Deep Learning Model for Shield Tunneling Advance Rate Prediction in Mixed Ground Condition Considering Past Operations

QIANG WANG1,2, XIONGYAO XIE1,2, AND ISAM SHAHROUR1,3

1School of Civil Engineering, Tongji University, Shanghai 200092, China
2Key Laboratory of Geotechnical and Underground Engineering, Ministry of Education, Tongji University, Shanghai 200092, China
3Laboratoire de Génie Civil et géo-Environnement, Lille 1 University, 59650 Villeneuve d’Ascq, France

Corresponding author: Xiongyao Xie (xiexiongyao@tongji.edu.cn)

This work was supported in part by the National Key Research and Department Program of China under Grant 2018YFC0809600 and Grant 2018YFC0809601, in part by the National Nature Science Funds of China under Grant 51778476, and in part by the Shanghai Science and Technology Development Funds under Grant 18DZ1205200.

ABSTRACT

The advance rate (AR) is a significant parameter in shield tunneling construction, which has a major impact on construction efficiency. From a practical perspective, it’s helpful to establish a predictive model of the AR, which takes into account the instantaneous parameters as well as the past operations. However, for shield tunneling in mixed ground conditions, most researches focused on the average values of AR per ring and neglect the influence of past operations. This article presents a long short-term memory (LSTM) recurrent neural network model, which was developed for the slurry shield tunneling in a mixed ground of round gravel and mudstone in Nanning metro. A temporal aggregated random forest is employed to rank the importance of the explanatory features. The model performances in different ground conditions are investigated. The results show that the LSTM model can be effectively implemented for the AR prediction. A high correlation is observed between predicted and measured AR with a correlation coefficient ($R^2$) of 0.93. The LSTM based AR predictive model is compared with the random forest (RF) model, the deep feedforward network (DFN) model, and the support vector regression (SVR) model. The comparison shows that the LSTM model has the best performances compared to other models. With one-fourth features, we can achieve a 95% prediction accuracy measured by the $R^2$ in the proposed model.

INDEX TERMS

Past operations, shield tunneling, advance rate prediction, deep learning, feature importance, mixed ground.

I. INTRODUCTION

Owing to the growing demand for efficient inter/intra-city transportation in China, shield tunneling in mixed ground conditions has been becoming increasingly common [1], [2]. It’s crucial to maintain the advance rate (AR) at a reasonable level in shield tunneling construction to ensure construction safety, control the shield attitude and improve the tunneling efficiency, especially in mixed ground condition. However, the complex geological conditions in the mixed ground resulted in great difficulties in shield tunneling control with severe consequences, such as excavation face instability [3] as well as high cutter wear [4]. In difficult ground conditions, such as the mudstone rich mixed ground, operators face difficulties to predict tunneling parameters’ change rules. They didn’t know whether their operations would lead to good or bad tunneling control. Therefore, to improve the quality of tunneling control, engineers need a good understanding of the time-varying rules of the instantaneous shield tunneling parameters. Thus, it’s necessary to establish an AR predictive model to achieve optimal operation conditions. Unfortunately, because of the complex shield-soil interaction, numerical methods are incapable of predicting the AR [5], consequently, engineers use simplified analytical approaches [6].

As a data-driven method, the machine learning (ML) based approach has been used in the AR prediction for earth pressure balanced (EPB) shield tunneling, slurry pressured balanced (SPB) shield tunneling, and hard rock tunneling boring machine (TBM) [7], [8].
Benardos et al. [9] employed the Artificial Neural Network (ANN) to predict the hard rock TBM AR in the Athens metro tunnel. They considered the rock properties as the input but ignored operation parameters. Yagiz [10] proposed a valuable dataset containing the rock mass properties and the penetration rate in Queens Water Tunnel. Data were used for the prediction of penetration rate ($PR = AR/rotation$ speed of cutterhead) via fuzzy inference system [11], ANN [12], particle swarm optimization (PSO) [13], extreme learning machine (ELM) [14], and support vector machine (SVR) [15]. Armaghani et al. [16] summarized researches about ML-based AR predictive models in hard rock TBM and applied the ANN model with optimization method to predict the AR in the Pahang- Selangor Raw Water Transfer (PSRWT) tunnel in Malaysia. Salimi et al. [17] employed an adaptive neuro-fuzzy inference system (ANFIS) and SVR to develop an AR predictive model in a hard rock TBM tunnel in Iran. Results indicated that the SVR model performed better than the ANFIS model. The ML-based AR predictive model has been rarely used in the SPB and EPB shield tunneling. Some statistical models were used in EPB shield tunneling [18], [19]. Li et al. [20] established a multiple nonlinear regression model for the AR predicting in the mixed ground condition using the orthogonal tests of tunneling parameters. The average values of each ring from an intercity rail transit shield tunnel in Guangdong Province were adopted in this predictive model. The model reached a correlation coefficient $R^2 = 0.693$. Sun et al. [21] proposed a random forest AR predictive model in mixed ground conditions containing clay, sand, and rock. The model employed the instantaneous values of the shield tunneling parameters. However, the RF AR predictive model achieved only $R^2 = 0.803$ in the test set. Elbaz et al employed ANFIS with genetic algorithm (GA) [22] and improved PSO [23] for the AR prediction of the shield in the tunnel section of Guangzhou Metro Line 9 in China, which performed well with $R^2 = 0.88$ in small data set (200 samples in totals).

Although great achievements have been made in the area of ML-based AR predictive methods for tunneling, engineers still face serious limitations: (i) data used for training and testing in the ML model is quite limited, which is usually averaged over one ring’s (EPB and SPB) [23], [24] or several cycles (hard rock TBM) in tunneling history and (ii) the temporal effect is rarely considered; only one step-back’s data is taken into consideration in the ML model, especially for the EPB and SPB. As a sperrmatic machine, hundreds of sensors on the shield record the shield-soil interactive responses via the programmable logic controller (PLC) every 1~10 seconds. These instantaneous values are fundamental for on-site operators to control the tunneling process. Consequently, it’s particularly interesting to establish an AR predictive model with instantaneous values considering a longer past operation effect.

