Tunnel wall convergence prediction using optimized LSTM deep neural network

Article *in* Geomechanics and Engineering · December 2022

DOI: 10.12989/gae.2022.31.6.545

7 authors, including:

Arsalan Mahmoodzadeh
Halabja University

Adil Mohammed
Cihan University

Hawkar Ibrahim
Salahaddin University - Erbil

Hanan Samadi
University of Tehran

Geomechanics and Engineering, *Vol. 31, No. 6 (2022) 545-556*
https://doi.org/10.12989/gae.2022.31.6.545

Tunnel wall convergence prediction using optimized LSTM deep neural network

Arsalan Mahmoodzadeh*[1], Mohammadreza Taghizadeh[2a], Adil Hussein Mohammed[3b], Hawkar Hashim Ibrahim[4c], Hanan Samadi[5d], Mokhtar Mohammadi[6e] and Shima Rashidi[7f]

[1]Rock Mechanics Division, School of Engineering, Tarbiat Modares University, Tehran, Iran
[2]Department of Civil engineering, Faculty of engineering, University of Kashan, Kashan, Iran
[3]Department of Communication and Computer Engineering, Faculty of Engineering, Cihan University-Erbil, Kurdistan Region, Iraq
[4]Department of Civil Engineering, College of Engineering, Salahaddin University-Erbil, 44002 Erbil, Kurdistan Region, Iraq
[5]School of Geology, College of Science, University of Tehran, Tehran, Iran
[6]Department of Information Technology, College of Engineering and Computer Science, Lebanese French University, Kurdistan Region, Iraq
[7]Department of Computer Science, College of Science and Technology, University of Human Development, Sulaymaniyah, Kurdistan Region, Iraq

(*Received April 27, 2022, Revised August 31, 2022, Accepted September 8, 2022*)

Abstract. Evaluation and optimization of tunnel wall convergence (TWC) plays a vital role in preventing potential problems during tunnel construction and utilization stage. When convergence occurs at a high rate, it can lead to significant problems such as reducing the advance rate and safety, which in turn increases operating costs. In order to design an effective solution, it is important to accurately predict the degree of TWC; this can reduce the level of concern and have a positive effect on the design. With the development of soft computing methods, the use of deep learning algorithms and neural networks in tunnel construction has expanded in recent years. The current study aims to employ the long-short-term memory (LSTM) deep neural network predictor model to predict the TWC, based on 550 data points of observed parameters developed by collecting required data from different tunnelling projects. Among the data collected during the pre-construction and construction phases of the project, 80% is randomly used to train the model and the rest is used to test the model. Several loss functions including root mean square error (RMSE) and coefficient of determination (R^2) were used to assess the performance and precision of the applied method. The results of the proposed models indicate an acceptable and reliable accuracy. In fact, the results show that the predicted values are in good agreement with the observed actual data. The proposed model can be considered for use in similar ground and tunneling conditions. It is important to note that this work has the potential to reduce the tunneling uncertainties significantly and make deep learning a valuable tool for planning tunnels.

Keywords: convergence estimation; deep neural network; LSTM; optimization; tunnel wall convergence

1. Introduction

Due to the rapid growth of the population, tunnels are becoming increasingly necessary for transportation and travel. To reduce hazards and improve predictability, tunnel construction is becoming increasingly popular among engineers around the world. Determination of geological characteristics along the tunnel route, the design features, the installed maintenance systems, as well as the construction expenditures are the most uncertainties in NATM or TBM tunnelling projects. Improper tunnel designs by engineers based on pre-construction studies can result in increased time consumption, cost overruns, and even the death of working personnel (Mahmoodzadeh *et al.* 2022a).

It is necessary to mention that tunnel wall convergence (TWC) is one of the most concern for tunnel construction. Possible risks can be reduced through performing the correct maintenance of tunnel and accurate prediction of this vital parameter. In recent decades, a few models including empirical, analytical, arithmetic, and soft computing methods have been proposed for estimating the TWC parameters of tunnelling projects.

Analytical methods (Fahimifar *et al.* 2010, Nomikos *et al.* 2011, Sterpi and Gioda 2009) and numerical methods (Debernardi and Barla 2009, Nadimi *et al.* 2011, Sharifzadeh *et al.* 2013) aim to model and solve the rheological and mechanical equations that monitor the

*Corresponding author, M.Sc.
E-mail: m.arsalan@modares.ac.ir
[a]M.Sc.
E-mail: reza96taghizadeh@gmail.com
[b]M.Sc.
E-mail: adil.mohammed@cihanuniversity.edu.iq
[c]M. Sc.
E-mail: hawkar.ibrahim@su.edu.krd
[d]M. Sc.
E-mail: samadi.hanan1996@ut.ac.ir
[e]Ph. D.
E-mail: mukhtar@lfu.edu.krd
[f]M. Sc.
E-mail: shima.rashid@uhd.edu.iq

ISSN: 2005-307X (Print), 2092-6219 (Online)

TWC during certain times. However, despite their considerable benefits, analytical and numerical methods have trouble being implemented in the field because they depend on precise parameters. Finding the proper parameters are also difficult because of the small scale of the laboratory compared to the field (Sterpi and Gioda 2009) and/or the time and expense of the tests needed to classify materials' rheological behavior (Guan *et al.* 2009).

The empirical methods can be considered as a powerful alternative strategy that might recreate the actual convergence patterns using rudimentary models just based on limited factors (Kontogianni *et al.* 2006, Sakurai 1978). Based on the mentioned techniques, a software platform can be developed to define and prediction TWC and also assess the convergence of actual tunnels by using experimental data and curve fitting techniques (Asadollahpour *et al.* 2014, Vu *et al.* 2013). Nevertheless, these models of methodological convergence are not yet practical for application, and selecting appropriate parameters for their use is challenging, making their use difficult when measuring TWC. It needs a more comprehensive study to decide this (González and Jiménez 2011).

With the development of soft computing methods, the use of deep learning and neural networks in the fields of tunneling and geomechanics engineering have expanded in recent years (Mahmoodzadeh *et al.* 2022b, Mirzaeiabdolyousefi et al. 2022, Fakhri *et al.* 2022, Bai *et al.* 2021, Luat *et al.* 2020, Liu *et al.* 2021a, b, Xiang *et al.* 2021). These methods have also shown good accuracy in the prediction of TWC. Researchers have used a variety of machine learning (ML) techniques in order to accomplish this aim. These techniques include adaptive neuro-fuzzy inference system (ANFIS) (Adoko and Wu 2012), artificial neural networks (ANN) (Adoko *et al.* 2013, Rafiai and Moosavi 2012), support vector machines (SVM) (Mahdevari *et al.* 2013).

Since many ML algorithms have different hyperparameters, it is crucial to optimize them to explore the robustness of the method for TWC prediction. This is because the ML algorithms have distinct hyperparameters. This study aims to employ long-short term memory (LSTM) algorithm for TWC prediction. According to the literature review (Adoko *et al.* 2013, Mahdevari *et al.* 2013, Feng *et al.* 2019, Hajihassani *et al.* 2019), this is the first comprehensive study of the use of LSTM to predict the TWC in the field of geotechnical engineering. The model is trained using 550 previous tunnel data points. The LSTM model was implemented for various hyper-parameter varieties and hidden layer structures to enhance its accuracy and performance. In order to avoid overfitting, the dropout method was used. K-nearest neighbors (KNN), generalized prediction and reweighting (GPR), support vector regression (SVR), and a decision tree are all compared to the LSTM model's results. Testing the accuracy of mentioned models is performed. The results of this study can be used to develop early warning systems for TWC. In addition, this research has the potential to be used as a foundation for future applications of LSTM to other issues in geotechnical fields.

