Research Article

Prediction of HFMD Cases by Leveraging Time Series Decomposition and Local Fusion

Ziyang Wang,[1] **Zhijin Wang** (ID)**,**[1] **Yingxian Lin** (ID)**,**[1] **Jinming Liu,**[1] **Yonggang Fu,**[1] **Peisong Zhang,**[2] **and Bing Cai**[1]

[1]*Computer Engineering College, Jimei University, Xiamen 361021, China*
[2]*School of Science, Jimei University, Xiamen 361021, China*

Correspondence should be addressed to Zhijin Wang; zhijin@jmu.edu.cn and Yingxian Lin; yxlin@jmu.edu.cn

Hand, foot, and mouth disease (HFMD) is an infection that is common in children under 5 years old. This disease is not a serious disease commonly, but it is one of the most widespread infectious diseases which can still be fatal. HFMD still poses a threat to the lives and health of children and adolescents. An effective prediction model would be very helpful to HFMD control and prevention. Several methods have been proposed to predict HFMD outpatient cases. These methods tend to utilize the connection between cases and exogenous data, but exogenous data is not always available. In this paper, a novel method combined time series composition and local fusion has been proposed. The Empirical Mode Decomposition (EMD) method is used to decompose HFMD outpatient time series. Linear local predictors are applied to processing input data. The predicted value is generated via fusing the output of local predictors. The evaluation of the proposed model is carried on a real dataset comparing with the state-of-the-art methods. The results show that our model is more accurately compared with other baseline models. Thus, the model we proposed can be an effective method in the HFMD outpatient prediction mission.

1. Introduction

Hand, foot, and mouth disease (HFMD) is a common infection caused by a group of viruses. It is likely to occur to children under 5 years old. HFMD causes a serious threat to children's health. Especially in developing Asian countries, this disease is more likely to cause big damage. China is a country with a large population and vast territory, and the development of different regions is uneven. Under this situation, it is difficult to control infectious diseases spread in China. HFMD has been a nationally notifiable disease since 2008. The new cases should be reported in 24 hours. However, the situation is still worsening. According to the data from the Chinese Centre for Disease Control and Prevention (CCDC) [1], nearly 2 million cases were reported in China in 2019, with an incident rate of over 137/100,000. Although most HFMD patients are self-limiting, HFMD can still be fatal. Thus, the prevention and control of HFMD are very important. And if health authorities had anticipated the situation before the outbreak, a lot of unnecessary damage could have been avoided.

Many methods have been proposed to predict HFMD cases. ARIMA is one of the most general time series models, which is already used in HFMD prediction work [2]. ARIMAX is the ARIMA with external parameters added, and study showed that the ARIMAX has better performance than ARIMA [3]. With the increase of computer computing power, multiple learning models are utilized in HFMD prediction, such as LSTM [4], RNN, and CNN-RNN [5]. These methods often attempt to learn the law of the disease spread trend based on a global predictor.

However, on the one hand, the HFMD outpatient data is nonlinear and nonstationary. On the other hand, the spread of HFMD is affected by complex and diverse external factors, such as climate, living habits, and living conditions. These two characteristics make it difficult to improve performance based on a global predictor. The relationship between target data and external factors provides a new idea to researchers,

and many studies focus on prediction using external factors to enhance the model performance have been down. The data about external factors is named exogenous data to distinguish it from target data. In this paper, we use new ideas to improve the accuracy of prediction: time series decomposition and local fusion.

Essentially, decomposition is the process of dividing a complex problem into subproblems that can be easily solved. In our experiments, a classical method named Empirical Mode Decomposition (EMD) is used to decompose the HFMD outpatient data. This method decomposes a time series into several subseries named Intrinsic Mode Function (IMF) and a residual. Each IMF contains a local feature. In addition, in our study, the residual is also treated as an IMF. Each IMF is treated equally by local predictors in the experiment.

In this paper, we propose a Concurrent Autoregression with Decomposition (CARD) model for HFMD prediction. We try to improve the accuracy of prediction as much as possible without exogenous data. CARD generates predicted value by fusing the output of the local predictors. The method utilizes two linear autoregression predictors to process the past outpatient data and the IMFs, respectively. Then, a fusion component fuses the outputs of two linear predictors. Finally, a global predictor is introduced to generate the predicted result. In a word, we propose an effective time series decomposition and local fusion method, which can catch a higher accuracy than several general methods that only use history outpatient data.

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

(1) We propose a novel prediction model, which applied time series decomposition and local fusion to the prediction of outpatient cases of HFMD

(2) A classical decomposition method named EMD is introduced to decompose the HFMD outpatient time series. Compare with several other decomposition methods, EMD is simpler and more efficient in this study

(3) The proposed method applies a linear weighted module to fuse the output of two local predictors. Each local predictor predicts an output result independently. Then, the fusion module trains to generate the final predicted value of the output of local predictors

The rest of the paper is organized as follows. Section 2 introduces related work. The CARD model we proposed is explained in detail in Section 3. Section 4 illustrates the experiment design. Section 5 analyzes the experimental results. Finally, the whole research is summarized in Section 6.

2. Related Work

This section introduces several most commonly used decomposition methods and fusion methods related to our research.

2.1. Decomposition Methods. A time series can be decomposed into several subseries via decomposition methods. For time series decomposition, the following methods are widely used: wavelet transform [6], Robust Seasonal-Trend Decomposition (RobustSTL) [7, 8], EMD [9], and Ensemble Empirical Mode Decomposition (EEMD) [10].

Wavelet transform [6] inherits and develops the idea of localization short-time Fourier transform. Wavelet transform is a local transform not only the frequency but also time can be obtained. The method replaces the basis of Fourier transform. For a signal that has been processed by wavelet transform, both frequency part and specific position in time can be obtained. Compared with Fourier transform, it has good time-frequency localization characteristics and can extract information on signals more effectively.

RobustSTL [7, 8] is a robust method for decomposing complex time series into trend, seasonality, and remainder components. This method allows for multiple seasonal, cyclic components, and multiple linear regressors with constant, flexible, seasonal, and cyclic influence.

