Neural Processing Letters (2023) 55:2321–2339
https://doi.org/10.1007/s11063-022-10795-9

HFMD Cases Prediction Using Transfer One-Step-Ahead Learning

Yaohui Huang[1] · Peisong Zhang[3] · Ziyang Wang[2] · Zhenkun Lu[1] · Zhijin Wang[2]

Accepted: 28 February 2022 / Published online: 5 April 2022

Abstract
Hand, foot and mouth disease (HFMD) is a susceptible viral infectious disease to infants and children, which led to millions of cases and hundreds of deaths annually in China. Existing predictive methods commonly learn the development patterns from historical observations. However, almost all these methods are neglect the immediate impact of exogenous factors on HFMD transmission. To solve the limitation, we consider the approximately unidirectional influences from temperature to confirmed cases and then propose a transfer one-step-ahead learning (Tr-OSH) method to establish their association. The Tr-OSH method first extract the unidirectional representation from temperature observations, and subsequently transfer the obtained representation for HFMD cases prediction. Moreover, we notice the independent correlation of each time step and period, and generate the independent representation by the relevance to upcoming values. Intensive experiments on real-world HFMD datasets demonstrate that our Tr-OSH method much efficaciously improves prediction accuracy.

Keywords Transfer learning · Prediction · HFMD · Time series · Representation learning

1 Introduction

Hand, foot and mouth disease (HFMD) is a prevalent infectious disease that has been listed as a statutorily notifiable category-C infectious disease since May 2008 in China [30]. It is caused by enteroviruses and primarily harms infants and young children. Due to the severe complications of HFMD, such as central nervous system (CNS) disorders and pulmonary [12], the mean annual severe illness rate was 11.8 per one million person-years, and the

✉ Zhenkun Lu
lzk06@sina.com

✉ Zhijin Wang
zhijinecnu@gmail.com

1 College of Electronic Information, Guangxi University for Nationalities, Daxue East Road 188, Nanning 530006, China

2 Computer Engineering College, Jimei University, Yinjiang Road 185, Xiamen 361021, China

3 School of Science, Jimei University, Yinjiang Road 185, Xiamen 361021, China

yearly mortality rate was 0.308 per one million person-years in China [31]. Besides, HFMD also caused colossal losses in Asian Pacific countries, inducing a substantial disease burden for governments and healthcare facilities all over the world [11]. Therefore, establishing an effective model to monitor and track the spread of HFMD received extensive attention in recent years.

Mathematical modeling, such as the susceptible infectious recovered (SIR) model and SEIR model, is regarded as a practical method to infer the transmission of HFMD [33]. However, these methods require relatively accurate classification of the target population and are sensitive to initial values. Recently, the learning methods have been widely used in related infectious disease problems, which can model the development of outpatients by end-to-end architecture, and forecast the upcoming confirmed cases merely depending on the historical observations [24]. Wang et al. [26] proposed a dual-grained learning method to predict weekly outpatients incorporate the fine-grained HFMD data. [34] consider the incubation period of HFMD and employ the traditional tree method to generate predictions. On account of the plentiful stochastic fluctuations, improving the predictive accuracy of HFMD cases is still a major challenge.

To integrate several exogenous factors, such as search query [14] and temperature [33], seems a promising way to improve the HFMD cases predictive accuracy. Benefit from the record of historical incidence and related factors, [5, 24, 27] show well performance in forecasting upcoming outpatients. However, most studies merely consider the impact of historical patterns in exogenous data, while neglecting the immediate and unidirectional impact on the transmission of infectious diseases. The temperature has a tremendous unidirectional influence on the activation of causative pathogens and rapidly impacts the risk of HFMD infection [18, 20]. Nevertheless, the number of search queries may increase after the epidemic outbreak [14]. Therefore, the change of meteorological factors commonly precedes the fluctuation of HFMD incidence, while the fluctuation of the search index has a delayed reaction.

We conclude the causal relationship between temperature and HFMD epidemics as a unidirectional weak-feedback mechanism: temperature affects the number of patients, but not vice versa. Depending on the above mechanism, we assume that fusing the future temperature representation can enhance the predictive accuracy of HFMD cases. In order to establish an end-to-end predictor with the representation of prospective temperature, the transfer learning approach is integrated into our designed architecture. We regard the mean temperature historical observations as the source time series and would like to learn the representation of upcoming temperature states. After that, the obtained source time series representation is delivered into the training processing of the target time series to generate final predictive results.

We proposed a transfer one-step-ahead learning (Tr-OSH) method to effectively benefit the source and target time series. The Tr-OSH leverages the dual-side multi-head attention components for source time series to enhance the association with target data. Then, the CNNRNN is incorporated to extract the potential dependencies between each time step. Finally, a fully connected linear layer maps the inputted representation to the final generation linearly. For target time series, the Tr-OSH uses a relatively concise learning structure to reduce the computational cost and avoid overfitting, which consists two directional transformation units (DTU) that highlight the critical time steps and periods to upcoming cases. At last, the Tr-OSH incorporates the source time series representation to make predictions.

The main contributions of this work can be summarized as follows:

(1) We utilized the unidirectional weak-feedback mechanism between temperature and HFMD cases.

(2) We proposed a transfer one-step-ahead learning (Tr-OSH) method to benefit the HFMD prediction with prospective temperature observations.
(3) The Tr-OSH method effectively captures the potential association in time steps and periods.
(4) The experiments on actual HFMD data collections show the effectiveness of the proposed Tr-OSH.

The remainder of this paper is organized as follows. Section 2 provide the literature review. Section 3 gives problem formulation on related time series problem. Section 4 illustrates the proposed Tr-OSH method. Section 5 gives the configurations of the experiment environment. Section 6 shows the experimental results and analyses. Finally, we conclude in Sect. 7.

2 Related Work

This section introduces related prediction methods. According to the relationship with transfer learning, these methods are divided into the traditional learning methods in time series prediction and transfer learning methods in time series prediction.

2.1 Traditional Learning Methods in Time Series Prediction

Traditional learning methods are categorized into mathematical, stochastic, and deep learning methods.

Mathematical methods. [33] employed a mathematical approach modeling the dynamic behaviors of HFMD with the temperature-dependent latent period. [21] considered the effect of vaccination and used mathematical simulating the spread of HFMD epidemics. However, these methods are sensitive to the initial parameters and show relatively weak robustness.

Stochastic methods. Autoregressive Integrated Moving Average (ARIMA) and its different variations (e.g., ARMA, GlobalAR [13] and VectorAR [22]) are common stochastic methods, which based on the famous Box-Jenkins principle [1, 25]. Nonetheless, the non-stationary and non-linear characteristics exist in the transmission of HFMD epidemics, which reduce the stochastic methods predictive accuracy [16].

Deep learning methods. Deep learning methods regard the historical observations as time series input vectors, and then learning the non-linearity mapping to the upcoming values [13]. Recurrent Neural Networks (RNN) have been widely applied in many infectious disease problems [2], it establishes connections between hidden units and can preserve memories of recent events [19]. Due to the vanilla RNNs prone to vanishing gradient or exploding gradient, two modified RNN architectures, long and short-term memory (LSTM) and gated recurrent unit (GRU), are proposed to obtain time-long dependence of time series data [29]. Wang et al. [26] captures the potential information from dual-grained HFMD data based on the GRU component and shows relatively well prediction performance than others benchmarks. Otherwise, efficient convolutional architecture for epidemiological predictions are also widely employed, Wu et al. [28] used RNNs to extract the long-term dependency from the inputs, and then integrated CNN to merge the potential information from different source data. However, few studies consider fusing the independent correlations of time steps and periods, which is a useful manner to improve the reasoning abilities of the predictive methods.

2.2 Transfer Learning Methods in Time Series Analysis

Transfer learning is the process of extracting the knowledge from one task to perform another target task [3]. In recent years, numerous studies designed many effective architectures to verify the feasibility in transfer learning on time series problem [35]. Transfer learning in time series analysis can be divided into two aspects: transfer learning via representation extraction and transfer learning via fine-tuning.