2. Literature review

Adoko and Wu (2012) used an ANFIS approach to predict the convergence of tunnel that was excavated using the new Austrian tunneling method (NATM). In their study, they collected a database containing 1057 data sets, and the ANFIS output demonstrated high level of accuracy based on predicted values. Their model was proposed as a predictor network for estimating of convergence in tunnels that was excavated by NATM method.

The prediction of TWC was performed using an ANN method by Mahdevari *et al.* (2012). Four parameters including cohesion (C), angle of internal friction (φ), elasticity modulus (E), and uniaxial compressive strength (UCS), and one geomechanical parameter of uniaxial tensile strength (σ_t), were found to have the greatest and least significant effects on the TWC, respectively, Out of the 12 effective input parameters considered in their study.

In order to foretell the TWC excavated by a tunnel boring machine (TBM), Mahdevari *et al*. (2013) used two methods of machine learning including ANN and SVM algorithms. Despite the fact that both of them were capable of accurate prediction, the SVM had a better performance.

The convergence of lined circular tunnels was predicted using a neural network (ANN)-based solution method by Rafiai and Moosavi (2012). They explained the constraints of the solution.

To predict the TWC with a bullet train, Adoko *et al.* (2013) used a mixed-model ANN and Markov analysis of relative speeds and distances. Finally, the study revealed high precision in the TWC model predictability compared with MARS.

Mahdevari *et al*. (2013) developed a dynamic predictor model for prediction of TWC on-site by using SVM network. Their model was flexible enough to be updated during the tunnel construction, as well as used data from newly excavated ring sections as part of the model's training. The updating process improved the accuracy of the results in every excavated rings.

For better prediction results of the convergence phenomenon, Feng *et al.* (2019) proposed a Bayesian approach to update the model with new data monitored during the tunnel construction.

Using the gene expression programming (GEP) method, Hajihassani *et al.* (2019) proposed a model and demonstrated its efficiency in predicting TWC.

Eight different ML techniques were employed by Torabi-Kaveh and Sarshari (2020) to predict the TWC rate: SVR, MLP-ANN, GPR, RBF-ANN, MNR, RT, MLR, and ET. MLP-ANN had the highest predictive accuracy for TWC out of the eight models used, while RT and GPR had the lowest accuracy.

3. Methodology

In spite of the many benefits of ML approaches, there is no universally applicable model of ML, according to the No-Free-Lunch theorem. In order to complete various optimization tasks, researchers have tried to evaluate the

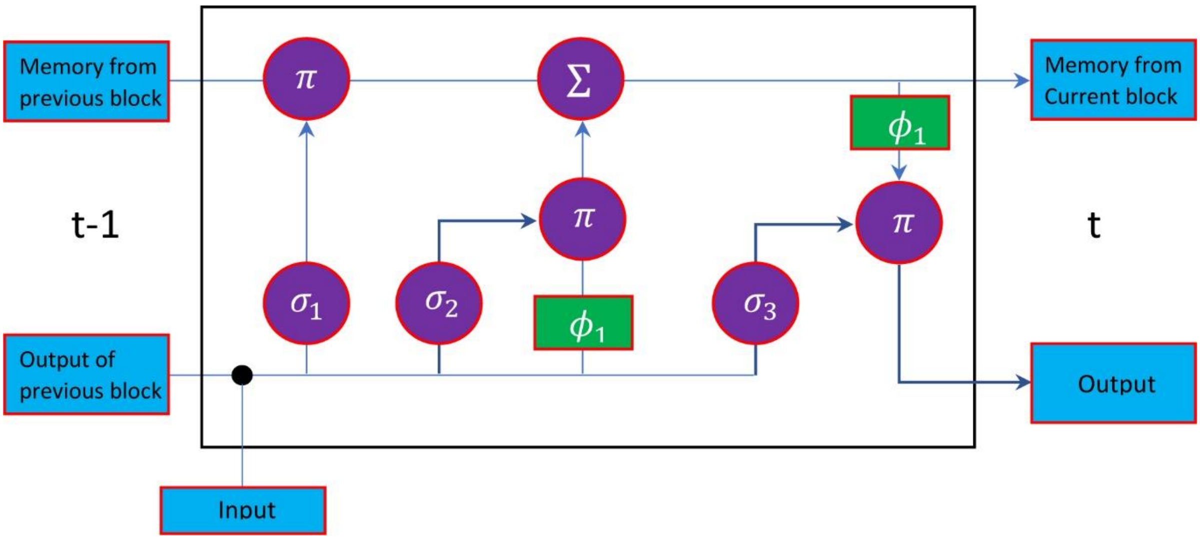


Fig. 1 Structure of LSTM network

effectiveness of different machine learning methods. The following are some of the key features of the LSTM model that motivate its use by the authors:

The present problem's data has a time dependency characteristic. To address this, we use recurrent neural network (RNN)-based models that can take into account temporal context. LSTM solves the problem of regulating the distribution and combination of inputs according to qualifying weights by adding knobs. In this way, LSTM gives us the most agency, and hence results that are the best. This, however, results in higher operational costs and complexity. Therefore, LSTM is chosen instead of simple RNN to obtain maximize accuracy. Evaluating the collected database is an important factor to identify the appropriate algorithm. Some artificial intelligence networks such as SVM and MLP can be performed by using small dataset, while others like RNNs can handle big database. While a number of algorithms including simple RNN, GRU, ANN, and LSTM prefer numeric input, others such as convolutional neural network (CNN) can handle categorical data.

3.1 Long-short term memory

Deep learning (DL) is a standard form of neural network with many layers. The knowledge included within DL networks is easier to remember than knowledge contained in simple neural networks. An RNN is a type of neural network that combines a loop network and a neural network. Across these loops, information persists in the networks. This network collects inputs and data from previous networks, executes specified procedures, and outputs the next network output. Only present knowledge may be required for certain applications, while extra information from the past may be needed for others. There is a learning delay in deep neural networks by which the gap between identified and required input is improved. However, LSTM networks are a special type of RNNs that are able to study these situations. Such networks are designed in such a way that they can reduce the problem of long-term dependence in permanent networks.

Recalling knowledge over a long period of time is a vital function of LSTM. This is a common choice because the accuracy of the model always depends on the sum of prior knowledge. The basic LSTM module, also known as the repeating module, interacts with four layers of neural networks, as presented in Fig. 1. In the module, there are three gates with activation functions of σ_1, σ_2, and σ_3, and it contains two output activation functions such as Φ1 and Φ2, as shown in Fig. 1. To depict addition and multiplication element-by-element, the Σ and π symbols are used. A bullet (•) symbol is representative of concatenating mathematical procedures. The cell state is the main component of LSTM, in which the current block memory (S_t-1) is produced from the prior block memory. Later in the process, the information flow is promptly approved. The first layer (S_t) in the network determines the sum of the prior knowledge to flow, and Eq. (1) shows the activity that is performed by this layer.

In the network, the two layers and the communication between them affect the storage of new data in the cell state. Eq. (2) shows a sigmoid layer (σ_2) that determines the I_t value to be updated, and Eq. (3) shows a Φ_1 tanh layer that generates a new candidate value ($\tilde{S}_t$). To create an item, both of them must be added with the state. The cell state is finally replaced with Eq. (4).

$$c_{ft} = \sigma_1(W_{cf} \cdot [O_{t-1}, x_t] + b_{cf}) \quad (1)$$

$$I_t = \sigma_2(WI \cdot [O_{t-1}, x_t] + bI) \quad (2)$$

$$\tilde{S}_t = tanh(WS \cdot [O_{t-1}, x_t] + bS) \quad (3)$$

$$S_t = (c_{ft} \times S_{t1}) + (I_t \times \tilde{S}_{t-1}) \quad (4)$$

In this paper, a new TWC dataset consisting of 550 samples is collected from several tunneling projects during the construction period of tunnels in different locations in Iran. Our investigation shows 12 effective input parameters

Table 1 Statistical analysis of the defined factors (input parameters).

