Received December 10, 2019, accepted February 7, 2020, date of publication February 14, 2020, date of current version March 4, 2020.

Digital Object Identifier 10.1109/ACCESS.2020.2974058

Prediction Model of Shield Performance During Tunneling via Incorporating Improved Particle Swarm Optimization Into ANFIS

KHALID ELBAZ[1], SHUI-LONG SHEN[1,2], WEN-JUAN SUN[3], ZHEN-YU YIN[4], AND ANNAN ZHOU[2]

[1]Key Laboratory of Intelligence Manufacturing Technology, Ministry of Education, Department of Civil and Environmental Engineering, College of Engineering, Shantou University, Shantou 515063, China
[2]Discipline of Civil and Infrastructure, School of Engineering, Royal Melbourne Institute of Technology, Melbourne, VIC 3001, Australia
[3]Department of Civil and Environmental Engineering, ATLSS Engineering Research Center, Lehigh University, Bethlehem, PA 18015, USA
[4]Department of Civil and Environmental Engineering, The Hong Kong Polytechnic University, Hong Kong

Corresponding author: Shui-Long Shen (shensl@stu.edu.cn)

This work was supported in part by the Research Funding of Shantou University for New Faculty Member under Grant NTF19024-2019, in part by the National Basic Research Program of China through 973 Program under Grant 2015CB057806, and in part by the Innovative Research Funding of the Science and Technology Commission of Shanghai Municipality under Grant 18DZ1201102.

ABSTRACT This paper proposes a new computational model to predict the earth pressure balance (EPB) shield performance during tunnelling. The proposed model integrates an improved particle swarm optimization (PSO) with adaptive neurofuzzy inference system (ANFIS) based on the fuzzy C-mean (FCM) clustering method. In particular, the proposed model uses shield operational parameters as inputs and computes the advance rate as the output. Prior to modeling, critical operational parameters are identified through principle component analysis (PCA). The hybrid model is applied to the prediction of the shield performance in the tunnel section of Guangzhou Metro Line 9 in China. The prediction results indicate that the improved PSO-ANFIS model shows high accuracy in predicting the EPB shield performance in terms of the multiobjective fitness function [i.e. root mean square error ($RMSE$) = 0.07, coefficient of determination (R^2) = 0.88, variance account (VA) = 0.84 for testing datasets, respectively]. The good agreement between the actual measurements and predicted values demonstrates that the proposed model is promising for predicting the EPB shield tunnel performance with good accuracy.

INDEX TERMS Earth pressure balance shield, principle component analysis, improved PSO-ANFIS, fuzzy C-mean, advance rate.

I. INTRODUCTION

With the progress of manufacturing technology, larger and increasingly complex tunnel projects are being constructed in many Chinese cities [1]–[4]. In tunnel excavation projects, one of the main aims is to optimise the performance of the drilling system. Therefore, accurate performance of the tunnel boring machine (TBM) can be employed to reduce the risks associated with high costs and time consumed during the tunnelling process [5]. Conversely, overestimating can be a negative effect for the utilization of project resources [6]. Thus, if the tunnelling process is addressed in an appropriate manner, the risks related to tunnelling projects will be decreased considerably [7]–[12]. In general, the TBM performance can be represented by the penetration rate and advance rate. The penetration rate is the linear distance between the excavation faces per unit time when the machine is in production. The advance rate is the rate of the machine face advancing forward, including both the production time and downtime [13]. As the advance rate determines the total construction time and the overall cost of a tunnelling project, one of the most essential efforts in tunnel construction design is to estimate the advance rate.

The performance analysis of the TBM and the development of accurate prediction models have been the ultimate goal and are still under development in several studies. In most of the previous studies, both empirical and theoretical approaches were developed for predicting the TBM performance. Typical input parameters can be categorised as follows: i) geological conditions [e.g. uniaxial compression strength (UCS) and

The associate editor coordinating the review of this manuscript and approving it for publication was Jenny Mahoney.

geological strength index (GSI)], and ii) operational parameters [e.g. thrust force (TF), cutterhead torque (CT), and tunnel diameter (D)] [14]–[17]. Owing to large uncertainties in geological environments and construction processes, empirical and theoretical approaches cannot display the nonlinear and dynamic behaviour of the TBM performance. By utilizing a large amount of field data, artificial intelligence (AI) models may overcome this limitation.

AI-based models emerged two decades ago to serve as an acceptable solution to several geotechnical problems; many comprehensive reviews have summarised the effectiveness of using AI models in widespread applications. To estimate the TBM performance, some AI models have been developed, including artificial neural networks (ANN) [18]–[20], fuzzy logic (FL) [21], and adaptive neurofuzzy inference systems (ANFIS) [22], [37]. ANN and non-linear multiple regression models have been used for estimating tunnel boring machine performance as a function of rock properties. However, to date, studies have mostly used only geological data in the prediction of TBM performance. Yin et al. [20] conducted a comparative study for identifying the soil parameters using different optimisation techniques such as genetic algorithms, particle swarm optimization, simulated annealing, differential evolution, and the artificial bee colony. Results showed that the differential evolution had the highest search ability but the slowest convergence speed.

On the other hand, the dynamic features of construction make the tunnelling process a nonlinear problem with large uncertainties, which challenges construction management and makes accurate predictions difficult. In this respect, ANN and FL can be used to address such challenges. However, there has been an argument as to whether AI models can yield reasonable solutions with robustness when addressing nonlinear problems with uncertainties [23], [24]. The other argument is that AI models may give distorted and/or inadequate explanations for problems owing to problems with the local minima and inferior generalization. As a result, hybrid models have been developed by incorporating the optimization algorithms with AI-based models. Salimi et al. [11] discussed the applicability of artificial intelligence to design sewage transfer system. The obtained results illustrated that the ANFIS had better performance for estimating water hammer phenomenon in the UPVC pipes while the PSO-ANFIS was found to be more suitable in metal pipes. Azad et al. [8] performed a comparative study for optimizing the performance of ANFIS model in simulating monthly rainfall magnitudes using different algorithms such as genetic algorithms, particle swarm optimization, differential evolution, and the artificial bee colony. The results showed that the hybrid models had better accuracy than the simple ANFIS model in escaping local optima [15]. So far, the basic idea of the hybrid approaches is to address the shortcomings of single approach and generate the effect of synergy in prediction, which have become the predominant approaches in recent years. For example, Elbaz et al. [25] proposed a hybrid model of a multiobjective genetic algorithm with ANFIS for predicting the shield performance, demonstrating better prediction accuracy than that from the traditional ANFIS technique. In spite of available hybrid optimization techniques, attempts to propose new ones are still ongoing.

This study aims to propose a hybrid multiobjective optimisation model for the prediction of shield machine performance during the tunnelling process. The proposed model is constructed using a fuzzy rule-based system optimised by an improved PSO algorithm, which simultaneously adjusts both antecedent and consequent variables. Principal component analysis is applied to examine the effect of different parameters on the advancement rate of the shield machine. To evaluate the performance of the proposed model, the prediction results are compared with the results of the ANFIS-FCM model.

The remaining content is organised as follows. Section II presents the AI technique and the proposed model. To verify the effectiveness of the proposed model, it is applied for predicting the tunnelling performance of a tunnel section in Guangzhou. Section III presents the real-time field monitoring data. Section IV describes a shield tunnelling performance database and presents the principal component analysis. Section V presents prediction results with a technical discussion. The last section concludes the study.

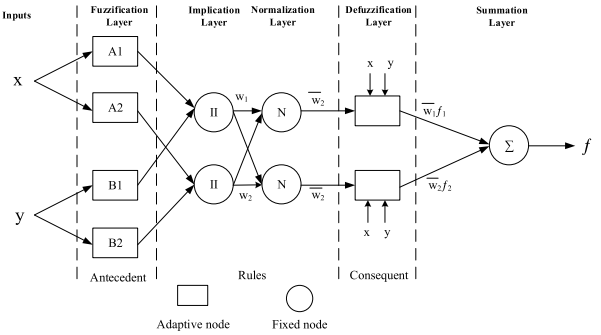


FIGURE 1. Structure of ANFIS model with two input parameters.

II. ARTIFICIAL-INTELLIGENCE BASED MODELING

A. ANFIS MODEL

ANFIS, developed by Jang [26], is a multilayer adaptive network-based fuzzy inference that maps relations between inputs and outputs. ANFIS is useful for solving complex problems with large uncertainties by creating a fuzzy inference system (FIS) with adjusted parameters of the membership function (MF). In particular, it uses neuro-adaptive learning methods to adjust membership function parameters until reaching the optimal solution. In this way, ANFIS combines the reasoning capacities of fuzzy logic principles with the learning capabilities of the ANN system to solve complicated and nonlinear issues. Fig. 1 shows the ANFIS architecture, with two input parameters (x, y) and one output parameter (f), using the Takagi-Sugeno fuzzy inference system.

The following content briefly describes the five layers of the ANFIS model. Further details of ANFIS can be found in

other literature such as [26]. In Layer 1, each node (i) has an MF of a linguistic variable. The output of each node is calculated according to the following equation:

$$Q_i^1 = \mu_{A_i}(x) = \frac{1}{1 + \left[\left(\frac{x - v_i}{\sigma_i}\right)^2\right]^{b_i}} \tag{1}$$

where x is the input value of node i, A_i is the linguistic variable associated with this node, and σ_i, v_i, and b_i are function parameters with $b_i > 0$. The parameters in this layer are defined as premise parameters.