Transfer learning via representation extraction. In [17], the representations are extracted from electronic health records by unsupervised pre-trained method, and then evaluated the generations through forecasting the health state of patients. Gupta et al. [7] leveraged deep RNN pre-trained for representations extraction and subsequent target tasks in the healthcare domain. Harutyunyan et al. [9] considered learning generic representations from clinical time series the multitask training method and shown well performance. We also consider the to benefit exogenous time series representation to capture the correlation between mean temperature and HFMD confirmed cases.

Transfer learning via fine-tuning. Zhuang et al. [35] trained a generate time series predictor to release the time-consuming in multi-channels anomaly detection. Ye et al. [32] pre-trained a fully convolutional network (FCN) to learn the time series representation and then fine-tuned two sub-networks parameters to make anomaly detect. Ma et al. [15] addressed the long-interval consecutive missing values imputation problem in air pollution time series by fine-tuning source domain trained LSTM to the target domain. However, the fluctuations of HFMD epidemics are more dramatic than mean temperature, transferring the parameters trained by mean temperature time series and fine-turning to the target domain cannot improve accuracy.

In this work, we proposed a method to transfer the upcoming representation on mean temperature tasks to the HFMD cases prediction tasks. To our best knowledge, there is no research on HFMD cases prediction that has effectively utilized transfer learning architecture on this problem before.

3 Problem Formulation

This section gives problem definitions and major notations. Table 1 lists the main notations used in this paper.

Time series. The consecutive historical observations of HFMD confirmed cases in a specific state could be described as a univariate time series (UTS). The symbol $\boldsymbol{Y} = (y_1, y_2, \ldots, y_L) \in \mathbb{R}^{L\times 1}$ represents the cases in L consecutive days, the symbol $\boldsymbol{S} = (s_1, s_2, \ldots, s_L) \in \mathbb{R}^{L\times 1}$ denotes the exogenous observations in L consecutive days as well.

Time series prediction problem. The prediction based on past target values is formulated as:

$$\hat{y}_{T+1,1} \leftarrow F(\boldsymbol{Y}_{1:T,1}), \tag{1}$$

where the symbol T is the look-back window size, which indicates the past T consecutive intervals. $\hat{y}_{T+1,1}$ represents the prediction of confirmed cases in the upcoming day. $F(\cdot)$ represents a mapping.

Commonly, as shown in Fig. 1a, the prediction using exogenous time series is defined as:

$$\hat{y}_{T+1,1} \leftarrow F([\boldsymbol{Y}_{1:T,1}; \boldsymbol{S}_{1:T,1}]), \tag{2}$$

where [;] indicates the concatenate operation for multiple group inputs.

Table 1 Major notations

Symbol	Notation
L	The time steps of time series data
T	The look-back window size for time series data
N	The number of variables
B	The batch size
I	The input feature vector, $\boldsymbol{I} \in \mathbb{R}^{L\times 1}$
H, h	The h-th head of attentions heads H
$\mathcal{X}$	The input matrix to learning method, $\mathcal{X} \in \mathbb{R}^{T\times N}$
$\mathcal{Y}$	The output matrix to learning method, $\mathcal{Y} \in \mathbb{R}^{1\times 1}$
$y_{T+1,1}$	The actual confirmed cases of the next day
$\hat{y}_{T+1,1}$	The confirmed cases prediction of the next day
$\boldsymbol{Y}$	The matrix of target time series, $\boldsymbol{Y} \in \mathbb{R}^{T\times 1}$
$\boldsymbol{S}$	The matrix of exogenous time series, $\boldsymbol{S} \in \mathbb{R}^{T\times 1}$
$F(\cdot)$	The mapping function
$\mathbf{h}$	The hidden state of recurrent neural network
$\mathbf{W}$	The learnable weight matrix
$\mathbf{b}$	The learnable bias term
$\boldsymbol{M}$	The output representation of multi-head attention
$\mathcal{R}$	The output representation of DTU component
$\mathcal{A}$	The linear representation of inputs

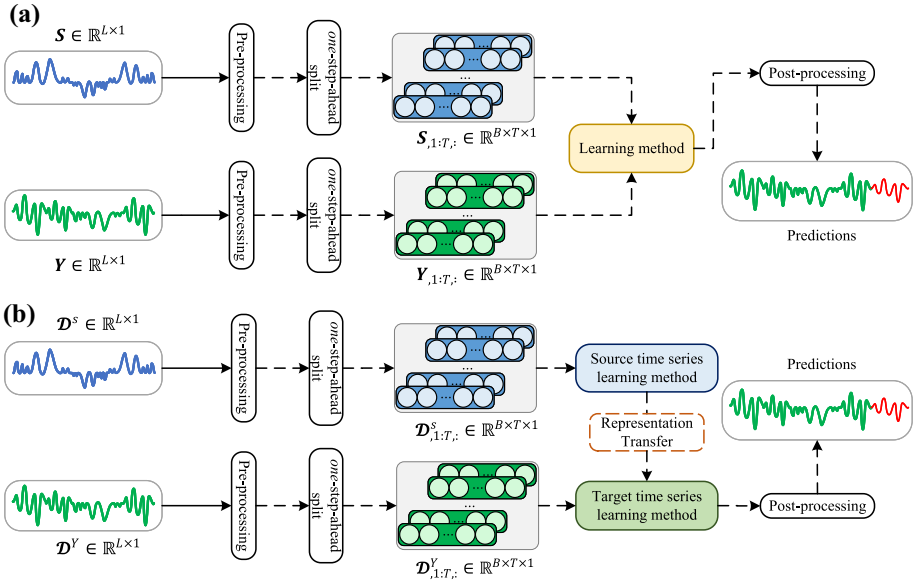


Fig. 1 The overview of time series learning architecture with exogenous inputs. **a** Traditional learning architecture. **b** The proposed transfer one-step ahead learning architecture

One-step-ahead split. The one-step-ahead split approach is applied to split time series data into data pairs further to organize the original input data into supervised data. Then, training the predictive model with the obtained supervised data. The one-step-ahead split approach is formulated as:

$$\begin{bmatrix} \boldsymbol{x}_{1,:} & \boldsymbol{x}_{2,:} & \cdots & \boldsymbol{x}_{T,:} \\ \boldsymbol{x}_{2,:} & \boldsymbol{x}_{3,:} & \cdots & \boldsymbol{x}_{T+1,:} \\ \cdots & \cdots & \cdots & \\ \boldsymbol{x}_{K-T-1,:} & \boldsymbol{x}_{K-T-1,:} & \cdots & \boldsymbol{x}_{K-1,:} \end{bmatrix} \rightarrow \begin{bmatrix} y_{T+1,1} \\ y_{T+2,1} \\ \cdots \\ y_{K,1} \end{bmatrix}, \tag{3}$$

where the left part is the input matrix, and the right part is the output results of the predictive model. To simplify the notation, let $\mathcal{X} \in \mathbb{R}^{(L-T-1)\times T\times N}$ and $\mathcal{Y} \in \mathbb{R}^{(L-T-1)\times 1\times 1}$ denote input matrix and output matrix. Each sample in $(\mathcal{X}, \mathcal{Y})$ is denoted by $(\boldsymbol{x} \in \mathbb{R}^{T\times N}, y \in \mathbb{R}^{1\times 1})$.

Transfer one-step-ahead learning problem. The symbols $\mathcal{D}^S$ and $\mathcal{D}^Y$ correspond to source and target time series, respectively. We consider the source time series $\mathcal{D}^S = \{(\mathcal{X}^S_{t:t+T,N}, \mathcal{Y}^S_{t+T+1,1})\}_{t=1}^{L-T-1}$ represents the exogenous observations, where t is the time steps and N represents the number of time series instances. And the target time series $\mathcal{D}^Y = \{(\mathcal{X}^Y_{t:t+T}, \mathcal{Y}^Y_{t+T+1})\}_{t=1}^{L-T-1}$ denotes the HFMD confirmed cases observations.

