Smart Structures and Systems, *Vol. 30, No. 6 (2022) 601-612*
https://doi.org/10.12989/sss.2022.30.6.601

DATCN: Deep Attention fused Temporal Convolution Network for the prediction of monitoring indicators in the tunnel

Bowen Du [1], Zhixin Zhang [1], Junchen Ye*[1], Xuyan Tan [2,3], Wentao Li [1] and Weizhong Chen [2,3]

[1] *SKLSDE and BDBC Lab, Beihang University, Beijing 100191, China*
[2] *State Key Laboratory of Geomechanics and Geotechnical Engineering, Institute of Rock and Soil Mechanics, Chinese Academy of Sciences, Wuhan 430071, China*
[3] *University of Chinese Academy of Sciences, Beijing 100049, China*

(*Received March 30, 2022, Revised August 30, 2022, Accepted September 18, 2022*)

Abstract. The prediction of structural mechanical behaviors is vital important to early perceive the abnormal conditions and avoid the occurrence of disasters. Especially for underground engineering, complex geological conditions make the structure more prone to disasters. Aiming at solving the problems existing in previous studies, such as incomplete consideration factors and can only predict the continuous performance, the deep attention fused temporal convolution network (DATCN) is proposed in this paper to predict the spatial mechanical behaviors of structure, which integrates both the temporal effect and spatial effect and realize the cross-time prediction. The temporal convolution network (TCN) and self-attention mechanism are employed to learn the temporal correlation of each monitoring point and the spatial correlation among different points, respectively. Then, the predicted result obtained from DATCN is compared with that obtained from some classical baselines, including SVR, LR, MLP, and RNNs. Also, the parameters involved in DATCN are discussed to optimize the prediction ability. The prediction result demonstrates that the proposed DATCN model outperforms the state-of-the-art baselines. The prediction accuracy of DATCN model after 24 hours reaches 90 percent. Also, the performance in last 14 hours plays a domain role to predict the short-term behaviors of the structure. As a study case, the proposed model is applied in an underwater shield tunnel to predict the stress variation of concrete segments in space.

Keywords: machine learning; mechanical behaviors; monitoring; prediction; tunnel

1. Introduction

The increasing growth of urban areas brings serious pressure on efficient transportation, which can be effectively alleviated by making full use of underground space as a public transportation (Shahrour *et al*. 2020, Mahmoodzadeh *et al*. 2020, Tan *et al*. 2020b). As an important project in underground transportation system, tunnel is often attacked by external load and complex environment, and sometimes is also affected by extreme events such as earthquake, collision and explosion. With the service life of the tunnel grows, more anomalies occur, resulting in cost overruns and even threatening public safety (Mahmoodzadeh *et al*. 2021, Tan *et al*. 2020a). Nowadays, structural health monitoring (SHM) has been widely recognized as a reliable technique to prevent disasters of tunnel structures (Wang and Ni 2020, Spencer *et al*. 2004). It is not only used to identify abnormal conditions, but also predict the future state of the structure. The prediction work for structural mechanical behavior is important to ensure the long-term stability of structures. Because if the protective measures are taken after the occurrence of abnormal, the damage has been formed and will affect the stability of the structures (Yu *et al*. 2021). Therefore, this study focused on the accurate prediction of structural future mechanical behaviors driven by the monitoring data and novel machine learning algorithm. The traditional research on the prediction of tunnel mechanical behavior mainly focuses on the analytical solution and the numerical solution (Fahimifar *et al*. 2010, Sharifzadeh *et al*. 2013). Due to the complexity of the solution process, the analytical and numerical results are difficult to be directly applied in engineering applications. In addition, the mechanical parameters in the actual engineering are quite different from those in the laboratory, so that the scale effect affects the calculation accuracy (Sterpi and Gioda 2009). In recent years, researchers have been interested in using data-driven models to predict the mechanical behavior of structures (Lin *et al*. 2017, Zheng *et al*. 2021, Prakash *et al*. 2018, Mei *et al*. 2016, Farahani and Penumadu 2016, Sajedi and Liang 2020). The data-driven models mainly include three categories, statistical models such as Bayes, time series models such as autoregressive (AR), and deep learning models developed on the basis of neural networks. The traditional researches focus on time series models, which predict future behavior base on historical behavior as an independent variable. Although these models are effective in many fields, they are not suitable for non-stationary time series. Also, time series models cannot incorporate newly acquired observations to update the models and provide advanced forecasts without reconstruction. In order to solve this problem, a large number of statistical models are

*Corresponding author, Ph.D.,
E-mail: yjchen@buaa.edu.cn

ISSN: 1738-1584 (Print), 1738-1991 (Online)

developed for prediction research, among which the most typical one is Bayesian model. Undesirable, Bayesian models cannot be applied to multivariate time series prediction problems, which makes the computational efficiency decrease when the amount of data is large. Recently, deep learning algorithms have attracted the attention of many scholars (Zhu *et al*. 2020, Mahdevari and Torabi 2012). Based on neural network, many algorithms have been developed, such as Convolutional Neural Network (CNN) and Recurrent Neural Network (RNN). Especially the RNN algorithm, the previous application cases showed that this algorithm has obvious advantages in processing time series (Cao *et al*. 2020, Chen *et al*. 2018, Carbonneau *et al*. 2008). CNN algorithm is widely used in processing digital images, but in many time series prediction problems, its prediction ability is more prominent than RNN algorithm (Xu *et al*. 2021, Yu *et al*. 2021, Hou and Qu 2021).

In summary, although a great deal of prediction works has been done by scholars, the prediction research on mechanical behavior of tunnel structure is seldom applied. The existing research is still in the implementation stage of machine learning algorithm, ignoring the features of monitoring data and the actual application conditions, which results in the decline of the prediction accuracy. Specifically, take underground engineering structure as a stress body, there is a certain spatial correlation between the monitoring data recorded by the measuring points at different positions, which cannot be ignored in the process of data analysis. In addition, in practical application, the current research can only predict the structural behavior in a continuous time interval, but not across time, which increases the calculation burden and low prediction accuracy when predicting the long-term mechanical behavior of the structure. Along this line, a deep attention fused temporal convolution network (DATCN) is developed in this paper to predict the spatial mechanical behaviors of structure, which integrates both the temporal effect and spatial effect and aimed to realize the cross-time prediction for underground structure. The structure of this study is organized as follows. Firstly, the pre-analysis for the feature of monitoring data obtained from a typical underwater shield tunnel is carried out. Then, the framework and methodology of DATCN is performed centering on the characteristics of monitoring data. Also, the data experiments are conducted to demonstrate the proposed DATCN model outperforms the other baselines. The involved parameters represented the observation time window and the prediction ability of different time scale are also discussed. As an important application, the presented model is used to predict the stress variation of concrete segments in space.