When we consider the instantaneous values of shield tunneling parameters, one big challenge is to deal with massive data, which exceeds the capacity of traditional ML methods [25] (e.g. ANN with two or three layers, SVR). Deep learning (DL) based methods are suitable for predicting the instantaneous AR, because of their ability to work with big data. DL model uses multiple layers to progressively extract higher-level features from the raw input[26]. In the last few years, deep learning researches focused on both academia and industry [26] to process massive data with limited computing resources, extract high-dimensional features, and strong nonlinear fitting ability[27]–[29]. In civil engineering, DL based methods have been widely adopted in crack identification [30], [31], micro-seismic event detection and location in underground mines [32] as well as safety management in construction sites [33]–[35]. To learn more patterns from the past time series, the latest scholars started to use a more advanced type of new network, namely, the long short-term memory (LSTM) neural network. LSTM can remember information for much longer periods due to its recurrent structure and gating mechanisms and is regarded as a state-of-the-art method for time series related problems, such as the prediction of landslide displacement [36], [37], traffic flow [38], [39], short-term water demand [40], stock market prices [41] and soil moisture [42]. Camila Correa-Jullian et al. [43] proposed LSTM based deep learning model to predict the performance of a solar hot water system, whose model employed stacked LSTM incorporating an MLP as hidden layers before the output layer to obtain the desired predicted value. The LSTM based deep learning model outperforms other ML-based models, such as ANN, Bayesian Ridge, Gaussian Process, and Linear Regression. He et al. [44] designed a framework named Traffic Pattern centric Segment Coalescing Framework (TP-SCF) to predict riding time during bus journey travel via LSTM based model, which notably outperforms the state-of-the-art approaches for all the scenarios considered. Li et al. [45] proposed a hybrid deep learning model combining wavelet packet decomposition (WPD) and long short-term memory (LSTM) network to predict the Photovoltaic (PV) power, which used the past PV series as input. Jang et al. [46] employed a deep learning model using LSTM to predict the business failure of the construction market, whose prediction model had superior prediction performance from a long-term perspective when the construction market and macroeconomic variables were used in addition to accounting variables. Karimi et al. [47] adopted an LSTM based model to predict the wastewater flow in the smart sewer management. They harbor the idea that LSTM is suitable for prediction when the dynamic of the system is changing. However, although LSTM based deep learning methods can obtain good results in time series prediction, the implementation of these methods requires a sufficient set of past data. Otherwise, these networks can’t learn the patterns behind the time series and therefore provide limited performance. Meanwhile, Li et al. [48] presented the LSTM neural network model for downscaling long-term daily temperature, but the model has difficulty in predicting extreme values for temperature due to the insufficient data and same variable type of input and output.

Q. Wang et al.: Deep Learning Model for Shield Tunneling AR Prediction
Few kinds of research conducted numerical experiments on large scale data set when they employed the deep learning methods in shield tunneling parameters prediction. For example, Gao et al. [49] proposed an RNN (including LSTM) model to predict the shield tunneling parameters including the cutterhead torque (TOR), cutterhead total thrust, and AR, which outperforms traditional ML regression methods (such as SVR, RF, and Lasso[50]) in most cases. However, only 3000 samples from a tunnel section in Shenzhen metro were put forward. The feasibility of RNN was not demonstrated (including LSTM) in dealing with the huge data set of instantaneous shield tunneling parameters. Besides, in Gao’ model, the AR at time $t-1$ is used as an input variable to predict the AR at time $t$. This use may result in that the output AR (time $t$) value simply mimics the input AR (time $t-1$) due to the short interval (usually one to ten seconds in PLC system) between time $t-1$ and time $t$. Moreover, Gao’s model only compares different ML-based model performance but ignores the influences of input tunneling parameters. A total of 43 kinds of PLC data for input seems redundant, especially for guidance of shield tunneling driving.

Motivated by recent advances in deep learning methods and their performance in time series prediction, we presented an LSTM based deep learning model for the AR prediction on large data set of a tunnel section in the mixed ground condition. The AR changed frequently and remained at a low level due to clogging problem in the mixed ground condition. The frequently-changed AR cannot be controlled by the shield operators. As a result, we developed the AR prediction model to help operators to enhance the understanding of the AR time-varying rules. Our contributions can be summarized as follows: Firstly, we developed an AR prediction model with multivariable tunneling parameters (not including the AR) at time $t-1$ to $t-k$ as inputs and the AR at time $t$ as output. Secondly, we proposed a temporal aggregated random forest to rank the importance of the explanatory features. We investigated the influence of the input PLC data on AR prediction and using one-third features to achieve a 97% prediction accuracy measured by the $R^2$ in the proposed model. Thirdly, we investigated the model performance in a tunnel section with different ground conditions, 665 rings and 411,404 samples. To overcome the difficulties of AR prediction in mudstone rich areas, we developed a deep learning model, which improved the prediction accuracy in the mudstone rich areas.

This article firstly introduces the data from the tunnel project in Nanning metro in section 2, then it proposed an LSTM based AR predictive model. The prediction results will be discussed in different ground conditions as well as the performances of different models.

## II. MATERIAL AND DATA

### A. BRIEF INTRODUCTION

Data were collected from monitoring a section of line 1 of Nanning metro. The section concerns the left line of Bailing station to Railway station (BR section). It includes 806 rings with a ring width of 1.5 m. Figure 1 shows the plan view and longitudinal geological profile of the tunnel. Figure 1 (a) shows adjacent buildings including 14 railway tracks, which constitute the main challenge in this section. Figure 1 (b) illustrates the soil distribution along the tunnel. 36 boreholes were drilled in the vicinity of the BR section as part of the geological site investigation. It could be observed that the round gravel and mudstone are frequently encountered. According to the geological survey report, about 524 rings (65% of total rings) are located in the round gravel and 282 rings (35% of total rings) are located in the mixed ground, which is composed of gravel, mudstone, and sand.

Figure 1 (b) shows that the mudstone area is around ring # 120 to ring # 220, and ring # 283 to ring # 470. The tunnel is located below the underground water, and the distance between the underground water table to the tunnel crown varies between 1.5m and 9m. Figure 2 (a) and (b) show the typical borehole results of sand, round gravel, and mudstone. The borehole depth in Figure 2 (a) and (b) varies between 15m and 20m, and between 20m and 25m, respectively. Table 1 summarizes the soil properties. The sand and round gravel are similar, while the mudstone is different, in particular in cohesion and permeability.

The tunnel was excavated using a slurry pressure balanced shield machine, whose diameter is 6.28 m. Figure 2 (c) illustrates the front view of the cutterhead at the achievement of the construction of the previous section. The design parameters of the shield machine are shown in Table 2. The shield is equipped by cutters, scrapers, and tearing cutters to operate in both round gravel and mudstone. The opening ratio of the cutterhead is 35%. 32 thrust cylinders (16 double-cylinder) divided into four groups (A, B, C, and D) were employed (Figure 2 (d)), which provide a maximum total thrust of 42,575 MN (350 bar) with a maximum cylinder stroke of 2200 mm. The rated torque of the shield machine cutterhead is 5.488 MN·m while the breakout one is 6.619 MN·m. The maximum rotation speed of the cutterhead is 3 rpm, and the maximum AR of the shield machine is 50 mm/min.