EMD [9] is a Fourier transform-based signal decomposition method, which can process any nonlinear and nonstationary signal adaptively. Compared with most of the decomposition methods, EMD is easy to use, since EMD decomposes data based on the local feature of the data, so this method is adaptive and does not require setting up extra parameters in advance.

EEMD [10] is a variant of EMD. For EMD, the extremum points of the signal will affect IMFs, and mode mixing will occur if the distribution of the extremum is uneven. EEMD is proposed to solve the mode mixing problem of EMD. This method using the advantage of uniform distribution of white noise spectrum, the white noise is added to the signal to be analyzed so that the signals of different time scales can be automatically separated to the corresponding reference scales. This method is mainly to add white noise to the signal to supply some missing scale which has good performance in signal decomposition.

In recent years, there are some time series prediction works using time series decomposition in several search areas. A regression model combined with wavelet transform is proposed to forecast the future value of the S&P 500 [11]. EMD is used for electricity load forecasting [12]. Besides, time series decomposition has been applied to disease prediction work. An ensemble model for chickenpox forecast utilizes the STL decomposition to generate the input of the model. Wavelet-ARIMA model got a good performance in COVID-19 case prediction [13]. An improved EEMD algorithm is used to decompose the diarrhea time series [14]. A TDDF model utilizes heterogeneous data to predict the HDMD cases [15].

The HFMD outpatient time series data is applied in our study. The spread of HFMD is easily affected by many external factors. Thus, the processing of the time series is difficult. But the adaptive nature of EMD overcomes this problem. In this paper, we introduce EMD to process our input data.

2.2. Fusion Methods. Time series forecasting has been a subject of interest in several different research areas including

disease control and prevention. In the practical problems of nature, things are not isolated from each other but inextricably connected. The same goes for HFMD. Many studies have fused exogenous data to improve the accuracy of prediction.

The spread of HFMD is influenced by many external factors, such as meteorological factors including temperature, humidity, rapid climate change, local policies, air quality, and population [15–17]. Besides, making good use of some data can help researchers to predict, for instance, the search engine query data [18, 19].

Several methods using exogenous data are collected, and these models can be classified into two categories—stochastic methods and learning methods. Stochastic methods usually combine the past data and exogenous data by a linear method and then learn a linear function to get prediction results [16, 17, 19–23]. The main differences between these methods are the regression of target variables, functions on exogenous data, and the decomposition of exogenous data. In the past few years, exogenous data has been widely used in learning methods. These methods can be roughly divided into the following three categories:

(1) *Traditional Learning Methods Using Exogenous Inputs*. The most common models are multiple linear regression (MLR), support vector regression (SVR), and neural network. For these methods, exogenous data is treated as an input dimension, just like the past data, in which each element of the inputs is equally treated. To prevent data jitter, these methods need to be validated.

(2) The learning methods focus on temporal, which inputs of different categories are differential treatment, such as [24–26]. For these methods, the temporal dynamics of input data is captured to use RNN structures, and a nonlinear mapping from inputs to the target is learned from training data. To differently treat exogenous inputs and target inputs, the encoder-decoder structure is employed to do time series prediction tasks. The encoder-decoder framework consists of two RNN layers and maps input sequence to output sequence [27]

(3) *Temporal Attention Learning Methods*. The attention mechanism is fused into sequential models to predict future values, such as TPA-LSTM [28], DA-RNN [29], HRHN [30], and LSTNet [31]. These models have strong memory abilities in keeping numerous samples. Especially for small-scale infection data, the training loss value would be very small, but the accuracy would be worse than the general methods and is not general enough.

Though the exogenous data can help to improve accuracy, it still has some unavoidable defects. That is, the exogenous data requires a mass of energy to collect and organize and it is unavailable sometimes. Therefore, it is not always wise to do prediction relying on exogenous data, especially in real-time systems. It is almost impossible to integrate required data into the model dynamically. Considering the drawbacks of exogenous data, we discussed above, our attention focuses on target data itself and we do not utilize the exogenous data.

3. The Proposed CARD

This section formulates the problem and illustrates our approach.

Figure 1 shows an overview of the proposed model. The model consists of 3 stages: data preprocessing (left), Concurrent Autoregression with Decomposition (upper right), and data postprocessing (bottom right). For any module in Figure 1, if it makes any changes in the input data, then this module will be connected to the following modules using dotted lines.

In the data preprocessing stage, the input data is the HFMD outpatient. The outpatient data is normalized and then further segmented. Finally, they are decomposed into finite IMFs and residual by EMD. In the CARD, softmax function is introduced to avoid unfairness in feature extraction. Two linear autoregression components are used to mine the sequence feature details and enhance the feature representation of input data. At last, the output of two linear components is fused and another linear component is applied to generate the predicted value. In the data postprocessing stage, the final result is generated and evaluated after denormalization.

3.1. Problem Formulation and Notations. The main notations are explained in Table 1.

Windows size T. A window is a subsequence of the original data. T is the length of the subsequence. And the subsequence is the data in a certain interval be observed to predict the value of future time point.

IMF. If we do not have a termination, the EMD algorithm will loop an infinite number of times. In our experiments, we set a max number of the IMF which is symbolized as K to stop the decomposition.

The problem of this paper can be addressed as the problem of time series prediction missions. A time series is a list of continuous history observation values with equal time intervals. Our goal is to get a predicted value of the outpatient value of the next day.

It is a mapping from the history observation time series and IMFs to the future outpatient value. The symbol $y_t \in \mathbb{R}^1$ is the value at time t. The history observation values with window size T are symbolized as $[y_1, y_2, \cdots, y_T]$. And $\mathbf{D}(y_1, y_2, \cdots, y_T)$ is the matrix obtained by decomposing the windowed time series. $\widehat{y}_{T+1}$ denotes the predicted value at time $T + 1$. The mapping process can be formulated as follows:

$$\widehat{\mathbf{y}}_{T+1} = F(y_1, y_2, \cdots, y_T, \mathbf{D}(y_1, y_2, \cdots, y_T)). \qquad (1)$$

In this study, $[y_1, y_2, \cdots, y_T]$ denotes the HFMD outpatient window size T.

