 1, and those compared to LSTM predictions are provided in Table 2. It can be concluded that the accuracy is obviously improved with the increase of incremental steps, and the numerical predictions by LSTM-FEM program can captures the main deformation features of rocks.

5. Conclusions

In this study, a deep-learning network has been proposed to predict the constitutive relations of rock materials, especially for the strain-softening effects. Considering that the strain–stress behavior of rocks is affected by the loading history, the LSTM method is adopted to calculate the deformation in conventional triaxial compression tests. The volumetric strain ε_v^{t-1}, deviatoric stress q^{t-1} and lateral strain ε_3^{t-1} at the previous moment and the axial strain ε_1^t and lateral stress σ_3^t at the current moment are taken as inputs, and the volumetric strain ε_v^t, deviatoric stress q^t and lateral strain ε_3^t at the current moment are taken as outputs to train the model. Compared with traditional methods, neural network method can accurately predict the mechanical properties of materials, even with strong strain-softening characteristics, without introducing any elastoplastic parameters or constitutive laws.

Two applications of the constructed LSTM model on the analytical constitutive model of granite and experimental tests of sandstone have been provided. The comparisons demonstrate that the LSTM-based method can capture the main mechanical features of rocks, such as strong strain-softening, even if the materials' properties are unknown. Besides, the materials' degradation during loading process can also be described by the LSTM model, indicating the evolution of material's nature. The application of proposed LSTM model in simulating the behaviors under loading and unloading conditions has been also carried out, which verifies the accuracy of the model in predicting rock deformation under different loading paths. In the end, the constructed LSTM model has been implemented in FEM program as a constitutive program, and successfully applied to the prediction of the deformation of sandstone.

In the out look, to expand the LSTM-FEM program to the structural computation of common geotechnical constructions, such as underground tunnel and slope, will be a challenging and practical task.