### TABLE 1. Soil properties in BR section.

| Soil           | Density/g/cm³ | Poisson ratio | Cohesion/kPa | Friction angle/° | Deformation modulus/MPa | Permeability coefficient/m/d |
|---------------|---------------|---------------|--------------|------------------|------------------------|-----------------------------|
| Sand          | 2.02          | 0.30          | 0.0          | 28.0             | 16.0                   | 18                          |
| Round gravel  | 2.05          | 0.27          | 0.0          | 35.0             | 25.0                   | 90                          |
| Mudstone      | 2.15          | 0.20          | 90.0         | 21.0             | 30.0                   | 0.01                         |

This article firstly introduces the data from the tunnel project in Nanning metro in section 2, then it proposed an LSTM based AR predictive model. The prediction results will
B. DESCRIPTION OF PLC DATA
The presence of both round gravel and mudstone in the soil has a significant influence on tunneling parameters. Indeed, the mudstone can cause severe clogging problems, which could result in great fluctuations in tunneling parameters [1]. Since the AR is the target value, we plot the instantaneous AR in different ground conditions. Data is collected via the PLC system at a recording interval of 10 seconds. In the following section, the tunneling parameters will be named as PLC data. Figure 3 illustrates the AR in the mixed ground condition. It does not include periods related to working stop and ring building, where the AR remains zero. The time of 247 minutes represents the time cost that the shield machine achieved the 1.5m driving in the ring # 175.

Figure 3 shows that the AR in the mixed ground is smaller than that in the mudstone and round gravel. The tunneling time of ring # 175 is about twice longer than that of ring # 420 and nearly four times longer than that of ring # 680. The AR in the mixed ground varies from 2 mm/min to 23 mm/min, while it remains around 10 to 15 mm/min in the mudstone condition. The highest values of the AR are in
TABLE 2. Partial design parameters of shield machine in BR section.

| Contents                  | Parameters                        |
|---------------------------|-----------------------------------|
| Type                      | Slurry pressure balanced          |
| Total length (m)          | About 83                          |
| Number of thrust cylinders| 32 (16 double-cylinder)            |
| Thrust (MN)               | 42.575 (Maxium)                   |
| Torque (MN-m)             | 5.488 (Rated), 6.619 (Maxium)     |
| Opening ratio of cutterhead| 35%                               |
| Cutter number             | 16                                |
| Cutter diameter (in)      | 17                                |
| Scraper number            | 46                                |
| Tearing cutter number     | 36                                |

FIGURE 3. AR in the mixed ground (ring # 175), mudstone (ring # 420), and round gravel (ring # 680).

the round gravel condition. The AR in the mixed ground has higher fluctuations than in the mudstone and round gravel, which brings out difficulties in AR prediction. To meet the prediction requirements in mixed ground conditions, the AR predictive model should capture these patterns.

Since the DL model is good at dealing with higher-level features, we used the available input PLC parameters to obtain a good prediction. Then we will conduct the feature selection on the input parameters by the temporal aggregated random forest approach to reduce the inter-dependence and provide better guidance for shield operators. Based on existing research [1], [21] and field experience, the AR is influenced by the slurry pressure in the excavation chamber (SPE), slurry pressure in the working chamber (SPW), the cutterhead torque (TOR), total thrust(THR), rotation speed(RS), pump pressure of thrust cylinders(PP) including the pumps of Group A (PPA), the pumps of Group B (PPB), pumps of Group C (PPC), pumps of Group D (PPD), and flow rate in the feed (FRF) and return (FRR) line. These parameters are selected as input parameters in the AR predictive model. According to the conclusion of Mokhtari and Mooney [51], the influence of the geological parameter on the AR has been included in the input tunneling data, so we only use the PLC data as input features. Table 3 summarizes the abbreviations, units, and statistics for the input and output parameters. Figure 4 shows the time series of the above-mentioned PLC data in the mixed ground. Clogging occurred in ring # 175, because of the strong fluctuation of SPE. Such fluctuations are observed in the time series of torque, total thrust, and slurry flow rate in the return line. The pump pressure of thrust cylinders in group C is larger than that in group A. It indicates that the operators tried to prevent the shield knocking head in the mixed ground. The pump pressure of group D is larger than that of group B because the shield is turning right in this area (Figure 1 (a)).

III. DL BASED AR PREDICTIVE MODEL

The AR predictive model can is described in (1) and (2) using the PLC data $X_{t-k}$ at time $t-k$, where $k \in [1, n]$. $Y_t$ designates the AR at time $t$.

$$X_{t-k} = (SPE_{t-k}, SPW_{t-k}, TOR_{t-k}, THR_{t-k}, RS_{t-k}, PP_{t-k},$$
$$PPA_{t-k}, PPB_{t-k}, PPC_{t-k}, PPD_{t-k}, FRF_{t-k}, FRR_{t-k})$$

(1)

$$Y_t = F(X_{t-1}, X_{t-2}, \ldots, X_{t-k}, \ldots, X_{t-n})$$

(2)
A time window with width $n$ is employed for the following reasons: (i) the spatial distribution of sensors in the shield machine, which makes the AR at time $t$ influenced by the past PLC data and (ii) the complicated shield-soil interaction, e.g., the slurry infiltration and filter cake formation [52]. Similar to the natural language process (NLP), the previous operation will influence the current AR. Hence, the AR predictive model is supposed to have the ability to consider past data. According to the plentiful applications of LSTM in the area of the NLP[53],

**FIGURE 4.** PLC data related to AR of ring # 175 in the mixed ground, AR is the model output and other PLC data are the model input.
the LSTM based neural network is suitable for the AR prediction.

A. LSTM BASED NEURAL NETWORK

Figure 5 shows the typical architectures of the feedforward network and the recurrent neural network (RNN). The major difference between RNN and feedforward network is the feedback in the hidden layers of RNN[26]. If we unfold the RNN in the time dimension, the architecture will turn to that illustrated in Figure 5 (c). It’s clear that the output at time \( t \) is not only affected by the current input \( x_t \), but also the previous hidden layer \( h_{t-1} \). That’s the reason that RNN can have a ‘memory’ of the past inputs. However, while it is easy to implement the forward pass calculation, it’s difficult to conduct the backpropagation during the training of the RNN due to the gradient vanishing problem[54]. Therefore, the LSTM neural network [55], [56] is put forward whose hidden layer has an additional cell state, \( c \), as illustrated in Figure 5 (d).