In Layer 2, every node computes the firing strength for each rule by multiplying the received signals:

$$Q_i^2 = w_i = \mu_{Ai}(x)\mu_{Bi}(y), \quad i = 1, 2 \tag{2}$$

In Layer 3, every node computes the ratio of the i^{th} rule's firing strength to its sum for all rules. The outputs are normalized firing strengths.

$$Q_i^3 = \overline{w}_i = \frac{w_i}{w_1 + w_2}, \quad i = 1,\ 2 \tag{3}$$

Layer 4 contains the adaptive node:

$$Q_i^4 = \overline{w}_i f_i = \overline{w}_i(p_i x + q_i y + r_i), \quad i = 1, 2 \tag{4}$$

where w_i is the output of layer 3; p_i, q_i, and r_i are the consequent parameters.

Layer 5 calculates the summation of all input signals as the overall output:

$$Q_i^5 = \textit{overall output} = \sum\nolimits_i \overline{w}_i f_i = \frac{\sum_i w_i f_i}{\sum_i w_i} \tag{5}$$

B. FUZZY C-MEANS CLUSTERING

By assigning a set of data into groups, fast and robust data clustering is essential to extract beneficial structures from large data. Fuzzy C-means (FCM) clustering is a powerful algorithm for clustering overlapped datasets. In FCM, the grade of a data point belonging to a cluster is identified by a membership. The membership shows a large value for data near the cluster centre and a small value for data far away from the cluster centre. FCM divides the selection of n vectors x_i $(i = 1, 2, \ldots, n)$ into fuzzy sets and determines the cluster centre for each set to minimise the fitness function.

The FCM clustering method works in the following procedure. Given n data points $(x_1, x_2, x_3, \ldots, x_n)$, the centre of the i^{th} cluster is randomly chosen as c_i $(i = 1, 2, \ldots, C)$, where C is the total number of clusters $(C \leq n)$.

The membership matrix U can be calculated as follows:

$$\mu_{ij} = \frac{1}{\sum_{k=1}^{C} \left(\frac{d_{ij}}{d_{kj}}\right)^{\frac{2}{m-1}}} \tag{6}$$

where $d_{ij} = ||c_i - x_j||$ is the Euclidean distance between the i^{th} cluster centre and the j^{th} data point, μ_{ij} is the coefficient of membership matrix U, and m is the fuzziness index.

The objective function can be computed as follows:

$$J(U,\ c_1, \ldots, c_2) = \sum_{i=1}^{c} J_i = \sum_{i=1}^{c} \cdot \sum_{j=1}^{n} \mu_{ij}^m d_{ij}^2 \tag{7}$$

Finally, a novel c fuzzy cluster centre C_i $(i = 1, 2, \ldots, C)$ can be calculated by utilizing the following equation:

$$C_i = \frac{\sum_{j=1}^{n} \mu_{ij}^m x_j}{\sum_{j=1}^{n} \mu_{ij}^m} \tag{8}$$

C. IMPROVED PSO

The PSO algorithm initialises a set of particles randomly scattered in the space of the objective function. Then, it updates generations to find the optima of all possible solutions (so-called particles). Each particle is defined by two positions and velocity values based on the two best fitness values: pbest and gbest. pbest is the best fitness solution of each particle fulfilled so far, whereas gbest is the global best solution gained by any particle in the population tracking by PSO. According to pbest and gbest values, all particles update their velocities and positions until the optimal solution is reached. As an optimization method, PSO is easy to understand and implement. It is computationally efficient and maintains the diversity of the swarm.

Assuming the position $x_i^t = (x_{i1}^t, x_{i2}^t, \ldots, x_{in}^t)$ and the velocity $v_i^t = (v_{i1}^t, v_{i2}^t, \ldots, v_{in}^t)$ of the i^{th} particle in the t^{th} iteration, the particle optimises its location in the $(t+1)^{th}$ iteration by utilizing the following equation [27]:

$$v_i^{t+1} = w.v_i^t + c_1.r_1.\left(p_i^t - x_i^t\right) + c_2.r_2.\left(g^t - x_i^t\right) \tag{9}$$

$$\textit{with} \quad -v_{\max} \leq v_i^{t+1} \leq v_{\max}$$

$$x_i^{t+1} = \left(x_i^t + v_i^{t+1}\right) \tag{10}$$

where p_i^t is the best location of particle i^{th} in iteration t^{th}, g^t is the global best location up to the t^{th} iteration, r_1 and r_2 are random values in the range of [0, 1], w is the inertia weight where $0 \leq w \leq 1$, and parameters c_1, and c_2 are the cognitive acceleration rate and social coefficient, respectively.

Inertia weight w greatly influences the contribution rate of the velocity from the previous step to the velocity at the next step. A traditional strategy of improving the inertia weight is applied as follows:

$$w = w_{final} + (w_{initial} - w_{final})(1 - \frac{T}{G_{\max}}) \quad \textit{if } g^t \neq x_i^t$$

$$w = w_{final} \quad \textit{if } g^t = x_i^t \tag{11}$$

where T is the iteration number, $T \in [0, G_{\max}]$; $G_{\max}$ is the maximum number of iterations; $w_{initial}$ is the initial inertia weight; and w_{final} is the development value at the maximum iteration.

Because the feature vector usually has high dimensions, PSO particles may easily be trapped in local optima rather than reach global optima [28]. Therefore, Clerc [29] added

a constriction factor k into PSO to verify the best convergence as follows:

$$v_i^{t+1} = k\left[v_i^t + c_1 r_1(p_i^t - x_i^t) + c_2 r_2(g^t - x_i^t)\right] \quad (12)$$

A rule of thumb is that the constriction factor should be a convex function in precocious iterations to avoid the early convergence to local minima, and a concave function in late iterations to change slowly until reaching a global optimum. Based on this rule, the constriction factor function is built as follows [30]:

$$k = \frac{\cos\left((\pi/G_{\max}) \times T\right) + 2.5}{4} \quad (13)$$

Because the inertia weight influences the degree of the particle velocity and the constriction factor affects the convergence performance of PSO, the following content explains the improvement of synchronously using the inertia weight and the constriction factor.

The improved PSO has both the inertia weight and constriction factor varying synchronously. Integrating equations (12) and (13), we get the following equation:

$$v_i^{t+1} = \left(\frac{\cos(\pi \times T/G_{\max}) + 2.5}{4}\right) \times\left[w \times v_i^t + c_1.r_1.\left(p_i^t - x_i^t\right) + c_2.r_2.\left(g^t - x_i^t\right)\right] \quad (14)$$

After a number of iterations, the particle may be close to the global optima; the inertia weight becomes smaller to allow the particle to retain its original speed and search for the optima in a smaller range. If the particle does not reach the accurate minima, then the inertia weight becomes greater to retain its original velocity for the global optima search.

Derivation details of this approach can be found in Lu et al. [44]. By solving these derivations, we can get the equation

$$w \geq \frac{c_1 + c_2}{2} - \frac{1}{k} \quad (15)$$

According to this equation, the inertia weight w should be no less than the maximum value of the right-hand side. On the right-hand side of Eq. (15), $-1/k$ reaches the maximum value of $-1/[(\cos(\pi/G_{max} \times 0) + 2.5)/4]$ when $T = 0$, based on Eq. (13). Therefore, the inertia weight $w_{final} = (c_1 + c_2)/2 - 1/[(cos(\pi/G_{max} \times 0) + 2.5]/4]$. When the values of c_1 and c_2 are equal to 2, then $w_{final} = 2 - (1/(3.5/4)) \sqcup 0.857$.

As the inertia weight w ranges from [0, 1], this study sets the initial value of the inertia weight $w_{initial} = 1$. Thus, the inertia weights can be presented as follows:

$$w = 0.857 + (1 - 0.857)\left(1 - \frac{T}{G_{\max}}\right) \quad if\ g_i^t \neq x_i^t$$
$$w = 0.857 \quad if\ g_i^t = x_i^t \quad (16)$$

D. STOPPING CRITERIA

Stopping criteria are specified as the conditions required to terminate the iterative search algorithm when there is no obvious improvement over the number of iterations. As usual, termination criteria include the expected value of accuracy and the maximum iteration number. To determine the appropriate number of iterations, a useful approach suggested by Zielinski and Rainier [24], which is based on comparing the results of various iteration numbers, is applied. In this work, the maximum number of iterations is set as a termination criterion. To determine an appropriate iteration number, we conducted trial computations by varying the population size of the improved PSO model based on the root mean square error (*RMSE*) [49].