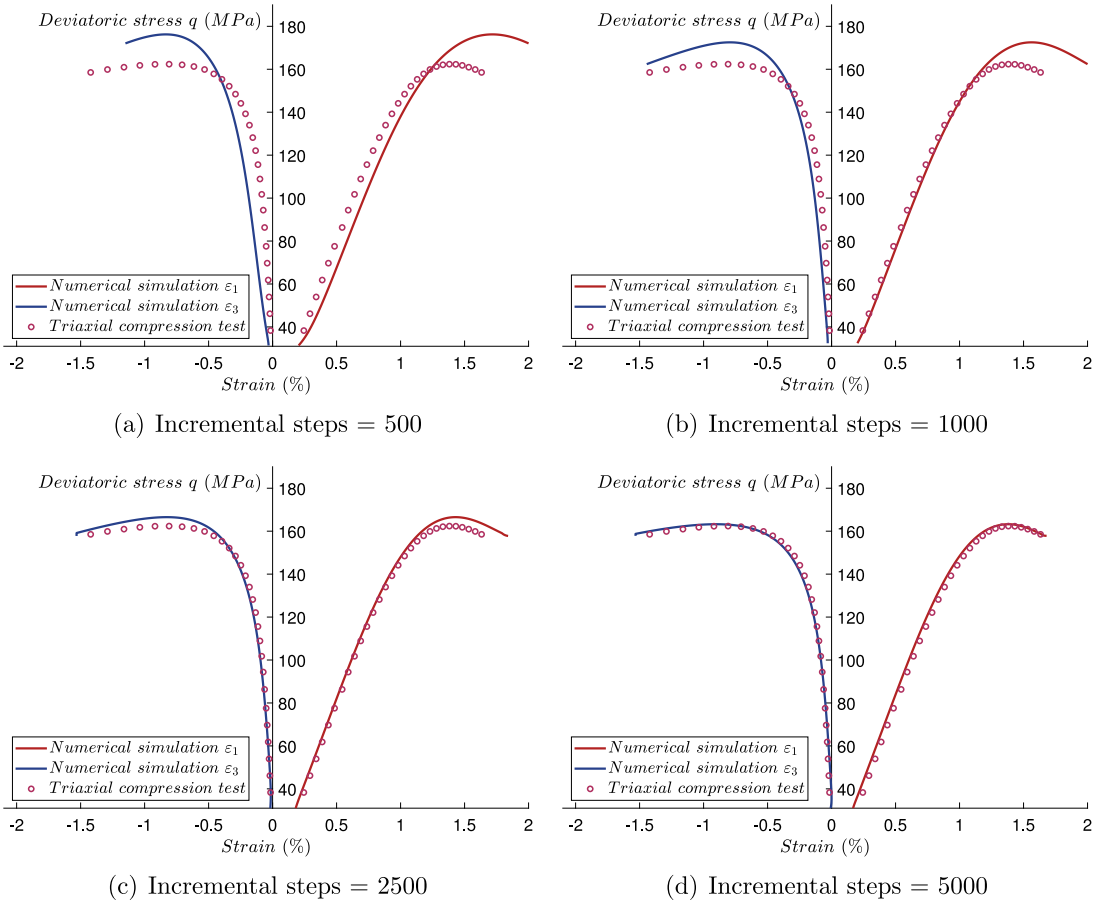


Fig. 14. Numerical simulation by LSTM-FEM program of the mechanical behavior of red sandstone with confining pressure of 20 MPa

CRediT authorship contribution statement

Shi L.L.: Data curation, Software, Formal analysis, Writing – original draft. **Zhang J.:** Methodology, Conceptualization, Writing – review & editing. **Zhu Q.Z.:** Methodology, Supervision, Funding acquisition. **Sun H.H.:** Software, Visualization.

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.

Acknowledgments

The authors would like to thank the National Natural Science Foundation of China (Grant No. 11902111) and the Fundamental Research Funds for the Central Universities (Grant No. B210202045).

Appendix. Brief recall of the micromechanical friction-damage model in Zhu et al. (2015)

The existence of microcracks leads to the discontinuity of displacement field in rock. To describe the behaviors of rock, the total strain $\boldsymbol{\varepsilon}$ is divided into elastic and inelastic components:

$$\boldsymbol{\varepsilon} = \boldsymbol{\varepsilon}^m + \boldsymbol{\varepsilon}^c \tag{A.1}$$

and the constitutive relation can be simply described by the generalized Hooke's law:

$$\boldsymbol{\sigma} = \mathbb{C}^m : (\boldsymbol{\varepsilon} - \boldsymbol{\varepsilon}^c) \tag{A.2}$$

where $\mathbb{C}^m$ is the initial fourth-order stiffness tensor of materials.

Mohr–Coulomb yield criterion based on local stress is expressed as,

$$f_s(\boldsymbol{\sigma}^c) = \| \boldsymbol{s}^c \| + \eta p^c \leqslant 0 \tag{A.3}$$

where $\boldsymbol{s}^c$ is deviatoric stress of $\boldsymbol{\sigma}^c$, defined by $\boldsymbol{s}^c = \mathbb{K} : \boldsymbol{\sigma}^c$. And p^c denotes mean stress of $\boldsymbol{\sigma}^c$, defined by $p^c = \mathrm{tr}\boldsymbol{\sigma}^c/3$.

After a series of derivation, under conventional triaxial loading, $\| \boldsymbol{s}^c \|$ and p^c can be derived as,

$$\| \boldsymbol{s}^c \| = -\sqrt{\frac{2}{3}}(\sigma_1 - \sigma_3) - \frac{2\mu^m}{\eta_2}\frac{\Lambda^c}{\omega} \tag{A.4}$$

$$p^c = \frac{1}{3}(\sigma_1 + 2\sigma_3) - \frac{k^m \eta}{\eta_1}\frac{\Lambda^c}{\omega} \tag{A.5}$$

in which Λ^c represents cumulative inelastic variable, and η_1 and η_2 are the constants only related to Poisson's ratio μ^m.

To simplify the equation, with $\chi = \frac{k^m \eta^2}{2\eta_1} + \frac{\mu^m}{\eta_2}$, the friction-damage strength criterion can be changed to the form below:

$$f_s = -\sqrt{\frac{2}{3}}(\sigma_1 - \sigma_3) + \frac{1}{3}\eta(\sigma_1 + 2\sigma_3) - 2\chi\frac{\Lambda^c}{\omega} \leqslant 0 \tag{A.6}$$

The damage criterion $f_\omega = (\frac{k^m\eta^2}{2\eta_1} + \frac{\mu^m}{\eta_2})(\frac{\Lambda^c}{\omega})^2 - R(\omega) = 0$ is adopted to obtain the relation:

$$\frac{\Lambda^c}{\omega} = \sqrt{\frac{R(\omega)}{\chi}} \tag{A.7}$$

With the pressure as positive, using Eq. (A.7) into Eq. (A.6), we obtain:

$$f_s = \sigma_1 - \frac{\sqrt{6}+2\eta}{\sqrt{6}-\eta}\sigma_3 - \frac{6\sqrt{R(\omega)\chi}}{\sqrt{6}-\eta} \leqslant 0 \tag{A.8}$$