The forward pass calculation of the LSTM hidden layer is realized with the gate layers, namely, the forget gate, the input gate, and the output gate, respectively, as illustrated in Figure 6. The gate layer is a kind of feedforward network with one hidden layer, whose output is a real number between 0 to 1 to decide which information to remember or to forget. The detailed calculation in Figure 6 can be expressed by (3).

\[
\begin{align*}
\tilde{f}_t & = \sigma(W_f \cdot [\vec{h}_{t-1}, \vec{x}_t] + \vec{b}_f) \\
\tilde{i}_t & = \sigma(W_i \cdot [\vec{h}_{t-1}, \vec{x}_t] + \vec{b}_i) \\
\tilde{c}_t' & = \tanh(W_c \cdot [\vec{h}_{t-1}, \vec{x}_t] + \vec{b}_c) \\
\tilde{c}_t & = \tilde{f}_t \odot \vec{c}_{t-1} + \tilde{i}_t \odot \tilde{c}_t' \\
\tilde{o}_t & = \sigma(W_o \cdot [\vec{h}_{t-1}, \vec{x}_t] + \vec{b}_o) \\
\vec{h}_t & = \tilde{o}_t \odot \tanh(\vec{c}_t)
\end{align*}
\]

B. MODELING PROCESS WITH LSTM

Figure 7 shows the flowchart to establish the DL based AR predictive model with the aid of the LSTM neural network. In the beginning, the PLC data selection process, as some data files are missing. Next, the data preprocessing is carried out before the data are fed to the LSTM layer. It includes the data cleaning, the conversion from the time-series data to supervised learning data, and data normalization. The DL model starts training with the data set randomly split into training, validation, and test set. With the hyper-parameters tuning based on experience, the model will obtain good generalization performances. In this study, the Keras (a high-level neural networks application program interface (API) written in Python) [57] is used to train the LSTM based AR predictive model and four Nvidia GeForce GTX 1080 Ti graphics processing units (GPU) are used to accelerate the computation. The GPU-accelerated deep learning frameworks make the training process much faster than the traditional central processing unit CPU[26] when dealing with massive samples.
For the demonstration, 665 complete rings with 411,404 samples (one sample is one sequence) are selected from the left line in the BR section. Data cleaning work is conducted by removing the outliers of the tunneling data according to the measurement range of different sensors. For example, the maximum value of AR is designed as 50 mm/min, so the AR measured by the PLC system at time $t$ is larger than 50 mm/min, all the tunneling data at time $t$ will be removed.

Data are divided into a training set, a validation set, and a test set with a ratio of 0.8:0.1:0.1[23], [51]. The 12 kinds of PLC data are regarded as input parameters, as shown in Table 3. When the time-series data has been transferred to the supervised learning data (from a sequence of numbers ordered by a time index to a problem comprised of input patterns ($X$) and output patterns ($y$)), the 0-1 normalization is performed on the input parameters to eliminate the units of measurement for PLC data from different sensors, as illustrated in (4).

$$X_{new} = \frac{X - X_{min}}{X_{max} - X_{min}} \tag{4}$$

Figure 8 demonstrates the proposed DL based AR predictive model in the mixed ground, which has one input layer, one LSTM layer, two fully connected layers, and one output layer. In this article, 5 minutes of past operation are used in the input layer, which means the time step is 30. We will discuss the influence of the time step on model performance in Section 4.3. The whole time series regression problem is transferred to a short time series prediction problem, used the previous 30 PLC data (AR not included) to predict the AR at next time. The input PLC data fed to the LSTM layer is a three-dimensional tensor. There are 192 neurons in the LSTM layer that are determined by the numerical experiments, which will be detailed in the following part. After the LSTM layer, a dropout layer is conducted to prevent the overfitting problem[58]. The key idea is to drop some neurons (set output to zero) randomly during the training process with some probability, which helps to prevent complex co-adaptations on training data. According to the experience, the dropout ratio adopted here is 0.2. Then we stack two fully connected layers with different neurons (dense layers in Keras) after the dropout layer. This operation is beneficial to extract more
features and transit features from high dimensions to low dimensions. At last, a dense layer with one neuron performed as the output layer is presented. The activation functions used in the dense layers are all Relu function, which has the advantage of biological plausibility, better gradient propagation, and efficient computation [59], as shown in (5). A batch size of 256 is used with the loss function of mean square error (MSE). The correlation coefficient \( R^2 \) is used to evaluate the predictive model performance.

\[
 f(x) = \begin{cases} 
 0 & \text{for } x < 0 \\
 x & \text{for } x \geq 0
\end{cases}
\]  

(5)

Figure 9 shows the numerical experiments of different LSTM layer neurons, and batch size in the proposed model. The selection for the appropriate values of LSTM layer neurons and batch size is based on \( R^2 \) value in the validation set. Tests were carried with the LSTM layer neurons of 16, 32, 64, 128, 192, 256, 320, and 384. Figure 9 (a) shows that the highest \( R^2 \) value is achieved with 192 neurons in the validation set. Meanwhile, the batch size of the training set of 32, 64, 128, 256, 512, and 1024 are compared in Figure 9 (b). It can be seen that the batch size of 256 is appropriate. We also evaluate the influence of the time step by taking the time steps 1 (1/6 min), 3 (1/2 min), 6 (1 min), 12 (2 min), 18 (3 min), 30 (5 min), 48 (8 min), and 60 (10 min). Figure 9 (c) shows that, the \( R^2 \) gradually increases with the increase of the time step, which indicates that the more temporal information is utilized, the better predictive performance can be achieved. But the longer training time and computing resources are needed when longer time steps are employed. As the PLC system stored the data every 10 seconds, we adopt 5 minutes of past inputs (time step = 30) in the proposed model.

It’s crucial to choose an appropriate optimization and initialization method in the training process of deep neural networks. Stochastic gradient descent (SGD) and its variants are usually used in the artificial neural network [60]. In this article, the Nesterov-accelerated adaptive moment estimation (Nadam[61]) optimization method with the learning rate of \( 2 \times 10^{-3} \) and the decay value of \( 4 \times 10^{-3} \) is employed. Nadam is Adam with Nesterov momentum, which makes the optimization algorithm converged much faster. The glorot_normal initialization[62] method is adopted, which draws samples as initial weights from a truncated normal distribution centered on zero with a standard deviation as illustrated in (6). The fan_in and fan_out are the number of input and output units in the weight tensor.

\[
 \text{std} = \sqrt{\frac{6}{\text{fan}_\text{in} + \text{fan}_\text{out}}}
\]  

(6)

With 109 epochs training, the variations of the loss function and \( R^2 \) are shown in Figure 10. The training time is 9s per epoch, and the total training time of the proposed model is 16.35 minutes. The early-stop technology [63] is conducted for preventing overfitting via monitoring the loss function of the validation set with the following criterion: if the loss function value doesn’t decrease after 10 epochs (patience), the training process will stop and the current model will be taken as the final model. As can be seen from Figure 10 (a), when the epochs are larger than 100, the validation loss remains unchanged, while the training loss is still decreasing. We believe it’s necessary to stop training so as to prevent the overfitting problem. As illustrated in Figure 10 (b), the \( R^2 \) values increased as the training epochs increased. The proposed model achieved \( R^2 = 0.9312 \) in the test set.
can be utilized as follows:

The permutation feature importance technology for the temporal aggregated RF-based feature tree). This article presents the permutation feature importance technology for indirectly verifying the deep learning model for the passenger demand prediction approach to the spatio-temporal work in LSTM based prediction model [39]. Inspired by the intricate connections in hidden states and memory cells of LSTM, it seems difficult to conduct the feature importance work in LSTM based prediction model [39]. Inspired by the feature selection approach employed in the spatio-temporal deep learning model for the passenger demand prediction [55], we use the temporal aggregated random forest (RF) based feature evaluation approach to indirectly verify the importance of the input PLC parameters on AR prediction.