3.2. Data Preprocessing. Normalization. The normalization operation scales the data in a specified range. In order to
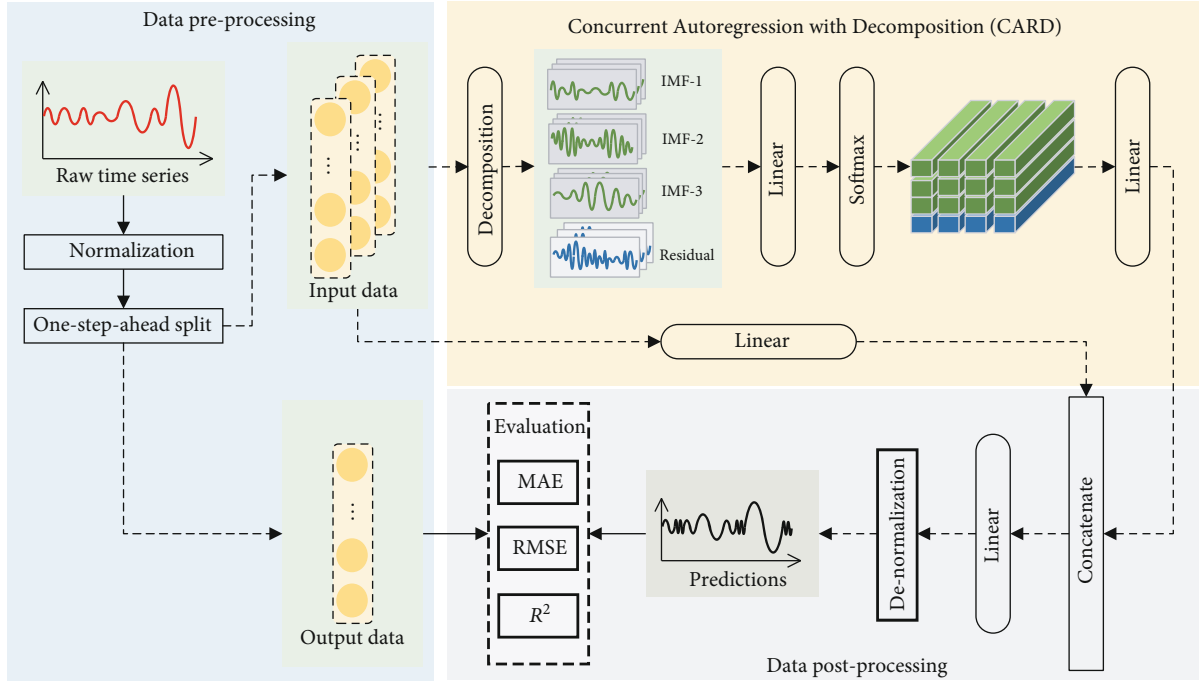
Figure 1: The scheme illustration of the proposed CARD.

Table 1: Notation and semantic.

Notation	Semantic
K	Number of points in the time series
T	Window size
k	Number of IMF
$\mathbf{O}$	Input matrix, $\mathbf{O} \in \mathbb{R}^{1 \times K}$
$\mathbf{Y}$	Matrix after split, $\mathbf{Y} \in \mathbb{R}^{(K-T-1) \times T}$
$\mathbf{X}$	Matrix after decomposition, $\mathbf{X} \in \mathbb{R}^{(K-T-1) \times T \times k}$
$\widehat{\mathbf{Y}}_{T+1}$	Output matrix, $\widehat{\mathbf{Y}}_{T+1} \in \mathbb{R}^{1 \times (K-T)}$

avoid large data dominance caused by the difference of data magnitude, normalization is essentially requisite.

Min-max normalization (0-1 normalization) is a widely use method in time series normalization. It is a linear transformation of the original data, making the result fall into the interval of (0,1). The original data can maintain the difference of value after the linear transformation. Thus, Min-max is suitable to normalize the outpatient time series in our study. The formula of the Min-max normalization is expressed as follows:

$$\mathbf{x}' = \frac{\mathbf{x} - \min{(\mathbf{x})}}{\max{(\mathbf{x})} - \min{(\mathbf{x})}}, \qquad (2)$$

where $\mathbf{x}$ denotes a sample of observed samples, $\mathbf{x}'$ is the normalization result, $\min{(\mathbf{x})}$ is the smallest value in the samples, and $\max{(\mathbf{x})}$ is the biggest.

Segmentation. The purpose of segmentation is to transform time series data into supervise data. For a given time series $\mathbf{Y}$ with K points, the segmentation formula is as follows:

$$\begin{bmatrix} y_1 & y_2 & \cdots & y_T \\ y_2 & y_3 & \ddots & y_{T+1} \\ \vdots & \ddots & \ddots & \vdots \\ y_{K-T-1} & \cdots & \cdots & y_{K-1} \end{bmatrix} \rightarrow \begin{bmatrix} y_{T+1} \\ y_{T+2} \\ \vdots \\ y_K \end{bmatrix}, \qquad (3)$$

where the left matrix is the input data and the right matrix is the output data.

Empirical Mode Decomposition. In this paper, we use EMD [13] to do data decomposition. We perform time series decomposition on the supervision data generated by segmentation. For each sequence, we decompose it into 3 IMFs and a residual.

An IMF must satisfy the requirements as follows:

(1) In any local time scale, the number of extrema and the number of points cross zero must be equal or the difference is 1

(2) At any point, the mean value of the upper envelop defined by the local maxima and the lower envelope defined by local minima is close to 0

The procedures of EMD algorithm are shown in Algorithm 1.

