

# A Novel Hybrid Approach for Diarrhea Prediction

Yongming Wang

Department of Computer Science & Technology  
East China Normal University  
Shanghai, China  
ymwang819@gmail.com

Junzhong Gu

Department of Computer Science & Technology  
East China Normal University  
Shanghai, China  
jzgu@ica.stc.sh.cn

**Abstract**—Accurate and reliable forecasts of diarrhea incidences are necessary for the health authorities to ensure the appropriate action for the control of the outbreak. In this paper, a novel hybrid model known as EEMD-GRNN is proposed to forecast the diarrhea incidences. The proposed approach first uses Ensemble Empirical Mode Decomposition (EEMD), which can adaptively decompose the complicated raw time series data into a finite set of intrinsic mode functions (IMFs) and a residue, which have simpler frequency components and higher correlations. The IMF components and residue are then modeled and forecasted using GRNN and the final prediction result can be obtained by these prediction results using the principle of ensemble. The proposed hybrid method is examined by predicting the monthly diarrhea cases number of children and adult located in Shanghai of China. The experimental results indicate that the proposed EEMD-GRNN model provides more accurate forecasts compared to the other ARIMA, single GRNN models and hybrid model (EMD-GRNN). Overall, the proposed approach was effective in improving the prediction accuracy.

**Keywords**—Diarrhea prediction; Ensemble empirical mode decomposition; Generalized regression neural network; Hybrid approach

## I. INTRODUCTION

An accurate and timely diarrhea prediction is crucial for predicting future health events or situations such as demands for health services and healthcare needs. As a kind of common and important infectious disease, diarrhea has a serious threat to human health and leads to one billion disease episodes and 1.8 million deaths each year (WHO, 2008). Hence, a robust prediction model for diarrhea facilitates preventive medicine and health care intervention strategies, by pre-informing health service providers to take appropriate mitigating actions to minimize risks and manage demand [1].

Over the past couple of decades, there have been wide attempts to capture the relationship between the available information using some straightforward linear regression assumptions, for example, the autoregressive integrated moving average (ARIMA). However, currently there is no evidence to support the assumption that the relationship between the past and future of diarrhea is a perfectly linear one. Many recent studies focus on the use of machine learning techniques, such as artificial neural networks (ANNs), to build a prediction model. Unlike traditional statistical models, ANNs are data-driven models. They do not require strong model assumptions and can map any nonlinear function without a

priori assumption about the properties of the data, even though the underlying relationships are unknown or hard to describe. Related works have shown that machine learning techniques outperform many traditional models.

In this paper we develop predictors using generalized regression neural networks (GRNNs) [2], a special type of neural networks. GRNN has only a single design parameter and is simple and fast in training. When using GRNN for diarrhea prediction, the observed original values of prediction variables are usually directly used for building prediction models. However, many factors underlie the diarrhea such as seasonal variations. Due to the complexity of the diarrhea incidence, it is difficult to capture its non-stationary property and accurately describe its moving tendency.

Empirical Mode Decomposition (EMD) [3] is a kind of adaptive signal decomposition technique using the Hilbert-Huang transform and can be applied with nonlinear and non-stationary time series. However, EMD suffers from an intrinsic drawback—the frequent appearance of mode mixing. Fortunately, there exists an improved method called Ensemble EMD (EEMD) which makes up for the deficiency of EMD. Different from other traditional decomposition methodologies such as wavelet decomposition, EEMD is an empirical, intuitive, direct and self-adaptive data processing method created especially for non-linear and non-stationary signal sequences. Therefore, the EEMD has been widely used in many fields [4-6]. However, existing literatures regarding diarrhea prediction have not adopted EEMD processes, and this study will be to fill this gap.

In this paper, we introduce EEMD and GRNN to predict the monthly number of diarrhea cases. A novel hybrid prediction algorithm called EEMD-GRNN is proposed. The proposed approach was compared with the EMD-GRNN, single GRNN approaches and traditional time series models, such as ARIMA, thus demonstrating that the proposed model is substantially featured with an excellent prediction capacity. Moreover, in order to evaluate the performance of the proposed approach, the real world diarrhea datasets are used as an illustrative example.