As an ensemble learning method [64], RF achieves great success in the regression problem [65] by integrating bootstrap aggregating (bagging) and random subspace (decision tree). This article presents the permutation feature importance technology for the temporal aggregated RF-based feature evaluation. The permutation feature importance technology can be utilized as follows:

Step 1- For a trained RF model \( f \) with input feature matrix \( \tilde{X} = (X^1, X^2, \ldots, X^m) \) and output vector \( \tilde{Y} \). The error measure can be described as \( \zeta(\tilde{Y}, f) \), for example, the mean squared error.

Step 2- Estimate the original model error \( \zeta^{\text{original}} = \zeta(\tilde{Y}, f(\tilde{X})) \).

Step 3- For the \( j \)th feature in \( \tilde{X} \), generate feature matrix \( \tilde{X}^{\text{perm}} \) by permuting \( j \)th feature in the input \( \tilde{X} \), which breaks the association between \( j \)th feature and true outcome \( \tilde{Y} \). Then estimate error \( \zeta^{\text{perm}, j} = \zeta(\tilde{Y}, f(\tilde{X}^{\text{perm}, j})) \) using the permuted input matrix.

Step 4- Calculate permutation feature importance \( F^I = \zeta^{\text{perm}, j} - \zeta^{\text{original}} \), then sort features by descending \( FI \).

Step 5- For a look-back time window \( k \), \( k \in [1, n] \), generate the input feature matrix at time \( t-k, \tilde{X}_{t-k} \) as expressed in (7).

\[
\begin{align*}
\tilde{X}_{t-k} &= (X^1_{t-k}, X^2_{t-k}, \ldots, X^m_{t-k}) \\
&= \begin{cases} \\
X^1_{t-k} &= (X^1_{t-k-1}, X^2_{t-k-1}, \ldots, X^m_{t-k-1}) \\
X^2_{t-k} &= (X^1_{t-k-2}, X^2_{t-k-2}, \ldots, X^m_{t-k-2}) \\
&\vdots \\
X^m_{t-k} &= (X^1_{t-m}, X^2_{t-m}, \ldots, X^m_{t-m}) 
\end{cases} \\
&= [X^1_{t-1}, X^2_{t-1}, \ldots, X^m_{t-1}] 
\end{align*}
\]

Step 6- Establish a standard random forest \( f_{t-k} \) for look-back time window \( k \), \( k \in [1, n] \), and calculate the \( j \)th permutation feature error by \( \zeta_{t-k} = \zeta(\tilde{Y}, f_{t-k}(\tilde{X}^{\text{perm}, j}_{t-k})) \).

Step 7- The feature importance for the \( j \)th variable at time \( t-k \) can be expressed as \( F^I_{t-k} = \zeta_{t-k} - \zeta_{t-k}^{\text{original}} \), which refers to the feature importance of the \( j \)th variable at \( t-k \) in the problem of predicting the AR at time \( t \).

Step 8- Normalize all feature importance to percentage via dividing each feature importance by the sum of all feature importance.

We will examine the feature importance partitioned by category first, such as \( \sum_{k=1}^{n} F^I_{t-k} \), and then investigate the feature importance partitioned by category and look-back time window \( k \). Combined with the time step evaluation in LSTM predictive model, we can explore which PLC data has a major influence on the AR prediction.

C. FEATURE IMPORTANCE OF AR PREDICTION MODEL

The identification of major features that affect the model prediction is crucial to explore how a model could be improved and to enhance the understanding of the shield tunneling process modeling. It is also significant in terms of input feature selection because it can reduce storage requirements and training times during the shield tunneling process. We need to identify the relative importance of AR prediction variables to improve shield tunneling AR control. However, due to the intricate connections in hidden states and memory cells of LSTM, it seems difficult to conduct the feature importance work in LSTM based prediction model [39]. Inspired by the feature selection approach employed in the spatio-temporal deep learning model for the passenger demand prediction [55], we use the temporal aggregated random forest (RF) based feature evaluation approach to indirectly verify the importance of the input PLC parameters on AR prediction.

As an ensemble learning method [64], RF achieves great success in the regression problem [65] by integrating bootstrap aggregating (bagging) and random subspace (decision tree). This article presents the permutation feature importance technology for the temporal aggregated RF-based feature evaluation. The permutation feature importance technology can be utilized as follows:

Step 1- For a trained RF model \( f \) with input feature matrix \( \tilde{X} = (X^1, X^2, \ldots, X^m) \) and output vector \( \tilde{Y} \). The error

\[\text{VOLUME 8, 2020} \quad 215319\]
various change models of AR. However, the proposed DL-based AR predictive model can capture these change models, as detailed in Figure 12 (c), (d), and (e). In most cases, the predictive model captured the patterns; only some values were badly estimated.

In ring # 420 in the mudstone ground (Figure 13), the value of $\Delta AR$ varies between $-8$ mm/min and $5$ mm/min, as shown in Figure 13 (a). Figure 13 (b) shows a special pattern, which is related to frequent stop-start operations. The stop-start operation helps to mitigate the clogging problem as the shield operators stopped the shield machine and increased the slurry circulation time to remove more mucks adhered to the cutterhead and submerged wall. The proposed model operated well in predicting the AR from around $11$ mm/min to zero, and then from zero to around $11$ mm/min.

In ring # 680 in the round gravel ground (Figure 14), the value of $\Delta AR$ ranges from $14$ mm/min to $-13$ mm/min. The variation tendency has been captured by the proposed model; the model prefers to estimate higher or lower AR at the peak values (Figure 14 (b), (c), and (e)), which has brought about an RMSE value of $2.67$ mm/min, the largest one among these rings.
The time-series form is helpful to observe the variation of AR in the tunneling process, as demonstrated in Figure 12 to Figure 14. Moreover, the joint plot is employed in Figure 15 to quickly visualize and analyze the relationship between the measured and predicted AR in the selected rings and describe their distributions. As illustrated...
in Figure 15 (a), (b), and (c), the scatters composed of the measured and predicted AR are around the line of \( y = x \), which indicates that the proposed predictive model performs well in predicting AR. Moreover, the distributions of predicted AR in the mixed ground (ring # 175) and in the mudstone ground (ring # 420) are similar to measured ones, while in the round gravel (ring # 680), there are some differences between the distributions of predicted AR and measured AR, especially in the low measured AR areas.