2. Background

2.1 Numerical simulation procedure

To demonstrate the application of presented DATCN model, a typical underwater shield tunnel is selected as

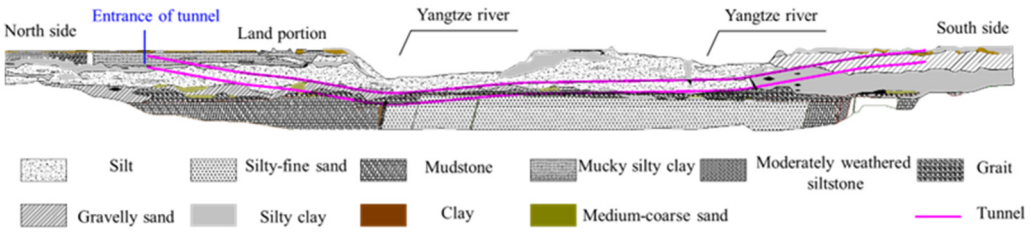


Fig. 1 Characteristics of stratigraphic distribution in site

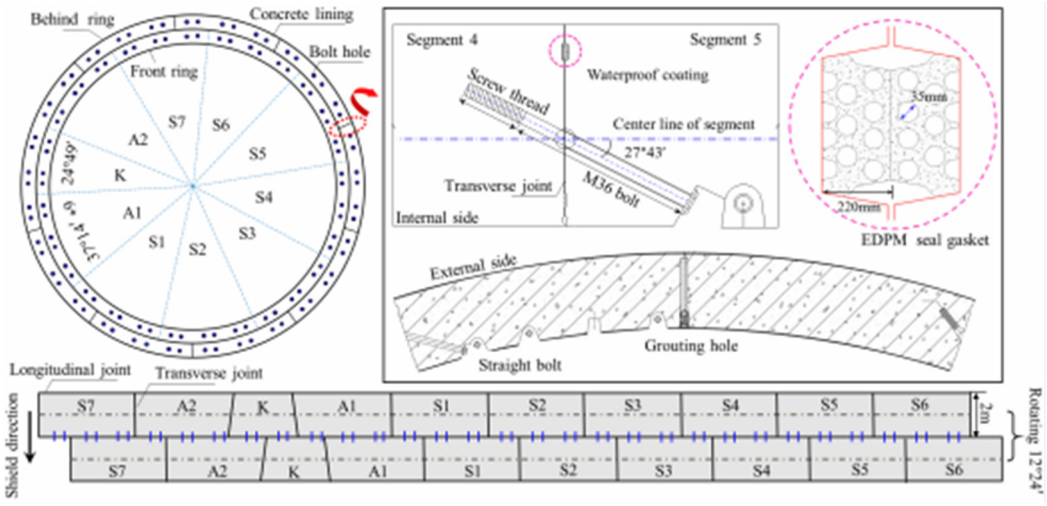


Fig. 2 Schematic diagram of tunnel cross section

Table 1 Dataset statics

Sensor_ID	Min	Max	Mean	Std
FSS_S01_B1_003	-15.137	-10.761	-13.039	1.089
FSS_S01_B1_004	-27.516	-17.032	-22.614	2.246
FSS_S01_B2_005	-22.199	-17.739	-20.15	1.105
FSS_S01_B2_006	7.302	15.747	11.389	1.887
FSS_S01_B3_007	253.718	259.827	256.293	1.415
FSS_S01_L1_002	-2.771	4.306	2.004	1.687
FSS_S01_L2_018	636.329	643.202	640.433	1.718
FSS_S02_B1_003	-18.609	-12.453	-15.648	1.545
FSS_S02_B1_004	-22.574	-14.965	-19.023	1.706
FSS_S02_B2_005	-16.312	-11.454	-13.739	1.126
FSS_S02_B2_006	-23.021	-15.097	-19.503	1.647
FSS_S02_L1_001	-17.135	-11.475	-14.521	1.33
FSS_S02_L1_002	-23.484	-14.019	-18.961	2.061
FSS_S03_B1_003	-7.388	-4.893	-6.151	0.465
FSS_S03_B6_013	-3.585	1.22	-1.466	1.408
FSS_S03_B7_015	4.408	9.358	6.611	1.402
FSS_S04_B6_013	-9.33	-4.929	-7.383	1.225
FSS_S04_B7_015	-5.508	-2.774	-4.347	0.714
FSS_S05_B1_003	-839.403	-836.41	-837.728	0.756
FSS_S05_B2_005	-859.417	-855.768	-856.996	0.864
FSS_S05_B4_009	-840.937	-837.842	-838.905	0.673
FSS_S05_B5_011	-841.877	-836.941	-839.285	1.171
FSS_S05_L2_017	-395.26	-391.506	-392.969	0.877
FSS_S06_B1_003	-180.023	-176.434	-177.675	0.933
FSS_S06_B2_005	-171.94	-168.099	-169.371	1.003
FSS_S06_B6_013	-179.007	-173.096	-175.652	1.588
FSS_S06_B7_015	-189.785	-183.11	-186.024	1.823
FSS_S06_L1_001	-231.285	-223.739	-226.236	2.08
FSS_S07_B1_003	-8.134	-5.03	-6.677	0.761
FSS_S07_B1_004	-7.496	-0.62	-4.616	1.57
FSS_S07_B2_005	-1.425	1.036	-0.414	0.509
FSS_S07_B2_006	-24.42	-16.573	-20.883	1.998
FSS_S07_B3_008	19.88	27.879	23.298	2.087
FSS_S07_B5_011	-4.483	-1.908	-3.141	0.548
FSS_S07_B5_012	-5.458	-0.368	-3.247	1.015
FSS_S07_L1_001	640.189	643.589	642.03	0.717

a case study: in which a SHMS has been employed for real-time monitoring.

The Nanjing Yangtze River tunnel is constructed in 2016 as a highway tunnel to link the traffic between two sides of the river. The highest water pressure applied on this project is about 720 KPa and multiple geological layers are crossed by tunnel section, such as silty-fine sand, clay, silty clay, medium-coarse sand, weathered siltstone, and pebbles, as shown in Fig. 1. The tunnel dimensions are 14.5 m in external diameter and 7014 m in length. The project is constructed using shield tunneling technology. The segment dimensions are 0.6 m in thickness and 2 m in width. Also, the tunnel cross section is displayed in Fig. 2. It denotes that each ring consists of ten segments, which are connected by bolts in both circumferentially and longitudinally.

The SHMS of Nanjing Yangtze River tunnel has been operation since July 2016. Fiber optical sensor is the primary instrument used for field monitoring. Based on the geological characteristics, ten monitoring sections in this project are determined to install sensors. Twenty stress sensors, four water pressure sensors, and two temperature sensors are pre-embedded in segments of every monitoring section. All monitoring data recorded by the sensors are transferred to central database and express to users. The

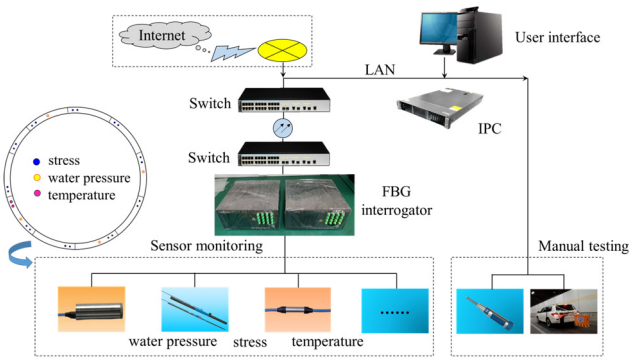


Fig. 3 Monitoring scheme in the study site

monitoring scheme and system networking are displayed in Fig. 3.