Let $\mathbf{X} = [\text{IMF}_1, \text{IMF}_2, \cdots, \text{IMF}_k, \text{residual}]$ be the decomposed data.

```
Input: The original signal x(t), max-IMF k
Output: k IMFs and a residual
1. i = 1;
2. while i < k do
3. Upper(t) = spline local maxima x(t);
4. Lower(t) = spline local minima x(t);
5. Avg(t) = 1/2(Upper(t) + Lower(t));
6. h(t) = x(t) – Avg(t);
7. if h(t) meets two requirements of IMF do
8. IMF_i = h(t);
9. i + +;
10. x(t) = Avg(t);
11. Residual = x(t)
```

ALGORITHM 1: EMD

3.3. Concurrent Autoregression. The processing of IMF.

We utilize the softmax function to process the IMF future. The softmax function is an extension of the logistic function. This function maps a k-dimensional vector containing any real number to another k-dimensional real-valued vector. Such that each element is in the interval $(0, 1)$, and the sum of all elements is 1. After the process of softmax, the largest value is highlighted and the other components that are far below the maximum value are suppressed. The formula of softmax function is expressed as follows:

$$\boldsymbol{\omega}_i = \frac{e^{x_i}}{\sum_{j=1}^{k} e^{x_j}}, \qquad (4)$$

where $\mathbf{x_i}$ is the output value of the i-th input vector. $\boldsymbol{\omega}$ is the weight matrix. k is the number of output elements. The generation of input memory $\mathbf{a}_i$ is based on the input vector $\mathbf{X}$ and the weight matrix $\boldsymbol{\omega}$. The formula is expressed as follows:

$$\mathbf{a}_i = \boldsymbol{\omega} \times \mathbf{X}. \qquad (5)$$

CARD employs a linear layer to receive a regression result of IMF. The formula is expressed as follows:

$$\mathbf{e}_q = \sum \boldsymbol{\omega}_q \times \mathbf{a}_i + \mathbf{b}_q, \qquad (6)$$

where $\mathbf{e}_q$ is the weighted IMF feature matrix, $\boldsymbol{\omega}_q$ is the weight corresponding to the input dimension, and $\mathbf{b}_q$ is a bias value.

The processing of the HFMD outpatient data. The processing of outpatient data is essentially the same as that of IMFs. The difference is the softmax function is not use for normalization. Only a linear component is applied to analyze the trends in outpatient data. The formula is expressed as follows:

$$\mathbf{e}_y = \sum \boldsymbol{\omega}_y \times \mathbf{y}_i + \mathbf{b}_y, \qquad (7)$$

where $\mathbf{e}_y$ is the weighted outpatient feature matrix, $\boldsymbol{\omega}_y$ is the weight corresponding to the input dimension, and $\mathbf{b}_y$ is a bias value.

Concatenation. The CARD model combines the output of two concurrent working components by the cat function in PyTorch. The data is treated as the input for last linear module, and finally, this module generates a predicted value $\widehat{y}_{T+1}$. The generation of $\widehat{y}_{T+1}$ is formulated as follows:

$$\widehat{\mathbf{y}}_{T+1} = \Phi \big(\boldsymbol{\omega} \big[\mathbf{e}_y; \mathbf{e}_q \big] + \mathbf{b} \big), \qquad (8)$$

where $[\mathbf{e}_y; \mathbf{e}_q]$ is the concatenated vector of dual side outputs, $\boldsymbol{\omega}$ is the weight of outputs from dual represented sources, $\widehat{\mathbf{y}}_{T+1}$ is the predicted value of the outpatient number in the next day, $\mathbf{b}$ is a bias value, and Φ represents the activation function.

3.4. Data Postprocessing. Denormalization.

Denormalization is an inverse procedure of normalization, and the denormalization formula is applied to generate the final prediction results acquired from our model. The formula is expressed as follows:

$$\mathbf{x} = \mathbf{x}' \times (\max(\mathbf{x}) - \min(\mathbf{x})) + \min(\mathbf{x}). \qquad (9)$$

The detail steps of the proposed CARD are shown in Algorithm 2.

4. Experimental Setup

This section configures our experiments. Section 4.1 introduces the dataset we use. Section 4.2 gives three evaluation metrics. And Section 4.3 presents the implementation of our model and the baseline models for comparison. All experiments are proceeding with the real-world HFMD outpatient case time series data.

4.1. Data.

The real dataset we applied in our experiments is HFMD outpatient case data which is collected from the Xiamen Center for Disease Control and Prevention (XCDC). This dataset is the daily record data from January 1, 2012, to December 30, 2018. A total of 2555 sample points are included. In Figure 2, the time series is shown at one-year intervals.

4.2. Metric.

To measure the performance of our proposed model and compare our model with the selected baseline models, 3 widely used standard methods are adopted in our experiments, and the formulas are defined as follows:

$$\text{MAE} = \frac{1}{T} \sum_{t=1}^{T} (|y_t - \widehat{y}_t|), \qquad (10)$$

$$\text{RMSE} = \sqrt{\frac{1}{T} \sum_{t=1}^{T} (y_t - y\wedge_t)^2}, \qquad (11)$$

$$R^2 = 1 - \frac{\sum_{t=1}^{T} (y_t - y\wedge_t)^2}{\sum_{t=1}^{T} (\bar{y}_t - y\wedge_t)^2}. \qquad (12)$$

In these equations, the parameter y_t is the real

Input: Observed HFMD outpatient time series $\mathbf{O} \in \mathbb{R}^{1 \times K}$, window size T, max-IMF k
Output: Prediction value for future cases $\widehat{\mathbf{Y}}_{T+1} \in \mathbb{R}^{1 \times (K-T)}$
1. $\mathbf{O}' \leftarrow$ normalization $\mathbf{O}$ using Equation (2);
2. $\mathbf{Y} \leftarrow$ Split using Equation (3);
3. $\mathbf{X} \leftarrow$ Time series decomposition using Algorithm 1;
4. for each sample $\mathbf{y}$ in $\mathbf{Y}$ do
5.   for $i \leftarrow 1$ do
6.   for $j \leftarrow 1$ do
7. $\mathbf{e}_y \leftarrow \mathbf{y}_{i,j}$ using Equation (7);
8. for each sample $\mathbf{x}$ in $\mathbf{X}$ do
9   for $i \leftarrow 1$ do
10   for $j \leftarrow 1$ do
11.   for $n \leftarrow 1$ do
12. $\boldsymbol{\omega} \leftarrow$ softmax $\mathbf{x}_{i,j}$ using Equation (4);
13. $\mathbf{A} \leftarrow \boldsymbol{\omega}$ and $\mathbf{X}$ using Equation (5);
14. $\mathbf{e}_x \leftarrow \mathbf{A}$ using Equation (6);
15. $\widehat{\mathbf{y}}'_{T+1} \leftarrow \mathbf{e}_x$ and $\mathbf{e}_y$ using Equation (8);
16. $\widehat{\mathbf{y}}_{T+1} \leftarrow$ denormalization $\widehat{\mathbf{y}}'_{T+1}$ using Equation (9);
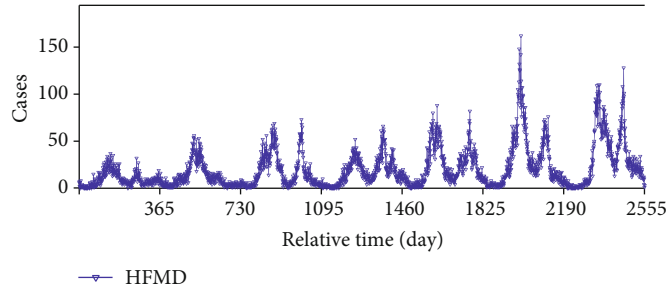
ALGORITHM 2: CARD