	GL [m]	C [kPa]	φ [°]	E [MPa]	ν	UW [kN/m^3]	H [m]	W [m]	CA [m^2]	X [m]	TC [mm]
mean	41.625	21.527	23.645	52.592	0.2965	19.983	141.34	10.236	85.6213	24.6000	27.314
STD	62.081	8.7795	15.278	21.075	0.1069	5.6418	141.89	2.0742	31.0371	14.5955	33.659
min	0.0000	2.0000	2.0000	12.000	0.1000	5.0000	9.0000	4.5500	16.2515	0.0000	0.0000
25%	5.0000	14.000	12.000	37.000	0.2200	15.000	43.000	8.5400	57.2513	12.0000	4.4000
50%	13.000	22.000	22.000	46.000	0.3300	21.000	65.000	11.210	98.6463	24.5000	11.350
75%	45.000	26.000	32.000	66.000	0.3600	24.000	234.00	11.887	110.930	37.0000	42.000
max	345.00	55.000	79.000	127.00	0.4900	34.000	665.00	14.000	153.860	50.0000	178.40

Table 2 The frequency table of the nominal parameters of EM and CS.

Parameter	Type	Frequency
EM	TBM	29%
	EPB-TBM	24%
	NATM	28%
	DB	19%
CS	Circular	47%
	Horseshoe	53%

on TWC based on the experience gained from the different types of tunnel construction, available information on the considered tunnels, and the previous studies in this field. The input parameters are groundwater level (GL), cohesion (C), friction angle (ϕ), elasticity modulus (E), Poissons' ratio (ν), unit weight (UW), tunnel depth (H), tunnel width (W), tunnel cross-sectional area (CA), tunnel cross-sectional shape (CS), distance from tunnel face (X) and finally excavation method (EM). Therefore, defined dataset has 12 input parameters and only one TWC output factor.

4. Experimental results

This section describes the tests performed with the optimized algorithms and the obtained results.

4.1 Collected dataset

To ensure that the considered system is applicable to a variety of tunnel designs, two characteristics including the shape of the tunnel section and the excavation method are defined. For the cross-sectional shape, two circular and horseshoe modes, and four modes including tunnel boring machine (TBM), earth pressure balance machine (EPB-TBM), NATM, and drilling blasting method (DB) are considered for the excavation method. Applying such features within the database can increase the flexibility and accuracy of the DL forecasting models.

Table 1 presents a quick overview of collected and gathered dataset. In this table, the statistical analysis including the mean, standard deviation (STD), minimum, maximum, and first, second, and third quartiles of each parameter are reported. In addition, for nominal parameters, the frequency table is reported in Table 2.

The TWC can be considered in different directions. In this study, the horizontal distance between the right and left walls of the tunnel is considered. In Fig. 2, the correlation between input and output parameters is presented. Fig. 3 shows that the estimation of TWC by using linear and statistical methods has a lot of error and low accuracy. Therefore, it is necessary to use non-linear methods (deep learning) to find patterns between inputs and output (TWS).

4.2 Evaluation metrics

The system's performance is measured using K-fold cross-validation (K=5). In cross-validation the samples are randomly subdivided into K subsamples of equal size. One subsample (K-1) is utilized for training, while the remaining (K) samples are used for testing the validation of model. All of the experiment data is used for both training and validation subsets, which is an advantage over methods that repeatedly use random subsamples. A verification experiment is only ever used once for verification purposes.

Different loss functions, including the coefficient of determination (R^2) and the root mean square error (RMSE), are used to evaluate the accuracy of the suggested model.

5. Development of neural networks

5.1 Normalization method

One of the main functions of machine learning is data standardization (Mahmoodzadeh *et al*. 2022a, b). As a result of various sources of information, outputs may not always conform to the same standard sizes and measurement systems. As a result, normalizing data during training, can help models converge more quickly with lower errors. All

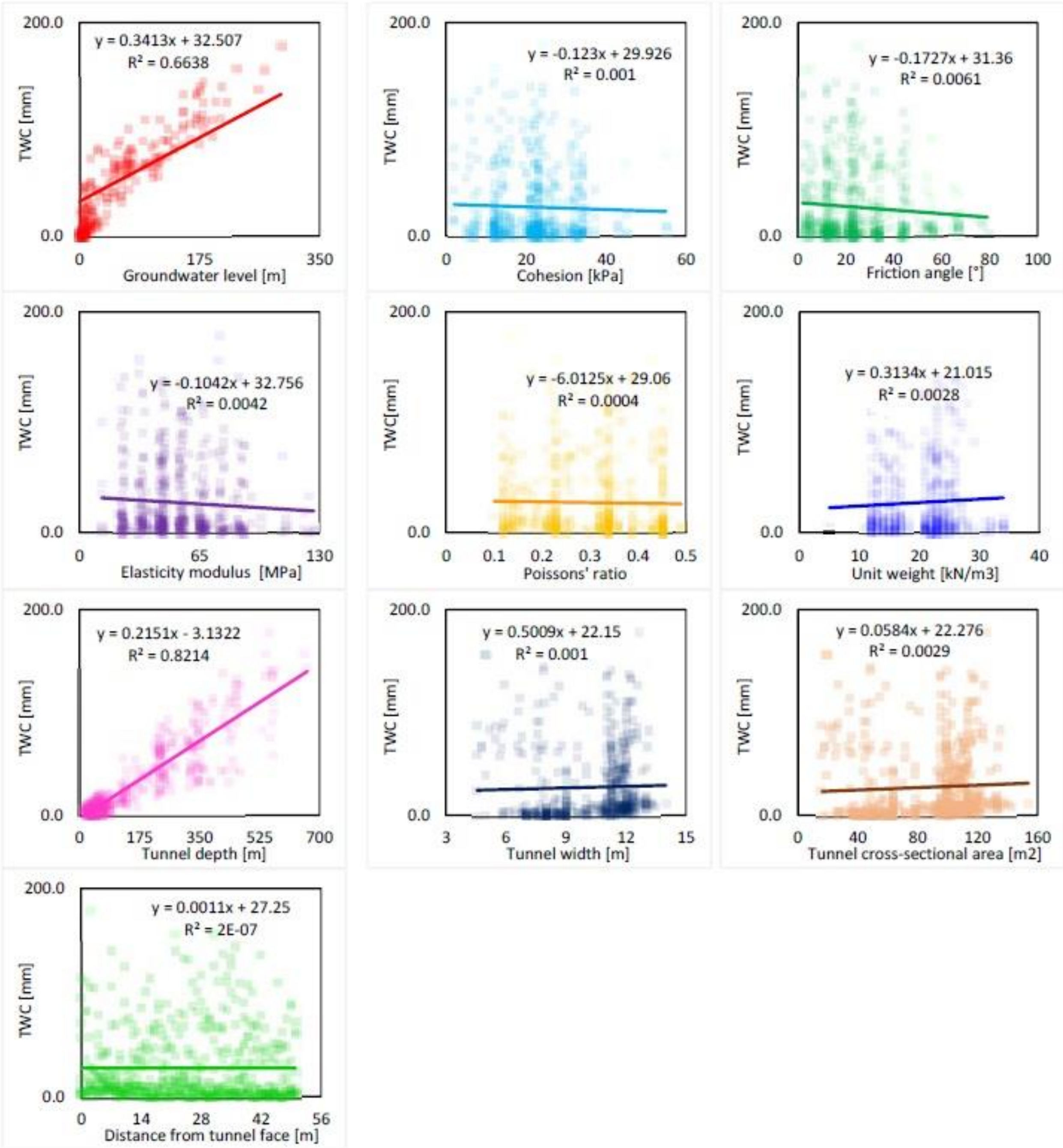


Fig. 2 Correlation between input parameters and TWC factor

data in this work are normalized to a range of zero to one using the Min-Max normalization method (see Eq. (5)).

$$x_{(0,1)} = \frac{x - Min}{Max - Min} \quad (5)$$

where x and $x_{(0,1)}$ denotes the original and normalized data respectively; Min and Max are represents the minimum and the maximum values of the entire dataset, including the training and testing data.

