Since January 2020 Elsevier has created a COVID-19 resource centre with free information in English and Mandarin on the novel coronavirus COVID-19. The COVID-19 resource centre is hosted on Elsevier Connect, the company's public news and information website.

Elsevier hereby grants permission to make all its COVID-19-related research that is available on the COVID-19 resource centre - including this research content - immediately available in PubMed Central and other publicly funded repositories, such as the WHO COVID database with rights for unrestricted research re-use and analyses in any form or by any means with acknowledgement of the original source. These permissions are granted for free by Elsevier for as long as the COVID-19 resource centre remains active.
A new hybrid prediction model of cumulative COVID-19 confirmed data

Guohui Li, Kang Chen, Hong Yang

PII: S0957-5820(21)00591-7
DOI: https://doi.org/10.1016/j.psep.2021.10.047
Reference: PSEP3126

To appear in: Process Safety and Environmental Protection

Received date: 9 June 2021
Revised date: 25 October 2021
Accepted date: 27 October 2021

Please cite this article as: Guohui Li, Kang Chen and Hong Yang, A new hybrid prediction model of cumulative COVID-19 confirmed data, Process Safety and Environmental Protection, (2021) doi:https://doi.org/10.1016/j.psep.2021.10.047

This is a PDF file of an article that has undergone enhancements after acceptance, such as the addition of a cover page and metadata, and formatting for readability, but it is not yet the definitive version of record. This version will undergo additional copyediting, typesetting and review before it is published in its final form, but we are providing this version to give early visibility of the article. Please note that, during the production process, errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

© 2021 Published by Elsevier.
A new hybrid prediction model of cumulative COVID-19 confirmed data

Guohui Li*, Kang Chen, Hong Yang*

School of Electronic Engineering, Xi’an University of Posts and Telecommunications, Xi’an, Shaanxi 710121, China

Correspondence should be addressed to Guohui Li, Hong Yang.

lghcd@163.com, uestcyhong@163.com

Abstract

Establishing an accurate and efficient prediction model is of great significance for governments and other social organizations to formulate prevention and control policies and curb the explosive spread of the pandemic. To improve prediction accuracy of cumulative COVID-19 confirmed data, a new hybrid prediction model based on gradient-based optimizer variational mode decomposition (GVMD), extreme learning machine (ELM), and autoregressive integrated moving average (ARIMA), named GVMD-ELM-ARIMA, is proposed. To solve the problem of selecting the $k$ value and the penalty factor $\alpha$ in variational mode decomposition (VMD), this paper proposes gradient-based optimizer variational mode decomposition (GVMD), which realizes the self-adaptive determination of $k$ value and $\alpha$ value. Firstly, GVMD decomposes the cumulative COVID-19 confirmed data into some intrinsic mode functions (IMFs) and a residual component (IMFr). Secondly, IMFs are predicted by ELM. Then, IMFr is predicted by ARIMA. Finally, the final prediction results are obtained by reconstructing the prediction result of IMFs and IMFr. The cumulative COVID-19 confirmed data of the United States, India and Russia is used to verify its effectiveness. Taking the United States as an example, compared with the average MAPE, RMSE and MAE of the single model, the average MAPE of the hybrid model is reduced by 47.27%, the average RMSE is reduced by 44.50%, and the average MAE is reduced by 55.34%. Compared with GVMD-ELM-ELM, GVMD-ELM-ARIMA proposed in this paper reduces the MAPE by 60%, the RMSE by 56.85%, and the MAE by 61.61%. The experimental results show that GVMD-ELM-ARIMA has best prediction accuracy, and it provides a new method for predicting the cumulative COVID-19 confirmed data.

Key Words:
COVID-19; cumulative confirmed data; prediction; variational mode decomposition

1. Introduction

As a new type of RNA virus, corona virus disease 2019 (COVID-19) causes the whole world into this viral crisis because of its strong transmissibility. Establishing an accurate and efficient prediction model is of great significance for governments and other social organizations to formulate prevention and control policies and curb the explosive spread of the pandemic. At present, the prediction of pandemic data
has been carried out pretty deeply [1-4], which can be roughly divided into three models: epidemic model, single model prediction, and hybrid prediction model.

Epidemic models used to study COVID-19 are mainly divided into three types: susceptible infectious (SI) model [5], susceptible infectious recovered (SIR) model [6-8], and susceptible exposed infectious recovered (SEIR) model [9,10]. Epidemic model can better predict the number of people infected and susceptible in the future, but their establishment process is complex, and the transmission coefficient needs to be obtained by a large number of experiments. Therefore, they can only roughly estimate the trend of the epidemic, but cannot achieve accurate prediction.

In addition to the epidemic model, single model prediction is the most common method to study COVID-19. Specifically, the single model can be subdivided into statistical theoretical model and artificial intelligence model. Statistical theoretical model mainly includes the logistic regression model [11], linear regression model [12], and multiple linear regression model [13], etc. The above statistical theoretical model has the advantages such as fast training speed, few reference factors, simple and easy to understand. However, its accuracy is low. It is difficult to deal with problems such as high complexity data. Different from statistical theoretical model, single artificial intelligence model can usually process signal well and reach the expected prediction accuracy. Therefore, some researchers used artificial intelligence model to predict COVID-19 such as support vector machine (SVM) [14], artificial neural network (ANN) [15,16], long short-term memory (LSTM) [17,18]. These models have a certain degree of prediction effect, but they all have some certain limitations in varying degrees.

In order to improve the COVID-19 confirmed data prediction accuracy and avoid the defect of the single intelligent model as much as possible, some researchers began to propose some hybrid models. By combining the advantages and disadvantages of different models, they expected to achieve the desired effect. Hybrid models are mainly divided into two categories: epidemic hybrid model with combined artificial intelligence and hybrid model based on decomposition. The former combined artificial intelligence technology with epidemic model to predict the future development of the epidemic in many aspects [19,20]. The epidemic hybrid model combined artificial intelligence can optimize the parameters and improve the prediction accuracy to a certain extent. However, the nature of epidemic model remains unchanged. After the introduction of artificial intelligence, there are too many parameters, and the inherent characteristics of data are limited. Another hybrid model is based on decomposition, which can decompose data into components of different complexity by introducing a mode decomposition algorithm. Therefore, this kind of model can better capture the inherent characteristics of data and reduce the complexity of original data, so as to achieve a better prediction effect [21, 22]. The hybrid model based on decomposition has been applied in many ways [23-25]. In the field of COVID-19 confirmed data prediction, some researchers have proposed hybrid prediction model based on decomposition. Singh [26] et al. combined discrete wavelet decomposition (DWD) and autoregressive integrated moving average (ARIMA) model to predict death data in Italy, Spain, and other countries. The result shows that the error of prediction after wavelet decomposition is greatly reduced and the prediction accuracy is high. But the essence of wavelet decomposition is Fourier transform, which cannot describe the local characteristics of the signal in the time domain, and it is not suitable for the decomposition of the non-stationary signal. Qiang et al. [27] proposed a hybrid prediction model for COVID-19 outbreak based on ensemble empirical mode decomposition (EEMD). The hybrid model decomposes the real data into several components of different frequency by EEMD, and then predicts these components by ARIMA respectively. This hybrid model has strong adaptability to each country and good prediction accuracy. However, EEMD and its improved algorithm such as complementary ensemble
empirical mode decomposition (CEEMD), and complete ensemble empirical mode decomposition with adaptive noise (CEEMDAN) have two shortcomings. The first is the lack of strict mathematical theoretical basis [28]. The second is the existence of modal aliasing [29-31]. It often separates similar frequencies into multiple patterns, which can lead to significant error in the prediction.