The learning processing in source time series is formulated as:

$$\hat{\boldsymbol{p}}_{T+1,1} \leftarrow F(\mathcal{X}^S; \mathcal{Y}^S), \tag{4}$$

where $\hat{\boldsymbol{p}}_{T+1,1}$ is the generation of source learning method. Subsequently, the generation, as a part of the input, is integrated into the target learning method makes further training:

$$\hat{y}_{T+1,1} \leftarrow F([\mathcal{X}^Y; \hat{\boldsymbol{p}}_{T+1,1}]; \mathcal{Y}^Y), \tag{5}$$

where the $\hat{y}_{T+1,1}$ indicates the predictive result. Hence, the target function in transfer one-step-ahead learning is formulated as follows:

$$\mathcal{L}^Y = \arg\min_{\boldsymbol{\omega},\boldsymbol{\phi}} \sum_t \| y_t - (\boldsymbol{\omega} \cdot \mathcal{X}^Y_t + \boldsymbol{\phi} \cdot \hat{\boldsymbol{p}}_t) \|_2, \tag{6}$$

where $\mathcal{L}^Y$ indicates the target time series prediction loss, $\boldsymbol{\omega}$ and $\boldsymbol{\phi}$ are the learnable parameters, $\| \cdot \|_2$ represents the ℓ_2-norm.

4 The Proposed Tr-OSH

This section illustrates the proposed Tr-OSH method and makes the description in three stages: data processing, source time series representation learning, and target time series representation learning. The schematic illustration of the proposed method ploted in Fig. 2. The pseudo-code for training Tr-OSH in the source domain and target domain are shown in Algorithms 1 and 2, respectively.

4.1 Data Processing

Normalization and de-normalization. To release the numerical difference between source time series and target time series, the normalization techniques introduced in the pre-processing stage.

The normalization operation can scale the original input data range to the other range. Normalization techniques are majorly divided into two categories: Min–Max normalization

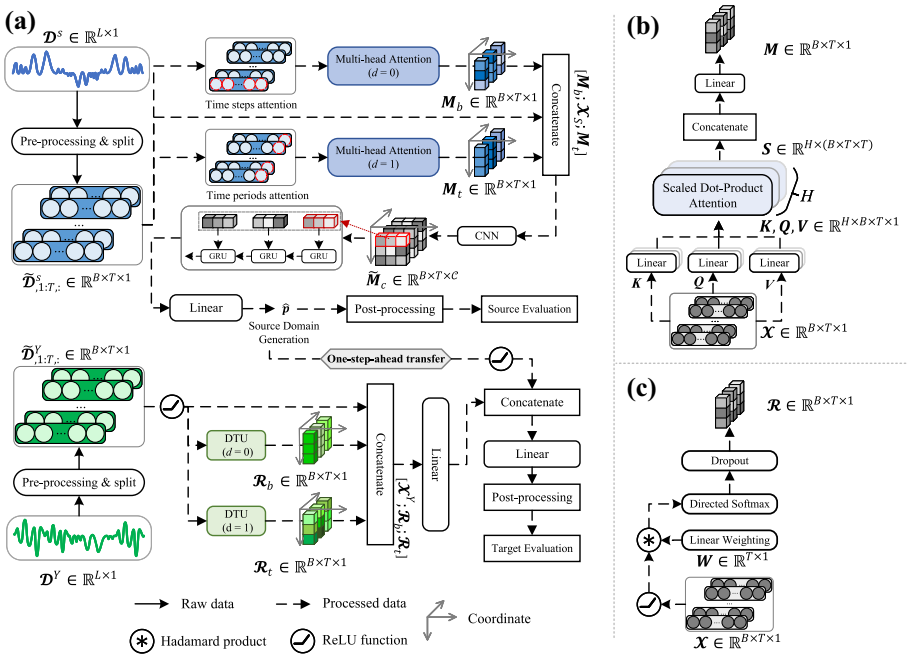


Fig. 2 The schematic illustration of the proposed transfer one-step-ahead learning method (Tr-OSH). **a** The architecture of proposed Tr-OSH, the upper and lower parts represent the processing process of source time series and the target time series, respectively. **b** The processing of multi-head attention mechanism. **c** The processing of directional transformation unit (DTU)

and z-score normalization. The z-score normalization scale the original inputs to conform to the standard normal distribution. However, this technique cannot maintain the original relative difference of each time step, so we use the Min–Max normalization to build a linear mapping to the original data and scale the inputs into the range of [0, 1]. The normalize and recovery processing of Min–Max normalization is formulated as:

$$\boldsymbol{I} = \frac{\boldsymbol{I} - \min(\boldsymbol{I})}{\max(\boldsymbol{I}) - \min(\boldsymbol{I})}, \tag{7}$$

$$\boldsymbol{I} = \boldsymbol{I}' \cdot (\max(\boldsymbol{I}) - \min(\boldsymbol{I})) + \min(\boldsymbol{I}), \tag{8}$$

where $\max(\cdot)$ is the maximal value of $\boldsymbol{I}$, and $\min(\cdot)$ is the minimal value of $\boldsymbol{I}$. $\boldsymbol{I} \in \mathbb{R}^{L\times 1}$ denotes a feature vector of the inputs, $\boldsymbol{I}'$ is the normalized data, L is the number of input samples.

4.2 Source Time Series Representation Learning

The upper part of Fig. 2a demonstrates the source time series representation learning processing.

To effectively distribute the attention on source time series, two multi-head attention components are employed to highlight the relatively important time step and periods for the upcoming values. As shown in Fig. 2b, the multi-head attention component linearly projects the sequential input vectors into different subspaces, and maps the query and key-value

Algorithm 1: Pseudo code for training Tr-OSH (source domain).

Input: Mean temperature observations $\mathcal{D}^S \in \mathbb{R}^{L\times 1}$, the number of attention heads H and window size T
Output: The representation $\hat{p} \in \mathbb{R}^{1\times 1}$ of source time series

$\tilde{\mathcal{D}}^S \leftarrow$ normalize $\mathcal{D}^S$ using (7);
$(\mathcal{X}^S, \mathcal{Y}^S) \leftarrow$ one-step-ahead split $\mathcal{D}^S$ using (3);
// Feed forward and backward gradient updating
foreach *sample* $(\tilde{\boldsymbol{x}}^S, \tilde{y}^S)$ *in* $(\tilde{\mathcal{X}}^S, \tilde{\mathcal{Y}}^S)$ **do**
  for $h \leftarrow 1$ *to* H **do**
    $Q_h, K_h, V_h \leftarrow$ mapping $\tilde{\boldsymbol{x}}^S$ using (9), (10) and (11);
    $S_{h,b}, S_{h,t} \leftarrow$ calculate attention scores using (12), (13);
  end
  $\boldsymbol{M}_b, \boldsymbol{M}_t \leftarrow$ summarize $\boldsymbol{S}_b$ and $\boldsymbol{S}_t$ using (14);
  $\tilde{\boldsymbol{M}}_c \leftarrow \boldsymbol{M}_b, \boldsymbol{M}_t$ and $\tilde{\mathcal{D}}^S$ using (15);
  for $t \leftarrow 1$ *to* T **do**
    $\boldsymbol{r}_t \leftarrow \tilde{\boldsymbol{M}}_c$ and $\boldsymbol{h}_{t-1}$ using (16);
    $\boldsymbol{z}_t \leftarrow \tilde{\boldsymbol{M}}_c$ and $\boldsymbol{h}_{t-1}$ using (17);
    $\tilde{\boldsymbol{h}}_t \leftarrow \tilde{\boldsymbol{M}}_c, \boldsymbol{r}_t$ and $\boldsymbol{h}_{t-1}$ using (18);
    $\boldsymbol{h}_t \leftarrow \tilde{\boldsymbol{h}}_t, \boldsymbol{z}_t$ and $\boldsymbol{h}_{t-1}$ using (19);
  end
  $\hat{p} \leftarrow$ linear projecting $\boldsymbol{h}_t$ using (20);
  Loss $\mathcal{L}^S \leftarrow \tilde{y}^S$ and $\hat{p}$ using MSE;
  Backward using Adam optimizer;
end
return $\hat{p}$

Algorithm 2: Pseudo code for training Tr-OSH (target domain).