To control the overfitting, a global validation strategy was implemented according to the definition by Mitchell [45]. We assume that c_j^* and $c_j^{*'}$ are the best-performing candidate groups found by computing the error rate ε for every element of $P(c)$ in the optimization set (op) and in the validation set (υ), respectively. $P(c)$ represents the powerset of classifiers $c = \{c_1, c_2, \ldots, c_n\}$ determining the population of all potential candidates c_j. The ranking error of the optimization set is denoted by $\varepsilon(v, c_j^*)$, and the ranking error of the validation set is denoted by $\varepsilon(v, c_j^{*'})$. c_j^* is considered as overfitting on op if there is an alternate candidate $c_j^{*'} \in P(c)$ that can be found such that $\varepsilon(v, c_j^*) > \varepsilon(v, c_j^{*'})$. In this way, overfitting is defined as

$$\text{Overfitting} = \varepsilon(v, c_j^*) - \varepsilon(v, c_j^{*'}) \quad (17)$$

E. HYBRID MODEL OF IMPROVED PSO-ANFIS

In order to predict the tunnelling performance with good accuracy, this study introduces an improved PSO-ANFIS model. In this hybrid model, the aforementioned improved PSO helps to tune and achieve the optimal values of ANFIS parameters through training. Fig. 2 shows a flowchart of the improved PSO-ANFIS model.

The improved PSO-ANFIS model works using the following procedure. Initially, all datasets are reprocessed for the training model, including the operational shield parameters and the corresponding advance rate. With postprocessed data, the initial ANFIS method is produced with all parameters randomly initialised. To achieve accurate prediction, the ANFIS model needs to be supported by an appropriate number of clusters. The initial ANFIS method utilises the FCM clustering approach to optimise the result by extracting a set of rules that model the datasets and form the FIS. Then, premise parameters (σ_i, v_i, b_i) and consequent parameters (p_i, q_i, r_i) of the ANFIS method are extracted in this step to estimate the dimensions of every particle for setting up the PSO algorithm in the next step.

The corresponding parameters for each MF are extracted iteratively to form a vector. In this vector, the parameters constitute the variables to be optimised by PSO; therefore, the length of every particle in PSO can be determined. Once the PSO parameters are specified, the initial population is generated. After initializing all particles, the improved PSO updates the velocity and the position of each particle in the swarm until a convergence is obtained to get the optimal values of the variables. The objective function of each particle is computed, and the best new values are updated accordingly.

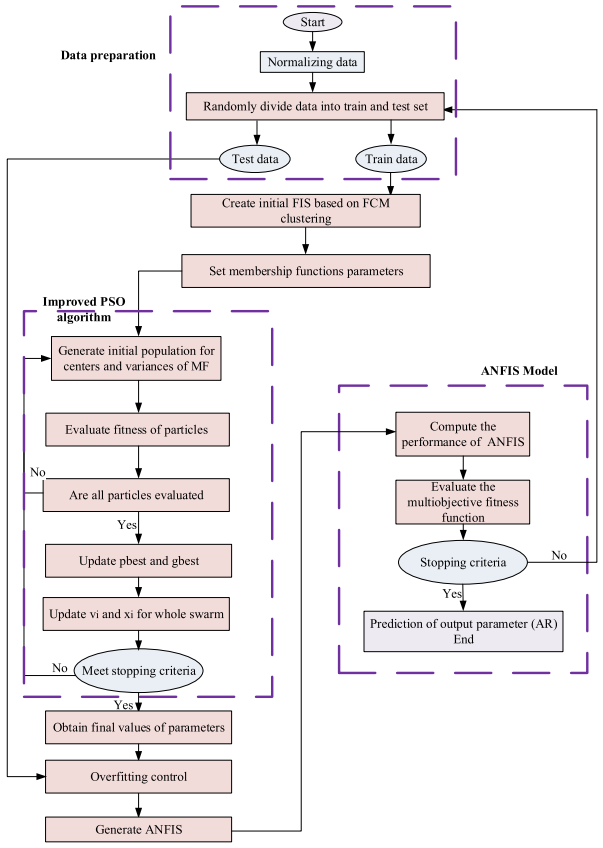


FIGURE 2. Workflow of improved PSO-ANFIS model.

The last step assigns these optimal values as antecedent and consequent parameters to the final ANFIS model.

III. PROJECT DESCRIPTION

As one type of TBM, an earth pressure balance (EPB) shield machine is suitable for digging tunnels in unstable ground such as clay, silt, and sand. In EPB shield tunnelling, there are an increasing number of computational models for predicting the cutting rate [31], torque and thrust [32], and advance speed [11], [25]. However, no studies have integrated improved PSO with ANFIS to predict the EPB shield performance.

This study completes this work by applying the proposed model of improved PSO-ANFIS to a field tunnelling project in Guangzhou, China [33], [34]. In the Guangzhou metro tunnels, an EPB shield machine with a diameter of 6.25 m was used to excavate the tunnel section between Maanshan Park Station and Liantang Station for Guangzhou Metro Line No. 9 [35], [36]. This case study is selected to verify the applicability of the proposed model. Also, this case represents a new project in the urban area of Guangzhou city, which needs to be carefully considered based on the existing infrastructures. The main specifications of the utilised machine are summarised in Table 1.

Fig. 3 shows a plan view of the construction site. The tunnel alignment is approximately 1280 m in length, with a burial depth varying from 7.0 to 10.0 m, as shown in Fig. 4. This study collects inputs of operational parameters and geological conditions from the monitoring and testing results along the tunnel alignment.

TABLE 1. Main features of utilised machine.

TBM type	EPB
External cutterhead diameter (m)	6.25
Inner lining diameter (m)	5.40
Outer lining diameter (m)	6.0
TBM length (m)	8.90
Cutterhead power (kW)	600
Disc cutter diameter (mm)	432
Disc cutter nominal spacing (mm)	90
Cutter tip nominal width (mm)	14.5

FIGURE 3. Location map of study area.

The advancement of the shield machine usually encountered silt clay soil at the studied section. The properties are listed in Table 2.

Field investigations showed that the void ratio ranges between 0.7 and 0.85, and the maximum cohesion value is 40 kPa. The silty clay soils have a plasticity of over 10 and a uniaxial compressive strength (UCS) of less than 2. The soils have a consistency index of less than 1.0, categorised as low-plasticity clay (CL) according to Casagrande's plasticity chart. In a standard penetration test (SPT), N values of the soils are over 10.

IV. DATA PREPROCESSING

In this project, the EPB shield machine has a built-in data acquisition system in which the actual data are collected by the sensors of every subsystem. The collected data are stored in the shield machine computer and transferred to the laboratory server over a fibre-optic network. This system is

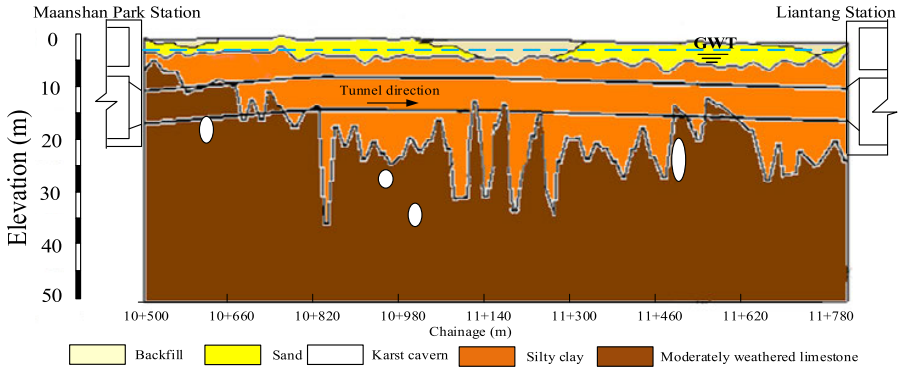


FIGURE 4. Longitudinal profile of guangzhou tunnel alignment.

TABLE 2. Geotechnical properties of encountered soil during tunnelling.

Soil layer	Properties	Unit	Average	Range
Silt clay	Density	(g/cm^3)	1.79	(1.7- 1.9)
	Water content	(%)	30	(25- 45)
	Void ratio		0.8	(0.7- 0.85)
	Cohesion	(kPa)	30	(20- 40)
	Friction	(°)	18	(10- 20)
	Hydraulic conductivity	(m/d)	0.08	(0.05- 0.15)

provided to simplify data collection during the tunnelling process and serves as a decision-making tool for tunnel engineers. Prior to the data analysis, raw data from the shield tunnelling are preprocessed based on the dimensional data. It is noteworthy that the data acquisition system records a wide diversity of shield operating data related to the tunnel performance, including the thrust force, cutterhead torque, and soil and grouting pressure. To adjust the dimensions of the monitoring data for the selection of shield parameter data, the following two criteria are adopted:

(1) The monitoring data should have meaningful values and be collected in the daily reports. Engineers refer to the daily reports to analyse the shield performance during the tunnelling process.
(2) The shield tunnelling parameters are examined and selected by tunnel experts so that the selected parameters can reflect the actual relationship of the shield tunnelling performance between different tunnel parameters.