Compared with EEMD, CEEMD and CEEMDAN, variational mode decomposition (VMD) not only has a solid theoretical basis, but also overcomes the phenomenon of modal aliasing [32-34] and can better capture the inherent characteristics of data. However, this predictive thinking is just beginning to be applied to the COVID-19 confirmed data prediction. Therefore, in order to avoid the shortcomings of other decomposition algorithms and further improve the prediction accuracy of COVID-19 confirmed data, VMD is introduced in this paper. Unfortunately, the VMD mode number \( k \) and penalty factor \( \alpha \) are set artificially, which are mainly dependent on subjective consciousness and are not convincing [35,36]. Gradient-based optimizer (GBO) is a primal heuristic algorithm proposed by Ahmadianfar et al. [37] in 2020, which has strong detection, development, convergence and effective ability to avoid local optimality. Therefore, this paper proposes gradient-based optimizer variational mode decomposition (GVMD) to determine mode number \( k \) and penalty factor \( \alpha \), which improve its adaptability. Although GVMD can optimally decompose real data to obtain intrinsic mode functions (IMFs), there are still some small components that cannot be decomposed, which are called residual component (IMFr). In previous studies, since the IMFr component is small, its influence is often ignored. However, in the study of COVID-19, the impact of IMFr can not be ignored, so we refer to Reference [38] to determine the treatment of IMFr. The IMFs obtained by GVMD have the characteristics of flat data and mainly low-frequency components. And IMFr is mainly composed of high-frequency components, and the data fluctuates greatly. Extreme learning machine (ELM) has the advantage of fast speed and strong versatility [39,40], while ARIMA can deal with high-frequency components, few parameters and simple application. According to the different characteristics of IMFs and IMFr, ELM is used to predict IMFs, and ARIMA is introduced to analyze and predict IMFr to improve the prediction accuracy.

To improve prediction accuracy of cumulative COVID-19 confirmed data, a new hybrid prediction model based on gradient-based optimizer variational mode decomposition (GVMD), extreme learning machine (ELM), and autoregressive integrated moving average (ARIMA), named GVMD-ELM-ARIMA, is proposed. The main work of this paper is as follows:

1. At present, most of the hybrid prediction models for COVID-19 are based on the combination of epidemiology and artificial intelligence, and the prediction model of decomposition and integration is just beginning. Considering the shortcoming of EEMD and its improved decomposition algorithm, in order to improve the prediction accuracy, this paper introduces VMD to decompose the cumulative COVID-19 confirmed data.

2. To solve the problem of selecting the \( k \) value and the penalty factor \( \alpha \) in variational mode decomposition (VMD), this paper proposes gradient-based optimizer variational mode decomposition (GVMD), which realizes the self-adaptive determination of \( k \) value and \( \alpha \) value.

3. Although the accuracy of the hybrid model has been high, there is still some error. In order to reduce the error and improve the prediction performance, this paper analyzed the IMFr, and innovatively introduced ARIMA to predict it, which further improving the prediction accuracy.

4. GVMD-ELM-ARIMA is proposed. Firstly, GVMD decomposes the cumulative COVID-19 confirmed data into some intrinsic mode functions (IMFs), and residual component (IMFr) is obtained. Secondly, IMFs are predicted by ELM. Then, IMFr is predicted by ARIMA. Finally, the final prediction results are obtained by reconstructing the prediction result of IMFs and IMFr. The cumulative confirmed data of the
United States, India, and Russia is used for simulation experiment, and several comparative models are added. The experimental results show that GVMD-ELM-ARIMA has best prediction accuracy, and provides a new method for predicting the cumulative COVID-19 confirmed data.

(5) The GVMD-ELM-ARIMA method is used to predict cumulative COVID-19 confirmed data in the United States, India, and Russia in the future, and its effectiveness is tested.

The rest of this paper is organized as follows. In Section 2, basic theory such as VMD, GVMD, ARIMA and ELM are briefly introduced. In Section 3, the GVMD-ELM-ARIMA hybrid prediction model is proposed. In Section 4, three case studies are performed and the effectiveness of the proposed model is analyzed. Section 5 is the multi-step prediction of COVID-19 cumulative data by GVMD-ELM-ARIMA. Section 6 is discussion. The last Section is conclusion. For the convenience of reading, the nomenclature used in this paper is shown in Appendix A.

2. Basic Theory

2.1 Variational mode decomposition

Variational mode decomposition is a new signal processing method based on EMD, which decomposes the signal by solving the variational formula adaptively.

Firstly, each IMF component obtained by VMD is defined as a signal with adjustable frequency and amplitude:

\[ u(t) = A_k(t) \cos(\phi_k(t)) \]  

where, \( u(t) \) is the kth modal component, \( A_k(t) \) and \( \phi_k(t) \) are instantaneous amplitude and phase respectively. The principle of VMD is as follows [41]:

(1) Hilbert transform is performed on \( u(t) \), and unilateral spectrum of modal function is obtained by establishing analytical signal:

\[ (\delta(t) + \frac{j}{\pi t}) \cdot u(t) \]  

(2) Mix the analytical signal and center frequency of each IMF to realize the spectrum shift to the base band:

\[ [(\delta(t) + \frac{j}{\pi t}) \cdot u(t)] \cdot e^{-j\omega_k t} \]  

(3) Estimate the bandwidth of each IMF component.

(4) Constraint variational model:

\[ \min_{\{A_k\}} \sum_{k=1}^{K} \|f(t)\|_2 \left\{ (\delta(t) + \frac{j}{\pi t}) \cdot u(t) \right\} e^{-j\omega_k t} \]  

where, \( \{u_k\} \) and \( \{\omega_k\} \) represents the set of sub-signals and their center frequencies, \( K \) represents the total number of sub-signals, \( \delta(t) \) is the Dirac distribution.