FIGURE 2: The distribution of outpatient cases ranges from Jan 1, 2012, to Dec 30, 2018.

observation value at time t, and $\widehat{y}_t$ is the predicted value predicted by model at the same time.

MAE is a basic and universal metric in regression mission. Compared with MAE, RMSE has the same degree as the data. For R^2, the denominator is understood as the dispersion degree of the original data, and the molecule is the error between the predicted data and the original data. The division of the two can eliminate the influence of the dispersion degree of the original data. These three metrics can be used together to evaluate the performance of the model comprehensively and objectively.

4.3. Configuration. Parameter settings. In our experiments, the target data is divided into two parts: training set (80%) and test set (20%). The batch size is set to 32. A set of experiments are completed to find the best values of window size, and the results are shown in Figure 3, and as we can see, the best performance is achieved when $T = 10$. For each experiment, we chose the learning rate between 0.0005 and 0.002 for a step 0.0005 to acquire the best performance of every model. We repeat each set of experiments five times and take the average value to obtain the final result. Thus, the result is stable and has a high level of credibility.

Decomposition algorithm. RobustSTL is more suitable for long-time series processing, and the time series we used is too short for this method. Therefore, we only consider wavelet transform, EMD, and EEMD as the time series decomposition method candidates. There are four experiments for comparison that have been done, and the results are shown in Figure 4. "IMF3" means the original data will be decomposition into three IMFs and one residual. Both "db" and "sym" are commonly used wavelet basis functions. "db" is the abbreviation of Daubechies, and "db2" represents a wavelet of order 2. "sym" is symlets and "sym3" means a wavelet of order 3. As we have seen in Figure 4, the wavelet transform-based approach has a time advantage, while the EEMD algorithm consumes too much computational time. Although the EMD algorithm is at a time disadvantage, it takes the lead in three metrics. Thus, this configure is applied in the formal experiments.

Baseline models. To verify the effectiveness of the EMD function and local fusion, experiments based on multiple models are necessary. MLR [32], LSTM [33], GRU [34], ED
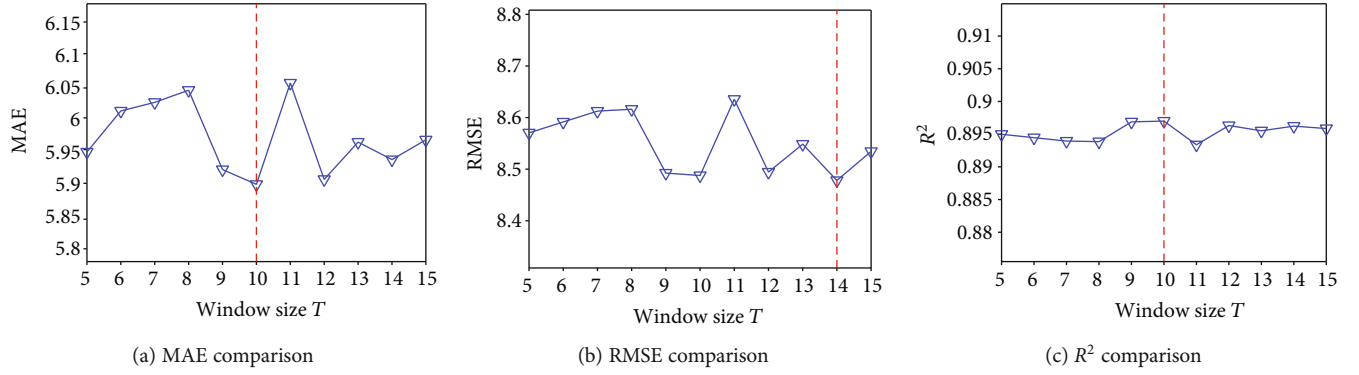
(a) MAE comparison

(b) RMSE comparison

(c) R^2 comparison

FIGURE 3: The comparison of CARD performance at different values of T in terms of MAE, RMSE, and R^2.