Operation data usually include a certain amount of outliers, which affect the quality of the data [48]. Zhao et al. find that the K-nearest neighbour (KNN)-based outlier detection method is appropriate for detecting the outliers from a large amount of data [47]. Inappropriate raw databases were screened as outliers based on the K-nearest neighbour algorithm proposed by [46]. This study adopted the distance-based method of the KNN technique summarised in Algorithm 1 to detect outliers. The reason is that the operating data usually include a certain number of outliers,

Algorithm 1 K-Nearest-Neighbour-Based Model [43]

Input: sample set $\{x_n\}_{n=1}^{N} \subset R^M$, and outlier percentage $s\% = 6\%$;
Output: A list of outliers;
1: **For** sample $x_1, \ldots, x_N$ in the process; **do**
2: Find the nearest neighbour K by setting $N_D(x_n, K)$ of the sample x_n, the set of K points $\in \{x_n\}_{n=1}^{N}$ are the nearest points to x_n based on the metric D;
3: Calculate the outlier score T for the neighbour K *of* x_n:
$$T(x_n, K) = \frac{\sum_{y \in N_D(x_n, K)} D(x_n, y)}{K};$$
4: **end for**
5: Sort the samples $\{x_n\}_{n=1}^{N}$ in the ascending order of $\{T(x_n, K)\}_{n=1}^{N}$, and then choose the last outlier samples $s\%$.
6: **Return**

which normally affect the quality of the data [48]. In this manner, a total of 200 operating datasets were selected for the prediction of tunnel performance.

A. SHIELD TUNNELLING PERFORMANCE DATABASE

Following the aforementioned data preprocessing, this study established databases of the shield tunnelling performance, focusing on shield machine specifications and operational parameters. The shield machine specifications were collected from manufacturer's documents. Operational parameters were directly extracted from a built-in data acquisition system in the EBP shield machine. In total, there are nine parameters: cutterhead torque (CT), thrust force (TF), soil pressure (SP), rotational speed of the screw rate (SC), cutterhead rotation speed (CR), grouting pressure (GP), grouting amount (GA), excavation depth (H), and advance rate (AR). Among them, AR is closely interrelated with the other eight parameters.

Fig. 5 presents frequency histograms of eight EPB shield-tunnelling parameters of the selected 200 datasets. Most of the operating parameters have wide distributions, which

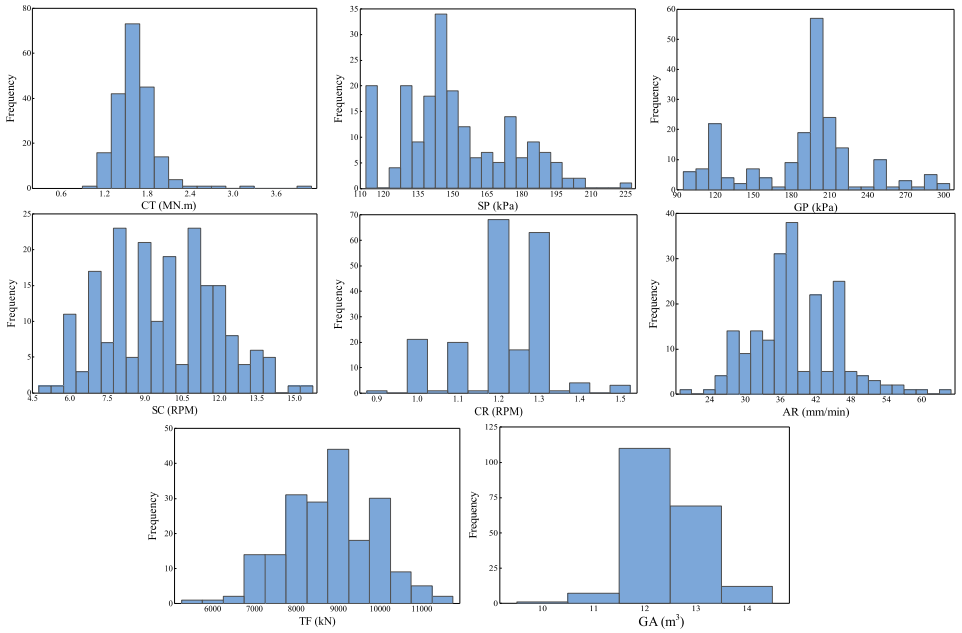


FIGURE 5. Histograms of different operating parameters in database.

result from frequent changes of the machine status during construction. Table 3 displays statistics of the nine parameters.

TABLE 3. Statistics of nine parameters in database.

Parameter	Unit	Type	Min.	Max.	Mean	St. deviation
TF	kN	Input	5600	11405	8821	1082
CT	MN.m	Input	1.0	4.0	1.58	0.327
SC	RPM	Input	5.0	15.5	9.76	2.229
CR	RPM	Input	0.9	1.5	1.21	0.106
GP	kPa	Input	100	300	188.9	45.44
SP	kPa	Input	113.3	223.3	151.4	22.99
H	m	Input	7.1	9.38	8.19	0.637
GA	m^3	Input	10	14	12.42	0.683
AR	mm/min	Output	20	63	38.54	7.12

B. PRINCIPLE COMPONENT ANALYSIS

Principle component analysis (PCA) is a conventional multivariate statistical approach used for classification and regression in various fields of study [37]–[39]. PCA can be applied to decrease the complex data form of forecasting variables to a lower dimension. During the analysis, the PCA can provide a few linear collections of the parameters that can be adjusted to summarize the data without losing much information. This method uses the orthogonal transformation to transform observations of possibly correlated variables to linearly uncorrelated variables. It reduces the dimensionality and keeps the informational value of the input data intact. PCA has been widely used for selecting independent variables and eliminating duplicate or highly associated variables. Fig. 6 shows a two-variable dataset, originally measured in the X-Y coordinate system.

In another coordinate system, the U axis refers to the principal direction of this dataset, and the V axis refers to the second most important direction. Usually, the V axis is orthogonal to the U axis; therefore, the covariance between the U and V variables is equal to zero. That means that all data are decorrelated by transforming from (X, Y) coordinates to (U, V) coordinates though an orthogonal transformation. PCA computes new variables as a linear combination of the original variables by calculating the covariance/correlation matrix of the data. When the variation of a dataset is caused by a natural property or a random experimental error, the variables are likely to follow normal distributions.

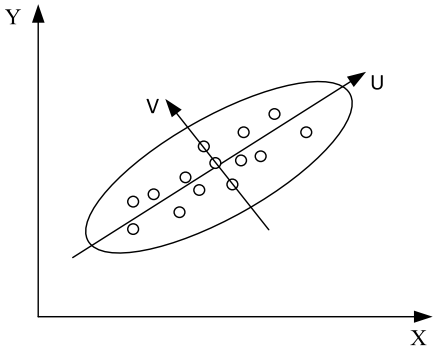


FIGURE 6. Principle components for data representation [38].

Linear transmutation transforms the input data into a set of components that are arranged according to their variance. The first principal component is the direction along which the data has the most variance. PCA projects the input data on a k-dimension eigenspace of k eigenvectors that are computed from a covariance matrix Σ of the data $N = [N_1, \ldots, N_n]$. N_i is i^{th} d-dimension data sample, and N refers to the number of samples. PCA chooses k, with $k < d$, eigenvectors having the largest eigenvalues that represent the main components

of the dataset. The selected eigenvectors are projected in a matrix and arranged into columns, where the first column corresponds to the largest eigenvalue. Eventually, PCA computes and determines the feature vector v from the data in the matrix [37].

Given the inputs of several parameters, this study uses PCA to identify critical input parameters that have the greatest impact on the advance rate (Fig. 7). This figure shows that the three inputs of CT, SC, and CR (i.e. the cases of 2, 3, and 4) are the most critical parameters, with the highest variance ratio of 93%. Therefore, CT, SC, and CR are selected as input parameters for developing the predictive model in this study, and the advance rate is considered as a function of these three parameters.

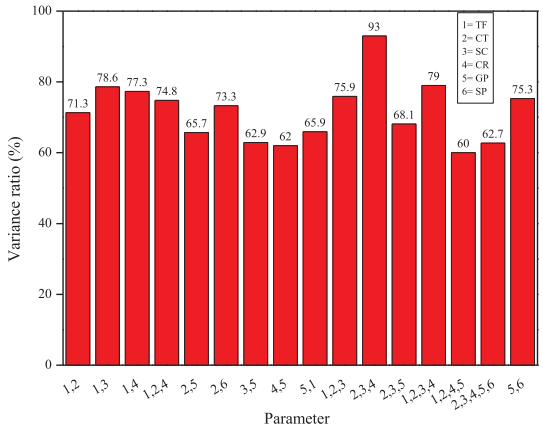


FIGURE 7. Principle components analysis for some parameters in this study.

V. RESULTS AND DISCUSSION

As previously mentioned regarding PCA, this study uses three input parameters (CT, SC, and CR) to predict AR within the hybrid improved PSO-ANFIS model. Additionally, an ANFIS model is established to compare the prediction accuracy of the hybrid model. For general computation procedures of the ANFIS-based FCM model and the PSO-ANFIS model, please refer to section II. Both ANFIS and PSO-ANFIS are implemented in MATLAB. This study has a total of 200 datasets, randomly divided into two subsets, of which 80% of the datasets are the training set and the other 20% are the testing set, following the recommendation of Swingler [40].

A. ANFIS-FCM MODEL

In the ANFIS model, all datasets are normalised to simplify the computational procedure using the following equation [41]:

$$X_n = \frac{(X - X_{min})}{(X_{max} - X_{min})} \tag{18}$$

where X and X_n are the measured and normalised data, respectively; X_{min} and X_{max} are the minimum and maximum data of X, respectively.

The ANFIS model in MATLAB requires users to determine the number and the type of membership functions (MFs). As there is no explicit method or formula to predict the necessary MF numbers [42], this study estimates the number of MFs by trial and error. The best estimates are obtained when using three Gaussian MFs. Table 4 lists the employed parameters in the developed model.