Min-Max normalization is a common and powerful technique in the research literature. If the range of values is expanded, this normalization approach may require redoing if the range of values is expanded after adding new data (Mahmoodzadeh *et al*. 2022a, b). It is suggested that the Max-Min range be expanded to cover a wider range of values than the extreme values of the original data in order to address the issue described above. In this manner, the new data will always fall inside this Max–Min range.

5.2 Model optimization

In order to coding the LSTM algorithm, the Tensor flow library in Python was used. This package offers a high-level programming interface for neural networks. The

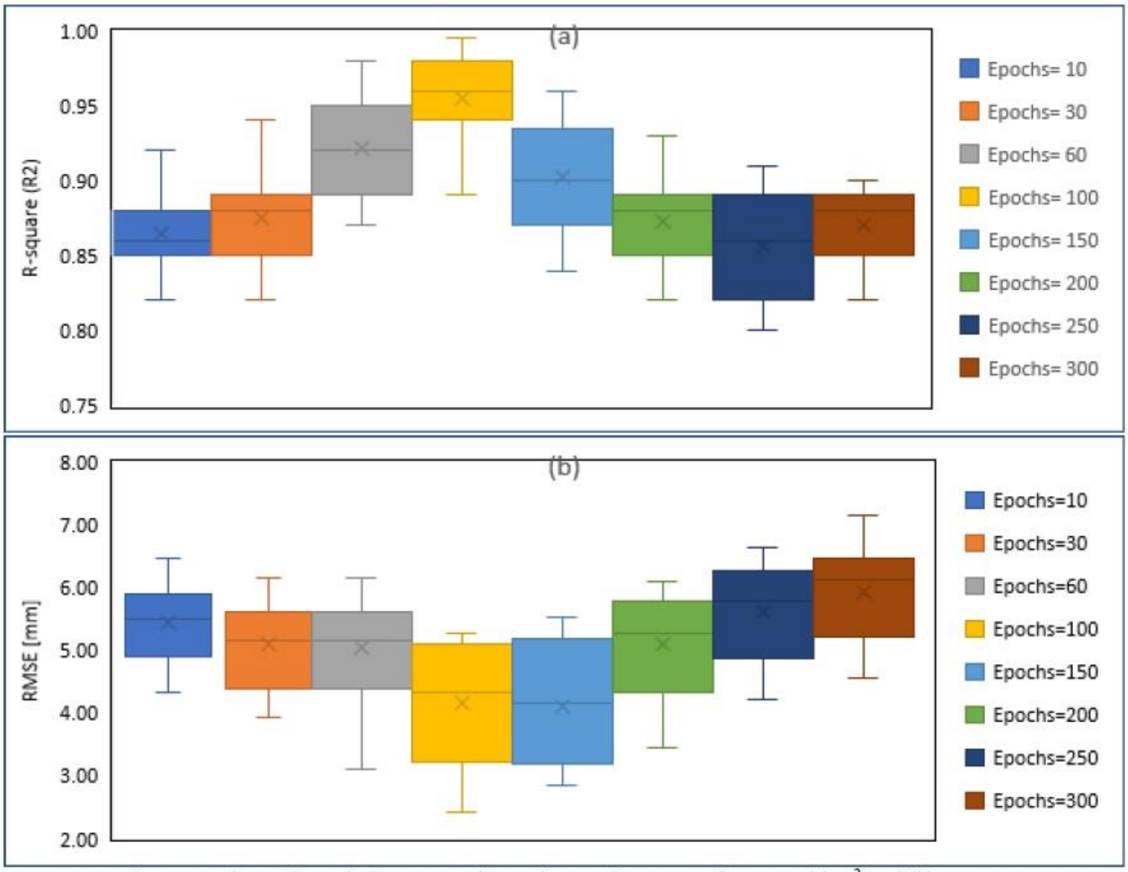


Fig. 3 Epoch number's influence on the testing performance of LSTM: (a) R^2 and (b) RMSE

computation was carried out on an Intel(R) Core (TM) i7-10750H CPU running at 2.60 GHz with 32 GB of RAM.

The establishment of hyperparameters is integral to optimizing training velocity and ensuring reliable predictions. The best settings for a given set of parameters may be found via parametric analyses. After some testing, acceptable ranges for variables likeepoch number [10, 30, 60, 100, 150, 200, 250, 300], hidden neurons [8, 16, 32, 64, 128], batch size [8, 16, 32, 64, 128], and dropout rate [0.1, 0.2, 0.3, 0.4, 0.5] are determined. It is possible to determine the best hyperparameters by analyzing the R^2 and RMSE values, which are used to evaluate the performance of a prediction model.

Due to the inherent unpredictability of the method, each iteration may provide a unique result. Therefore, the developed models are trained for 25 random tests and then two loss functions including R^2 and RMSE are apply during the testing period to ensure the accuracy of the prediction. The optimum values that were found are as follows: 64 hidden neurons, a 16-item batch size, a 0.3 percent attrition rate after 100 iterations, and ReLU activation function.

For illustration purposes, the procedure of choosing the optimum epoch number for the LSTM model is clarified in the following paragraph. As far as the number of epochs is concerned, epochs [10, 30, 60, 100, 150, 200, 250, 300] are chosen and compared. The box-plots of R^2 and RMSE for the testing period using the LSTM model with various training epochs are depicted in Fig. 3. At first, by increasing the number of epochs, R^2 rises while the RMSE falls. The accuracy starts to decline at this point, indicating an overfitting situation. The optimum number of epochs is 100. The Rectified Linear Unit (ReLU) $ReLu(x) = \begin{cases} x, & x > 0 \\ 0, & x \leq 0 \end{cases}$ activation function is adopted in the output layer. The Nesterov-accelerated adaptive moment estimation (Nadam) is applied as an optimization algorithm. When it comes to the activation functions in the output layer, varieties of commonly-used activation functions are analyzed, including linear, hard sigmoid, exponential, scaled exponential linear unit (SELU), sigmoid, tanh, ReLU. The outcomes of the testing period are summarized in Table 3. The largest R^2 of 0.96 and the smallest mean RMSE of 4.40 mm are yielded by The ReLU-based LSTM model. The ReLU-based LSTM model also outperforms almost all other activation functions in terms of mean time consumption (36.53 s). As a consequence, the ReLU-based LSTM model is very appropriate for this study.

6. Results and discussion

6.1 Comparison of the LSTM and other ML methods

The previous section investigates the performance of the neural network model and shows the usefulness of the network in predicting TWC. The results of the proposed