The stress sensors measure the resistance caused by the external loads on the structure. In this paper, we select the monitoring data from 36 stress sensors settled at different segments in the tunnel. The monitoring data is recorded every hour, a total of 1632 stress records were obtained from January 1st, 2020 to March 9th, 2020, each record contains the stress value of all 36 sensors. The statistics of the monitoring values of 36 sensors are shown in Table 1.

2.2 The necessity of spatial dependencies learning

Previous research usually trains a prediction model based on the monitoring data of each sensor, and predicts the trend of its future value. This approach not only wastes computing resources, but also ignores a fact that the monitoring data of different sensors may have similar data variation trends. These similarities can be used to predict the variation trend of multiple sensors simultaneously, and improve the accuracy of the prediction.

Table 2 Notations and corresponding comments

Notations	Comments
m	Total number of items recorded in the SHM system.
n	Number of sensors.
X	Raw monitoring data matrix.
p	Observation window.
g	The gap of prediction.
q	Prediction length.
$\mathcal{F}$	Prediction model to learn

Pearson Correlation Coefficient (PCC) is an important indicator which reflects the linear correlation among different variables. It is adopted to measure the similarities between the monitoring series of different sensors. The value of PCC varies in a range of -1 to 1, when PCC > 0, the trend of the two sensors is positively correlated, and when PCC < 0, it is negatively correlated, the larger absolute value means the stronger correlation between them. Given two monitoring series $\mathbf{x}$ and $\mathbf{y}$, the PCC value between them is calculated as

$$PCC(\mathbf{x},\mathbf{y}) = \frac{\sum_{i=1}^{n}(x_i-\overline{x})(y_i-\overline{y})}{\sqrt{\sum_{i=1}^{n}(x_i-\overline{x})^2\sum_{i=1}^{n}(y_i-\overline{y})^2}} \quad (1)$$

Calculate the PCC values of 36 sensors in the tunnel in pairs to obtain a PCC matrix and draw them on a heat map,

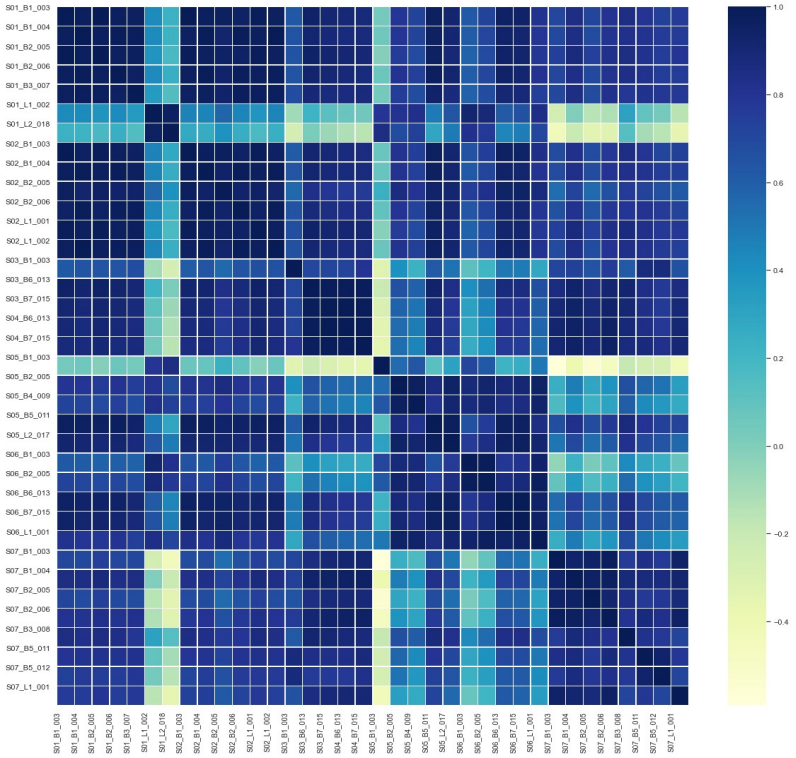


Fig. 4 The heat map of PCC values between different sensor pairs

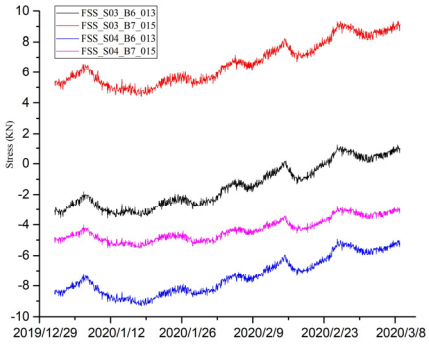


Fig. 5 Mesh grid of topographic model

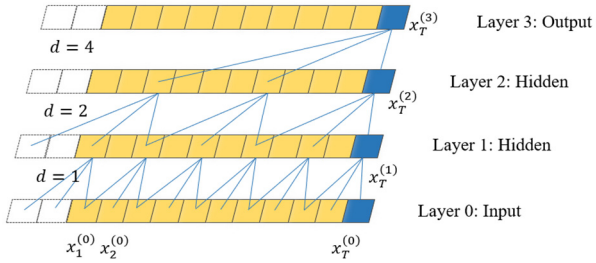


Fig. 6 Dilated convolution with dilation d = 1, 2, 4 and kernel size k = 3, zero padding is used to keep each hidden layer is the same length as the input layer

as Fig. 4 shows. The horizontal axis and vertical axis in the figure respectively represent 36 sensors. The color of the small square represents the PCC values between the monitoring data measured by the corresponding two sensors. The darker the color of the small square is, the higher the correlation is. Since the calculation of PCC value follows the commutative law, this matrix is symmetric, and the value on the diagonal is 1. It can be seen from the heat map that most of sensors have a PCC value larger than 0.8, some of them are larger than 0.95. The monitoring data of the 4 sensors with highest PCC value is plotted into a line graph, as Fig. 5 shows, although they are in different data ranges, they have very similar trends, so they can be used for simultaneous prediction.

2.3 Problem formalization

This section aims to formalize prediction problem of structural mechanical behaviours in the field of civil engineering. The structural health monitoring system collects a large amount of data generated from sensors during in-service process, the prediction task is to predict the value of the monitoring indicators in the future based on historical data. To describe the prediction problem, all notations which be used in this paper are listed in Table 2.

The total number of sensors is defined as n, and the total number of records collected from SHM system is defined as m, each record saves the values of n sensors at that moment, so the matrix of raw monitoring data is expressed as $X \in \mathbb{R}^{m \times n}$. Given a certain moment t, usually sampled every hour, the prediction task is to predict the value of monitoring data in all of sensors at next q hours after a prediction gap g hours, namely $X_{(t+g):(t+g+q-1)} = \{x_{t+g}, x_{t+g+1}, \ldots, x_{t+g+q-1}\}$. The gap g is set to allow the model to focus on learning the predicted interval q, reduce the amount of parameters, simplify the model and prevent the model from overfitting. The prediction uses the monitoring data in the past p hours before the moment t, where p is defined as observation window. The historical monitoring data from all of sensors in the range of observation window, constitutes a time series $X_{(t-p+1):t} = \{x_{t-p+1}, x_{t-p+2}, \ldots, x_t\}$.