TABLE 4. Main parameters of ANFIS and improved PSO-ANFIS models.

Specification	Characteristic/ Value	
	ANFIS	Improved PSO-ANFIS
MF type	Gaussian	Gaussian
Inputs/Output	3/1	3/1
Fuzzy structure	Takagi-Sugeno	Takagi-Sugeno
Output MF	Linear	Linear
Minimum improvement	$1x10^{-5}$	$1x10^{-4}$
Number of fuzzy rules	10	8
Number of epochs	200	
Initial step size	0.01	0.01
Step size decrease rate	0.9	0.9
Step size increase rate	1.1	1.1
Number of training dataset	160	160
Number of testing dataset	40	40
Maximum iteration number		300
Cognitive acceleration (c_1)		2
Social acceleration (c_2)		2
Number of epochs for each population		100

The Takagi-Sugeno method is applied as FIS owing to its high accuracy and good computational effectiveness in developing a systematic approach for constructing fuzzy rules from the input-output dataset. MATLAB with the genfis3 function is implemented to construct the initial FIS structure of the model. More ANFIS settings based on the FCM clustering are listed in Table 4.

To improve the model accuracy, four different cases were designed to evaluate the impact of using different numbers of clusters in the FIS function on the computational results with ANFIS. The four cases use 5, 7, 10, and 14 clusters, respectively.

Table 5 presents the computational results from different cases for the training set and the testing set in terms of the coefficient of determination (R^2), root mean square error ($RMSE$), and variance account (VA).

The variations in the statistics can quantify the impact of changing the cluster number on the network result for the ANFIS model. Small variations of R^2, $RMSE$, and VA indicate that the number of clusters in the ANFIS model only slightly affects the prediction accuracy. Among the four cases, the third case of the ANFIS model using 10 clusters has the best prediction accuracy. Therefore, the following ANFIS model uses 10 clusters in the FIS function to predict the tunnel performance.

Fig. 8 shows the correlations between the measured and predicted advance rates for the training set and the testing set. This figure shows a better correlation in the training set than in the testing set for the ANFIS model.

TABLE 5. Statistical analysis of different ANFIS models.

No.	Number of clusters	Network results					
		Training data			Testing data		
		R^2	*RMSE*	*VA*	R^2	*RMSE*	*VA*
1	5	0.8896	0.0978	0.8501	0.7661	0.1299	0.7101
2	7	0.8806	0.0960	0.7892	0.7601	0.1216	0.7218
3	10	0.8911	0.0804	0.8544	0.7799	0.1201	0.7501
4	14	0.8843	0.0898	0.8511	0.7712	0.1244	0.7217

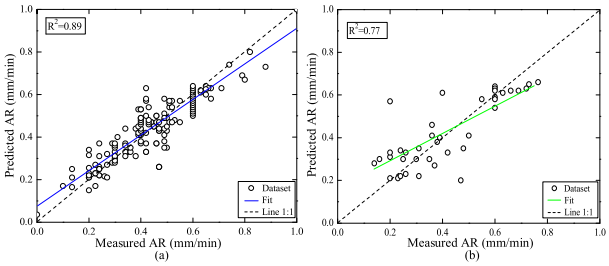


FIGURE 8. Comparison between measured and predicted AR from ANFIS-FCM model: (a) training set and (b) testing set.

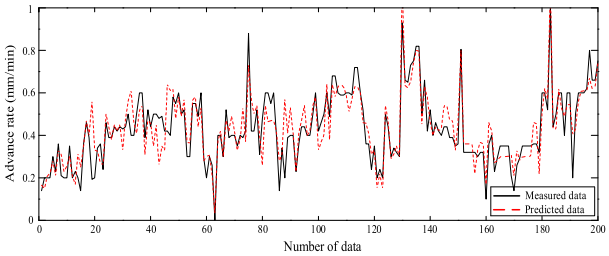


FIGURE 9. Comparison between measured and predicted AR from ANFIS-FCM model.

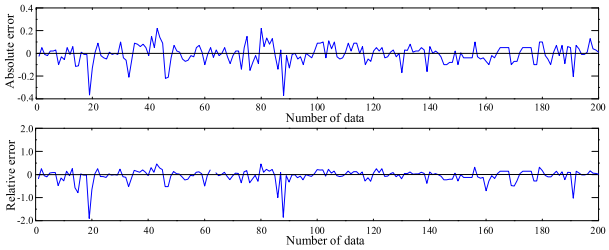


FIGURE 10. Absolute and relative error plots of ANFIS-FCM model.

In Fig. 9, it can be seen that the predicted values of the advance rate are relatively close to the measured values. In addition, the absolute and relative error indicators in Fig. 10 show that the predicted data can successfully follow the measured data with small discrepancies in the range of ±25%.

B. IMPROVED PSO-ANFIS MODEL

To develop the ANFIS model, the improved PSO is used to obtain optimum parameters for the ANFIS model. In this study, a Gaussian is applied as membership functions (MFs), as suggested by several researchers [14], [25]. In this hybrid model, PSO helps to establish closer relationships between the input and output. To determine the optimal PSO parameters, a trial-and-error approach is applied to

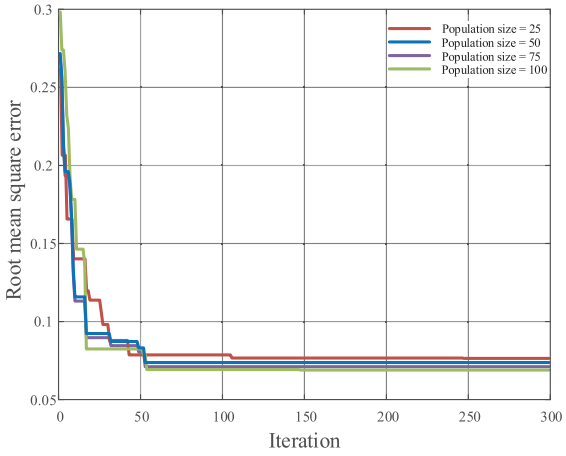


FIGURE 11. Convergence behaviour of improved PSO used to train ANFIS model.

find the maximum iteration numbers c_1 and c_2 [48]. These three parameters are eventually 300, 2, and 2, respectively (Table 4). As indicated in Fig. 11, the PSO algorithm converged the optimal fitness function after approximately 55 iterations, and then it settled at a constant level. This shows that PSO reached the optimal solution and that the search operation could be stopped.

The network performance results with the improved PSO-ANFIS model with different population sizes are displayed in Table 6. From this table, it can be concluded that the improved PSO-ANFIS model with a population size equal to 100 leads to the best prediction capacity. In the present study, the architecture of the fourth model (No. 4) was chosen as the best model to predict the tunnel performance, as shown in Table 6. In this study, the convergence speed of proposed model is considered [16]. Results showed that the 300 iterations required about 20 minutes to train the system. Otherwise, the computational volume of the proposed model is satisfied to achieve the accuracy from the predicted results.

TABLE 6. Comparison of analysis results from improved PSO-ANFIS model by varying the population size.

No.	Population size	Network results					
		Training data			Testing data		
		R^2	*RMSE*	*VA*	R^2	*RMSE*	*VA*
1	25	0.8915	0.0742	0.8612	0.8714	0.0779	0.8301
2	50	0.8942	0.0731	0.8661	0.8790	0.0753	0.8322
3	75	0.8997	0.0719	0.8649	0.8811	0.0712	0.8407
4	100	0.9011	0.0717	0.8798	0.8821	0.0701	0.8491
5	150	0.9012	0.0717	0.8788	0.8809	0.0701	0.8490

The comparison results in the estimation of the advance rate for the improved PSO-ANFIS model in the training set and the testing set are displayed in Fig. 12.

Scattered data in both plots are close to the line of equality (shown as a dashed line), demonstrating the good accuracy of the improved PSO-ANFIS model. To give a visual sense for the improved PSO-ANFIS model, Fig. 13 has been added

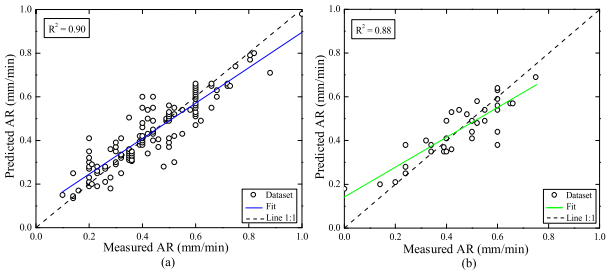


FIGURE 12. Comparison between measured AR and predicted AR from improved PSO-ANFIS model: (a) training set and (b) testing set.

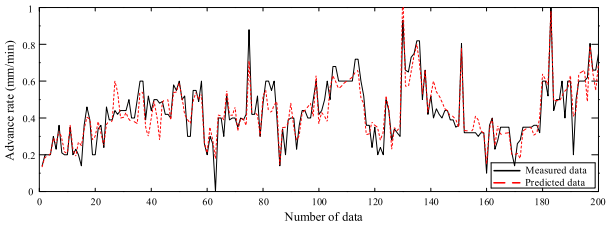


FIGURE 13. Comparison between measured AR and predicted AR from improved PSO-ANFIS model.