Considering that the mean stress is defined as $p = (\sigma_1 + 2\sigma_3)/3$, and deviatoric stress is $q = \sigma_1 - \sigma_3$. The final friction-damage strength criterion is:

$$f_s = q - \sqrt{\frac{3}{2}}\eta p - \sqrt{6R(\omega_c)\chi} \leq 0 \tag{A.9}$$

References

Andrieux, S., Bamberger, Y.M., 1986. JJ un modéle de materiau microfissure pour les roches et les betons. J. Méc. Théo. Et Appl. 5, 471–513.

Areias, P., Dias-da Costa, D., Pires, E., Barbosa, J.I., 2012. A new semi-implicit formulation for multiple-surface flow rules in multiplicative plasticity. Comput. Mech. 49 (5), 545–564.

Bahdanau, D., Cho, K., Bengio, Y., 2014. Neural machine translation by jointly learning to align and translate. arXiv preprint arXiv:1409.0473.

Baraldi, L., Grana, C., Cucchiara, R., 2017. Hierarchical boundary-aware neural encoder for video captioning. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. pp. 1657–1666.

Bezanson, J., Edelman, A., Karpinski, S., Shah, V.B., 2017. Julia: A fresh approach to numerical computing. SIAM Rev. 59 (1), 65–98.

Burden, R.L., Faires, J.D., Burden, A.M., 2015. Numerical Analysis. Cengage learning.

Carol, I., Bazant, Z.P., 1997. Damage and plasticity in microplane theory. Int. J. Solids Struct. 34 (29), 3807–3835.

De Sciarra, F.M., 2012. Hardening plasticity with nonlocal strain damage. Int. J. Plast. 34, 114–138.

Dormieux, L., Kondo, D., Ulm, F.J., 2006. A micromechanical analysis of damage propagation in fluid-saturated cracked media. C. R. Mec. 334 (7), 440–446.

Dragon, A., Mroz, Z., 1979. A continuum model for plastic-brittle behaviour of rock and concrete. Internat. J. Engrg. Sci. 17 (2), 121–137.

Fang, K., Pan, M., Shen, C., 2018. The value of SMAP for long-term soil moisture estimation with the help of deep learning. IEEE Trans. Geosci. Remote Sens. 57 (4), 2221–2233.

Furukawa, T., Yagawa, G., 1998. Implicit constitutive modelling for viscoplasticity using neural networks. Internat. J. Numer. Methods Engrg. 43 (2), 195–219.

Galerkin, B.G., 1915. Series solution of some problems of elastic equilibrium of rods and plates. Vestnik Inzhenerov I Tekhn. 19 (7), 897–908.

Ghaboussi, J., Garrett Jr., J., Wu, X., 1991. Knowledge-based modeling of material behavior with neural networks. J. Eng. Mech. 117 (1), 132–153.

Ghaei, A., Green, D., Taherizadeh, A., 2010. Semi-implicit numerical integration of Yoshida–Uemori two-surface plasticity model. Int. J. Mech. Sci. 52 (4), 531–540.

Graves, A., Mohamed, A.r., Hinton, G., 2013. Speech recognition with deep recurrent neural networks. In: 2013 IEEE International Conference on Acoustics, Speech and Signal Processing. Ieee, pp. 6645–6649.

Hansen, N., Schreyer, H., 1994. A thermodynamically consistent framework for theories of elastoplasticity coupled with damage. Int. J. Solids Struct. 31 (3), 359–389.

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

Horii, H., Nemat-Nasser, S., 1983. Overall moduli of solids with microcracks: Load-induced anisotropy. J. Mech. Phys. Solids 31 (2), 155–171.

Hu, M., Li, W., Yan, K., Ji, Z., Hu, H., 2019. Modern machine learning techniques for univariate tunnel settlement forecasting: A comparative study. Math. Probl. Eng. 2019.

Huang, L., Li, J., Hao, H., Li, X., 2018. Micro-seismic event detection and location in underground mines by using convolutional neural networks (CNN) and deep learning. Tunn. Undergr. Space Technol. 81, 265–276.

Hughes, T.J., 2012. The Finite Element Method: Linear Static and Dynamic Finite Element Analysis. Courier Corporation.

Jin, J., She, C., Shang, P., 2020. A strain-softening model of rock based on Hoek-Brown criterion. Rock Soil Mech. 41, 939–951.