The rest of this paper is organized as follows. Section 2 reviews related methods used in this paper which are EEMD and GRNN. The proposed model is described in Section 3. Section 4 presents the experimental results and the

effectiveness of the proposed methodology is discussed. Finally, Section 5 concludes the paper.

## II. METHODOLOGY

### A. Empirical Mode Decomposition

The basic idea of EMD is to identify the intrinsic oscillatory modes and to decompose original time series data into a finite and small number of oscillatory modes based on the local characteristic time scale by itself [3]. The decomposition is based on the following assumptions [6]: (1) the signal has at least two extreme-one maximum and one minimum; (2) the characteristic time scale is defined by the time lapse between the extreme; and (3) if the data are totally devoid of extreme but contain only inflection points, then they can be differentiated one or more times to reveal the extreme. Final results can be obtained by integration of the components. With the assumptions of decomposition, an original data series  $X(t)$  ( $t=1; 2, \dots, T$ ) can be decomposed in terms of the following sifting procedure. The detailed process of the EMD algorithm is shown as follows [3, 14-16]:

Step 1: Identify local extreme in the data  $\{x(t)\}$ .

All the local maxima are connected by a cubic spline line  $U(t)$ , which forms the upper envelope of the data. Repeat the same procedure for the local minima to produce the lower envelope  $L(t)$ . Both envelopes will cover all the data between them. The mean of upper envelope and lower envelope  $m_1(t)$  is given by:

$$m_1(t) = U(t) + L(t) / 2 \quad (1)$$

Subtracting the running mean  $m_1(t)$  from the original time series  $x(t)$ , we get the first component  $h_1(t)$ :

$$h_1(t) = x(t) - m_1(t) \quad (2)$$

The resulting component  $h_1(t)$  is an IMF if it is symmetric and have all maxima positive and all minima negative. An additional condition of intermittence can be imposed here to sift out wave forms with certain range of intermittence for physical consideration. If  $h_1(t)$  is not an IMF, the sifting process has to be repeated as many times as it is required to reduce the extracted signal to an IMF. In the subsequent sifting process steps,  $h_1(t)$  is treated as the data to repeat steps mentioned above:

$$h_{11}(t) = h_1(t) - m_{11}(t) \quad (3)$$

Again, if the function  $h_{11}(t)$  does not yet satisfy criteria for IMF, the sifting process continues up to  $k$  times until some acceptable tolerance is reached:

$$h_{1k}(t) = h_{1(k-1)}(t) - m_{1k}(t) \quad (4)$$

Step 2: If the resulting time series is an IMF, it is designated as  $c_1 = h_{1k}(t)$ . The first IMF is then subtracted from the original data, and the difference  $r_1$  given by:

$$r_1(t) = x(t) - c_1(t) \quad (5)$$

The residue  $r_1(t)$  is taken as if it were the original data, and we apply to it again the sifting process of Step 1.

Following the above procedures, we continue the process to find more intrinsic modes  $c_i$  until the last one. The final residue will be a constant or a monotonic function which represents the general trend of the time series. Finally we obtain:

$$x(t) = \sum_{i=1}^n c_i(t) + r_n, \quad (6)$$

$$r_{n-1}(t) - c_1(t) = r_i(t)$$

Where  $r_n$  is a residue. Thus, residue  $r_n(t)$  is the mean trend of  $x(t)$ . The IMFs  $\{c_1(t), c_2(t), \dots, c_n(t)\}$  include different frequency bands ranging from high to low. The frequency components contained in each frequency band are different and they change with the variation of time series  $x(t)$ , while  $r_n(t)$  represents the central tendency of time series  $x(t)$ .

### B. Ensemble Empirical Mode Decomposition

The EEMD [17] is the inheritor of the EMD. EEMD defines the true IMF components as the mean of the corresponding IMFs obtained via EMD over an ensemble of trials and generated by adding different realizations of white noise of finite variance to the original signal  $x[n]$ . The added white noise can help extract the true IMFs, and can offset them via ensemble averaging after serving their purpose [8]. Therefore, this can substantially reduce the chance of mode mixing and represent a significant improvement over the original EMD. The process of EEMD decomposition can be demonstrated by the following steps:

- 1) Add a white noise series to the original time series dataset;
- 2) Decomposition the data with added white noise into IFMs using the EMD procedure;
- 3) Repeat the step 1 and 2 iteratively, but use different white noise each time
- 4) Obtain the ensemble means of corresponding IMFs as the final results.