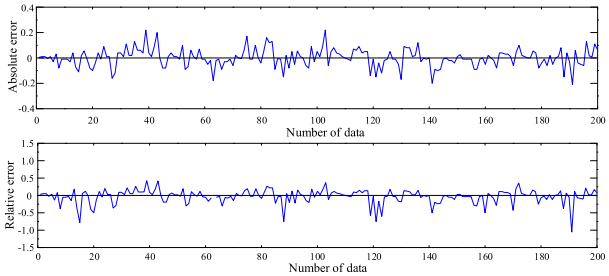


FIGURE 14. Absolute error and relative error of improved PSO-ANFIS model.

to show the relation between the measured and predicted AR for all databases. From Fig. 13, it can be seen that the AR predicted values are close to the measured values for almost all of the data. For more clarification, the absolute and relative errors of the outputs for the improved PSO-ANFIS model are plotted against the advance rate measured data, as depicted in Fig. 14.

The relative error of AR varies around zero, mostly in a smaller range (±15%) than the range (±25%) of the ANFIS-FCM model. This indicates that the improved PSO-ANFIS model has better accuracy in the prediction of tunnel boring machine performance when compared to the ANFIS-FCM model.

C. DISCUSSION

Applying computational techniques such as AI to the prediction of tunnelling performance has become increasingly popular. Previous studies on the prediction of TBM performance have mostly performed structural analyses under static load conditions far from actual working conditions [3]. Furthermore, predicting the TBM performance is a nonlinear and multivariable complex problem that cannot be accurately predicted using simple models.

In this respect, this study presented a hybrid multiobjective optimization technique to enhance the performance of TBM based mainly on dynamic operational factors. The dynamic operational factors, unlike the geological conditions, are controllable and thus can be manipulated by changing the machine orientation functions and optimal subsystems. Theoretically, the perfect prediction model is expected to have $RMSE = 0$, $R^2 = 1$, and $VA = 1$. A small value of the $RMSE$ and great values of the coefficient of determination R^2 and VA indicate a good prediction accuracy of the model. To assess the performance of the proposed model, this study uses a multiobjective fitness function with the objective of decreasing the $RMSE$ and increasing the coefficient of determination R^2 with a VA.

$$Minimize\ Fit = Z_1 \times RMSE - Z_2 \times R^2 + Z_3 \times VA \quad (19)$$

$$RMSE = \sqrt{\frac{\sum (x_{mea} - x_{pre})^2}{n}} \quad (20)$$

$$R^2 = 1 - \frac{\sum_{i=1}^{n} (x_{mea} - x_{pre})^2}{\sum_{i=1}^{n} (x_{mea} - x_m)^2} \quad (21)$$

$$VA = \left[1 - \frac{\mathrm{var}(x_{mea} - x_{pre})}{\mathrm{var}(x_{mea})}\right] \quad (22)$$

where x_{mea}, x_{pre}, and x_m are the measured, predicted, and mean of the x values, respectively; and n is the total number of datasets. Z_1, Z_2, and $Z_3 \in [0, 1]$, satisfying $Z_1 + Z_2 + Z_3 = 1$. To reach the optimal model, the values of Z_1, Z_2, and Z_3 are determined as 0.4, 0.31, and 0.29, respectively.

To understand the impact of the input parameters (CT, SC, and CR) on the response (AR) more fully, three-dimensional surface graphs are studied. Fig. 15 shows a surface graph of the improved PSO-ANFIS model to predict the advancement rate along with two input parameters while holding the third input parameter constant. As expected, the AR mainly follows a linear increasing trend with an increase in the SC, CT, and CR. However, there is a sharp decrease and a sudden increase in a local region ($0.4 < SC < 0.6$, $0 < CT < 0.2$). The fluctuations of AR values in the local region probably indicate either a sudden instability at the tunnel face or sudden changes in the geological features of this local region. This was validated by the TBM operator based on our discussion with him. The operator said that he clearly noticed sudden changes in local regions when operating the TBM machine. For instance, when finding an obvious variation of the extracted soil from the screw conveyor system or when the amount of soil extracted from the machine was very different from the estimated quantity, the TBM machine performed differently, presenting immediate changes in the operation parameters. Therefore, while the developed AI-based models should always find a good trade-off between the complexity of the model and the data

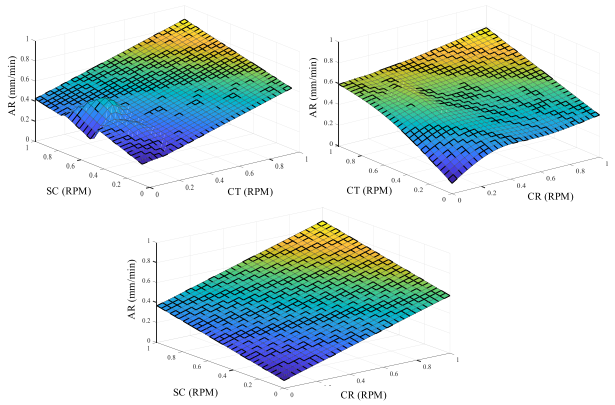


FIGURE 15. Surface graph of AR related to different parameters.

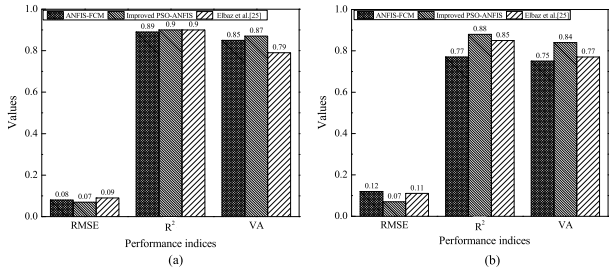


FIGURE 16. Comparison of performance indices for ANFIS-FCM, improved PSO-ANFIS, GA-ANFIS: (a) training set and (b) testing set.

dimensionality, we should also consider how to address the challenge of representing unexpected operations when operating the machine under different field conditions.

It can clearly be realised that the variation of the AR with CT, SC, and CR is found to be intuitive and in agreement with previous research. For instance, a similar model for the prediction of shield machine performance was developed by the author and his colleagues [12], [25] based on data from the Ma-Lian section of Guangzhou Metro Line No. 9. Their model integrates a genetic algorithm (GA) with the adaptive neurofuzzy inference system (ANFIS) based on a multiobjective fitness function. The results of their model are compared with the improved PSO-ANFIS, as shown in Fig. 16. This figure displays the assessment results from the ANFIS-FCM, improved PSO-ANFIS, and proposed model by [25]. Because of a smaller *RMSE* and greater values of R^2 and *VA*, the improved PSO-ANFIS model outperforms the GA-ANFIS and ANFIS-FCM models. The above analyses elucidate that the proposed model of PSO-ANFIS can predict the advance rate and represent their statistical features with reasonable accuracy.

In practical applications, the EPB shield machine can use shield parameters such as CT, CR, and SC as inputs to predict the shield tunnelling performance. It is noteworthy to mention that the proposed model can provide initial estimations of shield performance, especially for estimating the advance rate of the shield machine at the project planning stage. With the advance rate determined from the proposed model, project durations can be estimated, thus facilitating time allocations when developing construction plans. Briefly, the improved model in this research is expected to provide insightful suggestions to support engineers in the prediction of shield-tunnelling advancement, and can be used as intelligent selection to achieve an acceptable prediction for TBM performance.

VI. CONCLUSION

This study presented an AI-based model to predict the shield machine performance during the tunnelling process. In this regard, the most influential parameters were identified through PCA, and an improved PSO-ANFIS model was established to predict the advance rate of the EPB shield tunnelling. The proposed model was applied to a case study of the Guangzhou Metro Line 9 tunnelling project. For validation, prediction results from the improved PSO-ANFIS model were compared with the prediction results from an ANFIS-FCM model. Major conclusions were obtained as follows:

- The improved PSO-ANFIS model can predict the shield performance in terms of the advance rate, in good agreement with the measured advance rate for both the training and testing sets. The improved PSO-ANFIS model uses computation parameters tailored to the studied tunnel section for predicting the advance rate. Based on a multiobjective fitness function, the values of R^2, *RMSE*, and *VA* of 0.88, 0.07, and 0.84 for the testing datasets indicate that the proposed model is accurate.
- The proposed model demonstrates better prediction accuracy than the ANFIS and GA-ANFIS [25] models based on a multiobjective fitness function. Prediction results from this study can facilitate decision-makers in predicting the project duration and construction cost of EPB shield tunnels. This supports efficient construction management, particularly when developing construction plans.
- The absolute error of the improved PSO-ANFIS method was in an adequate range of $\pm 15\%$, whereas the ANFIS-FCM showed a wider error range of ± 0.25. This demonstrated the precise prediction of the improved model in the prediction of tunnel boring machine performance. Therefore, the improved model can be utilised to guide construction practices in a more meaningful way.
- The proposed model is general and can be used for analysing different tunnelling systems in other types of geological conditions. To improve the robustness, more tunnelling data should be collected for calibration and validation of the proposed model.

REFERENCES

[1] H.-M. Lyu, S.-L. Shen, J. Yang, and Z.-Y. Yin, "Inundation analysis of metro systems with the storm water management model incorporated into a geographical information system: A case study in Shanghai," *Hydrol. Earth Syst. Sci.*, vol. 23, no. 10, pp. 4293–4307, Oct. 2019, doi: 10.5194/hess-23-4293-2019.