Kardani, N., Zhou, A., Nazem, M., Shen, S.-L., 2021. Improved prediction of slope stability using a hybrid stacking ensemble method based on finite element analysis and field data. J. Rock Mech. Geotech. Eng. 13 (1), 188–201.

Khan, A.S., Xiang, Y., Huang, S., 1991. Behavior of Berea sandstone under confining pressure part I: Yield and failure surfaces, and nonlinear elastic response. Int. J. Plast. 7 (6), 607–624.

Kingma, D.P., Ba, J., 2014. Adam: A method for stochastic optimization. arXiv preprint arXiv:1412.6980.

Li, J., Li, P., Guo, D., Li, X., Chen, Z., 2021. Advanced prediction of tunnel boring machine performance based on big data. Geosci. Front. 12 (1), 331–338.

Li, Q., Zhao, Y., Yu, F., 2020. A novel multichannel long short-term memory method with time series for soil temperature modeling. IEEE Access 8, 182026–182043.

Liu, Y., Dai, F., 2018. A damage constitutive model for intermittent jointed rocks under cyclic uniaxial compression. Int. J. Rock Mech. Min. Sci. 103, 289–301.

Liu, Z., Li, L., Fang, X., Qi, W., Shen, J., Zhou, H., Zhang, Y., 2021. Hard-rock tunnel lithology prediction with TBM construction big data using a global-attention-mechanism-based LSTM network. Autom. Constr. 125, 103647.

Ma, L., Xu, H., Tong, Q., Dong, L., Zhang, N., Li, J., 2014. Post-yield plastic frictional parameters of a rock salt using the concept of mobilized strength. Eng. Geol. 177, 25–31.

Mahdevari, S., Shahriar, K., Sharifzadeh, M., Tannant, D.D., 2017. Stability prediction of gate roadways in longwall mining using artificial neural networks. Neural Comput. Appl. 28 (11), 3537–3555.

Martin, C., Chandler, N., 1994. The progressive fracture of Lac du Bonnet granite. Int. J. Rock Mech. Min. Sci. Geomech. Abstracts 31 (6), 643–659.

Nayak, G., Zienkiewicz, O., 1972. Elasto-plastic stress analysis. A generalization for various contitutive relations including strain softening. Internat. J. Numer. Methods Engrg. 5 (1), 113–135.

Olah, C., 2015. Understanding LSTM networks. URL http://colah.github.io/posts/2015-08-Understanding-LSTMs/.

Parisio, F., Samat, S., Laloui, L., 2015. Constitutive analysis of shale: A coupled damage plasticity approach. Int. J. Solids Struct. 75, 88–98.

Peng, Q., Chen, M., 2012. An efficient return mapping algorithm for general isotropic elastoplasticity in principal space. Comput. Struct. 92, 173–184.

Pourhosseini, O., Shabanimashcool, M., 2014. Development of an elasto-plastic constitutive model for intact rocks. Int. J. Rock Mech. Min. Sci. 66, 1–12.

Prat, P.C., Bažant, Z.P., 1997. Tangential stiffness of elastic materials with systems of growing or closing cracks. J. Mech. Phys. Solids 45 (4), 611–636.

Rumelhart, D.E., Hinton, G.E., Williams, R.J., 1986. Learning representations by back-propagating errors. Nature 323 (6088), 533–536.

Shao, J.F., Jia, Y., Kondo, D., Chiarelli, A.S., 2006. A coupled elastoplastic damage model for semi-brittle materials and extension to unsaturated conditions. Mech. Mater. 38 (3), 218–232.

Shao, J.F., Zhu, Q.Z., Su, K., 2003. Modeling of creep in rock materials in terms of material degradation. Comput. Geotech. 30 (7), 549–555.

Shen, W., Shao, J.-F., 2016. An incremental micro-macro model for porous geomaterials with double porosity and inclusion. Int. J. Plast. 83, 37–54.

Simo, J., Ortiz, M., 1985. A unified approach to finite deformation elastoplastic analysis based on the use of hyperelastic constitutive equations. Comput. Methods Appl. Mech. Engrg. 49 (2), 221–245.

Simo, J.C., Taylor, R.L., 1985. Consistent tangent operators for rate-independent elastoplasticity. Comput. Methods Appl. Mech. Engrg. 48 (1), 101–118.

Souley, M., Renaud, V., Al Heib, M., Bouffier, C., Lahaie, F., Nyström, A., 2018. Numerical investigation of the development of the excavation damaged zone around a deep polymetallic ore mine. Int. J. Rock Mech. Min. Sci. 106, 165–175.