(a) MAE comparison

(b) RMSE comparison

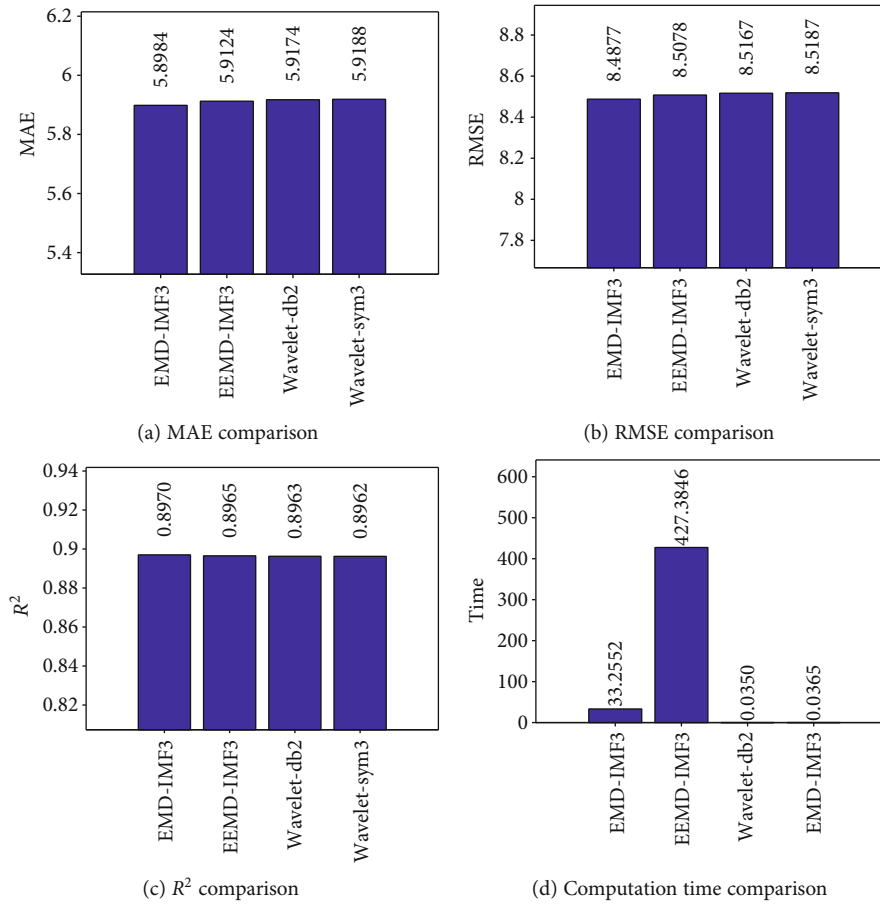(c) R^2 comparison

(d) Computation time comparison

FIGURE 4: The comparison of four methods in terms of MAE, RMSE, R^2, and computation time.

[35], CNN-1d, and CNN-RNN [36] are selected as the baseline models. To explain, MLR is a widely used regression model in many research areas. LSTM, GRU, and ED are improved neural networks based on RNN. CNN-RNN is a hybrid model of CNN and RNN.

Experiment process. All experiments could be divided into 3 groups. The main difference between the three groups of experiments is their input data. The input data of the first group of experiments contains only historical data, and the second group uses only the data after time series decomposition. The last group of experiments takes both historical and decomposed data as input. As a result, the final group has the best performance. Details are discussed in the next section.

5. Results and Analysis

This section gives prediction results, comparisons, and analyses.

5.1. Effects on Decomposition and Fusion. In this subsection, we investigate the effects of decomposition and fusion. As we can see in Figure 5, the result of three metrics shows that
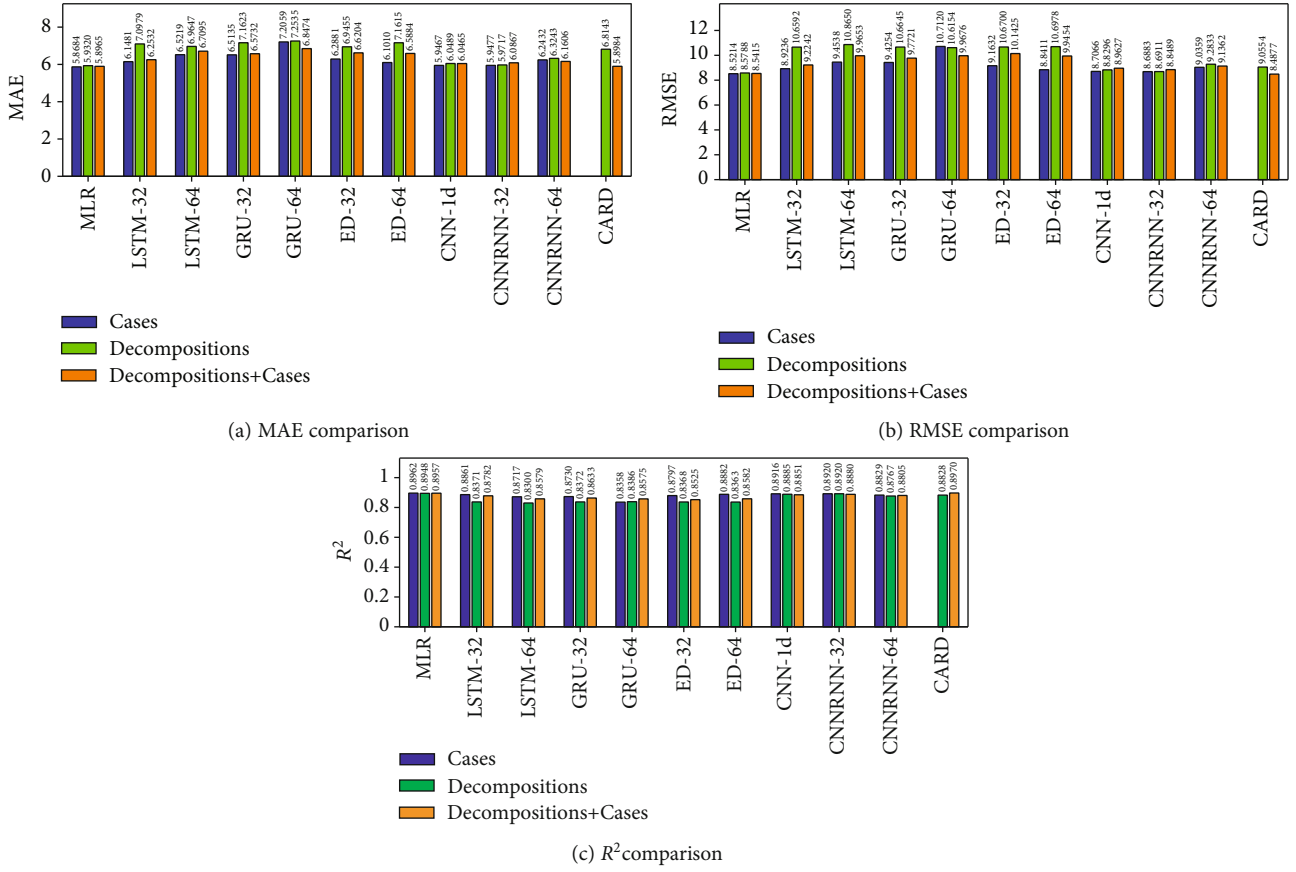
(a) MAE comparison

(b) RMSE comparison

(c) R^2 comparison

FIGURE 5: The comparison of eleven methods on three groups of inputs in terms of MAE, RMSE, and R^2. The windows size T is fixed at 10.

the CARD model we proposed has the ideal performance compared with baseline models.