### C. Generalized Regression Neural Network

The GRNN [7] is a kind of radial basis function networks which is based on a standard statistical technique called kernel regression. A typical GRNN is organized using four layers, namely the input layer, the pattern layer, the summation layer, and the output layer. The hidden layer has radial basis neurons, while neurons in the output layer have a linear transfer function. A typical architecture of the GRNN is presented in Fig. 1. Given a sufficient number of neurons, GRNN can approximate a continuous function to an arbitrary accuracy [8].

Given  $m$  input-output pairs  $\{X, Y\} \in \mathfrak{R}^n \times \mathfrak{R}^1$  and as the training samples, assume the original design of GRNN, that is, the number of hidden neurons is equal to the number of training samples. For a desired estimate of system output vectors  $Y$ , under the input vectors  $X$ , is achieved by a regression calculation. The procedure of the GRNN model can be represented as:

$$Y(X) = E[Y/X] = \frac{\int_{-\infty}^{\infty} Yf(Y, X)dX}{\int_{-\infty}^{\infty} f(Y, X)dX} \quad (7)$$

Where  $X$  is a  $d$ -dimensional input vector,  $Y$  is the predicted value of the GRNN model,  $E[Y/X]$  is the expected value of the output  $Y$ , given the input vector  $X$ ,  $f(Y, X)$  is the joint probability density function of  $X$  and  $Y$ . When probability density function adopt Gaussian function, the network output function of  $Y$  given the vector  $X$ :

$$Y(X) = \frac{\sum_{i=1}^n Y_i \exp(-D_i^2 / 2\sigma^2)}{\sum_{i=1}^n \exp(-D_i^2 / 2\sigma^2)} \quad (8)$$

Where  $D_i^2$  is defined as  $D_i^2 = (X - X_i)^T (X - X_i)$ ,  $\sigma$  denotes the smoothing parameter,  $X$  is the input variable of the network,  $X_i$  is a specific training vector of the neuron  $i$  in the pattern layer.

A good performance for GRNN method depends on smoothing factor  $\sigma$ , which is very important in using GRNN for prediction and determines the generalization capability of the GRNN. The smoothing factor is only free (adaptive) parameter, apart from the input and output layer, involved in the designing of the network.

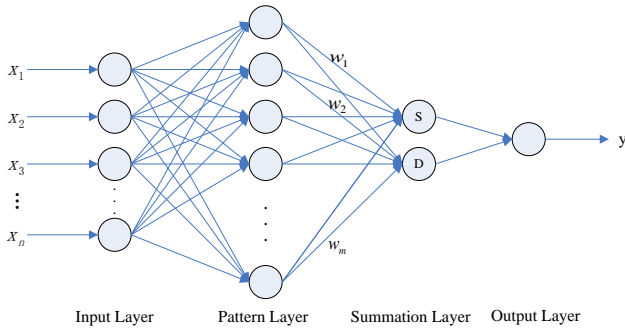


Fig. 1. Typical GRNN structure.

GRNN have several advantages [9], including: 1) it has one design parameter (smoothing factor); 2) it is easy to train since it is a one-pass algorithm; 3) it can accurately approximate functions from sparse and noisy data; 4) it can converge to the conditional mean surface by increasing the number of data samples; and 5) ability to model from a relatively small data set, and ability to handle outliers. It is these unique advantages that make us to choose GRNN as local models for each IFM.

From a time series prediction point of view, the purpose of GRNN is to define a function that produces outputs as close as possible to the actual values over the prediction horizon. Given a training set of  $T$  data points  $\{(x_i, y_i) | x_i \in \mathfrak{R}^d, y_i \in \mathfrak{R}\}_{i=1}^T$ , the GRNN try to construct a predictor function expressed by  $y=f(x)$ , where  $f(\cdot) : \mathfrak{R}^d \rightarrow \mathfrak{R}$  is the predictor function.