For solving the variational problem, the extended Lagrange function \( L\{u_k\}, \{\omega_k\}, \{\lambda\} \) can transform the constrained problem into a non-constrained problem to realize the solution of the variational problem [42], which is shown in formula (5).
\[ L(\{u_k\}, \{o_k\}, \theta) = C \prod_{k=1}^{K} \left[ \left( \delta(t) + \frac{j}{\pi} \right) u_k(t) \right] e^{-j\omega_k} \]
\[ + \left\| f(t) - \sum_{k=1}^{K} u_k(t) \right\|_2^2 + \left\langle \theta(t), f(t) - \sum_{k=1}^{K} u_k(t) \right\rangle \]  

where, \( C \) represents the quadratic penalty factor, \( \theta(t) \) represents the Lagrange multiplier and \( f(t) \) represents the original signal.

To solve formula (5), the alternating direction method of the multiplication operator is introduced. \( u_{k}^{n+1} \), \( o_{k}^{n+1} \) and \( \theta^{n+1} \) are updated alternately, where \( n \) represents the number of iterations. The specific process is as follows [43]:

\[ u_{k}^{n+1}(\omega) = \frac{\hat{f}(\omega) - \sum_{k=1}^{K} \hat{u}_{k}(\omega) + \hat{\theta}(\omega)}{1 + 2C(\omega - \omega_k)^2} \]  
\[ o_{k}^{n+1} = \frac{\int_{\omega} \omega |\hat{u}_{k}(\omega)|^2 d\omega}{\int_{0}^{\omega} |\hat{u}_{k}(\omega)|^2 d\omega} \]
\[ \hat{\theta}^{n+1}(\omega) = \hat{\theta}^{n+1}(\omega) + \tau \left[ f(t) - \sum_{k=1}^{K} \hat{u}_{k}^{n+1}(\omega) \right] \]

Determine whether to stop iteration convergence according to the accuracy \( e \). The conditions are as follows:

\[ \sum_{k=1}^{K} \left\| \hat{u}_{k}^{n+1} - \hat{u}_{k}^{n} \right\| < e \]  

2.2 VMD optimized by gradient-based optimizer

When VMD is used for decomposition, it is necessary to determine the VMD mode number \( k \) and penalty factor \( \alpha \) [44]. The selection of mode number \( k \) has a great influence on the decomposition effect. When \( k \) is too small, it may lead to insufficient decomposition of original data and large errors in the final prediction result. If \( k \) is too large, the original data may be over-decomposed. It results in modal aliasing and increases the algorithm complexity. \( \alpha \) affects the initial center constraint strength of the mode. It also has a great influence on the decomposition result.

In this paper, VMD optimized by gradient-based optimizer (GBO) is proposed, named GVMD. GBO is a primal heuristic algorithm proposed in 2020 [37]. The key to GVMD lies in the selection of its fitness function. The multiscale fuzzy entropy (MFE) reflects the complexity of the signal [45]. The more complex the signal is, the greater the multiscale fuzzy entropy is, and vice versa. If the IMF component has a strong regularity, the lower the noise, the lower the complexity of the signal, and the lower the multiscale fuzzy entropy. Therefore, when GBO is chosen to optimize VMD, the selected fitness function is multiscale fuzzy entropy. After calculating the fitness function values each time, the optimal combination \([k, \alpha]\) is determined by comparing and updating each other. The IMF component is obtained by replacing \([k, \alpha]\) into VMD. Its brief search principle is as follows:

1. Set the population size and initial position.
2. Calculate the multiscale fuzzy entropy corresponding to each point in the population, and get the best Xbest and worst Xworst.
(3) In order to achieve the balance between exploring the development process and seeking to improve the search ability, the gradient search rule (GSR) is introduced by using the best Xbest and worst Xworst in the population, and the GSR value is calculated. And the updating position of the population can be defined as follows:

$$X_{n+1} = X_n - GSR$$  \hspace{1cm} (10)$$

(4) In order to make better use of the region near \(X_n\), the movement direction (DM) is introduced, and a new point \(X_{1n}^*\) is generated by the value of DM. \(X_{1n}^*\) is defined as follows:

$$X_{1n}^* = X_n^* - GSR + DM$$ \hspace{1cm} (11)$$

(5) At the same time, two other special points \(X_{2n}^*\) and \(X_{3n}^*\) are generated, and the next iteration position is obtained by using the updated \(X_{1n}^*, X_{2n}^*, X_{3n}^*\):

$$x_{n+1}^* = r_e \times (r_e \times X_{1n}^* + (1 - r_e) \times X_{2n}^*) + (1 - r_e) \times X_{3n}^*$$ \hspace{1cm} (12)$$

where, \(r_e\) and \(r_s\) are two arbitrary numbers in \([0,1]\).

(6) In order to improve the performance of the optimization algorithm for solving complex problem, the local escaping operator (LEO) algorithm is introduced to solve the update point of the main position, the best and the worst point of the population. It is helpful to get out of the local optimum and accelerate the convergence speed. The specific optimization flow chart is shown in Figure 1.

\[\text{Figure 1. The specific optimization flow chart}\]

### 2.3 Autoregressive integrated moving average

Autoregressive moving average (ARMA), proposed by Box and Jenkins, is a model used to deal with random event sequences with fewer model parameters and simple application [46]. There are three basic types of ARMA such as autoregressive (AR) model, moving average (MA) model and autoregressive moving average (ARMA) model. The basic structure is as follows:

$$X_t = \phi_1 X_{t-1} + \phi_2 X_{t-2} + \cdots + \phi_p X_{t-p} + \theta_1 u_{t-1} + \theta_2 u_{t-2} + \cdots + \theta_q u_{t-q}$$ \hspace{1cm} (13)$$

where, \(\phi_1, \phi_2, \cdots, \phi_p\) is the autoregressive coefficient, \(\theta_1, \theta_2, \cdots, \theta_q\) is the moving average.
coefficient, \( P \) and \( Q \) are the order of the coefficient, \( u_t, u_{t-1}, \ldots, u_{t-P} \) are independent white noise sequence, denoted as ARMA\((p, q)\). If the original sequence is non-stationary and needs to be stationary after \( D \)-order difference, then the original sequence can be expressed as ARIMA \((p, d, q)\) sequence.

### 2.4 Extreme learning machine

ELM is a kind of machine learning system based on feedforward neural network (FNN), which is suitable for supervised learning and unsupervised learning problems [47]. ELM is regarded as a special FNN in the research, which is an improvement of FNN and its back-propagation algorithm. Its characteristic is that the weight of hidden layer nodes is random or artificially given, and does not need to be updated. The learning process only calculates the output weights [48]. Its network structure is shown in Figure 2. Suppose that there are \( N \) arbitrary samples \((x_i, t_i)\), \( X = [x_1, x_2, \ldots, x_N] \) and \( T = [t_1, t_2, \ldots, t_N] \) are the target matrix, namely the target vectors set of each individual sample, where, \( x_i = (x_{i1}, x_{i2}, \ldots, x_{in})^T \in \mathbb{R}^n, i = 1, 2, \ldots, N \) is the \( n \times 1 \) dimension vector, and \( t_i = (t_{i1}, t_{i2}, \ldots, t_{im})^T \in \mathbb{R}^m, i = 1, 2, \ldots, N \) is the \( m \times 1 \) dimension target.