Input: HFMD observations $\mathcal{D}^Y \in \mathbb{R}^{L\times 1}$, source time series representation $\hat{p}$ and window size T
Output: The prediction of next step HFMD confirmed cases $\hat{y} \in \mathbb{R}^{1\times 1}$

$\tilde{\mathcal{D}}^Y \leftarrow$ normalize $\mathcal{D}^Y$ using (7);
$(\mathcal{X}^Y, \mathcal{Y}^Y) \leftarrow$ one-step-ahead split $\mathcal{D}^Y$ using (3);
// Feed forward and backward gradient updating
foreach *sample* $(\tilde{\boldsymbol{x}}^Y, \tilde{y}^Y)$ *in* $(\tilde{\mathcal{X}}^Y, \tilde{\mathcal{Y}}^Y)$ **do**
  $\mathcal{R}_b, \mathcal{R}_t \leftarrow$ highlight $\tilde{\boldsymbol{x}}^Y$ using (21) and (22);
  $\mathcal{A} \leftarrow \mathcal{R}_b, \mathcal{R}_t$ and $\tilde{\boldsymbol{x}}^Y$ using (23);
  $\hat{O} \leftarrow \mathcal{A}$ and $\hat{p}$ using (24);
  Loss $\mathcal{L}^Y \leftarrow \tilde{y}^Y$ and $\hat{O}$ using MSE;
  Backward using Adam optimizer;
  $\hat{y} \leftarrow$ de-normalize $\hat{O}$ using (8);
end
return $\hat{y}$

pairs to the predictive target in each subspace, which shows better performance in extracting complex potential information than single-head attention.

First, given a set of queries and key-value pairs by the source time series observations, which can be defined as follows:

$$\boldsymbol{Q}_h = \mathbf{W}_{h,q} \cdot \tilde{\mathcal{X}}^S + \mathbf{b}_{h,q}, \tag{9}$$

$$\boldsymbol{K}_h = \mathbf{W}_{h,k} \cdot \tilde{\mathcal{X}}^S + \mathbf{b}_{h,k}, \tag{10}$$

$$\boldsymbol{V}_h = \mathbf{W}_{h,v} \cdot \tilde{\mathcal{X}}^S + \mathbf{b}_{h,v}, \tag{11}$$

where $\boldsymbol{Q}, \boldsymbol{K}, \boldsymbol{V} \in \mathbb{R}^{H*(B\times T\times 1)}$ represent the query, key and value, respectively. h denotes the h-th head of attentions heads, $\mathbf{W}_{h,:} \in \mathbb{R}^{B\times 1\times\gamma}$ and $\mathbf{b}_{h,:} \in \mathbb{R}^{B\times\gamma}$ is weight and bias of h-th head attention, γ is a constant indicates the hidden state dimensions, which used to keep the gradient stable. After that, the multi-head attention scores can be calculated as follows:

$$\boldsymbol{S}_{h,b} = \text{softmax}(\frac{(\boldsymbol{Q}_h \cdot \boldsymbol{K}_h^\tau)_{b,:,:}}{\sqrt{\gamma}}), d = 0, \tag{12}$$

$$\boldsymbol{S}_{h,t} = \text{softmax}(\frac{(\boldsymbol{Q}_h \cdot \boldsymbol{K}_h^\tau)_{:,t,:}}{\sqrt{\gamma}}), d = 1, \tag{13}$$

where $\boldsymbol{S}_h \in \mathbb{R}^{H*(B\times T\times T)}$ represents the multi-head temporal attention score, $\boldsymbol{S}_{h,b}$ and $\boldsymbol{S}_{h,t}$ denotes the score of the time step and period respectively. d indicates the attention direction of input vector. $d = 0$ denotes weighting on the each period, and $d = 1$ represents highlight the significant time steps.

Finally, summarize the attention context vectors from each head, and then generate the final attention representation by connecting the output of different heads.

$$\boldsymbol{M} = \mathbf{W}_m \cdot [\boldsymbol{S}_1 \times \boldsymbol{V}_1; \boldsymbol{S}_2 \times \boldsymbol{V}_2; \cdots; \boldsymbol{S}_H \times \boldsymbol{V}_H], \tag{14}$$

where $\boldsymbol{M} \in \mathbb{R}^{B\times T\times 1}$ denotes the final multi-head attention representation, which concatenate different transformation matrices in different head. $\mathbf{W}_m \in \mathbb{R}^{H\times\gamma\times 1}$ is linear projection weight matrix.

After collecting the transform information in different time intervals by the dual-direction multi-head attention mechanisms, the CNNRNN combination component is employed to extract further each time step's correlation of source time series. The CNN module is applied to generate indirect nonlinearity projecting for each time step.

$$\tilde{\boldsymbol{M}}_c = \text{Dropout}(\text{ReLU}(\mathbf{W}_c \otimes [\boldsymbol{M}_b; \boldsymbol{M}_t; \tilde{\mathcal{X}}^S] + \mathbf{b}_c)), \tag{15}$$

where the symbol $\otimes$ denotes the convolution operation, $\tilde{\boldsymbol{M}}_c \in \mathbb{R}^{B\times T\times K}$ indicates the product time series feature map, K is the output channel. $\boldsymbol{M}_b, \boldsymbol{M}_t \in \mathbb{R}^{B\times T\times 1}$ are the production of multi-head attention mechanism. $\mathbf{W}_c$ represents the kernel and $\mathbf{b}_c$ is the bias. Dropout and ReLU are introduced to prevent overfitting and gradient disappearance. The output feature vectors are then input into the GRU component:

$$\mathbf{r}_t = \sigma(\mathbf{W}_r \cdot [\tilde{\boldsymbol{M}}_c^t; \mathbf{h}_{t-1}] + \mathbf{b}_r), \tag{16}$$

$$\mathbf{z}_t = \sigma(\mathbf{W}_z \cdot [\tilde{\boldsymbol{M}}_c^t; \mathbf{h}_{t-1}] + \mathbf{b}_z), \tag{17}$$

$$\tilde{\boldsymbol{h}}_t = \tanh(\mathbf{W}_{\tilde{h}} \cdot [\tilde{\boldsymbol{M}}_c^t; \mathbf{r}_t \circ \mathbf{h}_{t-1}] + \mathbf{b}_{\tilde{h}}), \tag{18}$$

$$\mathbf{h}_t = (1 - \mathbf{z}_t) \circ \tilde{\boldsymbol{h}}_t + \mathbf{z}_t \circ \mathbf{h}_{t-1}, \tag{19}$$

where the $\mathbf{h}_t$ is the sequence hidden state at time t, $\tilde{\boldsymbol{M}}_c^t$ represents the input tensor at time t, $\mathbf{z}_t$, $\mathbf{z}_t$ and $\tilde{\boldsymbol{h}}_t$ are the reset, update, and new gates, respectively. σ is the sigmoid function,

and $\circ$ is the element-wise product. $\mathbf{W}_r$, $\mathbf{W}_z$ and $\mathbf{W}_{\tilde{h}}$ is the learnable weights, $\mathbf{b}_r$, $\mathbf{b}_z$ and $\mathbf{b}_{\tilde{h}}$ is bias for input and hidden states.

To obtain the upcoming state of source time series, the fully connected feed-forward layer is employed to converge the hidden state representations and make linear projecting to the forecast target:

$$\hat{\boldsymbol{p}} = \mathbf{W}_{sout} \cdot \mathbf{h}_t + \mathbf{b}_{sout}, \tag{20}$$

where $\hat{\boldsymbol{p}}$ represents the prediction and also the significant transfer parameter to the target method processing. $\mathbf{W}_{sout}$ is a weight parameter that needs to be learned, and $\mathbf{b}_{sout}$ is a bias term, and $\mathbf{h}_t$ is the hidden state which is represented from the source time series.