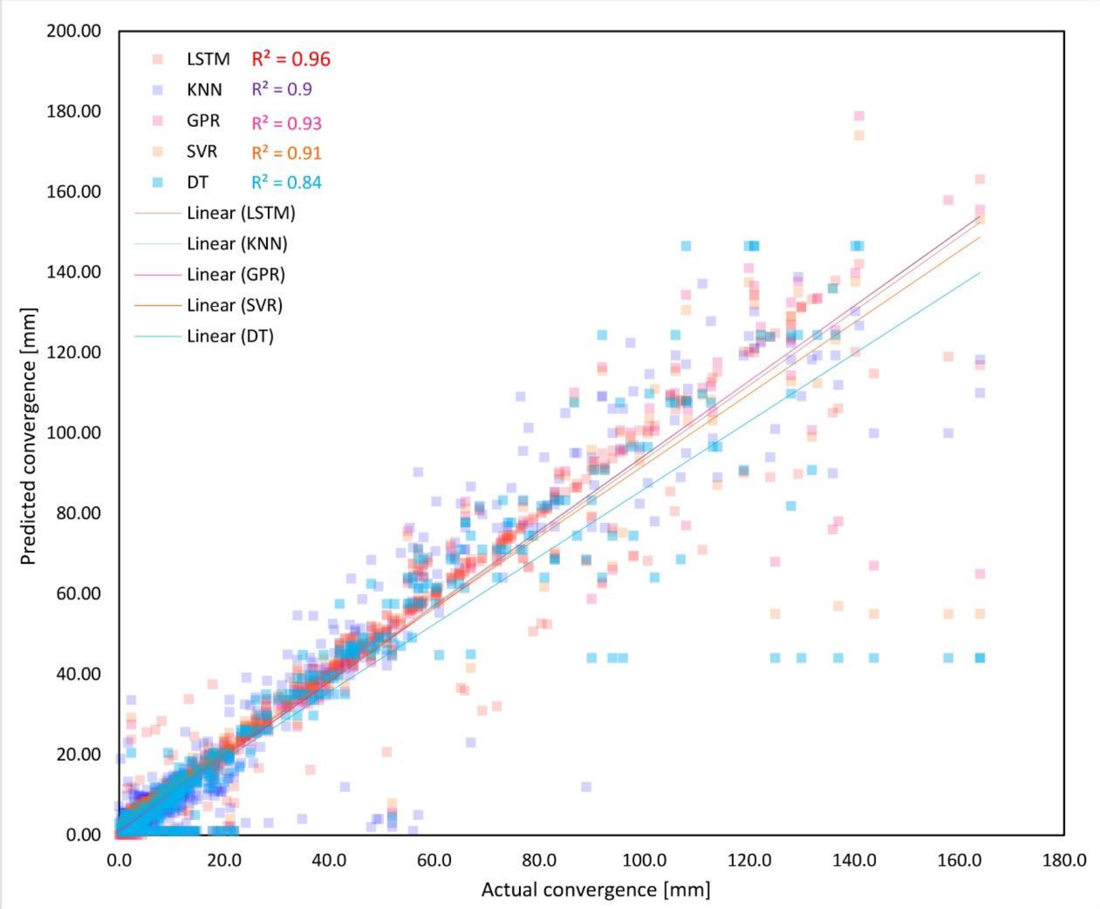


Fig. 4 Comparison of the ML models through their R^2 values

Table 4 Comparison of the LSTM model with other ML techniques

Statistical indices	LSTM	KNN	GPR	SVR	DT
R^2	0.96	0.90	0.93	0.91	0.84
RMSE	4.40	6.84	5.13	6.29	7.52

model should be compared with other ML methods to select the best and most flexible model. The four other ML models—GPR, SVR, DT, and KNN—all yielded the same outcomes as the optimized LSTM model. The models were fine-tuned by testing various hyper-parameters and architectures.

The best results from each models were used for the comparison. All five ML models' R^2 values are compared, as presented in Fig. 4. As can be seen the results of R^2 shows that the LSTM network is more accurate than the other models. As a second level of precision the GPR model has produced the high accuracy. Considering R^2=0.84, the DT model has minimum accuracy.

Table 4 presents the results of the statistical measures R^2 and RMSE, and it is clear that the improved LSTM model has the best accuracy. After the first approach, GPR is the second-most precise technique.

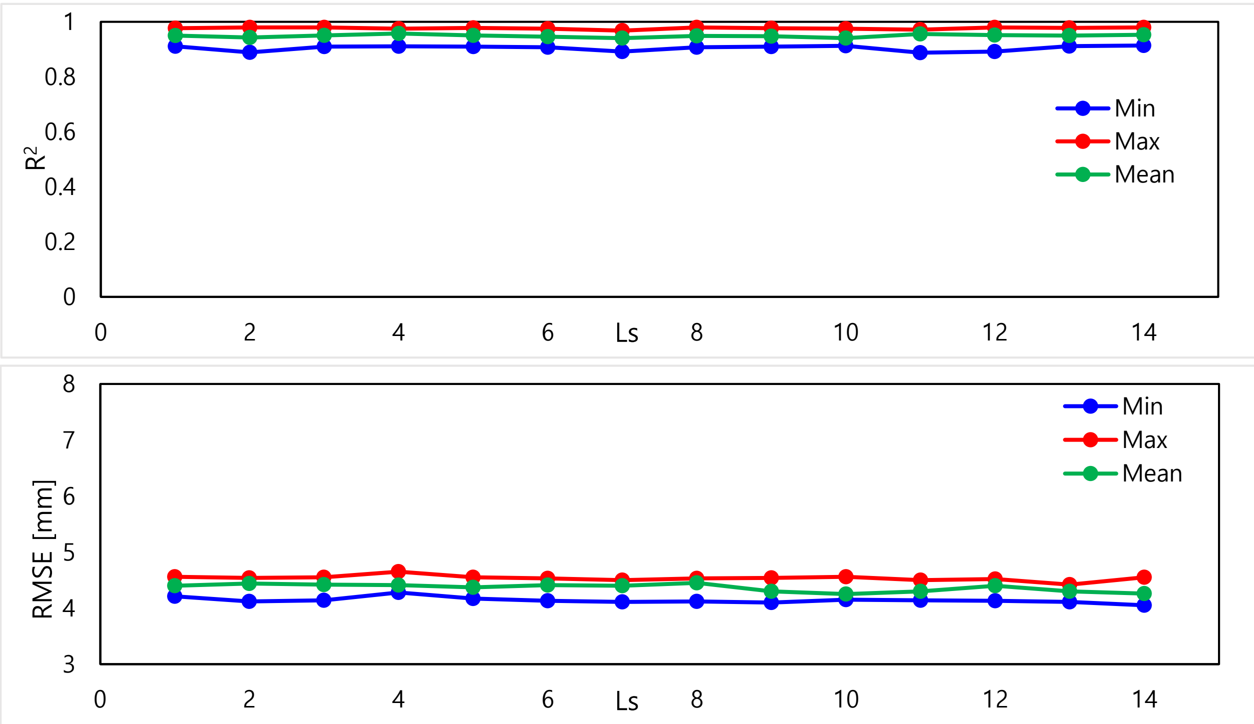
Thus, the LSTM model generated the most accurate TWCs compared to the other four ML models investigated in this research, given the datasets provided for this investigation. It is not important how the dataset is divided into training and testing sub-datasets, so that the other models not only has a good correlation, but also offer incredible accuracy.

Definitely, it's worth to mention that the prediction outcomes rely on kind of prediction model as well as the type and amount of factors examined in the database that all of the mentioned features play a vital role. Nevertheless, the approaches of this research show the highest accuracy of predictor networks and the appropriate selection of factors affecting TWC comparing with previous studies.

6.2 The influence of input time series length

The size of input time series data is determined by the length of the sliding window (denoted as L_s in this study). For example, if the frequency of recorded time series data is 1 meter and Ls = 10, the LSTM will be trained to predict the TWC at the next meter using data from every 10 meters.

The LSTM network is trained with various values of L_s so that the effect of input time series length (L_s) on the model performance could be studied. The R^2 and RMSE of models using different L_s for prediction are illustrated in Fig. 5. The maximum R^2 values for LSTM model prediction range from 0.97 to 0.98 (Fig. 5(a)) whereas the minimum RMSE values are between 4.05 mm and 4.28 mm (Fig.



(a)

(b)

Fig. 5 (a) R^2 and (b) RMSE values of the LSTM model with different lengths of input time series

5(b)). These results indicate that there is a satisfactory performance of the LSTM model. Therefore, with respect to these results, the LSTM model is resilient and its performance is independent of the input time series length.