![Figure 2. ELM network structure](image)

For an ELM network with \( L \) hidden layer node, the hidden layer node output is as follows:

\[
\begin{align*}
\mathbf{h}(x_i) &= [h_1(x_i), h_2(x_i), \ldots, h_L(x_i)] \\
h(x_i) &= g(w_i x_i + b_i)
\end{align*}
\]

(14)

where, \( g \) is the activation function, \( w_i = [w_{i1}, w_{i2}, \ldots, w_{im}] \) is the input weight, and \( b_i \) is the bias of the \( i \)th hidden layer unit. The output of ELM is:

\[
\begin{align*}
\mathbf{Y}(\mathbf{X}) &= [y(x_1), y(x_2), \ldots, y(x_N)] \\
y(x_i) &= \sum_{i=1}^{L} \beta_i h(x_i)
\end{align*}
\]

(15)

where, \( \beta_i \) is the weight between each hidden layer node and each output layer node.

In ELM, hidden layer node parameters such as \( w_i \) and \( b_i \) are randomly generated according to any continuous probability distribution, which are independent of training data. The goal of ELM learning is to minimize the error of the output. That is the error between the output result \( \mathbf{Y}(\mathbf{X}) \) and the matrix \( \mathbf{T} \).

\[
\min : ||\mathbf{Y}(\mathbf{X}) - \mathbf{T}||^2
\]

(16)

The minimum approximate square difference method is used to solve the weight, so that the solution with the minimum objective function \( \mathbf{Y}(\mathbf{X}) = \mathbf{T} \) is the optimal solution \( \beta^* \). The solution process of \( \beta^* \) is as follows:

\[
\mathbf{T} = \mathbf{Y}(\mathbf{X}) = \mathbf{H} \beta, \beta \in \mathbb{R}^{L \times M}
\]

(17)
\[
\begin{align*}
H &= [h(x_1),...,h(x_n)] = \begin{bmatrix} h_1(x_1),...,h_2(x_1) \\
h_1(x_2),...,h_2(x_2) \end{bmatrix} \\
T &= [t_1^T,...,t_n^T]^T
\end{align*}
\]  

(18)

In ELM, once the input weight \( w \) and hidden layer node bias \( b \) are determined, the output matrix \( H \) of the hidden layer is uniquely determined and the optimal solution \( \beta^* \) is obtained:

\[
\beta^* = H^+ T
\]

(19)

where, \( H^+ \) is the Moore-Penrose generalized inverse matrix of the matrix \( H \).

3 The proposed prediction model

Based on the theoretical basis of Section 2, a new hybrid prediction model based on gradient-based optimizer variational mode decomposition (GVMD), extreme learning machine (ELM), and autoregressive integrated moving average (ARIMA), named GVMD-ELM-ARIMA, is proposed. Its flow chart is shown in Figure 3. The detailed steps are as follows:

**Step 1: GVMD decomposition process**

The cumulative COVID-19 confirmed data are substituted into GVMD for optimization. The optimal \( k \) and \( \alpha \) of VMD are obtained through several iterations, and some intrinsic mode functions (IMFs) are obtained. The residual component \( IMFr \) is obtained by the difference between the original data and the sum of IMFs.

**Step 2: ELM prediction process**

ELM prediction model is established for the modal functions \([IMF_1,IMF_2,...,IMFn]\) obtained by GVMD, and the predicted results are denoted as \([IMF_1',IMF_2',...,IMFn']\).

**Step 3: ARIMA prediction process**

The ARIMA model is used to predict the high frequency residual component \( IMFr \). Firstly, the stability is tested. Then the optimal model is obtained by Akaike information criterion (AIC) delimiter. Finally, the prediction result \( IMFr' \) is obtained.

**Step 4: Hybrid reconstruction process**

\([IMF_1',IMF_2',...,IMFn']\) and \( IMFr' \) are reconstructed to complete the prediction and get the final output \( X_{\text{mixed}} \). The mixed output formula is as follows:

\[
X_{\text{mixed}} = \sum_{i=1}^{n} IMFi + IMFn'
\]

(20)
4 Simulation experiment

In this section, data source, comparison models, evaluation index of prediction result, detailed simulation experiment and result are given.

4.1 The experimental data source

The real cumulative COVID-19 confirmed data is used and carried out one-step prediction simulation experiment in this research. The real cumulative COVID-19 confirmed data comes from the United States, India and Russia. It includes data from January 22, 2020 to January 27, 2021, and the data is collected every day. The cumulative COVID-19 confirmed data can be obtained at website https://data.humdata.org/dataset/novel-coronavirus-2019-ncov-cases. Each country has 372 cumulative COVID-19 confirmed data, and the data of each country are simulated respectively. Firstly, the first 327 real cumulative COVID-19 confirmed data are used as training samples to obtain 45 predicted values, and then compared with the remaining 45 real cumulative COVID-19 confirmed data. The real cumulative COVID-19 confirmed data is shown in Figure 4.

![Figure 4. Cumulative confirmed cases](image)

4.2 Compared model description

In order to prove the performance of GVMD-ELM-ARIMA in this paper, five different models are selected for comparison. Among them, the single prediction model includes LSTM and ELM, and the hybrid model includes CEEMDAN-ELM, GVMD-ELM, GMD-ELM-ELM and GVMD-ELM-ARIMA. ELM and
LSTM refers to the direct prediction of cumulative COVID-19 confirmed data. GVMD-ELM is as follows: Apply the component decomposed by GVMD to the ELM for prediction, and finally reconstruct the result. GVMD-ELM-ELM is as follows: IMFs and IMFr obtained after GVMD decomposition are both predicted by ELM, and then reconstructed. GVMD-ELM-ARIMA is as follows: IMFs and IMFr obtained after GVMD decomposition are predicted by ELM and ARIMA respectively, and then reconstructed.

4.3 The evaluation index

In order to quantitatively describe the performance of the evaluation prediction model, mean absolute percentage error (MAPE), root mean square error (RMSE) and mean absolute error (MAE) are selected. Their formulas are as follows:

\[
MAPE = \frac{1}{n} \sum_{i=1}^{n} \left| \frac{x_i - \hat{x}_i}{\bar{x}} \right|
\]

(21)

\[
RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^{n} (x_i - \hat{x}_i)^2}
\]

(22)

\[
MAE = \frac{1}{n} \sum_{i=1}^{n} |x_i - \hat{x}_i|
\]

(23)

where, \( x_i \) is the ith actual value, \( \hat{x}_i \) is the ith predicted value, \( \bar{x} \) is the actual average, \( n \) is the total number of samples.