4.3 Target Time Series Representation Learning

As shown in the lower parts of Fig. 2a, to enhance the non-linearity and mitigate the gradient explosion and gradient disappearance problems, the ReLU layer is employed to rectify the delivery information of the target processing stage. After that, we proposed a directional transformation component to highlight the critical outbreak time steps or periods of target time series, which is defined as:

$$\mathcal{R}_b = \frac{\boldsymbol{\delta} \cdot \exp(\mathbf{W}_{db} \circ \tilde{\mathcal{X}}^Y_{b,:,:})}{\sum_b \exp(\mathbf{W}_{db} \circ \tilde{\mathcal{X}}^Y_{b,:,:}))}, d = 0 \tag{21}$$

$$\mathcal{R}_t = \frac{\boldsymbol{\delta} \cdot \exp(\mathbf{W}_{dt} \circ \tilde{\mathcal{X}}^Y_{:,t,:})}{\sum_t \exp(\mathbf{W}_{dt} \circ \tilde{\mathcal{X}}^Y_{:,t,:}))}, d = 1 \tag{22}$$

where $\exp(\cdot)$ represents the exponential function, $\mathcal{R}$ is viewed as the attention weight distribution on the d-th direction, $\mathbf{W}_{db}, \mathbf{W}_{dt} \in \mathbb{R}^{T\times 1}$ is the weighting matrix. $\boldsymbol{\delta}$ denotes a sign vector generated by the *Bernoulli* probability function. The processing is presented in Fig. 2c.

To establish the connection between the upcoming values and multiple representations, the linear global autoregression component is integrated to establish a linear projecting:

$$\mathcal{A} = \sum_T \mathbf{W}_a[\mathcal{R}_b; \mathcal{R}_t; \tilde{\mathcal{X}}^Y] + \mathbf{b}_a, \tag{23}$$

where $\mathcal{A} \in \mathbb{R}^{B\times 1\times 1}$ is the output potential representation of target time series observations, $\mathcal{R}_b, \mathcal{R}_t \in \mathbb{R}^{B\times T\times 1}$ denotes the transform representation of time period and time steps, respectively. $\mathbf{W}_a$ and $\mathbf{b}_a$ are the learnable parameters.

Eventually, the final predictive result is generated by fusion the potential representation of target and source time series:

$$\hat{O}_{T+1,1} = \mathbf{W}_o[\mathcal{A}; \hat{\boldsymbol{p}}] + \mathbf{b}_o, \tag{24}$$

where $\hat{O}_{T+1,1}$ is normalized prediction of the next step HFMD cases, $\hat{y}_{T+1,1}$ represents the de-normalize results. $\mathbf{W}_o$ is the weight of outputs from above two kind representations, and $\mathbf{b}_o$ is a bias term.

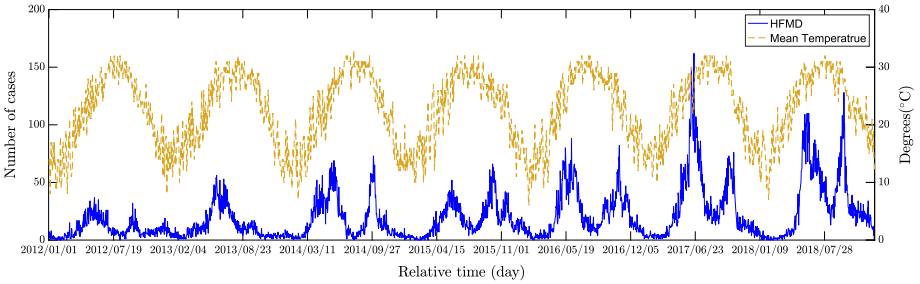


Fig. 3 The distribution of daily outpatient visit counts and mean temperature values from January 1, 2012 to December 31, 2018

Table 2 The basic statistical characteristics of source time series (mean temperature) and target time series (HFMD confirmed cases)

Characteristics	Mean temperature	HFMD
Min	6	0
Max	33	162
Mean	22.534	20.188
Median	23	13
STD	6.238	21.294

"STD" denotes standard deviation

5 Experimental Configurations

This section introduces data collections, performance measurements, comparable methods, and model configurations.

5.1 Data Collections

As Fig. 3 plotted, a total of 2556 days of observations were collected to conduct experiments and establish predictive methods.

The HFMD confirmed cases in Xiamen are obtained from the Chinese Center for Disease Control and Prevention (China CDC), which are statistic counts of patients diagnosed at health care facilities. All the clinical diagnoses of HFMD are conformed to the guidelines for diagnosing and treating hand-foot-mouth disease by the National Health Commission of China in 2018. All individual-level data is anonymized.

The mean temperature (°C) values were collected from Weather Underground,[1] which is a famous weather website with more than 33,000 private weather stations. Hence, it can provide comprehensive, timely, and reliable meteorological data from meteorological monitoring sites. The provided temperature information has been formatted in days.

The cases of HFMD outpatient and historical mean temperature record from January 1, 2012 to December 31, 2018 is arranged in chronological order. We used the front 80% samples to training models, and the rest 20% data are used to evaluate the predictive performance of models.

The basic statistical characteristics of mean temperature series and HFMD confirmed cases are listed in Table 2.

[1] http://www.wunderground.com.

5.2 Performance Measurements

Numerous measurements have been used to assess the model predict performance. To make a comprehensive evaluation, we combined four traditional metrics, i.e., MAE, $MAPE$, $RMSE$, and R^2, to estimate the predictive results of the corresponding models for univariate time series prediction. These metrics can be expressed in the following mathematical expressions:

(1) The Mean Absolute Error (MAE)

$$MAE = \frac{1}{\tau}\sum_{t\in\tau}(|y_{t+1} - \hat{y}_{t+1}|), \tag{25}$$

(2) The Mean Absolute Percentage Error ($MAPE$)

$$MAPE = \frac{1}{\tau}\sum_{t\in\tau}|\frac{y_{t+1} - \hat{y}_{t+1}}{y_{t+1}}|, \tag{26}$$

(3) The Root Mean Square Error ($RMSE$)

$$RMSE = \sqrt{\frac{1}{\tau}\sum_{t\in\tau}(y_{t+1} - \hat{y}_{t+1})^2}, \tag{27}$$

(4) The coefficient of determination (R^2)

$$R^2 = 1 - \frac{\sum_{t\in\tau}(y_{t+1} - \hat{y}_{t+1})^2}{\sum_{t\in\tau} {y_{t+1}}^2}. \tag{28}$$

where y_{t+1} and $\hat{y}_{t+1}$ are the ground truth values and predictions of HFMD confirmed cases at time $t + 1$. τ is the length of the test period. The performance with the smallest MAE, $MAPE$, $RMSE$, and the largest R^2 are considered the best model.

5.3 Competition Model

Many studies have been employed for infectious disease prediction and time series prediction. To demonstrate the performance of the proposed Tr-OSH method, we select some representative methods as competition models:

(1) *Autoregression with exogenous data (ARE)* [23] processes the target and exogenous time series simultaneously based on the traditional Autoregression (AR) method [8].
(2) *Long and short-term memory (LSTM)* [10] obtains time-long dependence of time-series data by several gate components.
(3) *Gated recurrent unit (GRU)* [4] using concise architecture than LSTM, which improves the computation efficiency.
(4) *Encoder–decoder* [4] adopt LSTM component in the encoding process and the decoding process, respectively.
(5) *Multivariate shapelet learning (MSL)* [23] learns the shapelets from historical observations and benefited these shapelets to generate the explainable predictions.
(6) *Convolution neural network (CNN1D)* [6] using convolution operation to explore deep feature representations of time series.
(7) *CNNRNN* [28] employs CNN to extract the feature representations, and then learns the relationship of each step through RNN.