**B. COMPARISONS WITH RF, DFN, AND SVR PREDICTIVE MODELS**

To verify the proposed DL based AR predictive LSTM model performance (LSTM model for short), three predictive models including the RF model [66], the deep feedforward network (DFN [67]) model, and the SVR model [66] are employed to analyze the model performances in predicting AR. The maximum tree depth in the RF model is 20 with the n_estimators of 100. The DFN model has a similar structure with the LSTM model, which consists of one input layer with 192 neurons, two hidden layers with 64 neurons, and 32 neurons, respectively, and one output layer with one neuron. All the activations functions used in the DFN model are Relu function, while the batch size is 256, and other parameters are the same with the proposed LSTM model. The RBF function is adopted as the kernel function in the SVR model, which has a scale of 25.574, a C value of 0.1829, and a bias of 0.1031. The \( \varepsilon \) value of the SVR model is 0.0183.

V shows \( R^2 \) values in the test set, the overall RMSE values, and the training time. The proposed LSTM based model shows the best performances in both \( R^2 \) and RMSE values. The RF model performs a little worse than the proposed LSTM based model, but the overfitting problem is the worst as the RF model achieves a high \( R^2 \) value of 0.9395 in the training set, but a low \( R^2 \) value of 0.8890 in the test set. The DFN model is also a kind of deep learning model, but the DFN model performance is worse than the LSTM model.
because the DFN model cannot consider past operations. The SVR model achieves the lowest $R^2$ value and the highest RMSE value, which seems unsuitable for prediction the instantaneous AR. For training time comparison, the LSTM and DFN model are the fastest due to the GPU acceleration, while the SVR model is the slowest one even though we used the CPU parallel computing.

To evaluate the predicted errors of these models, the RMSE values of each ring are plotted in Figure 16 (a). The SVR model achieves the largest RMSE values in almost all rings. The proposed LSTM model reaches the smallest RMSE values in the majority of rings, especially in the round gravel areas. The RF model achieves RMSE values as small as the proposed LSTM model in a few rings, mostly located in mudstone rings (around ring # 400). To check the distributions of RMSE values obtained by these four models, we carry out the violin plots- a hybrid of box plots and kernel density plots- in Figure 16 (b). The white dot in the middle of each violin means the median value of RMSE. It shows that the LSTM model has the smallest ones. The thick black bar in the center represents the interquartile range; the LSTM model is the smallest as well. The thin black line extended from the thick black bar represents the upper (max) and lower (min) adjacent values in the RMSE with a 95% confidence interval. The width of the violin is the density plot rotated and placed on each side, to show the distribution shape of the RMSE values. We can see that the RMSE values obtained by the LSTM model are mostly distributed between 1 mm/min to 2 mm/min, while the other methods have distributions between 1 mm/min to 3.5 mm/min.

The good performances of LSTM can be explained by the following two aspects: Firstly, from the perspective of the physical mechanism affecting the shield AR during the tunneling process, the interaction mechanism between the slurry and the soil is complicated, which leads to a hysteresis effect between the cutterhead and the soil on the excavation face. As a result, the current AR is affected by other PLC tunneling parameters measured in the past time. Secondly, from the data view, the input parameters of the LSTM model are 3D tensors, while the inputs of the other three models are only 2D matrices. From a data-driven perspective, the 3D tensor implies higher characteristics than the 2D matrices, so the LSTM model can learn the features needed from a higher feature space. Combining the above two factors, the proposed LSTM model performs better than other models.

C. PLC DATA INFLUENCE ON AR PREDICTION

With the aid of the temporal aggregated RF-based feature evaluation approach, we investigate the PLC data influence on the AR prediction. The relative feature importance of each kind of PLC data is shown in Figure 17. It can be observed that the THR, SPE, TOR, and the RS are the dominating factor in the AR prediction, followed by the FFR. However, the other PLC data, such as SPW, PP, etc., have lower contributions (less than 7%) on the AR prediction. The relative feature importance of different kinds of PLC data is consistent with the filed experience. Firstly, the cutterhead total thrust and torque are the governing factors of the shield machine advance rate. Secondly, when slurry shield tunneling operates in the mixed ground, the clogging problem resulted in great fluctuations of SPE. Consequently the advance rate had to decrease to reduce the tunnel face stability risk. During this process, the rotation speed and slurry flow rate in feedline changed frequently, thus have impacted the advance rate.

To explore the influence PLC data at the time $t - k$ on the AR at the time $t$, a heatmap of the relative feature importance of PLC data at different time steps is illustrated in Figure 18. We can see that the PLC data at time $t-1$ have that major influence on the AR at time $t$, especially for the SPE, and RS. Meanwhile, the THR, TOR, and FFR at time $t-4$ or
t—5 still have impact on current AR. However, the influence of time step in the LSTM neural network may be much more complicated than the results obtained by the random forest due to the gate mechanism in the hidden layer of the LSTM cell. For feature selection, we take the 5 minutes ahead THR, TOR, and RS as model input (as SPE can’t be controlled directly) and obtain an $R^2$ value of 0.883 in the test set, which means that we can use one-fourth features to achieve 95% prediction performance of the proposed model ($R^2 = 0.9312$).

V. CONCLUSION

This article presented a deep learning-based AR predictive model for shield tunneling construction using LSTM neural network. The model was used for the prediction of the AR in the mixed ground in Nanning metro. The model performances in different ground conditions were evaluated. The model was also compared to other models Analyses show that:

(1) The LSTM model considers well the influence of the past operations on the AR. The AR of SPB shield tunneling in the mixed ground condition of the round gravel and mudstone have great fluctuations, which can be captured by the LSTM model ($R^2 = 0.9312$ and RMSE $= 2.57 \text{ mm/min}$).

(2) The LSTM model performs better in the mixed ground condition and the mudstone condition than in the round gravel condition. Although proposed LSTM model can capture the variation patterns, it could misestimate the AR peak values.

(3) Compared with the RF model, the DFN model, and the SVR model, the LSTM model provides the best performances. The time hysteresis effect of the slurry and the soil, as well as the higher feature space of the LSTM model, may result in improving the performances of the proposed LSTM model.

(4) A tailored temporal aggregated random forest is employed to rank the importance of the explanatory features. The cutterhead total thrust, torque, and rotation speed have great influences on the advance rate prediction, and with the above-mentioned features, we can obtain a 95% performance compared with the proposed model measured by the $R^2$ value.