Based on the historical monitoring data, the prediction task is to train a model $\mathcal{F}$ to output the predicted value in the future, as Eq. (2) shows

$$X_{(t-p+1):t} \overset{\mathcal{F}}{\rightarrow} \hat{X}_{(t+g):(t+g+q-1)} \tag{2}$$

where $\hat{X}_{(t+g):(t+g+q-1)}$ is the predicted value calculated by the model $\mathcal{F}$ in the period of $[t+g, t+g+q)$.

3. Methodology

In this section, we first describe two building blocks of our framework, the temporal convolution block and the spatial convolution block, they work together to capture the spatial-temporal dependencies. Then we introduce the architecture of our framework. Finally, methods for training and metrics for evaluating are introduced.

3.1 Temporal dependencies learning: Temporal convolution network

Unlike in RNNs where the predictions for later timesteps must wait for their predecessors to complete, convolutions can be done in parallel since the same filter is used in each layer, which makes CNNs have higher training efficiency. Temporal Convolution Network (TCN), is a new type of algorithm that can be used to solve time series forecasting problem, some experiments have proved that TCN has higher prediction accuracy than RNNs, and it avoids the problem of exploding/vanishing gradients.

Each individual monitoring series will be processed separately by a temporal dependencies learning module. For a single historical monitoring series $x \in \mathbb{R}^p$, we use dilated convolution to capture the temporal dependencies. When kernel size is set to k, dilation coefficient is set to d, the output at time t is

$$h_t = \omega \underset{d}{*} x_t = \sum_{i=0}^{k-1} \omega_i * x_{t-i*d} \tag{3}$$

where ω is the weight of the convolution kernel, it is shared among each time step t. The entire time series x is processed by Eq. (4), then the length will decrease from p to $p-d$. The output of the last layer serves as the input of the next layer. As is common when using dilated convolutions, we increase d exponentially with the depth of the network.

$$h = \omega \underset{d}{*} x + b \tag{4}$$

Gating mechanisms are critical in recurrent neural networks. They have been shown to be powerful to control information flow through layers for temporal convolution networks as well. For a single historical monitoring series \$x\$, two different dilated convolution process produce two hidden state representation h_a and h_b, the output of gated convolution is

$$h_{gate} = \sigma(h_a) \odot tanh(h_b) \tag{5}$$

where σ is the sigmoid function which determines the ratio of information passed to the next layer, $\sigma(x) = \frac{1}{1+e^{-z}}$, $\odot$ is the element-wise product. In order to solve the problem of vanishing gradients, we can add a residual connection module to each layer of TCN, as the Eq. (6) shows. This effectively allows layers to learn modifications to the identity mapping rather than the entire transformation, which has repeatedly been shown to benefit very deep networks. The output of temporal dependencies learning module $\widetilde{h}$ will be fed to the spatial dependencies learning module at the same layer.

$$\widetilde{h} = W_{res} * (h_{gate} + x_{(t-p+d+1):t}) + b_{res} \tag{6}$$

3.2 Spatial dependencies learning: Self-attention mechanism

Suppose $\widetilde{h}_1, \widetilde{h}_2, \ldots, \widetilde{h}_n$ are temporal dependencies learning module's outputs of all n sensors, each hidden state sequence has the same length T, these sequences make up a hidden state matrix $\widetilde{H} \in \mathbb{R}^{n \times T}$. In this paper, we use a kind of attention mechanism called Dot-Product Self-Attention to capture the spatial correlations among monitoring data of different sensors in the tunnel. $\widetilde{H}_{:,t} \in \mathbb{R}^n$ is the slice of the hidden state matrix $\widetilde{H}$ at time step t, it stores the hidden state information of each sensor at this moment. Firstly, the spatial dependencies learning module takes the hidden state matrix as input to compute key K, query Q and value V, as Eq. (7) shows, where W^k, W^q, and W^v are the weights of transformation matrix.

$$\begin{aligned} K &= W^k \widetilde{H}_{:,t} \\ Q &= W^q \widetilde{H}_{:,t} \\ V &= W^v \widetilde{H}_{:,t} \end{aligned} \tag{7}$$

Secondly, perform matrix multiplication on Q and K to obtain the relation matrix A, a_{ij} indicates the contribution of sensor-j's value to the sensor-i.

$$A = QK^T \tag{8}$$

Then, we use softmax function to obtain a normalized relation matrix $\widetilde{A}$.

$$\begin{aligned} \widetilde{A} &= softmax(\frac{A}{\sqrt{n}}) \\ \widetilde{a}_{ij} &= softmax(a_{ij}, a_i) \\ &= \frac{\exp(a_{ij}/\sqrt{n})}{\sum_k \exp(a_{ik}/\sqrt{n})} \end{aligned} \tag{9}$$

At the end, compute the latent spatial representation $\widetilde{X}_{:,t}$, as Eq. (10) shows.

$$\widetilde{X}_{:,t} = \sigma(\widetilde{A}V) \tag{10}$$

W^k, W^q, and W^v are shared among different time step t, so that we can capture integrated information from multiple monitoring series, and then obtain spatial dependencies that are irrelevant to time. Every latent spatial representation are concatenated to obtain a matrix $\widetilde{X}$, which is a spatio-temporal representation of the input monitoring series X.

3.3 DATCN: Deep Attention fused Temporal Convolution Network

Combining the characteristics of TCN and Self Attention, a fusion prediction framework called Deep Attention fused Temporal Convolution Network (DATCN) is proposed in this section to predict the future monitoring indicators, the model architecture of DATCN is shown in Fig. 7.

The input monitoring series are extracted from the raw SHM data, then combine them into an input matrix $x^{(1)} \in \mathbb{R}^{p^{(0)} \times n}$, where $p^{(0)} = p$ is observation window. For layer $l \in \{1,2,\ldots,L\}$, $x^{(l)}$ is fed into the l-th Spatial-Temporal layer, and then obtain a latent temporal representation $\widetilde{h}^{(l)} \in \mathbb{R}^{p^{(l)} \times n}$ after a temporal dependencies learning module. Then, $\widetilde{h}^{(l)}$ will be used as input for spatial dependencies learning module at the same layer, obtain a latent spatial-temporal representation $\tilde{x}^{(l)}$, which will be fed into $l+1$ Spatial-Temporal layer, in other words, $x^{(l+1)} = \tilde{x}^{(l)}$. The dilation coefficient of the next layer is twice that of the previous layer, and its initial value of first layer is set to 2.

Spatial-Temporal layer is stacked L times, the length of $\tilde{x}^{(L)}$ varies as

$$p^{(L)} = p^{(0)} - \sum_{l=1}^{L} 2^l = p - 2^{L+1} + 2 \tag{11}$$

To achieve the predefined prediction length, a fully connected layer is set as Eq. (12) shows, where w^{fc} and b^{fc} are the parameters of fully connected layer, $\hat{y}$ is prediction result of the whole prediction model.

$$\hat{y} = W^{fc} \tilde{x}^{(L)} + b^{fc} \tag{12}$$

3.3 Model learning and evaluation

After the predicted value $\hat{y}$ is calculated by DATCN model, one loss function $\mathcal{L}$ called Mean Square Error (MSE), as Eq. (13) shows, is adopted to measure the distance between $\hat{y}$ and the ground truth y and guide the update process of model parameters. We train our model using Adam optimizer with an initial learning rate of 0.03, the training process can be complete in few minutes.