[2] H.-M. Lyu, S.-L. Shen, A. Zhou, and J. Yang, "Risk assessment of mega-city infrastructures related to land subsidence using improved trapezoidal FAHP," *Sci. Total Environ.*, Nov. 2019, Art. no. 135310, doi: 10.1016/j.scitotenv.2019.135310.

[3] W. Sun, M. Shi, C. Zhang, J. Zhao, and X. Song, "Dynamic load prediction of tunnel boring machine (TBM) based on heterogeneous in-situ data," *Automat. Construct.*, vol. 92, pp. 23–34, Aug. 2018.
[4] X.-X. Liu, S.-L. Shen, Y.-S. Xu, and Z.-Y. Yin, "Analytical approach for time-dependent groundwater inflow into shield tunnel face in confined aquifer," *Int. J. Numer. Anal. Methods Geomech.*, vol. 42, no. 4, pp. 655–673, Dec. 2017.
[5] K. Elbaz, J. S. Shen, A. Arulrajah, and S. Horpibulsuk, "Geohazards induced by anthropic activities of geoconstruction: A review of recent failure cases," *Arabian J. Geosci.*, vol. 9, no. 18, Nov. 2016, doi: 10.1007/s12517-016-2740-z.
[6] H.-M. Lyu, S.-L. Shen, A. Zhou, and J. Yang, "Perspectives for flood risk assessment and management for mega-city metro system," *Tunnelling Underground Space Technol.*, vol. 84, pp. 31–44, Feb. 2019.
[7] H. M. Lyu, W. J. Sun, S. L. Shen, and A. N. Zhou, "Risk assessment using a new consulting process in fuzzy AHP," *J. Construct. Eng. Manage.*, vol. 146, no. 3, 2020, Art. no. 04019112, doi: 10.1061/(ASCE)CO.1943-7862.0001757.
[8] A. Azad, M. Manoochehri, H. Kashi, S. Farzin, H. Karami, V. Nourani, and J. Shiri, "Comparative evaluation of intelligent algorithms to improve adaptive neuro-fuzzy inference system performance in precipitation modelling," *J. Hydrol.*, vol. 571, pp. 214–224, Apr. 2019, doi: 10.1016/j.jhydrol.2019.01.062.
[9] S.-L. Shen, Y.-X. Wu, and A. Misra, "Calculation of head difference at two sides of a cut-off barrier during excavation dewatering," *Comput. Geotechnics*, vol. 91, pp. 192–202, Nov. 2017, doi: 10.1016/j.compgeo.2017.07.014.
[10] D.-J. Ren, S.-L. Shen, A. Arulrajah, and H.-N. Wu, "Evaluation of ground loss ratio with moving trajectories induced in double-O-tube (DOT) tunnelling," *Can. Geotech. J.*, vol. 55, no. 6, pp. 894–902, Jun. 2018, doi: 10.1139/cgj-2017-0355.
[11] A. Salimi, H. Karami, S. Farzin, M. Hassanvand, A. Azad, and O. Kisi, "Design of water supply system from rivers using artificial intelligence to model water hammer," *ISH J. Hydraulic Eng.*, pp. 1–10, Apr. 2018, doi: 10.1080/09715010.2018.1465366.
[12] N. Zhang, S.-L. Shen, A. Zhou, and Y.-S. Xu, "Investigation on performance of neural networks using quadratic relative error cost function," *IEEE Access*, vol. 7, pp. 106642–106652, 2019, doi: 10.1109/ACCESS.2019.2930520.
[13] A. C. Adoko, C. Gokceoglu, and S. Yagiz, "Bayesian prediction of TBM penetration rate in rock mass," *Eng. Geol.*, vol. 226, pp. 245–256, Aug. 2017.
[14] M. Hasanipanah, A. Shahnazar, H. Arab, S. B. Golzar, and M. Amiri, "Developing a new hybrid-AI model to predict blast-induced backbreak," *Eng. Comput.*, vol. 33, pp. 349–359, Aug. 2016, doi: 10.1007/s00366-016-0477-7.
[15] A. Azad, H. Kashi, S. Farzin, V. P. Singh, O. Kisi, H. Karami, and H. Sanikhani, "Novel approaches for air temperature prediction: A comparison of four hybrid evolutionary fuzzy models," *Meteorol. Appl.*, vol. 27, no. 1, p. e1817, Jul. 2019, doi: 10.1002/met.1817.
[16] O. Kisi, A. Azad, H. Kashi, A. Saeedian, S. A. A. Hashemi, and S. Ghorbani, "Modeling groundwater quality parameters using hybrid neuro-fuzzy methods," *Water Resour. Manage.*, vol. 33, no. 2, pp. 847–861, Dec. 2018.
[17] 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. Civil Eng.*, vol. 22, pp. s130–s148, May 2017.
[18] S. Yagiz, C. Gokceoglu, 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. 808–814, Jun. 2009.
[19] S. Mahdevari and S. R. Torabi, "Prediction of tunnel convergence using artificial neural networks," *Tunnelling Underground Space Technol.*, vol. 28, pp. 218–228, Mar. 2012.
[20] Z.-Y. Yin, Y.-F. Jin, J. S. Shen, and P.-Y. Hicher, "Optimization techniques for identifying soil parameters in geotechnical engineering: Comparative study and enhancement," *Int. J. Numer. Anal. Methods Geomech.*, vol. 42, no. 1, pp. 70–94, Jul. 2017.
[21] O. Acaroglu, "Prediction of thrust and torque requirements of TBMs with fuzzy logic models," *Tunnelling Underground Space Technol.*, vol. 26, no. 2, pp. 267–275, Mar. 2011.
[22] D. Bouayad and F. Emeriault, "Modeling the relationship between ground surface settlements induced by shield tunneling and the operational and geological parameters based on the hybrid PCA/ANFIS method," *Tunnelling Underground Space Technol.*, vol. 68, pp. 142–152, Sep. 2017.
[23] R.-E. Precup and H. Hellendoorn, "A survey on industrial applications of fuzzy control," *Comput. Ind.*, vol. 62, no. 3, pp. 213–226, Apr. 2011.
[24] K. Zielinski and R. Laur, "Stopping criteria for a constrained single-objective particle swarm optimization algorithm," *Informatica*, vol. 31, no. 1, pp. 51–59, 2007.
[25] K. Elbaz, S. L. Shen, A. Zhou, D. J. Yuan, and Y. S. Xu, "Optimization of EPB shield performance with adaptive neuro-fuzzy inference system and genetic algorithm," *Appl. Sci.*, vol. 9, no. 4, p. 780, 2019, doi: 10.3390/app9040780.
[26] J. S. R. Jang, "ANFIS: Adaptive-network-based fuzzy inference system," *IEEE Trans. Syst., Man, Cybern.*, vol. 23, no. 3, pp. 665–685, May/Jun. 1993.
[27] R. Eberhart and J. Kennedy, "A new optimizer using particle swarm theory," in *Proc. MHS. Proc. 6th Int. Symp. Micro Mach. Hum. Sci.*, Oct. 1995, pp. 39–43.
[28] D. L., Z. T., W. L., and X. O., "A particle swarm optimization algorithm base on new information point," *Int. J. Digit. Content Technol. its Appl.*, vol. 6, no. 21, pp. 524–531, Nov. 2012.
[29] M. Clerc, "The swarm and the queen: Towards a deterministic and adaptive particle swarm optimization," in *Proc. Congr. Evol. Comput.*, vol. 3, Jul. 1999, pp. 1951–1957.
[30] L. D. S. Coelho, "A quantum particle swarm optimizer with chaotic mutation operator," *Chaos, Solitons Fractals*, vol. 37, no. 5, pp. 1409–1418, Sep. 2008.
[31] D.-J. Ren, S.-L. Shen, A. Arulrajah, and W.-C. Cheng, "Prediction model of TBM disc cutter wear during tunnelling in heterogeneous ground," *Rock Mech. Rock Eng.*, vol. 51, no. 11, pp. 3599–3611, Jul. 2018, doi: 10.1007/s00603-018-1549-3.
[32] U. Ates, N. Bilgin, and H. Copur, "Estimating torque, thrust and other design parameters of different type TBMs with some criticism to TBMs used in Turkish tunneling projects," *Tunnelling Underground Space Technol.*, vol. 40, pp. 46–63, Feb. 2014.
[33] Q.-L. Cui, H.-N. Wu, S.-L. Shen, Y.-S. Xu, and G.-L. Ye, "Chinese karst geology and measures to prevent geohazards during shield tunnelling in karst region with caves," *Natural Hazards*, vol. 77, no. 1, pp. 129–152, Jan. 2015.
[34] H.-M. Lyu, W.-J. Sun, S.-L. Shen, and A. Arulrajah, "Flood risk assessment in metro systems of mega-cities using a GIS-based modeling approach," *Sci. Total Environ.*, vol. 626, pp. 1012–1025, Jun. 2018.
[35] K. Elbaz, S. L. Shen, W. C. Cheng, and A. Arulrajah, "Cutter-disc consumption during earth pressure balance tunnelling in mixed strata," *Proc. Inst. Civil Eng.-Geotech. Eng.*, vol. 171, no. 3, pp. 363–376, 2018, doi: 10.1680/jgeen.17.00117.
[36] K. Elbaz, S.-L. Shen, Y. Tan, and W.-C. Cheng, "Investigation into performance of deep excavation in sand covered karst: A case report," *Soils Found.*, vol. 58, no. 4, pp. 1042–1058, Aug. 2018.
[37] F. A. Mousa, R. A. El-Khoribi, and M. E. Shoman, "A novel brain computer interface based on principle component analysis," *Procedia Comput. Sci.*, vol. 82, pp. 49–56, 2016, doi: 10.1016/j.procs.2016.04.008.
[38] 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.
[39] W. Chen, W. Xiong, J. Du, and J. Cheng, "Scale registration based on descriptor analysis and B-spline matching," in *Proc. TENCON Region Conf.*, Nov. 2017, pp. 1451–1456.
[40] K. Swingler, *Applying Neural Networks: A Practical Guide*, vol. 442. New York, NY, USA: Academic, 1996.
[41] H. Khamesi, S. R. Torabi, H. Mirzaei-Nasirabad, and Z. Ghadiri, "Improving the performance of intelligent back analysis for tunneling using optimized fuzzy systems: Case study of the Karaj subway line 2 in iran," *J. Comput. Civil Eng.*, vol. 29, no. 6, Nov. 2015, Art. no. 05014010.
[42] M. Rezakazemi, A. Ghafarinazari, S. Shirazian, and A. Khoshsima, "Numerical modeling and optimization of wastewater treatment using porous polymeric membranes," *Polym. Eng. Sci.*, vol. 53, no. 6, pp. 1272–1278, 2013.
[43] V. Hautamaki, I. Karkkainen, and P. Franti, "Outlier detection using k-nearest neighbour graph," in *Proc. 17th Int. Conf. Pattern Recognit. (ICPR)*, Aug. 2004, pp. 430–433.