Sutskever, I., Vinyals, O., Le, Q.V., 2014. Sequence to sequence learning with neural networks. Adv. Neural Inf. Process. Syst. 27.

Tan, X., Konietzky, H., Frühwirt, T., 2014. Laboratory observation and numerical simulation of permeability evolution during progressive failure of brittle rocks. Int. J. Rock Mech. Min. Sci. 68, 167–176.

Vinyals, O., Toshev, A., Bengio, S., Erhan, D., 2015. Show and tell: A neural image caption generator. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. pp. 3156–3164.

Walsh, J.B., 1980. Static deformation of rock. J. Eng. Mech. Div. 106 (5), 1005–1019.

Wang, K., Sun, W., 2018. A multiscale multi-permeability poroplasticity model linked by recursive homogenizations and deep learning. Comput. Methods Appl. Mech. Engrg. 334, 337–380.

Wawersik, W., Brace, W., 1971. Post-failure behavior of a granite and diabase. Rock Mech. 3 (2), 61–85.

Wu, R., Fujita, Y., Soga, K., 2020. Integrating domain knowledge with deep learning models: An interpretable AI system for automatic work progress identification of NATM tunnels. Tunn. Undergr. Space Technol. 105, 103558.

Xu, K., Ba, J., Kiros, R., Cho, K., Courville, A., Salakhudinov, R., Zemel, R., Bengio, Y., 2015. Show, attend and tell: Neural image caption generation with visual attention. In: International Conference on Machine Learning. PMLR, pp. 2048–2057.

Yuan, S.S., Zhu, Q.Z., Zhao, L.Y., Chen, L., Shao, J.F., Zhang, J., 2020. Micromechanical modelling of short-and long-term behavior of saturated quasi-brittle rocks. Mech. Mater. 142, 103298.

Zhang, W., Li, H., Li, Y., Liu, H., Chen, Y., Ding, X., 2021a. Application of deep learning algorithms in geotechnical engineering: A short critical review. Artif. Intell. Rev. 54 (8), 5633–5673.

Zhang, J., Shen, W., Oueslati, A., De Saxcé, G., 2017. Shakedown of porous materials. Int. J. Plast. 95, 123–141.

Zhang, N., Shen, S.L., Zhou, A., Jin, Y.F., 2021b. Application of LSTM approach for modelling stress–strain behaviour of soil. Appl. Soft Comput. 100, 106959.

Zhang, P., Yin, Z.Y., Jin, Y.F., 2021c. State-of-the-art review of machine learning applications in constitutive modeling of soils. Arch. Comput. Methods Eng. 28 (5), 3661–3686.

Zhou, J., Shi, X., Du, K., Qiu, X., Li, X., Mitri, H.S., 2017. Feasibility of random-forest approach for prediction of ground settlements induced by the construction of a shield-driven tunnel. Int. J. Geomech. 17 (6), 04016129.

Zhu, Q.Z., Kondo, D., Shao, J., 2008. Micromechanical analysis of coupling between anisotropic damage and friction in quasi brittle materials: Role of the homogenization scheme. Int. J. Solids Struct. 45 (5), 1385–1405.

Zhu, Q., Liu, H., Wei, W., Shao, J.F., 2015. A micromechanical constitutive damage model for Beishan granite. Chin. J. Rock Mech. Engi. 34 (003), 433–439.
Zhu, Q., Shao, J.F., 2015. A refined micromechanical damage–friction model with strength prediction for rock-like materials under compression. Int. J. Solids Struct. 60, 75–83.
Zhu, Q., Shao, J.F., Mainguy, M., 2010a. A micromechanics-based elastoplastic damage model for granular materials at low confining pressure. Int. J. Plast. 26 (4), 586–602.
Zhu, Q., Zhou, C., Shao, J.F., Kondo, D., 2010b. A discrete thermodynamic approach for anisotropic plastic–damage modeling of cohesive-frictional geomaterials. Int. J. Numer. Anal. Methods Geomech. 34 (12), 1250–1270.
Zienkiewicz, O.C., Taylor, R.L., 2005. The Finite Element Method for Solid and Structural Mechanics. Elsevier.
Zienkiewicz, O., Valliappan, S., King, I., 1969. Elasto-plastic solutions of engineering problems 'initial stress', finite element approach. Internat. J. Numer. Methods Engrg. 1 (1), 75–100.