6.3 Comparison of single-layer LSTM and multilayer LSTM

Multilayer models are typically explored to enhance prediction precision or address more complex issues. So far, the LSTM network has only used a single layer. The model is tested with two, three, and four hidden layers to determine how changing the number of hidden layers influences the results of prediction. Table 5 summarizes the outcomes of LSTM models with both single and multiple layers. Increasing the number of hidden layers from one to three yields better results in terms of loss functions such as R^2 and RMSE. The approaches obtained from the four-layer model are not more accurate and even the accuracy of the network is reduced to a small amount. It was decided to consider the three-layer model as a potential optimal solution for the LSTM. The comparison between the actual data and prediction results for both the single-layer and multi-layer LSTM models are displayed in Fig. 6. All predicted results are in line with the observed TWC records and have a good agreement with measured one. For this reason, the prediction cannot be improved upon beyond the three hidden layers. According to Table 5, the training times for single-layer, double-layer, three-layer, and four-layer models were 36.53 seconds, 71.38 seconds, 135.1 seconds, and 178.42 seconds, respectively. The amount of time required increases in proportion to the depth of obfuscation.

Table 5 Comparison of single-layer and multilayer LSTM models

Model		RMSE [mm]	R^2	CPU training time [s]
Single-layer LSTM	Mean	4.40	0.96	36.53
	Max	5.28	0.99	38.14
	Min	2.43	0.89	34.71
Double-layer LSTM	Mean	3.41	0.97	71.38
	Max	4.38	0.99	85.29
	Min	2.85	0.93	64.93
Three-layer LSTM	Mean	3.16	0.98	135.1
	Max	3.72	0.99	156.39
	Min	2.98	0.96	124.28
Four-layer LSTM	Mean	4.15	0.96	178.42
	Max	4.68	0.98	203.81
	Min	3.47	0.95	132.96

6.4 Effect of dropout technique

Overfitting is a risky issue in ANNs due to the vast number of input factors and the small amount of training dataset. Thus overfitting should be avoided if a reliable prediction is to be obtained. Prediction of TWC with ANNs is more difficult because of the risk of overfitting. The dropout method is an effective regularization technique for addressing the overfitting issue. The main purpose of dropout is to reduce the compatibility of neurons in networks and to prevent them from being unnecessarily dependent on certain neurons. For each training iteration,

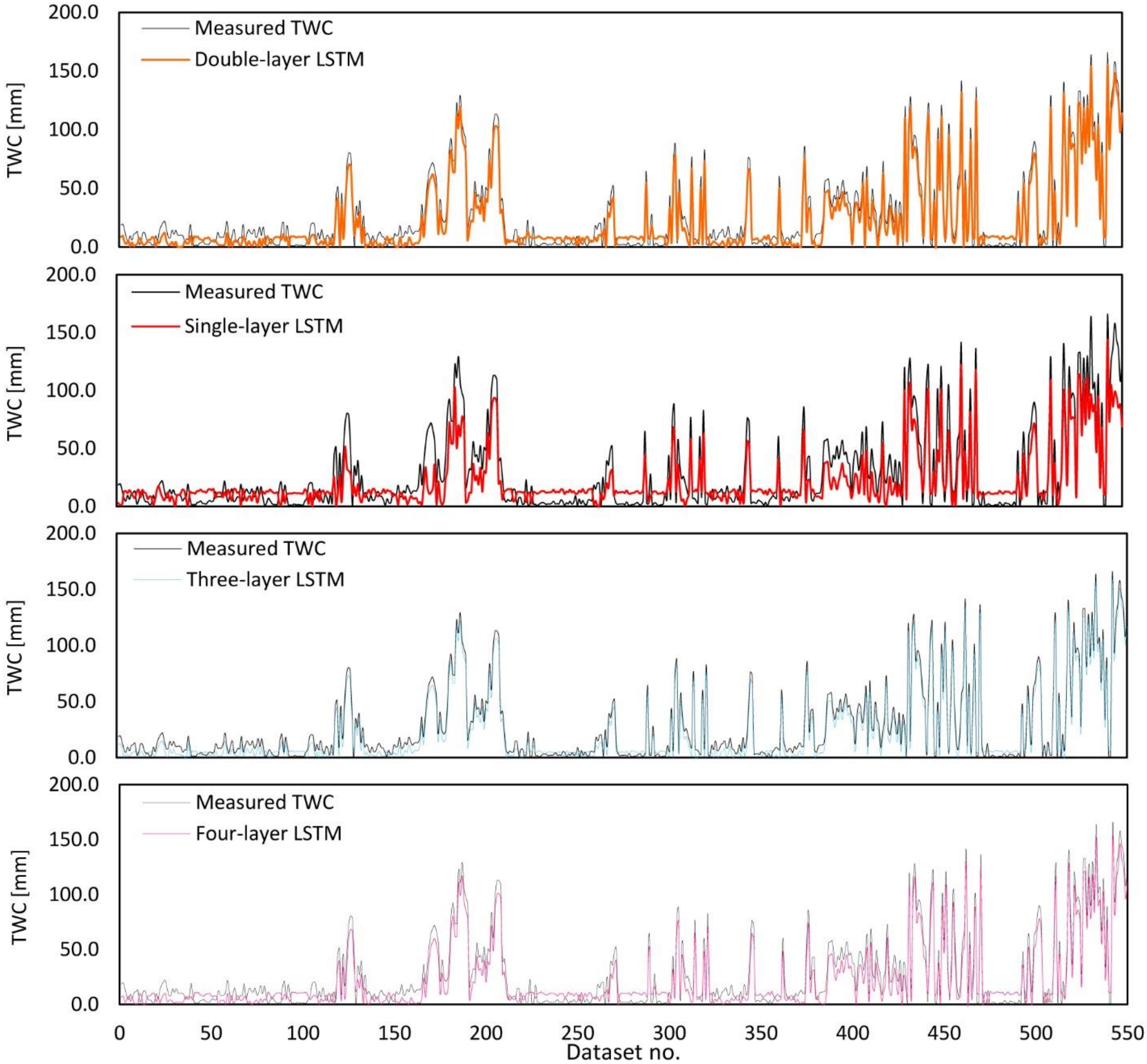


Fig. 6 Comparison of single-layer and multilayer LSTM models with actual TWC data.

the neurons are multiplied by a random variable with a Bernoulli distribution and probability p. The attrition rate agrees with (1-p). Fig. 7 shows the structural variation between models by considering dropout and without it. The following are the formulas:

Without dropout

$$\tilde{p}_t = w_\circ h_t + b_\circ \tag{6}$$

With dropout

$$\tilde{p}_t = w_\circ \tilde{h}_t + b_\circ = w_\circ (r_t \odot h_t) + b_\circ \tag{7}$$

where $\tilde{p}_t$ denotes the model output before it is processed by the active function at time t; h_t denotes the hidden layer's output vector as described in; The weight matrix and bias that connect the hidden and output layers are represented by $w_\circ$ and $b_\circ$, respectively. The output vector of the hidden layer following dropout is represented by $\tilde{h}_t$; r_t denotes a Bernoulli distribution-based random vector. The final model output is

$$p_t = f(\tilde{p}_t) \tag{8}$$

where p_t is the model output; $f()$ represents the output layer's activation function. The effect of the dropout strategy on the overfitting problem has been investigated using a three-layer LSTM model with and without dropouts. Considering all other constant parameters, the dropout rate is set at 0.5. The difference between the predictive models with and without dropout is illustrated in Fig. 8. As observed in the graph, the wave-like fluctuations occur with the lack of dropout. This should be noted that the fluctuation is a characteristic phenomenon of overfitting. Occasionally, the predicted values differ from the measured values. On the other hand, the LSTM model with dropout has a good correlation with the measured data and the R^2 and RMSE values are 0.99 and 2.25, respectively. Consequently, the dropout technique is an advantageous tool for mitigating overfitting.