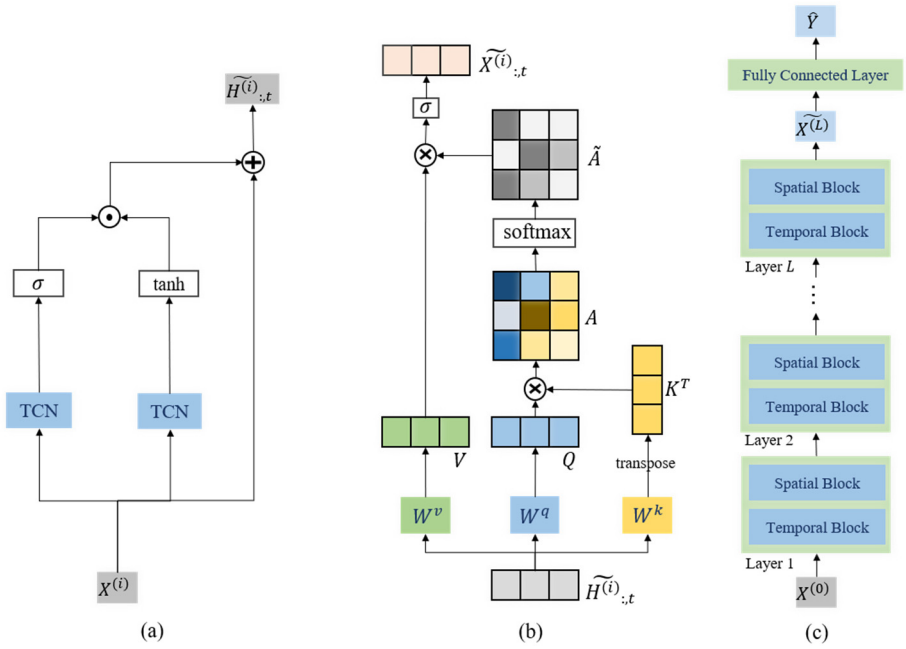


Fig. 7 The architecture of proposed DATCN model. (a) Temporal Block; (b) Spatial Block; (c) The overall architecture and data flow of DATCN

$$MSE(y,\hat{y}) = \frac{1}{m}\sum_{i=1}^{m}(y_i - \hat{y}_i)^2 \tag{13}$$

We use Root Mean Squared Errors (RMSE), Mean Absolute Percentage Errors (MAPE), and Pearson Correlation Coefficient (PCC) to measure the performances. RMSE is a widely used metric to reflect the distance between the predicted results and the ground truth, its value has a same magnitude with original monitoring series. Given ground truth $y = \{y_1, y_2, \ldots, y_n\}$, and its predicted values $\hat{y} = \{\hat{y}_1, \hat{y}_2, \ldots, \hat{y}_n\}$, RMSE is calculated as

$$RMSE(y,\hat{y}) = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2} \tag{14}$$

Since the values of monitoring series from different sensors may be in different magnitudes, we cannot measure the prediction accuracy among different sensors only using RMSE. MAPE measures the ratio of absolute error, excludes the influence of magnitude, so it can be used to evaluate the model's prediction accuracy among different sensors. MAPE is calculated as

$$MAPE(y,\hat{y}) = \frac{100\%}{n}\sum_{i=1}^{n}\left|\frac{y_i - \hat{y}_i}{y_i}\right| \tag{15}$$

Different to RMSE and MAPE, PCC is adopted to measure the similarity of variation trend between predicted results and target ground truth. A larger PCC means a stronger correlation. For MSE, RMSE and MAPE, lower values are better. For PCC, higher values are better.

4. Prediction of stress response of Nanjing Yangtze River tunnel

4.1 Preprocession of raw SHM data

The raw SHM data cannot be fed into DATCN directly, some preprocession operations are needed to make the raw data match the input format of the model. As mentioned in the Section 2.3, the raw SHM dataset X contains m records of n stress sensors, each record stores the average stress value of each sensor within 1 hour. For each hour t, the input series $\mathcal{X}_t = X_{t-p+1:t}$ and target series $\mathcal{Y}_t = X_{t+g:t+g+q-1}$ are produced according to the pre-chosen value of p, g and q. Then, concatenate all $\mathcal{X}_t$ and $\mathcal{Y}_t$ together into input matrix $\mathcal{X} \in R^{m\prime\times p\times n}$ and target matrix $\mathcal{Y} \in R^{m\prime\times q\times n}$, where m' is the number of valid records that can be used to produce series.

The monitoring data in different sensors might in different magnitudes, to avoid the adverse impact during model training, the min-max normalization will be applied. We use $\mathcal{X}_{:,:,i}$ to represent the data from sensor i, it will be transformed into $\mathcal{X}^{normal}_{:,:,i}$ as Eq. (16) shows. $\mathcal{X}^{normal}$ and $\mathcal{Y}^{normal}$ represent the normalized matrix after the min-max normalization.

$$\begin{aligned} x_{min} &= \min(\mathcal{X}_{:,:,i}) \\ x_{max} &= \max(\mathcal{X}_{:,:,i}) \\ \mathcal{X}^{normal}_{:,:,i} &= \frac{\mathcal{X}_{:,:,i} - x_{min}}{x_{max} - x_{min}} \end{aligned} \tag{16}$$

We split $\mathcal{X}^{normal}$ and $\mathcal{Y}^{normal}$ into a training set (70%) and test set (30%) in chronological order. The training set will be used to update the learnable parameters of the model, then test set will be fed into the trained model

Table 3 Experiment results of different prediction models

Model	RMSE↓	MAPE↓	PCC↑
LR	0.4924	5.0779	0.7402
SVR	0.4174	4.6636	0.7714
MLP	0.4915	4.6083	0.6851
RNN-Seq2Seq	0.4041	4.0424	0.7407
GRU-Seq2Seq	0.3988	4.0571	0.7591
LSTM-Seq2Seq	0.4155	4.0491	0.7151
DATCN	0.2830	2.5675	0.8523

to evaluate its prediction ability.

4.2 Performance of the proposed model

Some experiments are conducted to demonstrate the predictive ability of the DATCN model which proposed in this paper. The observation window p is selected as 24 hours, the gap of prediction g is selected as 10 hours, and the prediction length q is 6 hours. Some classical series prediction models commonly used in the field of structural health monitoring are used for comparative experiments, including (1) LR: Linear Regression assumes the change of stress with time follows a linear relationship; (2) SVR: Support Vector Regression is a branch of SVM; (3) MLP: Multiple Perceptron is a feedforward neural network, which can contain multiple hidden layers in addition to input layer and output layer; Some RNNs-based model are also used as baselines, which are trained and tested in a Seq2Seq form, including (4) RNN: Recurrent Neural Network takes sequence data as input, carries out recursion in the evolution direction of sequence and all nodes are linked by chain; (5) LSTM: Long Short-Term Memory is suitable for predicting important events with very long intervals and delays in time series; (6) GRU: Gated Recurrent Unit can solve the long term memory dependence of series prediction problems and the gradient problem in back propagation. In the three RNNs-based model, we set the hidden size as 20, 15 and 15 respectively, the learning rate is 0.003. In DATCN, the learning rate is 0.03. The epoch for the training phase is set to 100. The experiment results are obtained by averaging the prediction metrics of 36 stress sensors, listed in Table 3, it is obvious that DATCN has best performance among all listed prediction models. DATCN has much lower absolute and relative error than the other prediction models, which proves that it has very high prediction accuracy. Also, the high values of PCC indicates that the correlation coefficient between the predict value and the real value is strong, which means DATCN can capture the variation trend of stress more precisely.