[44] Y. Lu, M. Liang, Z. Ye, and L. Cao, "Improved particle swarm optimization algorithm and its application in text feature selection," *Appl. Soft Comput.*, vol. 35, pp. 629–636, Oct. 2015, doi: 10.1016/j.asoc.2015.07.005.
[45] T. M. Mitchell, *Machine Learning*. New York, NY, USA: McGraw-Hill, 1997, pp. 52–79.
[46] S. Ramaswamy, R. Rastogi, and K. Shim, "Efficient algorithms for mining outliers from large data sets," *ACM SIGMOD Rec.*, vol. 29, no. 2, pp. 427–438, Jun. 2000.
[47] J. Zhao, M. Shi, G. Hu, X. Song, C. Zhang, D. Tao, and W. Wu, "A data-driven framework for tunnel geological-type prediction based on TBM operating data," *IEEE Access*, vol. 7, pp. 66703–66713, 2019, doi: 10.1109/ACCESS.2019.2917756.
[48] Y. Chen, D. Miao, and H. Zhang, "Neighborhood outlier detection," *Expert Syst. Appl.*, vol. 37, no. 12, pp. 8745–8749, Dec. 2010.
[49] A. Shahnazar, H. Nikafshan Rad, M. Hasanipanah, M. M. Tahir, D. Jahed Armaghani, and M. Ghoroqi, "A new developed approach for the prediction of ground vibration using a hybrid PSO-optimized ANFIS-based model," *Environ. Earth Sci.*, vol. 76, no. 15, pp. 1–17, Aug. 2017, doi: 10.1007/s12665-017-6864-6.

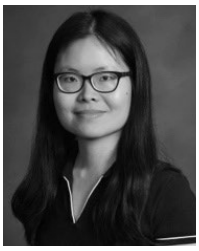
KHALID ELBAZ received the Ph.D. degree from the Department of Civil Engineering, School of Naval Architecture, Ocean and Civil Engineering, Shanghai Jiao Tong University, in 2019. He is currently a Research Associate with the School of Civil and Environmental Engineering, Shantou University.

His researches focus on developing and applying soft computing techniques, and heuristic optimization algorithms to solve and adjust issues related to tunnelling and deep excavation in the field. He has developed several computational models to help engineers in both planning and construction stages of shield tunnelling and assess the risk of the adjacent infrastructures.

SHUI-LONG SHEN received the B.Sc. degree in underground space technology from Tongji University, in 1986, the M.Phil. degree in structural engineering from Tongji University, in 1989, and the Ph.D. degree in geotechnical engineering from Saga University, Japan, in 1998.

He worked with the Institute of Lowland Technology (ILT) as a Lecturer from 1998 to 2001. He served as an Associate Editor of *Lowland Technology International*, an International Journal. From 2001 to 2003, he worked with the National Institute for Environmental Studies, Tsukuba, the Science City, Japan. In 2003, he joined the Department of Civil Engineering (DCE), Shanghai Jiao Tong University (SJTU) as a Faculty Member. In 2019, he joined the College of Engineering, Shantou University as the Dean. He has been keeping collaboration with other international organization, e.g., Saga University, Virginia Tech, The University of Kansas, The University of Hong Kong, the Suranaree University of Technology, Thailand, the Ecole Centrale de Nantes France, the Swinburne University of Technology, and RMIT University, Australia as a Guest/Visiting/Adjunct Professor.

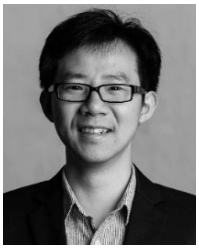
Dr. Shen also serves as an Editor/Editorial Board Member of four International Journals, e.g., the *Canadian Geotechnical Journal*, *Geotextiles and Geomembranes*, *Computers and Geotechnics* (ELSEVEIR), *Marine Georesources and Geotechnology* (Taylor and Francis), *Lowland Technology International*, and *Geotechnical Engineering—SEAGS* and domestic journals, e.g., the *Chinese Journal of Geotechnical Engineering*.



WEN-JUAN SUN received the Ph.D. degree in civil engineering from Virginia Polytechnic Institute and State University, in 2014. She has been in civil engineering with Lehigh University as a Research Associate since 2016. She has published more than 50 refereed articles in top-tier technical journals and international conferences. Her research focuses on developing and applying computational models to solve engineering problems for supporting sustainable and resilient infrastructures. At system-level, she has developed computational models to assess the risk and resilience of critical infrastructures, with considerations of different types of interdependencies and large uncertainties in terms of rigorous models. At fine levels, she has developed physics-based computational models to analyze failure mechanisms of infrastructure materials at multiple scales. She has also been serving as a member in the Objective Resilience Committee and the Nano-mechanics and Micro-mechanics Committee, American Society of Civil Engineers (ASCE). She has been serving as a Reviewer for more than ten journals.

ZHEN-YU YIN received the B.Sc. degree in civil engineering from Zhejiang University, in 1997, and the M.Sc. and Ph.D. degrees from the Ecole Centrale de Nantes, France, in 2003 and 2006, respectively.

Followed by 5 years, he was an Engineering Consultancy with the Zhejiang Jiahua Architecture Design Institute. Then, he has been working as Postdoctoral Researcher with the Helsinki University of Technology, Finland; the University of Strathclyde, Glasgow, U.K.; the Ecole Centrale de Nantes; and the University of Massachusetts at Amherst (Umass), Amherst, USA. In 2010, he joined Shanghai Jiao Tong University as Special Researcher and received the Professor of Exceptional Rank of Shanghai Dong-Fang Scholar. In 2013, he joined the Ecole Centrale de Nantes as an Associate Professor before moving to Hong Kong. He has been an Associate Professor of geotechnical engineering with The Hong Kong Polytechnic University, since 2018. He has published more than 100 articles in peer reviewed international journals. Since 2012, he has been a member of the Granular Materials Committee, American Society of Civil Engineers.



ANNAN ZHOU received the Ph.D. degree in geotechnical engineering from the University of Newcastle, in 2011, with an Australian Endeavor Scholarship. After his graduation, he joined RMIT as a tenured Lecturer in geotechnical engineering and was promoted to a Senior Lecturer in 2014 and an Associate Professor in 2018.

He has published more than 70 refereed journal articles with more than 1400 citations. He has successfully supervised 2 M.Sc. and 4 Ph.D. to their completion. The major research area of includes constitutive/numerical modelling of multiphase porous media and advanced laboratory testing on unsaturated soils. From 2014 to 2016, he was elected to be a RMIT University Research Committee Board Member to represent the ECR academics over the University. The quality of his publications has been recognized by being awarded several international awards like the ISSMGE ECR International Best Paper Award, in 2015, and the CGJ Editor's Choice, in 2016. He received the ARC DECRA Fellowship, in 2012, the AGS Hugh Trollope Medal, in 2014, the CGJ Outstanding Reviewer, in 2016, and the ISSMGE Bright Spark Lecturer Award, in 2018. As a Chief Investigator, he has secured several competitive research grants more than 1.5 millon, including ARC DE, DP, and LP. Since 2015, he has been serving as an Editorial Board Member for *Canadian Geotechnical Journal* (NRC), *Underground Space* (Elsevier), and *Advance in Civil Engineering* (Hindawi). He also is invited as an External Reviewer for several national grants (like ARC, ISF, and NSFC) and numerous flagship journals (like *Scientific Reports*, *Geotechnique*, and *Computers and Geotechnics*).