Despite the above-mentioned achievements, some improvements could be operated in this predictive model. Indeed, the proposed model can only predict the AR at the next moment, which seems insufficient in assisting the real shield tunneling operation. For better guidance in shield tunneling, it’s necessary to develop a multi-steps AR predictive model, which can inform the shield operator about the AR variation tendency in the following few minutes. In addition, the proposed AR predictive model can only inform the operators how the AR changes, but can’t provide suggestions on steering the tunneling parameters. If we want to realize the intelligent shield tunneling, it’s essential to build a cycle of prediction and control with intelligent methods, such as reinforcement learning and transfer learning. One possible application of the proposed AR predictive model can be the environmental model [68] in reinforcement learning, which can be employed to simulate the actual shield tunneling process. For the proposed temporal aggregated random forest feature selection approach, it’s better to verify with other methods, in particular for the inter-dependent input features.

ACKNOWLEDGMENT

The authors thank Dr. Zhangwei Ning (Sixense Inc.), and also thank the China Railway 16 Bureau Group Beijing Metro Engineering Construction Co., Ltd. for providing the data about the section from Bai-cang-ling station to Railway station of Nanning metro line 1.

REFERENCES

[1] X. Xie, Q. Wang, Z. Huang, and Y. Qi, “Parametric analysis of mixshield tunnelling in mixed ground containing mudstone and protection of adjacent buildings: Case study in nanning metro,” Eur. J. Environ. Civ. Eng., vol. 22, no. 1, pp. s130–s148, Aug. 2018.

[2] Z. Huang, C. Wang, J. Dong, J. Zhou, J. Yang, and Y. Li, “Conditioning experiment on sand and cobble soil for shield tunneling,” Tunnelling Underground Space Technol., vol. 87, pp. 187–194, May 2019.

[3] J. Zhao, Q. M. Gong, and Z. Eisenstein, “Tunnelling through a frequently changing and mixed ground: A case history in singapore,” Tunnelling Underground Space Technol., vol. 22, no. 4, pp. 388–400, Jul. 2007.

[4] J. Küpferle, Z. Ziska, B. Schoesser, A. Röttger, M. Alber, M. Thewes, and W. Theisen, “Influence of the slurry-stabilized tunnel face on shield TBM tool wear regarding the soil mechanical changes—Experimental evidence of changes in the tribological system,” Tunnelling Underground Space Technol., vol. 74, pp. 206–216, Apr. 2018.

[5] G. Gu, Y. Yan, C. Feng, and Q. Lin, “Analysis of advancing speed and force characteristics of shield cutterhead in composite strata,” Tunnel Construct., vol. 38, no. 11, pp. 1895–1902, 2018.

[6] J. Rostami, “Performance prediction of hard rock tunnel boring machines (TBMs) in difficult ground,” Tunnelling Underground Space Technol., vol. 57, pp. 173–182, Aug. 2016.

[7] S. Leng, J.-R. Lin, Z.-Z. Hu, and X. Shen, “A hybrid data mining method for tunnel engineering based on real-time monitoring data from tunnel boring machines,” IEEE Access, vol. 8, pp. 90430–90449, 2020.

[8] X. Wang, H. Lu, X. Wei, G. Wei, S. S. Behbahani, and T. Isley, “Application of artificial neural network in tunnel engineering: A systematic review,” IEEE Access, vol. 8, pp. 119527–119543, 2020.

[9] A. G. Benardos and D. C. Kaliampakos, “Modelling TBM performance with artificial neural networks,” Tunnelling Underground Space Technol., vol. 19, no. 6, pp. 597–605, Nov. 2004.

[10] S. Yagiz, “Utilizing data mining techniques for predicting TBM performance in hard rock condition,” Tunnelling Underground Space Technol., vol. 23, no. 3, pp. 326–339, May 2008.

[11] R. Mikaeil, M. Z. Naghadehi, and F. Sereshki, “Multifactorial fuzzy approach to the penetrability classification of TBM in hard rock conditions,” Tunnelling Underground Space Technol., vol. 24, no. 5, pp. 500–505, Sep. 2009.

[12] S. Yagiz, C. Gokeceoglu, E. Sezer, and S. Iplikci, “Application of two non-linear prediction tools to the estimation of tunnel boring machine performance,” Eng. Appl. Artif. Intell., vol. 22, nos. 4–5, pp. 818–824, 2009.

[13] S. Yagiz and H. Karahan, “Prediction of hard rock TBM penetration rate using particle swarm optimization,” Int. J. Rock Mech. Mining Sci., vol. 48, no. 3, pp. 427–433, Apr. 2011.

[14] C. Shao, X. Li, and H. Su, “Performance prediction of hard rock TBM based on extreme learning machine,” in Lecture Notes in Computer Science: Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), vol. 8103, no. 2, 2013, pp. 409–416.

[15] S. Mahdevari, K. Shahriar, S. Yagiz, and M. Akbarpour Shirazi, “A support vector regression model for predicting tunnel boring machine penetration rates,” Int. J. Rock Mech. Mining Sci., vol. 72, pp. 214–229, Dec. 2014.

[16] D. I. Armaghan, E. T. Mohammad, M. S. Narayanasamy, N. Narita, and S. Yagiz, “Development of hybrid intelligent models for predicting TBM penetration rate in hard rock condition,” Tunnelling Underground Space Technol., vol. 63, pp. 29–43, Mar. 2017.

[17] A. Salimi, J. Rostami, C. Moormann, and A. Delisio, “Application of non-linear regression analysis and artificial intelligence algorithms for performance prediction of hard rock TBMs,” Tunnelling Underground Space Technol., vol. 58, pp. 236–246, Sep. 2016.


[18] H. Wang and D. Fu, “Mathematical model and the related parameters for EPB shield tunneling,” Tunn. and Underground Space Technol., vol. 26, pp. 247–256, 2011.

[19] H. M. Zhang, X. G. Wu, and W. H. Zeng, “Study on tunneling experiment and mathematical model of EPB shield,” Tunn. and Underground Space Technol., vol. 26, pp. 247–256, 2011.

[20] J. Cao, Z. Li, and J. Li, “Financial time series forecasting model based on CEEMDAN and LSTM,” Phys. A, Stat. Mech. Appl., vol. 519, pp. 127–139, Apr. 2019.

[21] K. Fang, C. Shen, D. Kifer, and X. Yang, “Prolongation of SMA to spatiotemporally seamless coverage of continental U.S. using a deep learning neural network,” Geophys. Res. Lett., vol. 44, no. 21, p. 11,090–11,039, Nov. 2017.

[22] C. Correa-Jullian, J. M. Cardemil, E. L. Droguett, and M. Behzad, “Assessment of deep learning techniques for prognosis of solar thermal systems,” Renew. Energy, vol. 145, pp. 2178–2191, Jan. 2020.