### III. PROPOSED EEMD-GRNN APPROACH

Considering the aforementioned points in section II, in the current research the powerful combination of positive aspects

of EEMD and GRNN algorithm is presented to one-step-ahead diarrhea time series prediction problem. As shown in Fig. 2, the proposed EEMD-GRNN modeling framework is generally composed of the following three main steps:

#### Step 1: Decompose time series by EEMD

The original diarrhea time series are first decomposed into a finite and often a small number of intrinsic mode functions (IMFs) and a residue using EEMD technique (here the residual  $r_{n+1}(t)$  also be considered as an IMF).

#### Step 2: Local-GRNN predictor construction

After the components (IMFs and a residue) are adaptively extracted via EEMD, each IFM component is modeled by an independent GRNN which are used to generate local predictor to forecast the component series respectively.

#### Step 3: Muti-local GRNN predictor ensemble

The forecasts of all IFM components are aggregated using another independent GRNN model, which model the relationship among the IMFs and the residue, to produce an ensemble forecasts for the original time series.

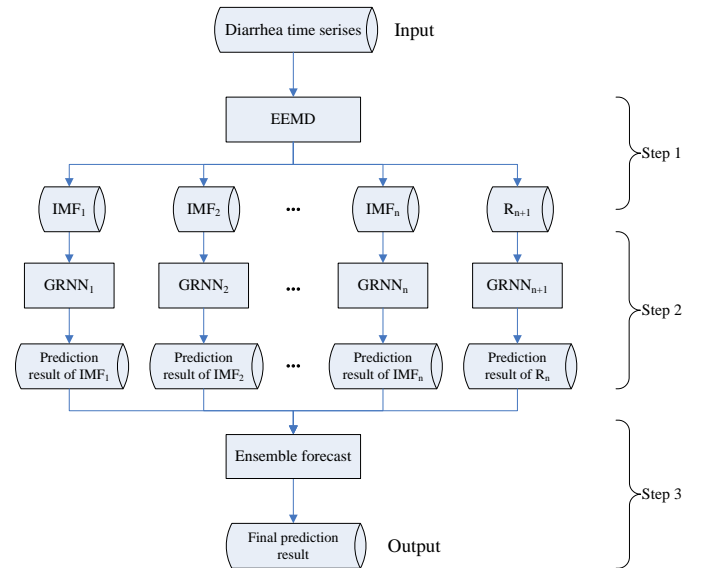


Fig. 2. The EEMD-GRNN modeling framework.

Several studies, for example [10, 11], have indicated that selecting model inputs is probably the most critical task for a time series prediction model, since it contains important information embedded in the data. The statistical approach to examine partial-auto-correlation function (PACF) of the time series was recognized as a good and parsimonious method in the determination of model inputs [12, 13]. So, in this study, the model inputs in the approach are mainly determined by the plot of PACF. After determining the relationship between input(s) and output(s), the input/output pairs can be constructed for each IFM component.

The performance of GRNN is mainly affected by the smoothing factor  $\sigma$ . There are no general rules for the choice of smoothing factor. In this study, the optimal smoothing factors for each local predictor are determined by the trial-and-error

method. Normalization required for neural network modeling in general is also included in our preprocessing. Thus, the inputs are normalized by the method of maximum and minimum normalization; after simulation, the corresponding estimate results are rescaled through the contrary process of the employed normalization method.

Through EEMD, different characteristics information of original time series can be displayed on different scales, and the proposed method can more fully capture the local fluctuations of raw data. Moreover, each IMF component has similar frequency characteristics, simple frequency components, and strong regularity, therefore allowing this model to reduce the complexity of local GRNN modeling and further improve GRNN prediction efficiency and accuracy.

#### IV. RESULTS AND DISCUSSION

In order to validate the effectiveness of the proposed EEMD-GRNN model, comprehensive experiments based on two real world diarrhea datasets were conducted. First, data description and performance criteria used in this study are presented and then the experimental results are reported. Finally, the result are compared and discussed.