4.3 Analysis of the proposed model DATCN

To figure out the spatial dependencies of monitoring data among different sensors, this paper uses Self-Attention mechanism to calculate the relation matrix $\tilde{A}$. $\tilde{a}_{ij}$ indicates the contribution of sensor-j to the value of sensor-i. For each sensor i, its relation weights with all n sensors

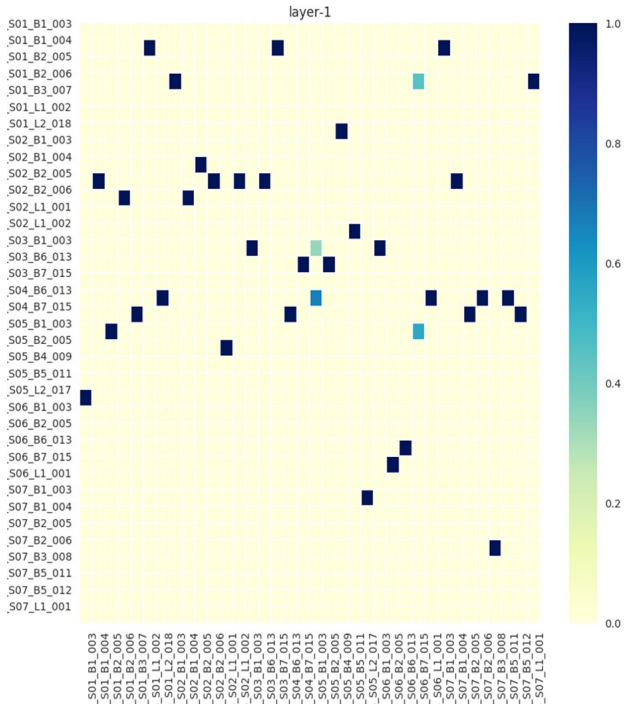


Fig. 8 The relation matrix among different sensors

add up to 1, in other words, $\sum_j \tilde{a}_{ij} = 1$. In order to explore the direct spatial dependence of model input and output, we use a one-layer model to do experiment. The observation window \$p\$ is set to 6, the gap of prediction g is set to 10, and prediction length q is set to 6, the relation matrix is as Fig. 8 shows. Softmax function can amplify the scores of positions with larger relation weights, making the obtained relation matrix very sparse. It can be seen from Fig. 8 that there are 4 very important sensors, S01-B1-004, S02-B2-005, S04-B6-013 and S04-B7-015.

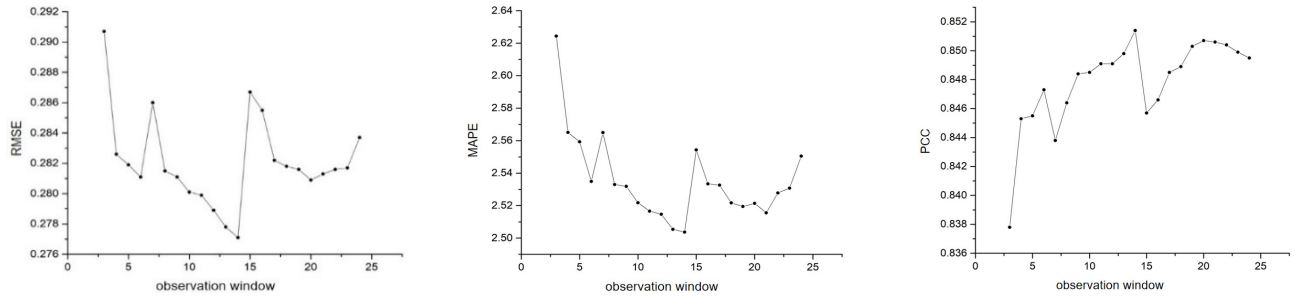
5. Parameters analysis

5.1 Effect of observation window on prediction capability

The length of observation window is an important parameter for the prediction problems, therefore it is necessary to explore the influence of observation window on stress prediction effectiveness.

The length of observation window p is set from 3 to 24. For each $p \in [3,24]$, experiments are conducted on the proposed DATCN model. The experiments are repeated multiple times to eliminate the influence of randomness on the experiment results. Three evaluation indicators RMSE, MAPE and PCC were recorded and drawn in a line diagram to reflect how does the predictive power of DATCN change varies with the increase of observation window, as Fig. 9 shows.

According to the Eq. (11), to make sure the output length of all Spatial-Temporal layers $p^{(L)}$ is positive, there is a constraint between the setting of parameter p and the depth of DATCN model L



(a) Variation trend of RMSE value (b) Variation trend of MAPE value (c) Variation trend of PCC values

Fig. 9 Evaluation results under different observation windows

$$L \leq \lfloor \log_2(p+1) \rfloor - 1 \quad (17)$$

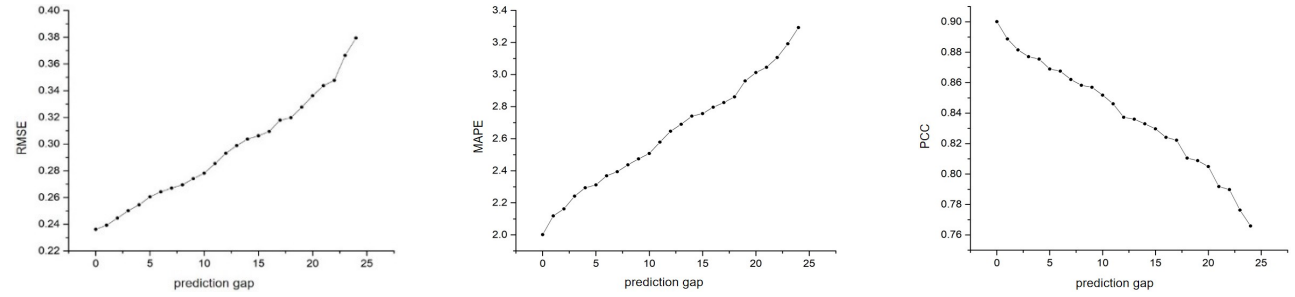
Therefore, when $p \in [3,6]$, set model depth L as 1, when $p \in [7,14]$, set model depth L as 2, when $p \in [15,24]$, set model depth L as 3. The plot shows that, for all three indicators, the curve's trend turns when the model depth changes. At interval [3,6] and [7,14], RMSE and MAPE decrease with the rise of observation window, but when $p = 7$, the prediction error increases suddenly, it's because the output length of Spatial-Temporal layers is shortened to 1, there is not enough latent spatio-temporal information in the output matrix. When $15 \leq p \leq 24$, the prediction error decrease in the first half, reaches the local minimum point at around 20, then gradually increase in the second half. The results of RMSE and MAPE values indicate that there is a global minimum point when $p = 14$. The variation trend of PCC is opposite to that of RMSE and MAPE. P CC value increase with the rise of observation window at interval [3,6] and [7,14], when $p = 7$, correlation coefficient between predicted value and true value decreases suddenly for the same reason as RMSE and MAPE vary. When $15 \leq p \leq 24$, the correlation coefficient between predicted values and ground truth increase in the interval [15,20], reaches the local maximum point when $p = 20$, then gradually decrease in the interval [20,24]. Therefore, it is inferred from Fig. 9 that the proposed prediction model DATCN achieves best performance when observation window is set as 14. Under this condition, the model has lowest prediction error and highest accuracy to capture the variation information of monitoring data. One possible explanation is that the monitoring data has a potential daily period, so that when the observation window p and the gap of prediction g interval set by the experiment add up to exactly 24 hours, the model can achieve the best performance.