[23] P. He, G. Jiang, S.-K. Lam, and Y. Sun, “Learning heterogeneous traffic patterns for travel time prediction of bus journeys,” Inf. Sci., vol. 512, pp. 1394–1406, Feb. 2020.

[24] P. Li, K. Zhou, X. Lu, and S. Yang, “A hybrid deep learning model for short-term PV power forecasting,” Appl. Energy, vol. 259, Feb. 2020, Art. no. 114216.

[25] Y. Jang, J. Jeong, and Y. K. Cho, “Business failure prediction of construction contractors using a LSTM RNN with accounting, construction market, and macroeconomic variables,” J. Manag. Eng., vol. 36, no. 2, Mar. 2020, Art. no. 04910939.

[26] H. S. Karimi, B. Natarajan, C. A. Ramsey, J. Henson, J. L. Tedder, and E. Kemper, “Comparison of learning-based wastewater flow prediction methodologies for smart sewer management,” J. Hydrodl., vol. 577, Oct. 2019, Art. no. 123977.

[27] X. Li, Z. Li, W. Huang, and P. Zhou, “Performance of statistical and machine learning ensembles for daily temperature downsampling,” Theor. Appl. Climatol., vol. 140, nos. 1–2, pp. 571–588, Apr. 2020.

[28] X. Gao, M. Shi, X. Song, C. Zhang, and H. Zhang, “Recurrent neural networks for real-time prediction of TBM operating parameters,” Autom. Construct., vol. 98, pp. 225–235, Feb. 2019.

[29] M. Li, M. Van De Geer, and P. Bühlmann, “The group lasso for logistic regression,” J. Roy. Stat. Soc. B, Stat. Methodol., vol. 70, no. 1, pp. 53–71, Jan. 2008.

[30] S. Mokhtari and M. A. Mooreey, “Predicting EPDM advance rate performance using support vector regression modeling,” Tunnelling and Underground Space Technol., vol. 104, Oct. 2020, Art. no. 103520.

[31] F. Min, W. Zhu, and X. Han, “Filter cake formation for slurry shield tunneling in highly permeable sand,” Tunnelling and Underground Space Technol., vol. 38, pp. 423–430, Sep. 2013.

[32] V. Makarenkov, L. Raksha, and B. Schapira, “Choosing the right word: Using bidirectional LSTM tagger for writing support systems,” Eng. Appl. Artif. Intell., vol. 84, pp. 1–10, Sep. 2019.

[33] I. Sutskever, “Training recurrent neural networks,” Univ. Toronto, Toronto, ON, Canada, Tech. Rep., 2013.

[34] S. Hochreiter and J. U. Schmidhuber, “Long short-term memory,” Neural Comput., vol. 9, no. 8, pp. 1735–1780, 1997.

[35] F. A. Geers, J. Schmidhuber, and F. Cummins, “Learning to forget: Continual prediction with LSTM,” Neural Comput., vol. 12, no. 10, pp. 2451–2471, Oct. 2000.

[36] F. Chollet, Keras. GitHub, 2015.

[37] A. Agarwal, S. Negahban, and M. J. Wainwright, “Noisy matrix deconvolution via convex relaxation: Optimal rates in high dimensions,” Ann. Statist., vol. 40, no. 2, pp. 1171–1197, Apr. 2012.

[38] X. Glorot, A. Bordes, and Y. Bengio, “Deep sparse rectifier neural networks,” in Proc. 14th Int. Conf. Artif. Intell. Statist., vol. 15, 2011, pp. 315–322.

[39] Y. Bengio, N. Boulanger-Lewandowski, and R. P. Pascanu, “Advances in optimizing recurrent networks,” in Proc. IEEE Int. Conf. Acoust., Speech and Signal Process., May 2013, pp. 8624–8628.

[40] T. Dozat, “Incorporating Nesterov momentum into Adam,” in Proc. 2016, no. 1, pp. 2013–2016, 2016.

[41] X. Glorot and Y. Bengio, “Understanding the difficulty of training deep feedforward neural networks,” in Proc. 13th Int. Conf. Artif. Intell. Statist. (AISTATS), vol. 9, 2010, pp. 249–256.

[42] L. Prechelt, Early Stopping—But When? (Lecture Notes in Computer Science: Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), vol. 7700, 1998, pp. 55–69.

[43] L. Breiman, Random Forests, Mach. Learn., vol. 45, no. 1, pp. 5–32, 2001.

[44] P. Zhang, R.-P. Chen, and H.-N. Wu, “Real-time analysis and regulation of EPB shield steering using random forest,” Autom. Construct., vol. 106, Oct. 2019, Art. no. 102860.
[66] Y. Zhou, S. Li, C. Zhou, and H. Luo, “Intelligent approach based on random forest for safety risk prediction of deep foundation pit in subway stations,” J. Comput. Civil Eng., vol. 33, no. 1, pp. 1–14, 2019.

[67] R. Pascanu, G. Montufar, and Y. Bengio, “On the number of response regions of deep feed forward networks with piece-wise linear activations,” pp. 1–17, 2013.

[68] J. Kober, J. A. Bagnell, and J. Peters, “Reinforcement learning in robotics: A survey,” Int. J. Robot. Res., vol. 32, no. 11, pp. 1238–1274, Sep. 2013.

QIANG WANG received the B.S. degree from the School of Civil And Transportation Engineering, Hohai University, China, in 2013, and the Ph.D. degree from the School of Civil Engineering, Tongji University, China, in 2018.

He is currently a Postdoctoral Researcher with the Department of Geotechnical Engineering, Tongji University. His research interests are automation in shield tunneling construction, including machine learning predictive model for tunneling parameters, intelligent sensing, and control for tunneling construction.

XIONGYAO XIE received the B.S. degree in applied geophysics from Jilin University in 1995 and the M.S. degree in geophysical exploration and the Ph.D. degree in civil engineering from Tongji University, China, in 1998 and 2001, respectively.

He is currently a Professor with the School of Civil Engineering, Tongji University, where he also serves as the Vice Dean. His research interests include nondestructive testing of underground structure and smart underground construction.

ISAM SHAHROUR received the B.S. degree in mechanics sciences from Lille 1 University in 1980 and the M.S. and Ph.D. degrees in civil engineering from the Ecole Nationale des Ponts et Chaussées, Paris, France, in 1982 and 1984, respectively.

He is strongly involved in the research, higher education, and partnership with the socio-economic sector. He was the Vice President of the Research and Doctoral Program with the Lille 1 University from 2007 to 2012 and the President of the Innovation Agency Lille Metropole Technopole for regional economic development via innovation. He is currently the Director of the Regional Research Laboratory Civil and Geo-Environment Engineering and the Head of the international master Urban Engineering and Habitat. He coordinates a large-scale smart city demonstrator SunRise. His research activities concern smart and sustainable cities and geotechnical and geo-environmental engineering.

* * *