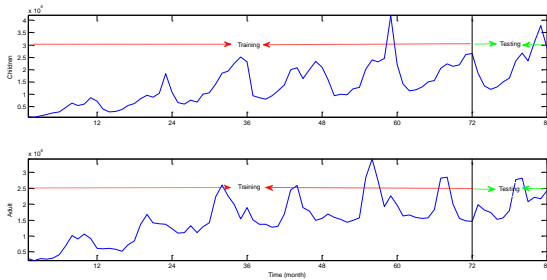


Fig. 3. Monthly diarrhea of children and adult from 2006(1)-2012(12).

##### A. Data Sets

In this study, the monthly diarrhea cases number data of children (0-15) and adult (>15) from 2006.01 to 2012.12 in Shanghai of China has been used. All data employed in this study are obtained from the Shanghai Municipal Center for Disease Control & Prevention. In particular, first, the data of the diarrhea of children are used to witness the whole process of the proposed method. In the same way, the corresponding prediction results of adult diarrhea are shown and further confirm the validity of the proposed method, accordingly.

There are in total 84 data points in each diarrhea dataset and the monthly series behavior is illustrated in Figure 3. The plot exhibits a permanent deterministic pattern of long-term upward trend with short-term fluctuations that are independent from one time period to the next. From Fig. 3, it can be seen that the series appear to be nonlinear, non-stationary in that the mean is increasing over time.

In order to testify the performance of the proposed prediction methods, the collected data is divided into two sets, training data and testing data. To achieve a more reliable and accurate result, a long period is served as the training period. Based on these considerations, the first 72 data points are used

as the training samples while the remaining 12 data points are used as the testing sample. The statistic characteristics of children and adult diarrhea in monthly time scale is tabulated in Table I.

To assess the forecast capacity of the EEMD-GRNN model, four indices for error forecast serve as the criteria to evaluate the prediction performance; they are mean absolute error (MAE), mean absolute percent error (MAPE), root mean square error (RMSE) and the coefficient of determination ( $R^2$ ). MAE, MAPE and RMSE are measures of the deviation between actual and predicted values. The models with the smallest MAE, MAPE, RMSE and the largest  $R^2$  are considered to be the best models.

TABLE I. STATISTICAL CHARACTERISTICS OF CHILDREN AND ADULT DIARRHEA FOR TRAINING AND TESTING DATA.

Indexes	Children		Adult	
	Training	Testing	Training	Testing
Max	41923.0000	37808.0000	34325.0000	28273.0000
Min	802.0000	11973.0000	2330.0000	15220.0000
Mean	13055.7500	21675.4167	14704.5556	20800.9167
SD	8069.2917	8214.7503	6892.3226	4296.5920

##### B. Prediction Results

According to the proposed hybrid EEMD-GRNN approach, in Stage 1, the original children diarrhea time series are decomposed into three independent IMFs (Illustrated in Fig. 4) and one residual employing EEMD technique, which exhibit a stable and regular variation. This means that the interruption and coupling between the different characteristics information embedded in the original data have been weakened to an extent. Thus, the local GRNN prediction model is easier to build.

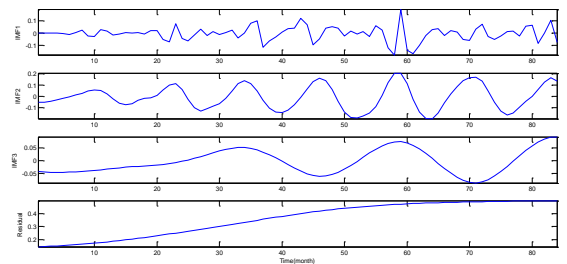


Fig. 4. EEMD decomposition result for children diarrhea.

After using EEMD to decompose the original children diarrhea data into three IMFs and a residue, these are then used to build the local GRNN prediction model for each IMF. In Stage 2, the relationship between the data of each IMF in the same frequency band should be identified prior to obtaining the prediction results of IMFs. The PACF is employed as a detector to determine the correlations between them. The lag orders of the autoregressive process of each IMF are two, four, five and five, respectively.