(8) *Dual grained representation (DGR)* [26] benefits the non-linear representations from target time series and exogenous data by two learning units.
(9) *Dual sides autoregression (DSAR)* [25] acquires linear representations from target observations and exogenous inputs independently, and fusion these representations in combination stage.

5.4 Model Configuration

For fair competition on all models, each method is trained using the Adam optimizer. Besides, the window size T is adjusted for all methods from 1 to 20, and the batch size is searched from {1, 2, 4, 8, 16, 32, 64, 128, 256}. The mean squared error (MSE) is chosen as the loss function. For LSTM and GRU methods, the number of hidden neurons is set to {32, 64}. For the Encoder–decoder method, the structures are based on LSTM, and the number of hidden neurons is set to 32. The training epoch and learning rate are tuned to the optimal states of each method.

6 Results and Analyses

In this section, we conduct several experiments to reveal the effectiveness of the Tr-OSH method. Firstly, we assess the parameter sensitiveness of the proposed method. Secondly, we compare the performance of Tr-OSH and competition methods. Then, the ablation analysis is introduced to demonstrate the substructure validity of the Tr-OSH.

6.1 Effects on *T* and *B*

The length of windowed size indicates utilizing past T days observations to predict the upcoming values, which is highly correlated with the incubation period of diseases. Therefore, we measure the performance in terms of MAE, $MAPE$, $RMSE$, and R^2 to investigate the optimal window size. While holding other parameters constant, the window size T gradually increases from 1 to 20, and the results are presented in Fig. 4. The optimal performance is found around $T = 4$, which conform to the incubation period of HFMD (averages around 3–5 days).[2]

Obviously, as window size increases, performance becomes more unstable and deteriorates. When $T = 6$, the second best MAP and $MAPE$ are shown, while other metrics are unsatisfactory. Similarly, when $T = 11$ displayed the second best $RMSE$ and R^2, but the MAE and $MAPE$ relatively worse. This phenomenon results from the characteristics that MAE and $MAPE$ are applied to evaluate the absolute error while the $RMSE$ and R^2 incline to express the squared loss, which magnifies the predictive error in the outbreak period. Hence, when $T = 4$, the overall forecast performance is relatively superior.

Considering the evaluation results of window size T, we fixed the $T = 4$ to assess the effects on B, and demonstrated the results in Fig. 5. Remarkably, the optimal values of all matrices are found when $B = 2$. The size of B is associated with the weighting grained of the consecutive periods. The smaller B indicates the finer-grained weighting for each inputted time period to next-step values, and more easy to obtain a better performance. However, the

[2] http://www.nhc.gov.cn/.

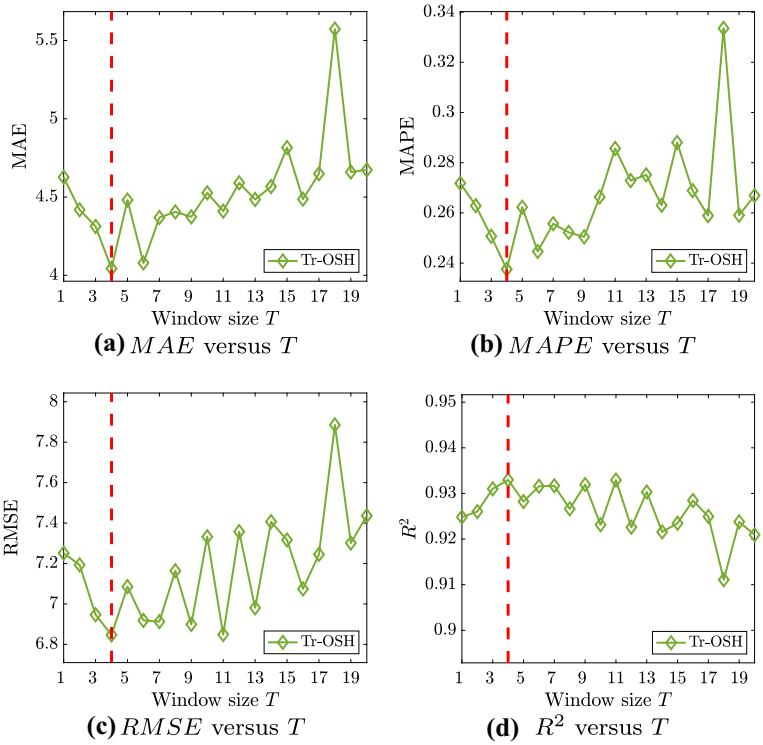


Fig. 4 The sensitiveness of window size T in terms of MAE, $MAPE$ $RMSE$ and R^2. The optimal values highlighted by the red dash lines. (Color figure online)

Tr-OSH method shows the worse performance when $B = 1$. A possible reason is that when $B = 1$, the proposed method tends to over-fitting.

6.2 Performance Comparison

According to the conducted experiments on window size T and batch size B, we learn the effects of the parameters to the Tr-OSH method with respect to the four metrics, the optimal values of these metrics are obtained around $T = 4$ and $B = 2$. Meanwhile, each competition method also uses the optimal results from different window size T. The experimental results are plotted in Table 3, which ARE, DSAR, DGR, and Tr-OSH methods utilize the mean temperature observation, and other methods merely use the HFMD observations.

Several important conclusions are summarized from the comparison experiments:

(1) The proposed Tr-OSH method has the best performance over four metrics.
(2) In terms of MAE, the CNNRNN has the second best performance.
(3) From the perspective of $MAPE$, the LSTM has the second best performance.
(4) The DGR has the second best performance in terms of $RMSE$ and R^2.
(5) Fusion the temperature series, the accuracy of prediction at outbreak period has been enhanced.
(6) Simple CNN1D structure shows the worst performance.

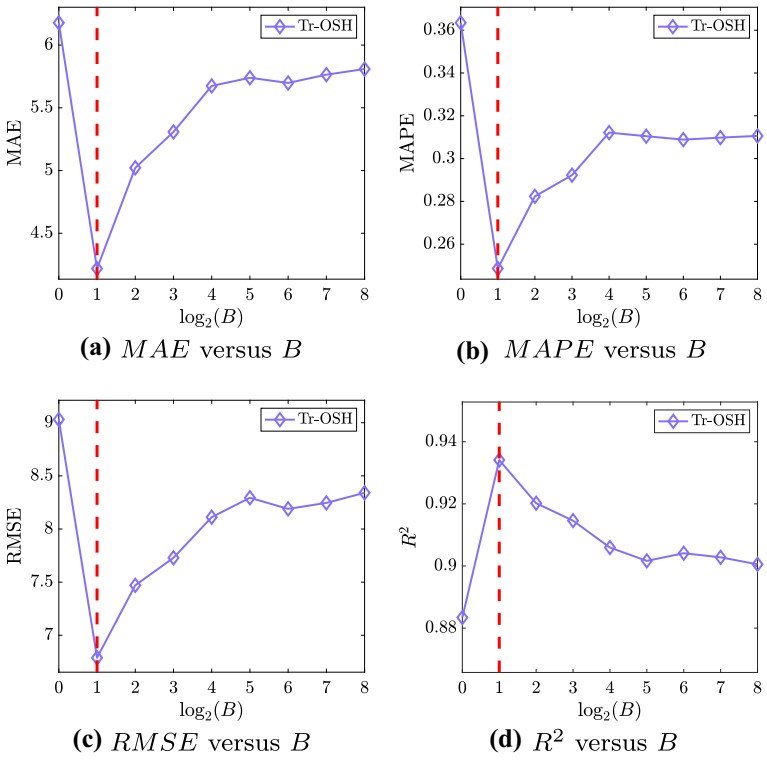


Fig. 5 The sensitiveness of batch size B in terms of MAE, $MAPE$ $RMSE$ and R^2. The optimal values highlighted by the red dash lines. (Color figure online)