4.4 American data simulation experiment

4.4.1 GVMD decomposition result

The cumulative COVID-19 confirmed data has been affected by uncertainties such as weather, temperature and government measures to combat the pandemic. Its fluctuation range is large. In order to predict it more accurately, GVMD initial optimization parameters setting are shown in Table 1.

| The main parameter                   | The parameter value |
|--------------------------------------|---------------------|
| The population size                  | 30                  |
| Maximum number of iterations         | 50                  |
| Regularization parameter optimization range | [1500,3500]         |
| Kernel function width optimization range | [3,9]               |

The cumulative confirmed data of 372 days in the United States are substituted into GVMD for optimization, and the optimal fitness function is obtained through 50 iterations. Finally, the optimal parameters are found as \( k = 3 \) and \( \alpha = 3369 \), which are decomposed into 3 modes IMF. GVMD decomposition result is shown in Figure 5.
VMD is the most advanced mode decomposition method. The problem is that the $k$ value and $\alpha$ value which are difficult to determine are overcome by GBO. To prove the superioriity of GVMD, CEEMDAN, which is the most advanced among the current improved EEMD, is selecte to compare with GVMD. CEEMDAN decomposition result is shown in Figure 6.

Figure 5. GVMD decomposition result

Figure 6 shows that IMF components obtained by CEEMDAN are disturbed by noise, and a relatively large swing appears at the right end, thus affecting the whole component sequence. Therefore, large errors will be caused and the decomposition result will be distorted. The component waveform distortion of GVMD is smaller than that of CEEMDAN, and VMD has better noise robustness. In addition, according to the decomposition result of the original data, only low-frequency components exist after the decomposition of GVMD, and there is no high-frequency component with high complexity. However, CEEMDAN has many high-frequency components, which is not conducive to the prediction.

4.4.2 ARIMA processing process
After GVMD, the residual component can be obtained as follows:

\[ \text{IMF}_r = Y - \sum_{i=1}^{k} \text{IMF}_i \]

where, \( Y \) represents the cumulative confirmed data. The residual component is shown in Figure 7.

![Residual component](image)

**Figure 7.** The residual component

Figure 7 shows that the residual component of the cumulative confirmed data in the United States after decomposition is relatively large, reaching about 2W personages. Therefore, it is necessary to deal with the residual component. In this paper, ARIMA model is used to deal with the residual component. The processing process is as follows:

1) Stationarity test

The stationarity test results such as \( H_p = 1, H_{p\text{-value}} = 1.000E-03 \) can be obtained. They conform to the characteristics of stationary sequence and do not need differential operation.

2) Correlation test

Test the autocorrelation function (ACF) and partial autocorrelation function (PACF) on the data to judge the ARMA model. The results are shown in Figure 8.

![Correlation test](image)

**Figure 8.** Correlation test between ACF and PACF
Figure 8 shows that both ACF and PACF have the phenomenon of trailing, so the mixed model of AR and MA is selected.

3) AIC order determination

It is difficult to determine the order according to ACF and PACF charts. In this paper, a more effective AIC is selected for delimiting, and the AIC order heat map is shown in Figure 9.

![AIC order heat map](attachment:image.png)

**Figure 9.** The AIC order heat map

4) Establish a model

AR =5 and MA =3 are determined according to the heat map, and the model ARIMA (5, 0, 3) is established, and the residual component $\text{IMFr}$ is predicted.

4.4.3 Prediction result of each model

This section presents in detail the prediction result and performance of GVMD-ELM-ARIMA proposed in this paper. In order to show the superiority and advancement of this model, the five models in Section 4.2 are compared and analyzed. The comparison diagram and error diagram of the prediction model are firstly shown to observe the prediction effect. Finally, the error box plot and the evaluation index in Section 4.3 are used for quantitative analysis. In this paper, the prediction result of each model is firstly presented in Figure 10. Then the prediction comparison result of each model is shown in Figure 11. Figure 12 further shows the prediction error distribution diagram.
Figure 10. The prediction result of ELM, LSTM, CEEMDAN-ELM, GVMD-ELM, GVMD-ELM-ELM and GVMD-ELM-ARIMA.
Figure 10 shows that the prediction effect of GVMD-ELM-ARIMA is better. Compared with other models, the prediction result of GVMD-ELM-ARIMA has the highest coincidence degree with the original data. Figure 11 shows that the prediction points of GVMD-ELM-ARIMA and CEEMDAN-ELM are the closest to the original data points and LSTM is the farthest from the original data points. According to the prediction error distribution diagram in Figure 12, the predicted value of GVMD-ELM-ARIMA is the closest to the true value. From the first day to the 35th day, the data value fluctuated around the base line. On day 35 to 40, the error fluctuates slightly. In single model, the prediction effect of ELM is better, and the prediction error is small from day 1 to day 10. However, the prediction error of the next 11 to 30 days is larger than the base line. In order to improve the prediction accuracy of qualitative analysis, the error box diagram and the evaluation index of each model are shown in Figure 13 and Table 2 respectively.
Figure 13. The error box diagram

Table 2. The evaluation index

| Model            | MAPE  | RMSE         | MAE     |
|------------------|-------|--------------|---------|
| ELM              | 0.0047| 1.41E+05     | 1.06E+05|
| LSTM             | 0.0063| 1.7557E+05   | 1.43E+05|
| CEEMDAN-ELM      | 0.0042| 1.53E+05     | 7.31E+04|
| GVMD-ELM         | 0.0032| 8.56E+04     | 7.01E+04|
| GVMD-ELM-ELM     | 0.0030| 7.88E+04     | 6.59E+04|
| GVMD-ELM-ARIMA   | 0.0012| 3.40E+04     | 2.53E+04|

Figure 13 shows that GVMD-ELM-ARIMA proposed in this paper has the best prediction effect, with the maximum error of 74703.42, the minimum error of -86186.61 and the median of -7387.99. Compared with the other 5 models, GVMD-ELM-ARIMA has the smallest overall fluctuation range and relatively uniform prediction error distribution, and its prediction performance is significantly better than other models.

According to the evaluation index in Table 2, it can be concluded that:

(1) Compared with the average MAPE, RMSE and MAE of the single model, the average MAPE of the hybrid model is reduced by 47.27%, the average RMSE is reduced by 44.50%, and the average MAE is reduced by 55.34%.

(2) Compared with the average MAPE mean, RMSE mean and MAE mean of the hybrid model with untreated residual component (CEEMDAN-ELM, GVMD-ELM), that of the hybrid model with treated residual component (GVMD-ELM-ELM, GVMD-ELM-ARIMA) decreased by 43.24%, 52.72% and 36.31% respectively.