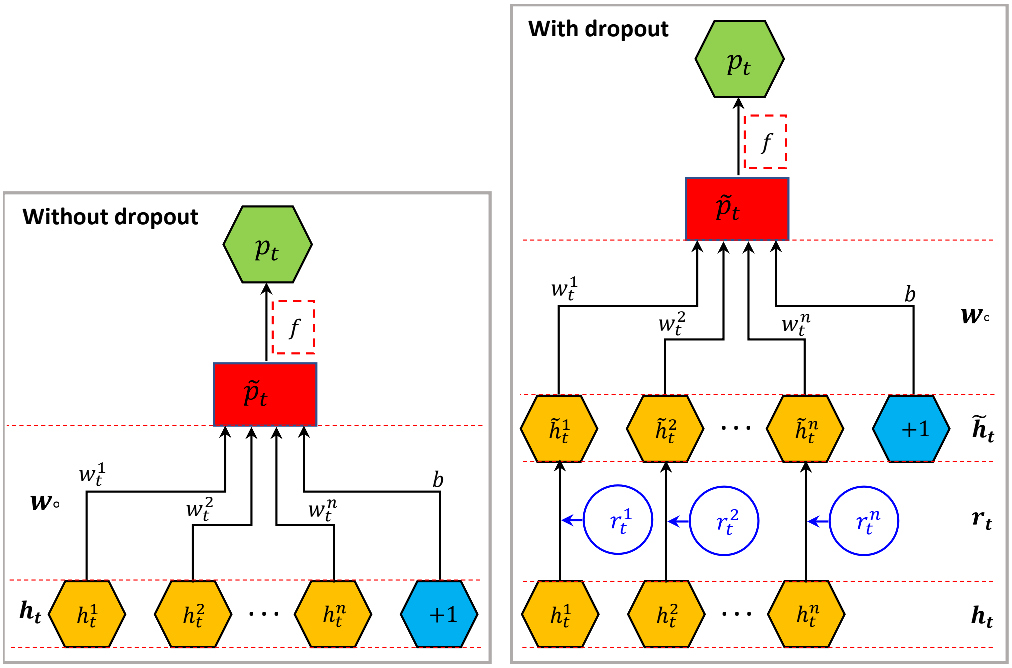


Fig. 7 Dropout technic for overfitting: (left) without dropout; (right) with dropout

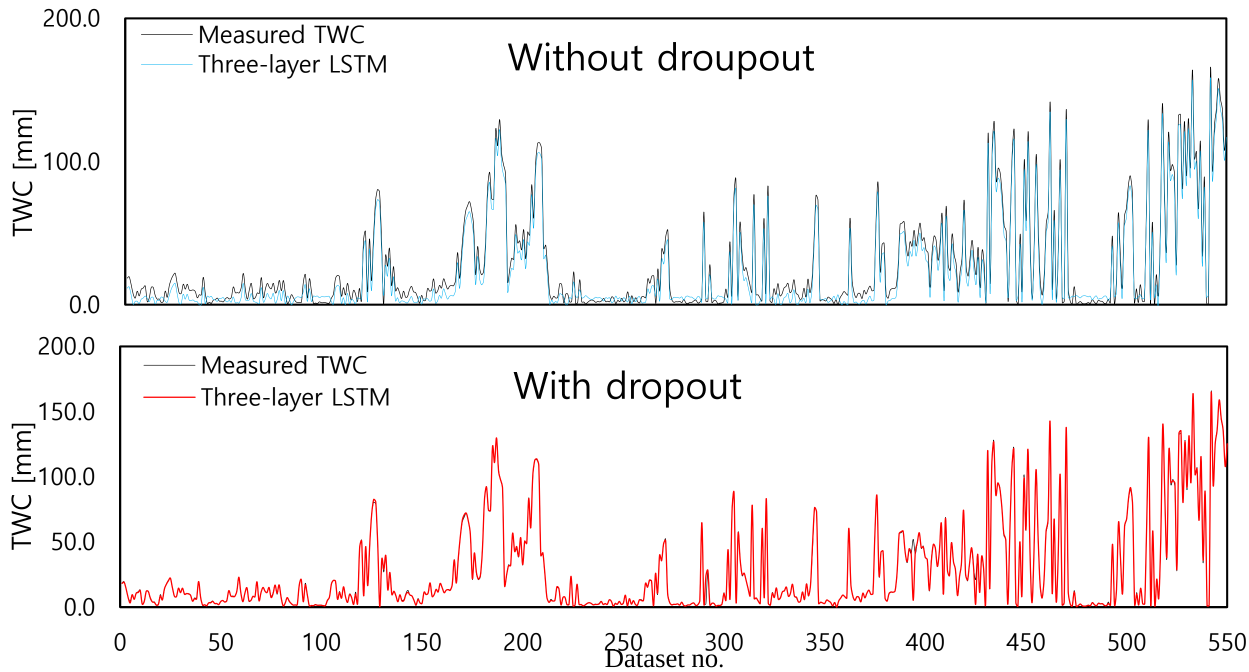


Fig. 8 Comparison of the prediction results for three-layer LSTM models with and without dropout

6.5 Limitations and future studies

Prediction of TWC is a difficult task because of the interconnected structure of influencing variables that are accompany with uncertainty. A big database can make a more accurate and powerful TWC predictor network. In order that the developed model to be used for TWC estimation in all kinds of tunneling projects, the determined parameters should be adjusted according to the quantity and complexity of the data.

7. Conclusions

The obtained results in this study led to the following conclusions:

- The LSTM model is superior to other RNN networks in identifying the statistical relationships between the defined inputs and the target parameter which is TWC.
- The LSTM model's output is sensitive to its hyperparameters, which must be tuned for specific datasets.

- The accuracy and precision of the developed model to predict the TWC was evaluated by using statistical indices such as R^2 and RMSE. The results demonstrated a good agreement between predicted and measured data, thus the developed network has a high potential to estimate TWC with high accuracy.
- Based on the defined database for this study, the LSTM model has better performance in the prediction of TWC than the other ML techniques such as GPR, SVR, KNN, and DT.
- Considering extra hidden layers in the structural network prolongs the training period based on the determined sub-dataset.
- The performance of developed network is affected by changing the number of hidden layers. In this research, three hidden layers were considered to obtain the optimum performance of the model.
- By using the dropout technic, overfitting can be avoided and accuracy can be improved.

Conflict of interest

There is no conflict of interest.

References

Adoko, A.C. and Wu, L. (2012), "Estimation of convergence of a high-speed railway tunnel in weak rocks using an adaptive neuro-fuzzy inference system (ANFIS) approach", *J. Rock Mech.. Geotech. Eng.*, **4**(1), 11-18. https://doi.org/https://doi.org/10.3724/SP.J.1235.2012.00011.

Adoko, A.C., Jiao, Y.Y., Wu, L., Wang, H. and Wang, Z.H. (2013), "Predicting tunnel convergence using multivariate adaptive regression spline and artificial neural network", *Tunn. Undergr. Sp. Tech*., **38**, 368-376. https://doi.org/https://doi.org/10.1016/j.tust.2013.07.023.

Asadollahpour, E., Rahmannejad, R., Asghari, A. and Abdollahipour, A. (2014), "Back analysis of closure parameters of Panet equation and Burger's model of Babolak water tunnel conveyance", *Int. J. Rock Mech. Min. Sci*., **68**, 159-166. https://doi.org/https://doi.org/10.1016/j.ijrmms.2014.02.017.

Bai, X.D., Cheng, W.C., Ong, D.E.L. and Ge, L. (2021), "Evaluation of geological conditions and clogging of tunneling using machine learning", *Geomech. Eng*., **25**(1), 59-73. https://doi.org/10.12989/gae.2021.25.1.059.

Debernardi, D. and Barla, G. (2009), "New viscoplastic model for design analysis of tunnels in squeezing conditions", *Rock Mech. Rock Eng*., **42**(2), 259. https://doi.org/10.1007/s00603-009-0174-6.