Table 3 The performance comparison in terms of MAE, $MAPE$, $RMSE$ and R^2

Model	MAE	$MAPE$	$RMSE$	R^2
AR	5.87195	0.31936	8.45458	0.89780
LSTM16	5.78652	0.31596	8.35753	0.90013
LSTM32	5.85146	0.31140	8.38398	0.89950
GRU16	5.83043	0.31946	8.37650	0.89968
GRU32	5.82171	0.32078	8.37669	0.89967
MSL	5.81452	0.31874	8.34186	0.90051
CNN1d	5.95201	0.32894	8.45965	0.89768
CNNRNN	5.76217	0.31690	8.32234	0.90097
Encoder–decoder	5.85165	0.32667	8.39041	0.89934
ARE	5.89654	0.33101	8.39652	0.89920
DSAR	5.87570	0.32117	8.41720	0.89870
DGR	5.87554	0.32599	8.31492	0.90115
Tr-OSH	4.04341	0.23754	6.84674	0.93298

Table 4 The ablation analysis in terms of MAE, $MAPE$, $RMSE$ and R^2

Model	MAE	$MAPE$	$RMSE$	R^2
Tr-OSH-Y	4.45224	0.2547	7.35699	0.92261
Tr-OSH-AR(Y)	5.88936	0.32481	8.53233	0.89591
Tr-OSH-AR(S)	4.34208	0.25587	7.10348	0.92785
Tr-OSH	4.04341	0.23754	6.84674	0.93298

We divide the competition methods into two categories: one that uses only the HFMD observations for training, and the other that utilize the exogenous temperature data.

The simple CNN has the worst performance for the method merely using target time series (HFMD). A possible reason is that the HFMD time series consists of numerous stochastic fluctuations. The CNN1D is hard to extract the comprehensive relationship between each time step. Hence, the CNNRNN further integrates the RNNs component and performs relatively well. The LSTM and GRU also demonstrate the effectiveness of recurrent architecture. Meanwhile, LSTM shows the second best performance in terms of $MAPE$. The MSL method can reduce the error in high-risk periods by learning the significant subsequence. But some general information is also neglected and leads to reduced accuracy.

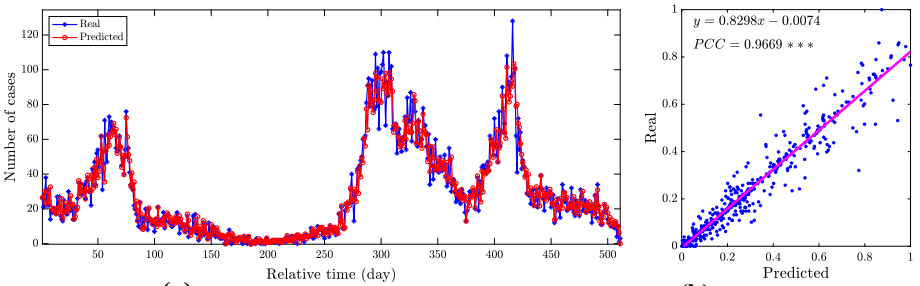
Fusion the source time series (mean temperature), the performance of ARE and DSAR are significantly reduced compared with others learning methods, which illustrates the vanilla linear-based methods incapable of fitting the complex fluctuations. The DGR approach uses the GRU unit to improve non-linearity and capture the associated characterization for both the target and source time series processing. Therefore, DGR obtains the second best performance in terms of $RMSE$ and R^2.

The proposed Tr-OSH has the best performance over all the inputs in four evaluation metrics. Considering the superiority of the RNN unit to extract potential correlation information for time series, we incorporate the CNNRNN component to enhance the non-linearity of source representations after being transformed by dual directional multi-head attention mechanism. Inspired by the unidirectional weak-feedback mechanism, the generation representation of the source time series is combined into the target training process. Besides, the critical time steps or periods of target inputted are also strengthened through the DTU component. The proposed Tr-OSH takes the local time steps and periodical correlation of sequences into account and fusion the non-linearity representations to product predictions. Consequently, the proposed method exhibits the best performance.

6.3 Ablation Analysis

In this section, we conduct an ablation analysis to demonstrate the usefulness of each designed module in our proposed Tr-OSH method. The performance comparison of three perspectives and proposed Tr-OSH are illustrated in Table 4.

To demonstrate the contribution of each training module in Tr-OSH, we replace the learning module in source time series or target time series with the simple AR component, respectively. The Tr-OSH-AR(S) uses AR to produce the representation of source time series, which slightly affects predictive performance. A possible reason is that the target time series representations have a more important place in forecasting upcoming HFMD cases. The performance of Tr-OSH-AR(Y) further verified this reason, which supplants the target learning unit with the AR, the accuracy drops significantly. Then, to show the effectiveness of



(a) Predictive results and real values **(b)** Correlation analyses

Fig. 6 The visualizations and correlation analyses between real values and predictions

the exogenous temperature information, we remove the representation of source time series in target processing in Tr-OSH-Y. The predictive accuracy is obviously decreasing. Therefore, we summarize the observations as follows: (1) Each design module plays a significant role in the proposed Tr-OSH method. (2) The representation structure of the target time series has a relatively important contribution. (3) It is useful to integrate the exogenous information to improve accuracy in Tr-OSH.

6.4 Analyses on Predictions

The visualization comparing actual values and predictions is presented in Fig. 6a. From the performance, we summarize the following information: (1) The development of HFMD epidemics can be well traced in the low-risk periods. (2) The performance in high-risk period with gentle trend also performs well, while the outbreak period prediction still room for improvement.

The normalized results and Pearson correlation of actual values and predictions is demonstrated in Fig. 6b. To estimate the relationship between numerical magnitude and accuracy, we use the first-order function to fitting the results, which draw the following conclusions: (1) The predicted value is significantly correlated to the actual value. (2) Consistent with the visualization of predictions, the predictive error will magnify as the increase of cases magnitude.

7 Conclusions

In this paper, we investigated a novel HFMD predictive transfer method Tr-OSH to fusion the representation of temperature depending on the unidirectional weak-feedback mechanism. Specifically, we incorporate the multi-head attention component and CNNRNN method to extract the non-linearity association from source time series and produce upcoming representation. Then, we transfer the source representation to the target processing, and combined the generations, which capture from the time steps and period relationship of target time series, to obtain the final results. The experiments on real-world HFMD historical observations demonstrate the effectiveness of the proposed method.

In the future, we will focus on investigating the cross-correlations phenomenon in multivariate time series prediction and multi-horizon prediction.

Acknowledgements This work was supported in part by the Natural Science Foundation of Fujian Province (CN) (Nos. 2021J01857 and 2021J01859). The authors would like to thank the editor and anonymous reviewers for their helpful comments in improving the manuscript quality.

Data Availability The data used to support the findings of this study are available from the corresponding author upon request.