(3) Compared with GVMD-ELM-ELM, GVMD-ELM-ARIMA proposed in this paper reduces MAPE by 60%, the RMSE by 56.85%, and MAE by 61.61%.

By comparing the evaluation index in Table 2 with the model data in the previous paragraph, we can draw the following conclusions:

(1) The prediction accuracy of the hybrid model is better than that of the single prediction model, which also demonstrates the advantages of the hybrid model based on mode decomposition in the prediction.

(2) The accuracy of the hybrid model after processing the residual component is higher than that of the hybrid model without processing the residual component. It shows that the processing of the residual component can further improve the prediction accuracy.
(3) GVMD-ELM-ARIMA has the minimum MAPE, RMSE and MAE, which proves that GVMD-ELM-ARIMA has better prediction performance and can effectively improve the prediction accuracy.

4.5 India simulation result

As the experimental process has been demonstrated in detail in Section 4.4, in order to save space, this section briefly presents the experimental process and result of the cumulative confirmed data in India. The prediction comparison result of each model is shown in Figure 14. The prediction error distribution diagram is shown in Figure 15, and the error box diagram is shown in Figure 16. The evaluation index of each model is shown in Table 3.

![Figure 14](image_url)

(a) Complete result

![Figure 15](image_url)

(b) Locally enlarged view

*Figure 14. The prediction comparison result of each model*
Figure 14, Figure 15 and Figure 16 show that compared with other models, GVMD-ELM-ARIMA has the best prediction effect. The coordinates of the predicted points are relatively close to those of the original data points, and the errors have been fluctuating slightly around the base line. The accuracy of the hybrid model is still higher than that of the single model, which is consistent with the results in Section 4.4. GVMD-ELM-ARIMA proposed in this paper still has the smallest MAPE, RMS and MAE. In conclusion, the experimental conclusion in this section is consistent with the result in section 4.4.

4.6 Russia simulation result

In order to save space, this section briefly presents the experimental process and result of the
cumulative confirmed data in Russia. The prediction comparison result of each model is shown in Figure 17. The prediction error distribution diagram is shown in Figure 18, and the error box diagram is shown in Figure 19. The evaluation index of each model is shown in Table 4.

Figure 17. The prediction comparison result of each model
The experimental conclusion in this section is the same as that in section 4.4. So it will not be repeated.

5 Multi-step prediction

The GVMD-ELM-ARIMA model proposed in this paper is used to predict the cumulative COVID-19 confirmed data in the United States, India and Russia, and its prediction effect is verified. It includes data from January 5, 2021 to October 18, 2021 and the data is collected every day. The cumulative COVID-19 confirmed data can be obtained at website https://data.humdata.org/dataset/novel-coronavirus-2019-ncov-cases.
Each country has 287 cumulative COVID-19 confirmed data, and the data of each country are simulated respectively. Considering the multi-step prediction and the influence of outliers on the experiment, this paper takes the average value of confirmed data every 7 days as the reference point. 266 confirmed data (38 sample points) from January 5, 2021 to September 27, 2021 are selected as the training set, and 21 data (3 sample points) from September 28, 2021 to October 18, 2021 are selected as the test set. Finally, a total of 10 values are predicted, and the first three predicted values are compared with the test set as the verification point of the prediction effect. In order to facilitate readers to understand the selection of data sets, Figure 20 is provided as a reference. Figure 21 represents the forecast results of COVID-19 cumulative data in the US, India and Russia respectively.

![Figure 20. Schematic diagram of data set selection](image)
Figure 21 shows that 10 points are predicted, and each point represents the average value of one week, which can represent the prediction data of 70 days in total. The first three predicted values are close to the real test data, and good results are obtained, which proves the effectiveness of the proposed model.

6 Discussion

Qiang et al. [27] proposed a hybrid prediction model for COVID-19 outbreak based on ensemble empirical mode decomposition (EEMD). CEEMDAN is an improved version of EEMD. CEEMDAN can solve the problem of modal aliasing and is more widely used in real life. However, this algorithm will add noise to the signal for many times and then carry out overall average every time when IMF is calculated, which increases the complexity of the algorithm. In practice, it can cause modal aliasing, intermittent and high
frequency anomalies, etc. However, VMD is essentially a set of adaptive Wiener filter, which has strong robustness to noise, so VMD is introduced in this paper.

The decomposition effect of VMD is better, but it has high requirements on mode number $k$ and parameter $\alpha$, and the setting of parameters will directly affect the decomposition result. GBO has a fast convergence rate and can avoid the local optimal problem. Therefore, this paper combines GBO and VMD to propose GVMD to determine the mode number $k$ and parameter $\alpha$. As shown in Figure 5 and Figure 6, compared with CEEMDAN decomposition result, GVMD decomposition result is relatively smooth and mainly consists of low-frequency components, indicating that GVMD is suitable for data decomposition in this experiment. GVMD has strong adaptability and can be applied to signal denoising, feature extraction, fault diagnosis and time series prediction.

ELM has the advantages of fast speed and strong universality. It can be seen from the results in Table 2, Table 3 and Table 4 that the prediction error of ELM alone in this experiment is small. Therefore, ELM is selected to predict IMFs, and GVMD-ELM is proposed. At the same time, it can be found that although the GVMD-ELM has achieved a better prediction effect, there are still some errors. ARIMA can process high-frequency components with fewer parameters and simple application. In order to further reduce the error, according to the characteristics of high-frequency IMFr, ARIMA is introduced to process IMFr, and GVMD-ELM-ARIMA is proposed. Comparing GVMD-ELM-ELM and GVMD-ELM-ARIMA with other prediction models, it can be found that the processing of IMFr can better improve the prediction accuracy. At the same time, the prediction result of GVMD-ELM-ELM and GVMD-ELM-ARIMA are compared. The comparison result shows that ARIMA has a better effect on IMFr than ELM, and MAPE, RMSE, MAE in Table 2, Table 3 and Table 4 are reduced to a certain extent.

At present, there are many studies on the prediction of the cumulative COVID-19 confirmed data, but most of them are based on a single model. In this paper, a variety of hybrid models are proposed to compare with the single model. The simulation result shows that the prediction accuracy of the hybrid model is all higher than that of the single model, which is also reflected in the experimental results in Section 4.4, Section 4.5 and Section 4.6. At the same time, it can be found that although the hybrid model has achieved a better prediction effect, there are still some errors. To solve this problem, IMFr is processed for the first time in this paper. Comparing GVMD-ELM-ELM and GVMD-ELM-ARIMA with other prediction models, it can be found that the processing of IMFr can better improve the prediction accuracy. At the same time, the prediction result of GVMD-ELM-ELM and GVMD-ELM-ARIMA are compared. The comparison result shows that ARIMA has a better effect on IMFr than ELM, and MAPE, RMSE, MAE in Table 2, Table 3 and Table 4 are reduced to a certain extent. Therefore, the ARIMA model is applied to the residual processing after the decomposition of the cumulative COVID-19 confirmed data.