Time series decomposition is an important part of this study. We decompose the HFMD outpatient time series into finite and multitime scale IMFs and a residual; then, each subsequence is modeled and predicted with a local linear predictor separately. The single IMF contains a specific physical meaning, such as seasonality and trend. Each sequence is treated equally in the model. Compared with the original data, each IMF can represent the local features by itself. This means that separate predictions for each sequence and then fusion may give better results than using only the raw data, and the experimental results proved this.

The prediction accuracy of all baseline models has increased after fusing the HFMD outpatient case data. This result shows the superiority of data fusion. A possible explanation is that existing models do not work well with complex time series like IMF, and some methods cannot capture the relation between different sequences. The data processing of the CARD can be divided into two stages. In the first stage, each sequence is predicted separately, and then, the results are fused. In the second stage, the predicted values are obtained from the fused data, which can analyze the relationship between each sequence. So, we get better results than other models. By the way, the IMFs may lose some features in the original data. These defects are more obvious with the short and complex time series data. However, the fusion of IMF and case data overcomes this shortage. That may

explain why all models have various degrees of improvement after fusion.

5.2. Comparison. The main results are shown in Figure 5. The major results can be observed, and the analysis of them is as follows:

Out of all the models, CARD performs the best. In detail, MLR is the second-best model. Compared to MLR, CARD is slightly behind in MAE and RMSE, and we are slightly ahead in R^2. In addition, we are at least 0.1, 0.4, and 0.1 ahead of the other models in three metrics. The advantages of our model are described in Section 5.1, and these should explain the leading position of our model.

In the experiments using only decomposed data as input, several baseline models showed various degrees of degradation in performance. And their performance is improved if the outpatient data is added to the dataset. However, as we can see in Figure 5, the best performance of these models is still obtained in the first set of experiments—the input is outpatient data. In contrast, the performance of CNN-1d and MLR shows only small fluctuations. One possible explanation is that the EMD algorithm filters peaks in the time series while CNN-1d and MLR are insensitive to peaks. Therefore, the accuracy of these two models is not affected much. LSTM, ED, and GRU study the dependence of time series, and since these models cannot capture the relationship between the series, the IMF may negatively affect the prediction accuracy. CARD performs weighting at each time point and predicts

each IMF separately. Finally, the model generates a result by fusion. Thus, CARD solves these problems and obtains better performance.

Although the CARD model does not make revolutionary advances, however, the model is much less computationally intensive compared to most neural network models. Therefore, the model has relatively low hardware requirements. Moreover, this model still has good predictive performance when using only historical data, which means that the data needed to run the model is easily available. This further lowers the threshold for practical using the model. Therefore, our proposed model has good prospects for practical applications.

6. Conclusions

Our experiment indicates that data decomposition and local fusion can improve prediction performance. In this paper, we propose a time series decomposition and local fusion model named CARD for HFMD outpatient case prediction. The main conclusions of this study are shown as follows:

(1) Compared with wavelet transform and EEMD, the EMD method has advantages in predicting accuracy in terms of HFMD outpatient prediction. Therefore, EMD is suitable for HFMD outpatient time series

(2) The fusion model we proposed is superior to the most general methods, which means that such a model still has great potential in infectious disease forecasting

Our study must go further research. In this paper, we do not test the predicting accuracy on the multistep prediction. In the next step, we can try to extend our model to multistep times series prediction and other diseases.

Data Availability

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

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

References

[1] C C for Disease Control, Prevention, "The general situation of notifiable infectious diseases in china in 2019," 2020, 2020, http://www.nhc.gov.cn/jkj/s3578/202004/b1519e1bc1a944fc8ec176db600f68d1.shtml.

[2] L. Liu, R. Luan, F. Yin, X. Zhu, and Q. Lu, "Predicting the incidence of hand, foot and mouth disease in Sichuan province, China using the ARIMA model," *Epidemiology and Infection*, vol. 144, pp. 144–151, 2016.

[3] Z. Du, L. Xu, W. Zhang, D. Zhang, S. Yu, and Y. Hao, "Predicting the hand, foot, and mouth disease incidence using search engine query data and climate variables: an ecological study in Guangdong, China," *BMJ Open*, vol. 7, no. 10, article e016263, 2017.

[4] J. Gu, L. Liang, H. Song et al., "A method for hand-foot-mouth disease prediction using GeoDetector and LSTM model in Guangxi, China," *Scientific Reports*, vol. 9, no. 1, p. 17928, 2019.

[5] Y. Wang, Z. Cao, D. Zeng, X. Wang, and Q. Wang, "Using deep learning to predict the hand-foot-and-mouth disease of enterovirus A71 subtype in Beijing from 2011 to 2018," *Scientific Reports*, vol. 10, no. 1, p. 12201, 2020.

[6] A. Grossmann and J. Morlet, "Decomposition of hardy functions into square integrable wavelets of constant shape," *SIAM Journal on Mathematical Analysis*, vol. 15, no. 4, pp. 723–736, 1984.

[7] Q. Wen, J. Gao, X. Song, L. Sun, H. Xu, and S. Zhu, "RobustSTL: a robust seasonal-trend decomposition algorithm for long time series," in *Proceedings of the 33rd International Conference on Artificial Intelligence, AAAI Press*, pp. 5409–5416, Honolulu, Hawaii, USA, 2019.

[8] C. Robert, C. William, and T. Irma, "STL: a seasonal-trend decomposition procedure based on loess," *Journal of Official Statistics*, vol. 6, no. 1, pp. 3–73, 1990.

[9] N. E. Huang, Z. Shen, S. R. Long et al., "The empirical mode decomposition and the Hilbert spectrum for nonlinear and non-stationary time series analysis," *Proceedings of the Royal Society of London. Series A: Mathematical, Physical and Engineering Sciences*, vol. 454, no. 1971, pp. 903–995, 1998.

[10] Z. Wu and N. E. Huang, "Ensemble empirical mode decomposition: a noise-assisted data analysis method," *Advances in Adaptive Data Analysis*, vol. 1, no. 1, pp. 1–41, 2009.