5.2 Prediction ability under different predict gaps

After getting the most optimal observation window, a series of experiments are carried out to explore the influence of different prediction gaps on the DATCN's prediction ability. The prediction gap g is set from 0 to 24, as in Section 5.1, three evaluation indicators RMSE, MAPE and PCC are recorded and drawn in a line diagram to reflect the characteristics of prediction error and prediction relevance varies with the increase of prediction gap, as Fig. 10 shows. It is obvious that the prediction error increase as g increases, the similarity of variation between predicted results and real value decrease as g increases, so the prediction ability becomes weaker as the prediction gap increases. This is because correlation between future value of stress and its current value becomes lower when the prediction gap goes larger. When $g \leq 10$, the PCC value is greater than 0.85, the model prediction accuracy is relatively high, the value of RMSE and MAPE are also in a acceptable range, so the proposed model is suitable for stress prediction when prediction gap is smaller than 10.

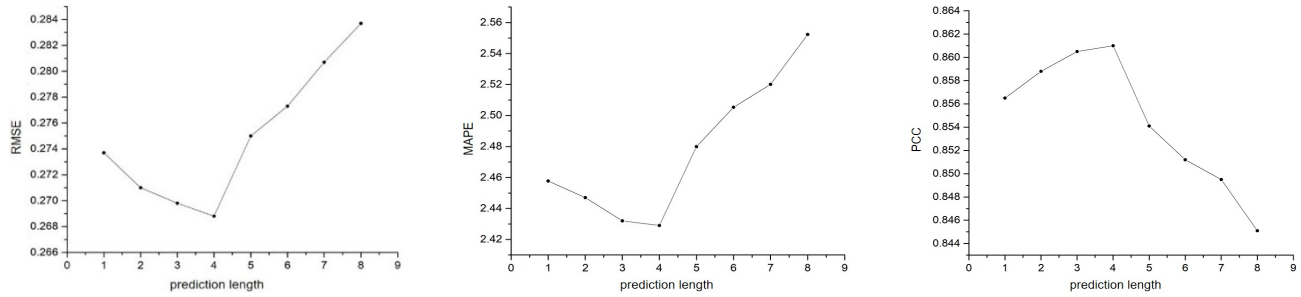
5.3 Prediction ability under different predict lengths

After the optimal observation window and prediction gap in the DATCN model are fixed, we conduct a series of experiments to verify the performance of the model under different prediction lengths. Set $p = 14$ and $g = 10$, experiments were carried out under $q \in [1,8]$, three



(a) Variation trend of RMSE value (b) Variation trend of MAPE value (c) Variation trend of PCC values

Fig. 10 Evaluation results under different prediction gaps



(a) Variation trend of RMSE value (b) Variation trend of MAPE value (c) Variation trend of PCC values

Fig. 11 Evaluation results under different prediction lengths

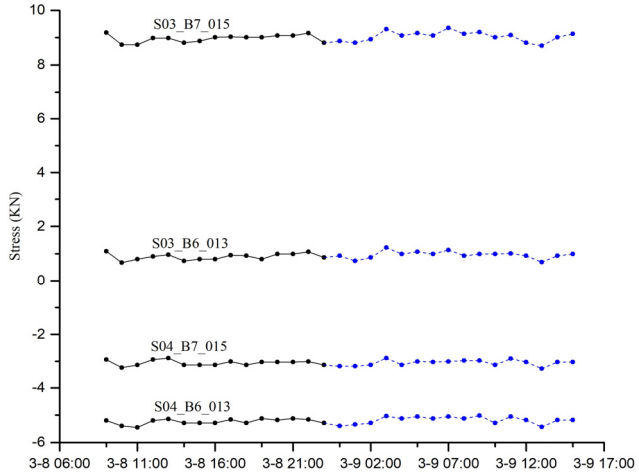


Fig. 12 The predicted stress data series of 4 sensors, which names are S03_B6_013, S03_B7_015, S04_B6_013 and S04_B7_015. The black solid lines represent the ground truth, and the blue dashed lines represent the predicted values

evaluation indicators RMSE, MAPE and PCC are recorded and drawn in a line diagram to reflect the characteristics of prediction error and prediction relevance varies with q, as Fig. 11 shows. As we can see from the figure, the values of RMSE and MAPE decrease when $q \in [1,4)$, increase when $q \in [4,8)$, and reach the minimum value 0.2688 and 2.4290 respectively when $q = 4$, the values of PCC increases when $q \in [1,4)$, decreases when $q \in [4,8)$, and reach the maximum value 0.8610 when $q = 4$. Therefore, when $p = 14$ and $g = 10$, choose the prediction length q as 4 will be the best practice.

6. Application: prediction of stress variation on multiple sensors

The effectiveness and superiority of the proposed DATCN model has been verified in Section 4, and its involved hyper-parameters have been discussed in Section 5. As a potential application, the DATCN model is employed in Nanjing Yangtze River tunnel to simultaneously predict the stress variation on multiple sensors.

After the hyper-parameters are fixed, and the model has been trained based on a certain scale of dataset, the set of model weights Θ are dumped and stored for the future usage. In this case, the prediction gap g is selected as 10, the prediction length q is selected as 6, the best observation window p is 14, which is obtained by experiment. By moving the window of input series, we can use this pre-trained model to predict the stress value for the next 16 hours at most. Specifically, given the present time \$t\$, the monitoring data in the interval of $[t-13, t]$ is putting into the model to generate the predicted value in the interval of $[t+11, t+16]$, the interval of $[t-19, t-6]$ is attached to the predicted value in the interval of $[t+5, t+10]$, the interval of $[t-25, t-12]$ is attached to the predicted value in the interval of $[t-1, t+4]$ but the predicted value in time $t-1$ and t are useless, so we can obtain the predicted value from $t+1$ to $t+16$ through moving the window of input series. The stress data from 22:00 on March 7 to 23:00 March 8 are inputted to the pre-trained model to generate the predicted trend of stress at 0:00 to 15:00 on March 9, only four sensors' curves are plotted in Fig. 12 due to space limitation. It can be seen from the curve that variation trends of stress value in four sensors are similarly, and they are all in a safe range. Therefore, the tunnel is in a safe state and there is no structural health risk for the next 16 hours.