Declarations

Conflict of interest The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

1. Alabdulrazzaq H, Alenezi MN, Rawajfih Y, Alghannam BA, Al-Hassan AA, Al-Anzi FS (2021) On the accuracy of arima based prediction of covid-19 spread. Results Phys 27:104509. https://doi.org/10.1016/j.rinp.2021.104509
2. Alfred R, Obit JH (2021) The roles of machine learning methods in limiting the spread of deadly diseases: a systematic review. Heliyon. https://doi.org/10.1016/j.heliyon.2021.e07371
3. Baireddy S, Desai SR, Mathieson JL, Foster RH, Chan MW, Comer ML, Delp EJ (2021) Spacecraft time-series anomaly detection using transfer learning. In: Proceedings of the IEEE conference on computer vision and pattern recognition workshops, IEEE, virtual, pp 1951–1960. https://doi.org/10.1109/CVPRW53098.2021.00223
4. Cho K, van Merrienboer B, Bahdanau D, Bengio Y (2014) On the properties of neural machine translation: encoder–decoder approaches. In: Proceedings of the 8th workshop on syntax, semantics and structure in statistical translation. Association for Computational Linguistics, Doha, pp 103–111. https://doi.org/10.3115/v1/W14-4012
5. Gao Q, Liu Z, Xiang J, Tong M, Zhang Y, Wang S, Zhang Y, Lu L, Jiang B, Bi P (2021) Forecast and early warning of hand, foot, and mouth disease based on meteorological factors: evidence from a multicity study of 11 meteorological geographical divisions in mainland china. Environ Res 192:11031. https://doi.org/10.1016/j.envres.2020.110301
6. Goodfellow I, Bengio Y, Courville A (2016) Deep Learning. MIT Press. http://www.deeplearningbook.org
7. Gupta P, Malhotra P, Narwariya J, Vig L, Shroff G (2020) Transfer learning for clinical time series analysis using deep neural networks. J Healthc Inf Res 4(2):112–137. https://doi.org/10.1007/s41666-019-00062-3
8. Hamilton JD (1994) Time series analysis. Princeton University Press, Princeton
9. Harutyunyan H, Khachatrian H, Kale DC, Steeg GV, Galstyan A (2019) Multitask learning and benchmarking with clinical time series data. Sci Data 6(1):1–18. https://doi.org/10.1038/s41597-019-0103-9
10. Hochreiter S, Schmidhuber J (1997) Long short-term memory. Neural Comput 9(8):1735–1780
11. Koh WM, Badaruddin H, La H, Chen MIC, Cook AR (2018) Severity and burden of hand, foot and mouth disease in Asia: a modelling study. BMJ Glob Health 3(1):e000442. https://doi.org/10.1136/bmjgh-2017-000442
12. Li H, Lao Q (2017) The pulmonary complications associated with ev71-infected hand-foot-mouth disease. Radiol Infect Dis 4(4):137–142. https://doi.org/10.1016/j.jrid.2017.01.001
13. Lim B, Zohren S (2021) Time-series forecasting with deep learning: a survey. Philos Trans R Soc A Math Phys Eng Sci 379(2194):20200209. https://doi.org/10.1098/rsta.2020.0209
14. Liu Y, Feng G, Tsui KL, Sun S (2021) Forecasting influenza epidemics in hong kong using google search queries data: A new integrated approach. Expert Syst Appl 185:115604. https://doi.org/10.1016/j.eswa.2021.115604
15. Ma J, Cheng JC, Ding Y, Lin C, Jiang F, Wang M, Zhai C (2020) Transfer learning for long-interval consecutive missing values imputation without external features in air pollution time series. Adv Eng Inform 44:101092. https://doi.org/10.1016/j.aei.2020.101092
16. Mills TC (2019) Chapter 4-arima models for nonstationary time series. In: Applied time series analysis. Academic Press, pp 57–69. https://doi.org/10.1016/B978-0-12-813117-6.00004-1

17. Miotto R, Li L, Kidd BA, Dudley JT (2016) Deep patient: an unsupervised representation to predict the future of patients from the electronic health records. Sci Rep 6(1):1–10. https://doi.org/10.1038/srep26094
18. Pearson D, Basu R, Wu XM, Ebisu K (2020) Temperature and hand, foot and mouth disease in California: an exploratory analysis of emergency department visits by season, 2005–2013. Environ Res 185:109461. https://doi.org/10.1016/j.envres.2020.109461
19. Perrusquia A, Yu W (2021) Identification and optimal control of nonlinear systems using recurrent neural networks and reinforcement learning: an overview. Neurocomput In Press. https://doi.org/10.1016/j.neucom.2021.01.096
20. Qi H, Li Y, Zhang J, Chen Y, Guo Y, Xiao S, Hu J, Wang W, Zhang W, Hu Y, Li Z, Zhang Z (2020) Quantifying the risk of hand, foot, and mouth disease (hfmd) attributable to meteorological factors in east china: a time series modelling study. Sci Total Environ 728:138548. https://doi.org/10.1016/j.scitotenv.2020.138548
21. Shi L, Zhao H, Wu D (2019) Modeling periodic hfmd with the effect of vaccination in mainland China. Complexity 2020:8763126. https://doi.org/10.1155/2020/8763126
22. Shih S, Sun F, Lee H (2019) Temporal pattern attention for multivariate time series forecasting. Mach Learn 108(8–9):1421–1441. https://doi.org/10.1007/s10994-019-05815-0
23. Wang Z, Cai B (2021) Covid-19 cases prediction in multiple areas via shapelet learning. Appl Intell 2021:1–12. https://doi.org/10.1007/s10489-021-02391-6
24. Wang Z, Huang Y, He B, Luo T, Wang Y, Lin Y (2019) TDDF: HFMD outpatients prediction based on time series decomposition and heterogenous data fusion in xiamen, china. In: Proceedings of the 15th international conference advanced data mining and applications, Dalian, China, pp 658–667. https://doi.org/10.1007/978-3-030-35231-8_48
25. Wang Z, Huang Y, Cai B, Ma R, Wang Z (2020a) Stock turnover prediction using search engine data. J Circuits Syst Comput. https://doi.org/10.1142/S021812662150122X
26. Wang Z, Huang Y, He B (2020b) Dual-grained representation for hfmd prediction within public health cyber-physical systems. Softw Pract Exp Early Access. https://doi.org/10.1002/spe.2940
27. Wang Z, Huang Y, He B, Luo T, Wang Y, Fu Y (2020) Short-term infectious diarrhea prediction using weather and search data in Xiamen, China. Sci Program 2020:1–12. https://doi.org/10.1155/2020/8814222
28. Wu Y, Yang Y, Nishiura H, Saitoh M (2018) Deep learning for epidemiological predictions. In: Proceedings of the 41st international ACM SIGIR conference on research and development in information retrieval, Ann Arbor, pp 1085–1088. https://doi.org/10.1145/3209978.3210077
29. Xiao Y, Yin H, Zhang Y, Qi H, Zhang Y, Liu Z (2021) A dual-stage attention-based conv-lstm network for spatio-temporal correlation and multivariate time series prediction. Int J Intell Syst 36(5):2036–2057. https://doi.org/10.1002/int.22370
30. Xing W, Liao Q, Viboud C, Zhang J, Sun J, Wu JT, Chang Z, Liu F, Fang VJ, Zheng Y, Cowling BJ, Varma JK, Farrar JJ, Leung GM, Yu H (2014) Hand, foot, and mouth disease in china, 2008–12: an epidemiological study. Lancet Infect Dis 14(4):308–318. https://doi.org/10.1016/S1473-3099(13)70342-6
31. Yang B, Liu F, Liao Q, Wu P, Chang Z, Huang J, Long L, Luo L, Leung G, Cowling B, Yu H (2017) Epidemiology of hand, foot and mouth disease in china, 2008 to 2015 prior to the introduction of ev-a71 vaccine. Eurosurveillance 22(50):1–10. https://doi.org/10.2807/1560-7917.ES.2017.22.50.16-00824
32. Ye H, Ma X, Pan Q, Fang H, Xiang H, Shao T (2019) An adaptive approach for anomaly detector selection and fine-tuning in time series. In: Proceedings of the 1st international workshop on deep learning practice for high-dimensional sparse data, New York, pp 1–7. https://doi.org/10.1145/3326937.3341253
33. Zhao H, Shi L, Wang J, Wang K (2021) A stage structure hfmd model with temperature-dependent latent period. Appl Math Model 93:745–761. https://doi.org/10.1016/j.apm.2021.01.010
34. Zhong R, Wu Y, Cai Y, Wang R, Zheng J, Lin D, Wu H, Li Y (2018) Forecasting hand, foot, and mouth disease in Shenzhen based on daily level clinical data and multiple environmental factors. Biosci Trends 12:450–455. https://doi.org/10.5582/bst.2018.01126
35. Zhuang F, Qi Z, Duan K, Xi D, Zhu Y, Zhu H, Xiong H, He Q (2021) A comprehensive survey on transfer learning. Proc IEEE 109(1):43–76. https://doi.org/10.1109/JPROC.2020.3004555