Based on the correlation between the data of each IMF, the input and the output pair vectors of the local GRNN model can be generated. Then the respective GRNN model is built and trained in terms of the input and the output vectors of the IMFs

and residue. For building the GRNN prediction model, the Matlab R2013a software package is adapted in this study and the optimal smoothing factor for each IFM is selected based on the corresponding minimum mean absolute error (MAE) on the out-of-sample testing samples. After that, the established GRNN model produces the one-step-ahead prediction results of each series of IMFs. In Stage 3, the final prediction results can be obtained using another independent GRNN model using the prediction results of each IFM as the input.

In order to reflect the model superiority, it is necessary to build other models to compare with the proposed model. Some other popular single prediction approaches recommended by recent works on time series prediction are selected as benchmarks. The benchmarks include time series techniques and artificial intelligence (AI) techniques. Amongst time series techniques, the autoregressive integrated moving average (ARIMA) models are adopted. For AI models, single GRNN is employed for the purpose. Furthermore, a hybrid learning approach with the EMD selected as decomposition method is also utilized. The simulation of this method is in general similar to the proposed model.

In the modeling of the single GRNN model like the EEMD-GRNN model, the input layer lags number has been determined using PACF and the smoothing factor are selected through the implementation of iterative optimization procedures.

In this study, the ARIMA model has three steps: model identification, parameter estimation, and diagnostic checking. The test time series data were processed by taking the first-order regular difference and the first seasonal difference to remove the growth trend and seasonality characteristics. We used the SPSS.19 statistical software to formulate the ARIMA model. Estimate the model parameters and utilize the Akaike Information Criterion (AIC) value to identify the best model. The model obtained from the training data set is ARIMA (2,1,1)(1,1,1)<sub>12</sub> model, the future one-head monthly cases number of children diarrhea can be obtained.

### C. Comparison and discussion

The comparisons of prediction models for the monthly number of children diarrhea are made between the ARIMA model, the single GRNN model, the hybrid EMD-GRNN model and the hybrid EEMD-GRNN model. The actual diarrhea cases number for children and predicted values of different models are illustrated in Fig. 6 and the prediction performances are shown in Table II. Through model comparisons, the proposed hybrid EEMD-GRNN model performs best. It can be observed from Fig. 6 that the predicted values obtained from the proposed EEMD-GRNN model are closer to the actual values than those obtained from the other models. This phenomenon signifies that the hybrid model can combine different advantages from EEMD and GRNN.

As seen from Table II and Fig. 6, it is clear that the hybrid EEMD-GRNN model performs much better than ARIMA model and single GRNN model, and outperforms the hybrid EMD-GRNN model. More precisely, the MAE, RMSE, MAPE and R<sup>2</sup> of the proposed EEMD-BPN model are, respectively, 664.361, 811.925, 3.1% and 0.991. That these values are smaller than other models. This indicates that there is a smaller

deviation between the actual and predicted values using the proposed EEMD-GRNN model. Thus, the proposed EEMD-GRNN model provides a better prediction result than the other models based on MAE, RMSE, MAPE and R<sup>2</sup>.

The possible reason is that the proposed hybrid model adequately makes use of the advantages of the decomposition methods and GRNN algorithm and integrates them well. In comparisons between EMD-GRNN and EEMD-GRNN, the decomposition method of EEMD is superior to EMD in terms of contribution to the prediction accuracy.

Similarly, the proposed EEMD-GRNN method also performs well in terms of predicting the diarrhea for adult. According to above steps, the ARIMA model generated from the data set is ARIMA (2,1,0)(1,1,1)<sub>12</sub>. Table III summarizes the diarrhea cases number for adult prediction results using the ARIMA, single GRNN, EMD-GRNN and EEMD-GRNN models. It can also be observed that the proposed EEMD-GRNN model has the smallest MAE, RMSE, MAPE and R<sup>2</sup> values in comparison with the single GRNN and ARIMA models and hybrid EMD-GRNN model. Thus, the proposed method produces lower prediction errors and outperforms other models with respect to predicting of diarrhea for adult.