Fahimifar, A., Tehrani, F.M., Hedayat, A. and Vakilzadeh, A. (2010), "Analytical solution for the excavation of circular tunnels in a visco-elastic Burger's material under hydrostatic stress field", *Tunn. Undergr. Sp. Tech*., **25**(4), 297-304. https://doi.org/https://doi.org/10.1016/j.tust.2010.01.002.

Fakhri, D., Khodayari, A., Mahmoodzadeh, A., Hosseini, M., Ibrahim, H. and Mohammed, A. (2022), "Prediction of mixed-mode I and II effective fracture toughness of several types of concrete using the extrema gradient boosting method and metaheuristic optimization algorithms", *Eng. Fract. Mech*., In press, https://doi.org/10.1016/j.engfracmech.2022.108916.

Feng, X., Jimenez, R., Zeng, P. and Senent, S. (2019), "Prediction of time-dependent tunnel convergences using a Bayesian updating approach", *Tunn. Undergr. Sp. Tech*., **94**, 103118. https://doi.org/https://doi.org/10.1016/j.tust.2019.103118.

González el Á lamo, J.A. and Jiménez, R. (2011), "Prediction of convergences in rock tunnels excavated by conventional methods", *Proceedings of the 12th ISRM Congress*, Beijing, China.

Guan, Z., Jiang, Y. and Tanabashi, Y. (2009), "Rheological parameter estimation for the prediction of long-term deformations in conventional tunnelling", *Tunn. Undergr. Sp. Tech*., **24**(3), 250-259. https://doi.org/https://doi.org/10.1016/j.tust.2008.08.001.

Hajihassani, M., Abdullah, S.S., Asteris, P.G. and Jahed Armaghani, D. (2019), "A gene expression programming model for predicting tunnel convergence", *Appl. Sci*., **9**(21), 4650. https://doi.org/10.3390/app9214650.

Kontogianni, V., Psimoulis, P. and Stiros, S. (2006), "What is the contribution of time-dependent deformation in tunnel convergence?", *Eng. Geol*., **82**(4), 264-267. https://doi.org/https://doi.org/10.1016/j.enggeo.2005.11.001.

Liu, J., Jiang, Y., Zhang, Y. and Sakaguchi, O. (2021a), "Influence of different combinations of measurement while drilling parameters by artificial neural network on estimation of tunnel support patterns", *Geomech. Eng*., **25**(6), 439-454. https://doi.org/10.12989/gae.2021.25.6.439.

Liu, L.L., Yang, C. and Wang, X.M. (2021b), "Landslide susceptibility assessment using feature selection-based machine learning models", *Geomech. Eng*., **25**(1), 1-16. https://doi.org/10.12989/gae.2021.25.1.001.

Luat, N.V., Lee, K. and Thai, D.K. (2020), "Application of artificial neural networks in settlement prediction of shallow foundations on sandy soils", *Geomech. Eng*., **20**(5), https://doi.org/10.12989/gae.2020.20.5.385.

Mahdevari, S. and Torabi, S.R. (2012), "Prediction of tunnel convergence using artificial neural networks", *Tunn. Undergr. Sp. Tech*., **28**, 218-228. https://doi.org/https://doi.org/10.1016/j.tust.2011.11.002 .

Mahdevari, S., Shirzad Haghighat, H. and Torabi, S.R. (2013), "A dynamically approach based on SVM algorithm for prediction of tunnel convergence during excavation", *Tunn. Undergr. Sp. Tech*., **38**, 59-68. https://doi.org/https://doi.org/10.1016/j.tust.2013.05.002.

Mahdevari, S., Torabi, S.R. and Monjezi, M. (2012), "Application of artificial intelligence algorithms in predicting tunnel convergence to avoid TBM jamming phenomenon", *Int. J. Rock Mech. Min. Sci*., **55**, 33-44. https://doi.org/https://doi.org/10.1016/j.ijrmms.2012.06.005.

Mahmoodzadeh, A., Nejati, H.R., Mohammadi, M., Mohammed, A., Ibrahim, H. and Rashidi, S. (2022a), "Numerical and Machine learning modeling of hard rock failure induced by structural planes around deep tunnels", *Eng. Fract. Mech*., **271**, 108648. https://doi.org/10.1016/j.engfracmech.2022.108648.

Mahmoodzadeh, A., Rashidi, S., Mohammed, A., Hama Ali, H. and Ibrahim, H. (2022b), Machine learning approaches to enable resource forecasting process of road tunnels construction. Communication Engineering and Computer Science, North America, mar. 2022. Available at: <https://conferences.cihanuniversity.edu.iq/index.php/COCOS/22/paper/view/718>. Date accessed: 21 Sep. 2022. http://doi.org/10.24086/cocos2022/paper.718.

Mirzaeiabdolyousefi, M., Mahmoodzadeh, A., Ibrahim, H., Rashidi, S., Majeed, M. and Mohammed, A. (2022), "Prediction of squeezing phenomenon in tunneling projects: Application of Gaussian process regression", *Geomech. Eng*., **30**(1), 11-26. https://doi.org/10.12989/gae.2022.30.1.011.

Nadimi, S., Shahriar, K., Sharifzadeh, M. and Moarefvand, P. (2011), "Triaxial creep tests and back analysis of time-dependent behavior of Siah Bisheh cavern by 3-Dimensional Distinct Element Method", *Tunn. Undergr. Sp. Tech*., **26**(1), 155-162. https://doi.org/https://doi.org/10.1016/j.tust.2010.09.002.

Nomikos, P., Rahmannejad, R. and Sofianos, A. (2011), "Supported axisymmetric tunnels within linear viscoelastic burgers rocks", *Rock Mech. Rock Eng*., **44**(5), 553-564. https://doi.org/10.1007/s00603-011-0159-0.

Rafiai, H. and Moosavi, M. (2012), "An approximate ANN-based solution for convergence of lined circular tunnels in elasto-plastic rock masses with anisotropic stresses", *Tunn. Undergr. Sp. Tech*., **27**(1), 52-59. https://doi.org/https://doi.org/10.1016/j.tust.2011.06.008.

Sakurai, S. (1978), "Approximate time-dependent analysis of tunnel support structure considering progress of tunnel face", *Int. J. Numer. Anal. Method. Geomech*., **2**(2), 159-175. https://doi.org/10.1002/nag.1610020205.

Sharifzadeh, M., Tarifard, A. and Moridi, M.A. (2013), "Time-dependent behavior of tunnel lining in weak rock mass based on displacement back analysis method", *Tunn. Undergr. Sp. Tech*., **38**, 348-356. https://doi.org/https://doi.org/10.1016/j.tust.2013.07.014.

Sterpi, D. and Gioda, G. (2009), "Visco-plastic behaviour around advancing tunnels in squeezing rock", *Rock Mech. Rock Eng*., **42**(2), 319-339. https://doi.org/10.1007/s00603-007-0137-8.

Torabi-Kaveh, M. and Sarshari, B. (2020), "Predicting convergence rate of namaklan twin tunnels using machine learning methods", *Arabian J. Sci. Eng*., **45**(5), 3761-3780. https://doi.org/10.1007/s13369-019-04239-1.

Vu, T.M., Sulem, J., Subrin, D., Monin, N. and Lascols, J. (2013), "Anisotropic closure in squeezing rocks: The example of Saint-Martin-la-Porte access gallery", *Rock Mec. Rock Eng*., **46**(2), 231-246. https://doi.org/10.1007/s00603-012-0320-4.

Xiang, G., Yin, D., Cao, C. and Yuan, L. (2021), "Application of artificial neural network for prediction of flow ability of soft soil subjected to vibrations", *Geomech. Eng*., **25**(5), 395-403. https://doi.org/10.12989/gae.2021.25.5.395.

CC