The GVMD-ELM-ARIMA proposed in this paper achieves the best prediction effect in the single-step prediction of the mixed model, and has the lowest MAPE, RMSE and MAE. In order to test the prediction effect of GBO-ELM-ARIMA, the cumulative COVID-19 confirmed data of various countries is predicted in multiple steps. Figure 20 shows the multi-step prediction results of GVMD-ELM-ARIMA for cumulative COVID-19 confirmed data. This prediction method has achieved certain prediction results and realized 40 days data prediction, and predicted the epidemic development under the condition of obtaining more data. It will provide a new method for cumulative COVID-19 confirmed data prediction.

The GVMD-ELM-ARIMA proposed in this paper provides a new method for the prediction of cumulative COVID-19 diagnostic data, but with limited energy and knowledge level, there are still some problems to be further studied.
(1) For the IMFs and IMFr components after data decomposition, only several methods are selected for prediction and comparison, and the optimal model is selected. In fact, more prediction models can be explored to study the effect of prediction.

(2) Due to the limitations of neural network, it needs a certain amount of data as support. When the amount of data is small, there will be some problems in the accuracy of prediction. Therefore, more other processing methods of medium and short-term prediction can be considered.

(3) If we can use the online real-time data to compare the predicted data with the actual data in time and realize error correction, it will be a good improvement direction.

(4) In multi-step prediction, the prediction accuracy will descend with the increase of prediction step, which is also a direction worthy of improvement.

7 Conclusions

To improve the prediction accuracy of the cumulative COVID-19 confirmed data, avoid the low prediction accuracy of single model and the deficiency of CEEMDAN, GVMD-ELM-ARIMA is proposed in this paper. The main conclusions and innovations in this paper are as follows:

(1) VMD is applied to the prediction of the cumulative COVID-19 confirmed data, which avoids the disadvantages of CEEMDAN such as modal aliasing and too long operation time, and improves the prediction accuracy.

(2) VMD needs to preset the value of mode number $k$ and parameter $\alpha$, which will directly affect the decomposition result. In order to solve this problem, GVMD is proposed. The multiscale fuzzy entropy is used as fitness value to determine the mode number $k$ value and parameter $\alpha$ value, which enhances the adaptability of VMD.

(3) In order to further improve the prediction accuracy of the cumulative COVID-19 confirmed data, ARIMA is used for the first time to analyze and predict the residual component after GVMD.

(4) GVMD-ELM-ARIMA is proposed. The five comparison models are brought into the cumulative COVID-19 confirmed data by the United States, India and Russia for one-step simulation test. The results show that the GVMD-ELM-ARIMA model has good prediction accuracy.

(5) The GVMD-ELM-ARIMA model is introduced into the latest epidemic data of the United States, India and Russia to carry out multi-step prediction, and its prediction results are verified. It provides a new method for predicting cumulative COVID-19 confirmed data.

Appendix A. The nomenclature used in this paper.

| List of nomenclature                          |
|----------------------------------------------|
| AIC  Akaike information criterion            |
| ANN  Artificial neural network               |
| AR   Autoregressive                          |
| ARIMA Autoregressive integrated moving average |
| EEMD Ensemble empirical mode decomposition   |
| CEEMD Complete ensemble empirical mode decomposition |
| CEEMDAN Complete ensemble empirical mode decomposition with adaptive noise |
| COVID-19 Corona virus disease 2019           |
| DWD  Discrete wavelet decomposition          |
| ELM  Extreme learning machine                |
| Abbreviation | Description |
|--------------|-------------|
| FNN | Feedforward neural network |
| GBO | Gradient-based optimizer |
| GSR | Gradient search rule |
| GVMD | Gradient-based optimizer variational mode decomposition |
| IMF | Intrinsic mode function |
| IMFr | Intrinsic mode function residual component |
| IMFs | Intrinsic mode functions |
| LEO | Local escaping operator |
| LSTM | Long short-term memory |
| MA | Moving average |
| MAE | Mean absolute error |
| MAPE | Mean absolute percentage error |
| MFE | Multiscale fuzzy entropy |
| RMSE | Root mean square error |
| SEIR | Susceptible exposed infectious recovered |
| SI | Susceptible infectious |
| SIR | Susceptible infectious recovered |
| SVM | Support vector machine |
| VMD | Variational mode decomposition |

**References**

[1] Santosh KC. COVID-19 prediction models and unexploited data. Journal of Medical Systems 2020, 44(9): 170(1-4).

[2] Bhapkar HR, Mahalle PN, Dey N, Santosh KC. Revisited COVID-19 mortality and recovery rates: Are we missing recovery time period?. Journal of Medical Systems 2020, 44(12): 202(1-5).

[3] Xiang Y, Jia, YH, Chen LL, Guo L, Shu, BZ, Long ES, COVID-19 epidemic prediction and the impact of public health interventions: A review of COVID-19 epidemic models. Infectious Disease Modelling 2021, 6: 324-342.

[4] Santosh KC. AI-driven tools for coronavirus outbreak: Need of active learning and cross-population train/test models on multidudinal/multimodal data. Journal of Medical Systems 2020, 44(5): 93(1-5).

[5] Cong W, Jie Y, Xu W, Min L. Analysis on early spatiotemporal transmission characteristics of COVID-19. Acta Physica Sinica 2020, 69(8): 080701(1-10).

[6] Cooper I, Mondal A, Antonopoulos CG. A SIR model assumption for the spread of COVID-19 in different communities. Chaos, Solitons and Fractals 2020, 139: 110057(1-19).

[7] Rafieenasab S, Zahiri AP, Roohi E. Prediction of peak and termination of novel coronavirus COVID-19 epidemic in Iran. International Journal of Modern Physics C 2020, 31(11): 2050152(1-23).

[8] Singh P, Gupta A. Generalized SIR (GSIR) epidemic model: An improved framework for the predictive monitoring of COVID-19 pandemic. ISA Transactions 2021. DOI: 10.1016/j.isatra.2021.02.016.

[9] Annas S, Pratama MI, Rifandi M, Sanusi W, Side S. Stability analysis and numerical simulation of SEIR model for pandemic COVID-19 spread in Indonesia. Chaos, Solitons and Fractals 2020, 139: 110072(1-17).

[10] Feng S, Feng ZB, Ling C, Chang C, Feng ZK. Prediction of the COVID-19 epidemic trends based on SEIR and AI models. Plos One 2021, 16(1): e0245101(1-15).
[11] Ibrahim MA, Al-Najafi A. Modeling, control, and prediction of the spread of COVID-19 using compartmental, logistic, and Gauss models: A case study in Iraq and Egypt. Processes 2020, 8(11): 1400(1-16).