TABLE II. PERFORMANCE OF THE FOUR MODELS FOR CHILDREN

Metrics	Models			
	ARIMA	GRNN	EMD-GRNN	EEMD-GRNN
MAE	3364.427	3330.324	2123.084	664.361
RMSE	4437.433	4212.306	3049.842	811.925
MAPE	17.1%	14.9%	14.3%	3.1%
R <sup>2</sup>	0.846	0.953	0.9666	0.991

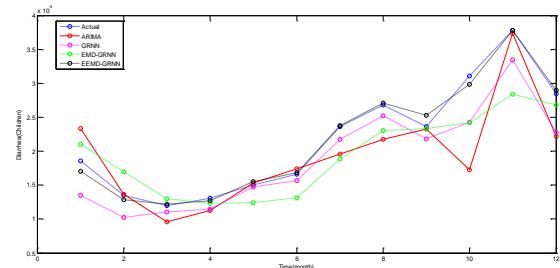


Fig. 6. Prediction results of diarrhea for children.

Fig. 7 depicts the actual diarrhea cases number for adult and the predicted values from the ARIMA, single GRNN, and EMD-GRNN, EEMD-GRNN models. From the Fig. 7, it can be observed that the proposed EEMD-GRNN model provides good prediction results. The predicted values of the proposed model are closer to the actual values than the other three models.

TABLE III. PERFORMANCE OF THE FOUR MODELS FOR ADULT.

Metrics	Models			
	ARIMA	GRNN	EMD-GRNN	EEMD-GRNN
MAE	2999.752	2644.084	1839.788	1110.087
RMSE	3737.192	3643.624	2476.043	1484.657
MAPE	14.09%	11.69%	10.51%	5.14%
R <sup>2</sup>	0.823	0.969	0.971	0.995

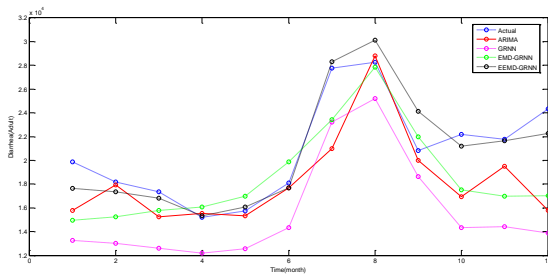


Fig. 7. Prediction results of diarrhea for adult.

## V. CONCLUSIONS

In this paper, a novel hybrid approach integrating the EEMD algorithm and the GRNN model is proposed to settle the diarrhea prediction problem. The main contribution of the paper is to propose a novel hybrid method for a stable prediction of nonlinear and non-stationary diarrhea time series data. The proposed method pre-processes the diarrhea time series data and decomposes them into more stationary and regular components (IMFs or residue) using the EEMD technique. Furthermore, the corresponding GRNN model for each divided component is easier to build. After the IMF components and residue are forecasted in the built GRNN model, the prediction values are then aggregated using another independent GRNN model as the final prediction results. This study compared the proposed method with the single GRNN, ARIMA models and hybrid EMD-GRNN model, using MAE, RMSE, MAPE and  $R^2$  as its criteria. Experimental results showed that the proposed EEMD-GRNN model is better and more efficient for prediction diarrhea in Shanghai areas.

There are several advantages of the proposed methodology. First, thanks to the non-linearity and non-stationary of diarrhea, hybrid the EEMD algorithm and GRNN model is a very wise practice for the diarrhea prediction. Moreover, it has rarely been mentioned in previous literature. Thus, applying this hybrid method to forecast diarrhea is very important for the future studies. Furthermore, from the simulation process and results, we can find this hybrid approach is useful in prediction diarrhea. Next, in terms of empirical results, it is a clear finding that the hybrid model can describe them comprehensively. The conventional single prediction models cannot do this very well. However, a hybrid method can integrate the advantages of other single models which conduce to boosting the model prediction ability and enhancing prediction efficiency. From this point of view, in terms of different criteria, it is unsurprising that the hybrid approach performs better than the single ARIMA and GRNN methods, and also superior to other hybrid models, for instance, EMD-GRNN model. Both statistical errors are reduced effectively in this hybrid model. Therefore, the proposed method is very suitable for prediction with nonlinear, non-stationary and strong complexity data, and is an efficient method for diarrhea prediction.