[11] K. Zhang, R. Gencay, and M. Ege Yazgan, "Application of wavelet decomposition in time-series forecasting," *Economics Letters*, vol. 158, pp. 41–46, 2017.

[12] Y. Yaslan and B. Bican, "Empirical mode decomposition based denoising method with support vector regression for time series prediction: a case study for electricity load forecasting," *Measurement*, vol. 103, pp. 52–61, 2017.

[13] S. Singh, K. S. Parmar, J. Kumar, and S. J. S. Makkhan, "Development of new hybrid model of discrete wavelet decomposition and autoregressive integrated moving average (ARIMA) models in application to one month forecast the casualties cases of COVID-19," *Chaos, Solitons & Fractals*, vol. 135, p. 109866, 2020.

[14] Y. Wang, J. Gu, Z. Zhou, and Z. Wang, "Diarrhoea outpatient visits prediction based on time series decomposition and multi-local predictor fusion," *Knowledge-Based Systems*, vol. 88, pp. 12–23, 2015.

[15] Z. Wang, Y. Huang, B. He, T. Luo, Y. Wang, and Y. Lin, "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*, pp. 658–667, Dalian, China, 2019.

[16] J. Wang, Y. Xiao, and R. A. Cheke, "Modelling the effects of contaminated environments on HFMD infections in mainland China," *Biosystems*, vol. 140, pp. 1–7, 2016.

[17] Z. Yang, J. Hao, S. Huang et al., "Acute effects of air pollution on the incidence of hand, foot, and mouth disease in Wuhan, China," *Atmospheric Environment*, vol. 225, p. 117358, 2020.

[18] P. Wang, W. B. Goggins, and E. Y. Y. Chan, "Hand, foot and mouth disease in Hong Kong: a time-series analysis on its relationship with weather," *PLoS One*, vol. 11, no. 8, article e0161006, 2016.

[19] D.-C. Huang and J.-F. Wang, "Monitoring hand, foot and mouth disease by combining search engine query data and meteorological factors," *Science of the Total Environment*, vol. 612, pp. 1293–1299, 2018.

[20] S. Chen, X. Liu, Y. Wu et al., "The application of meteorological data and search index data in improving the prediction of HFMD: a study of two cities in Guangdong Province, China," *Science of The Total Environment*, vol. 652, pp. 1013–1021, 2019.

[21] C. Song, X. Shi, Y. Bo, J. Wang, Y. Wang, and D. Huang, "Exploring spatiotemporal nonstationary effects of climate factors on hand, foot, and mouth disease using Bayesian spatiotemporally varying coefficients (STVC) model in Sichuan, China," *Science of the Total Environment*, vol. 648, pp. 550–560, 2019.

[22] Y. Yang, E. You, J. Wu et al., "Effects of relative humidity on childhood hand, foot, and mouth disease reinfection in Hefei, China," *Science of the Total Environment*, vol. 630, pp. 820–826, 2018.

[23] Z. Du, W. R. Lawrence, W. Zhang, D. Zhang, S. Yu, and Y. Hao, "Interactions between climate factors and air pollution on daily HFMD cases: a time series study in Guangdong, China," *Science of The Total Environment*, vol. 656, pp. 1358–1364, 2019.

[24] Z. Wang, Y. Huang, B. He et al., "Short-term infectious diarrhea prediction using weather and search data in Xiamen, China," *Scientific Programming*, vol. 2020, Article ID 8814222, 12 pages, 2020.

[25] Z. Wang, Y. Huang, B. Cai, R. Ma, and Z. Wang, "Stock turnover prediction using search engine data," *Journal of Circuits, Systems and Computers*, vol. 30, no. 7, 2020.

[26] Y. Wang, J. Li, J. Gu, Z. Zhou, and Z. Wang, "Artificial neural networks for infectious diarrhea prediction using meteorological factors in Shanghai (China)," *Applied Soft Computing*, vol. 35, pp. 280–290, 2015.

[27] I. Sutskever, O. Vinyals, and Q. V. Le, "Sequence to sequence learning with neural networks," in *Proceedings of 28th International Conference on Neural Information Processing Systems, Curran Associates, Inc. Palais des Congrès de Montréal*, pp. 3104–3112, Montréal Canada, 2014.

[28] S. Shih, F. Sun, and H. Lee, "Temporal pattern attention for multivariate time series forecasting," *Machine Learning*, vol. 108, no. 8-9, pp. 1421–1441, 2019.

[29] Y. Qin, D. Song, H. Chen, W. Cheng, G. Jiang, and G. W. Cottrell, "A dual-stage attention-based recurrent neural network for time series prediction," in *Proceedings of the 26th International Joint Conference on Artificial Intelligence, International*

Joint Conferences on Artificial Intelligence, pp. 2627–2633, Melbourne, Australia, 2017.

[30] Y. Tao, L. Ma, W. Zhang, J. Liu, W. Liu, and Q. Du, "Hierarchical attention-based recurrent highway networks for time series prediction," https://arxiv.org/abs/1806.00685.

[31] G. Lai, W. Chang, Y. Yang, and H. Liu, "Modeling long- and short-term temporal patterns with deep neural networks," in *Proceedings of the 41st International Conference on Research Development in Information Retrieval, ACM*, pp. 95–104, Ann Arbor, MI, USA, 2018.

[32] P. Wang and M. L. Puterman, "Mixed logistic regression models," *Journal of Agricultural, Biological, and Environmental Statistics*, vol. 3, no. 2, pp. 175–200, 1998.

[33] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural Computation*, vol. 9, no. 8, pp. 1735–1780, 1997.

[34] J. Chung, C. Gulcehre, K. Cho, and Y. Bengio, "Empirical evaluation of gated recurrent neural networks on sequence modeling," https://arxiv.org/abs/1412.3555.

[35] K. Cho, B. van Merrienboer, C. Gulcehre et al., "Learning phrase representations using RNN encoder-decoder for statistical machine translation," https://arxiv.org/abs/1406.1078.

[36] Y. Wu, Y. Yang, H. Nishiura, and M. Saitoh, "Deep learning for epidemiological predictions," in *The 41st International ACM SIGIR Conference on Research and Development in Information Retrieval, Association for Computing Machinery*, pp. 1085–1088, New York, NY, USA, 2018.