7. Conclusions

In this study, a novel data-driven model is proposed to predict the mechanical behaviors in near time horizon and early perceive the abnormal conditions of structure. The proposed model solves the problems of incomplete consideration of factors and difficult to predict the performance across time scale existing in previous studies. Specifically, the proposed prediction model, named DATCN, integrates both temporal effect and spatial effect through TCN and self attention algorithm. The data experiments are conducted based on the monitoring data recorded by a SHMS installed in an underwater shield tunnel. It can be found from comparing the predicted results obtained from DATCN with that obtained from some baselines including LR, SVR, MLP, and RNNs, that the prediction accuracy of DATCN is higher significantly. The parameters represented the observation time window and

the prediction ability of crossing different time scale are discussed, the experiments indicate that the performance in last 14 hours play a domain role to predict the short-term behaviors of structure, and the prediction ability would drop with the increase of crossing time scale. The prediction accuracy for the behaviors after 24 hours reaches 90 percent. As an important application, the presented model is employed to predict the spatial stress distribution of Nanjing Dinghuaimen tunnel in next 24 hours. This application is crucial to prevent disasters in engineering practice and can provide important reference to similar engineering.

Acknowledgments

We thank the reviewers for their constructive comments on this research work. This work is supported by the National Key R&D Program of China No. 2018YFB2101003, the National Natural Science Foundation of China under Grant No. 51991395, U1806226, 51778033, 51822802, 71901011, U1811463, 51991391, the Science and Technology Major Project of Beijing under Grant No. Z191100002519012.

References

Cao, B.T., Obel, M., Freitag, S., Mark, P. and Meschke, G. (2020), "Artificial neural network surrogate modelling for real-time predictions and control of building damage during mechanised tunnelling", *Adv. Eng. Software*, **149**, 102869. https://doi.org/10.1016/j.advengsoft.2020.102869

Carbonneau, R., Laframboise, K. and Vahidov, R. (2008), "Application of machine learning techniques for supply chain demand forecasting", *Eur. J. Operat. Res.*, **184**(3), 1140-1154. https://doi.org/10.1016/j.ejor.2006.12.004

Chen, T., Yin, H., Chen, H., Wu, L., Wang, H., Zhou, X. and Li, X. (2018), "Tada: trend alignment with dualattention multi-task recurrent neural networks for sales prediction", *Proceedings of 2018 IEEE International Conference on Data Mining* (*ICDM*), Singapore, November, pp. 49-58. https://doi.org/10.1109/ICDM.2018.00020

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", *Tunnell. Undergr. Space Technol.*, **25**(4), 297-304. https://doi.org/10.1016/j.tust.2010.01.002

Farahani, R.V. and Penumadu, D. (2016), "Full-scale bridge damage identification using time series analysis of a dense array of geophones excited by drop weight", *Struct. Control Health Monitor.*, **23**(7), 982-997. https://doi.org/10.1002/stc.1820

Hou, L. and Qu, H. (2021), "Automatic recognition system of pointer meters based on lightweight CNN and WSNs with on-sensor image processing", *Measurement*, **183**, p. 109819. https://doi.org/10.1016/j.measurement.2021.109819

Lin, S.W., Yi, T.H., Li, H.N. and Ren, L. (2017), "Damage detection in the cable structures of a bridge using the virtual distortion method", *J. Bridge Eng.*, **22**(8), 04017039. https://doi.org/10.1061/(ASCE)BE.1943-5592.0001072

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

Mahmoodzadeh, A., Mohammadi, M., Daraei, A., Faraj, R.H., Omer, R.M.D. and Sherwani, A.F.H. (2020), "Decision-making in tunneling using artificial intelligence tools", *Tunnell. Undergr. Space Technol.*, **103**, 103514. https://doi.org/10.1016/j.tust.2020.103514

Mahmoodzadeh, A., Mohammadi, M., Ibrahim, H.H., Rashid, T.A., Aldalwie, A.H.M., Ali, H.F.H. and Daraei, A. (2021), "Tunnel geomechanical parameters prediction using Gaussian process regression", *Mach. Learn. Applicat.*, **3**, 100020. https://doi.org/10.1016/j.mlwa.2021.100020

Mei, L., Mita, A. and Zhou, J. (2016), "An improved substructural damage detection approach of shear structure based on ARMAX model residual", *Struct. Control Health Monitor.*, **23**(2), 218-236. https://doi.org/10.1002/stc.1766

Prakash, G., Sadhu, A., Narasimhan, S. and Brehe, J.M. (2018), "Initial service life data towards structural health monitoring of a concrete arch dam", *Struct. Control Health Monitor.*, **25**(1), e2036. https://doi.org/10.1002/stc.2036

Sajedi, S.O. and Liang, X. (2020), "A data-driven framework for near real-time and robust damage diagnosis of building structures", *Struct. Control Health Monitor.*, **27**(3), e2488. https://doi.org/10.1002/stc.2488

Shahrour, I., Bian, H., Xie, X. and Zhang, Z. (2020), "Smart technology applications for the optimal management of underground facilities", *Undergr. Space*, **6**(5), 551-559. https://doi.org/10.1016/j.undsp.2020.12.002

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", *Tunnell. Undergr. Space Technol.*, **38**, 348-356. https://doi.org/10.1016/j.tust.2013.07.014

Spencer Jr, B., Ruiz-Sandoval, M.E. and Kurata, N. (2004), "Smart sensing technology: opportunities and challenges", *Struct. Control Health Monitor.*, **11**(4), 349-368. https://doi.org/10.1002/stc.48

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

Tan, X., Chen, W., Wang, L., Tan, X. and Yang, J. (2020a), "Integrated approach for structural stability evaluation using real-time monitoring and statistical analysis: Underwater shield tunnel case study", *J. Perform. Constr. Facil.*, **34**(2), 04019118. https://doi.org/10.1061/(ASCE)CF.1943-5509.0001391

Tan, X., Chen, W., Wu, G., Wang, L. and Yang, J. (2020b), "A structural health monitoring system for data analysis of segment joint opening in an underwater shield tunnel", *Struct. Health Monitor.*, **19**(4), 1032-1050. https://doi.org/10.1177/1475921719876045

Wang, Y. and Ni, Y. (2020), "Bayesian dynamic forecasting of structural strain response using structural health monitoring data", *Struct. Control Health Monitor.*, **27**(8), e2575. https://doi.org/10.1002/stc.2575

Xu, Y., Li, D., Xie, Q., Wu, Q. and Wang, J. (2021), "Automatic defect detection and segmentation of tunnel surface using modified Mask R-CNN", *Measurement*, **178**, 109316. https://doi.org/10.1016/j.measurement.2021.109316

Yu, A., Mei, W. and Han, M. (2021), "Deep learning based method of longitudinal dislocation detection for metro shield tunnel segment", *Tunnell. Undergr. Space Technol.*, **113**, 103949. https://doi.org/10.1016/j.tust.2021.103949

Zheng, X., Yi, T.H., Yang, D.H. and Li, H.N. (2021), "Stiffness estimation of girder bridges using influence lines identified from vehicle-induced structural responses", *J. Eng. Mech.*, **147**(8), 04021042. https://doi.org/10.1061/(ASCE)EM.1943-7889.0001942

Zhu, H., Wang, X., Chen, X. and Zhang, L. (2020), "Similarity search and performance prediction of shield tunnels in operation

through time series data mining", *Automat. Constr.*, **114**, 103178. https://doi.org/10.1016/j.autcon.2020.103178