Our study has the some limitations that need further research. First, future studies may aim at combining EEMD and other prediction tools, like support vector regression (SVR), in evaluating the ability of the proposed prediction scheme. Second, integrating GRNN and other time series processing

techniques, such as wavelet transformation and seasonal adjustment method (SAM), in further improving the prediction capabilities can also be investigated in future studies.

## ACKNOWLEDGMENT

This research is supported by the Fund of The Shanghai Science and Technology Development Foundation (Grant No. 13430710100). The authors are grateful to the editor and anonymous reviewers for their suggestions in improving the quality of the paper.

## REFERENCES

- [1] I. N. Soyiri, D. D. Reidpath, "An overview of health forecasting," *Environmental health and preventive medicine*, vol.18, pp. 1-9, 1998.
- [2] D. Specht, "A general regression neural network," *IEEE Transactions Neural Networks*, vol. 2, pp. 568-576, 1991.
- [3] N. E. Huang, Z. Shen, S. R. Long, M. C. Wu, H. H. Shih, Q. Zheng, "The empirical mode decomposition and the Hilbert spectrum for nonlinear and nonstationary time series analysis," In: *Proceedings of the royal society of London series a-mathematical physical and engineering sciences*, series A., vol. 454, pp. 903-995, 1998.
- [4] N. E. Huang, Z. Shen, S. R. Long, "A new view of nonlinear water waves: the Hilbert spectrum," *Annu. Rev. Fluid Mech*, vol. 31, pp. 417-457, 1999.
- [5] D. J. Yu, J. S. Cheng, Y. Yang, "Application of EMD method and Hilbert spectrum to the fault diagnosis of roller bearings," *Mech. Syst. Signal Process*, vol. 19, pp. 259-270, 2005.
- [6] Hu, Jianming, Jianzhou Wang, and Guowei Zeng, "A hybrid forecasting approach applied to wind speed time series," *Renewable Energy*, vol. 60, pp. 185-194, 2013.
- [7] D. F. Specht, "A general regression neural network," *IEEE Transactions on Neural Networks*, vol. 2, pp. 568-576, 1991.
- [8] J. D. Wu, J. C. Liu, "A forecasting system for car fuel consumption using a radial basis function neural network," *Expert Systems with Applications*, vol. 39, pp. 1883-1888, 2012.
- [9] W. Yan, "Toward automatic time-series forecasting using neural networks," *Neural Networks and Learning Systems*, *IEEE Transactions on*, vol. 23, pp. 1028-1039, 2012.
- [10] G. Q. Zhang, B. E. Patuwo, and M. Y. Hu, "Forecasting with artificial neural networks: The state of the art," *International journal of forecasting*, vol. 14, pp. 35-62, 1998.
- [11] G. J. Bowden, G. C. Dandy, H. R. Maier, "Input determination for neural network models in water resources applications. Part 1- background and methodology," *Journal of Hydrology*, vol. 301, pp. 75-92, 2005.
- [12] Ö. Kisi, "Constructing neural network sediment estimation models using a data-driven algorithm," *Mathematics and Computers in Simulation*, vol. 79, pp. 94-103, 2008.
- [13] K. P. Sudheer, A. K. Gosain, K. S. Ramasastri, "A data-driven algorithm for constructing artificial neural network rainfall-runoff models," *Hydrological Processes*, vol. 16, pp. 1325-1330, 2002.
- [14] P. Flandrin, G. Rilling, P. Goncalves, "Empirical mode decomposition as a filter bank," *IEEE Signal Process. Lett.*, vol. 2, pp. 112-114, 2004.
- [15] N. E. Huang, "Review of empirical mode decomposition," *Proc. SPIE*, pp. 71-80, 2001.
- [16] M. C. Wu, C. K. Hu, "Empirical mode decomposition and synchrogram approach to cardiorespiratory synchronization," *Phys. Rev. E*, vol. 73 51917, 2006.
- [17] Z. Wu, N. E. Huang, "Ensemble empirical mode decomposition: a noise-assisted data analysis method, center for ocean-land-atmosphere studies," *Tech Rep*, vol. 51, 2004.