[12] Ladh N, Bhardwaj P, Charan J, Mitra P, Goyal JP, Sharma P, Singh K, Misra S. Association of environmental parameters with COVID-19 in Delhi, India. Indian Journal of Clinical Biochemistry 2020. DOI: 10.1007/s12291-020-00921-6.

[13] Rath S, Tripathy A, Tripathy AR. Prediction of new active cases of coronavirus disease (COVID-19) pandemic using multiple linear regression model. Diabetes & Metabolic Syndrome: Clinical Research & Reviews 2020, 14(5): 1467-1474.

[14] Balli S. Data analysis of COVID-19 pandemic and short-term cumulative case forecasting using machine learning time series methods. Chaos, Solitons and Fractals 2021, 142: 110512.

[15] Moftakhar L, Seif M, Safe MS. Exponentially increasing trend of infected patients with COVID-19 in Iran: A comparison of neural network and ARIMA forecasting models. Iranian Journal of Public Health 2020, 49: 92-100.

[16] Hamadneh NN, Tahir M, Khan WA. Using artificial neural network with prey predator algorithm for prediction of the COVID-19: The case of Brazil and Mexico. Mathematics 2021, 9(2): 180(1-14).

[17] Shyam Sunder Reddy K, Padmanabha Reddy YCA, Mallikarjuna Rao Ch. Recurrent neural network based prediction of number of COVID-19 cases in India. Materials Today: Proceedings 2020. DOI: 10.1016/j.matpr.2020.11.117.

[18] Shahid F, Zameer A, Muneeb M. Predictions for COVID-19 with deep learning models of LSTM, GRU and Bi-LSTM. Chaos, Solitons and Fractals 2020, 140: 110212(1-9).

[19] Zhang H, Cui WT, Kang ZJ, Yang T, Lou B, Chi YT, Long H, Ma M, Yuan Q, Zhang SP, Zhang D, Xin JM. Predicting COVID-19 using hybrid AI mode. IEEE Transactions on Cybernetics 2020, 50(7): 2891-2904.

[20] Zhang GP, Liu XD. Prediction and control of COVID-19 spreading based on a hybrid intelligent model. Plos One 2021, 16(2): e0246360(1-12).

[21] Yang H, Gao LP, Li GH. Underwater acoustic signal prediction based on MVMD and optimized kernel extreme learning machine. Complexity 2020, 2020: 6947059(1-18).

[22] Wang RH, Li CS, Fu WL, Tang G. Deep learning method based on gated recurrent unit and variational mode decomposition for short-term wind power interval prediction. IEEE Transactions on Neural Networks and Learning Systems 2020, 31(10): 3814-3827.

[23] Shrivastava NA, Lohia K, Panighra BK. A multiobjective framework for wind speed prediction interval forecasts. Renewable Energy 2016, 87: 903-910.

[24] Wang DY, Wei S, Luo HY, Yue CQ, Grunder O. A novel hybrid model for air quality index forecasting based on two-phase decomposition technique and modified extreme learning machine. Science of the Total Environment 2017, 580: 719-733.

[25] Chu JW, Dong YC, Han XX, Xie J, Xu XY, Xie G. Short-term prediction of urban PM2.5 based on a hybrid modified variational mode decomposition and support vector regression model. Environmental Science and Pollution Research 2021, 28(1): 56-72.

[26] Singh S, Parmar KS, Kumar J. 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 and Fractals 2020, 135: 109866(1-8).

[27] Qiang XL, Aamir M, Naeem M, Ali S, Aslam A, Shao ZH. Analysis and forecasting COVID-19 outbreak in Pakistan using decomposition and ensemble model. Computers, Materials and Continua 2021,
[28] Li YX, Li Y A, Chen X, Yu J. Research on ship-radiated noise denoising using secondary variational mode decomposition and correlation coefficient. Sensors 2018, 18(1): 48(1-17).

[29] Jiang Y, Huang GQ, Yang QS, Yan ZT, Zhang CF. A novel probabilistic wind speed prediction approach using real time refined variational model decomposition and conditional kernel density estimation. Energy Conversion and Management 2019, 185: 758-773.

[30] Lang X, Rehman NU, Zhang YF, Xie L, Su HY. Median ensemble empirical mode decomposition. Signal Processing 2020, 176: 107686(1-8).

[31] Yang H, Li LL, Li GH, Guan QR. A novel feature extraction method for ship-radiated noise. Defence Technology 2021. DOI: 10.1016/j.dt.2021.03.012.

[32] Dragomiretskiy K, Zosso D. Variational mode decomposition. IEEE Transactions on Signal Processing 2014, 62(3): 531-544.

[33] Li GH, Chang WN, Yang H. A new hybrid model for underwater acoustic signal prediction. Complexity 2020, 2020: 5167469(1-19).

[34] Yang H, Cheng YX, Li GH. A denoising method for ship radiated noise based on Spearman variational mode decomposition, spatial-dependence recurrence sample entropy, improved wavelet threshold denoising, and Savitzky-Golay filter. Alexandria Engineering Journal 2021, 60(3): 3379-3400.

[35] Chen DN, Zhang YD, Yao CY, Lai BW, Lv SJ. Fault diagnosis method based on variational mode decomposition and multi-scale permutation entropy. Computer Integrated Manufacturing Systems 2017, 23(12): 2604-2612.

[36] Zheng XX, Zhou GW, Ren HH, Fu Y. A rolling bearing fault diagnosis method based on variational mode decomposition and permutation entropy. Journal of Vibration and Shock 2017, 36(22): 22-28.

[37] Ahmadianfar I, Bozorg-Haddad O, Chu XF. Gradient-based optimizer: A new metaheuristics optimization algorithm. Information Sciences 2020, 540: 131-159.

[38] Tang ZP, Zhang TT, Wu JC, Du XX, Chen KJ. Multistep-ahead stock price forecasting based on secondary decomposition technique and extreme learning machine optimized by the differential evolution algorithm. Mathematical Problems in Engineering 2020, 2020: 2604915(1-13).

[39] Kasun LLC, Zhou HM, Huang GB, Vong CM. Representational learning with ELMs for big data. IEEE Intelligent Systems 2013, 28(6): 31-34.

[40] Chorowski J, Wang J, Zurada JM. Review and performance comparison of SVM- and ELM-based classifiers. Neurocomputing 2014, 128: 507-516.

[41] Yang H, Li LL, Li GH. A new denoising method for underwater acoustic signal. IEEE Access 2020, 8: 201874-201888

[42] Chen JX, Zhu Z, Zhang XD. Feature cognitive model combined by an improved variational mode and singular value decomposition for fault signals. Cognitive Computation and Systems 2020, 2(2): 66-71.

[43] Li GH, Zheng MT, Yang H. Cycle analysis method of tree ring and solar activity based on variational mode decomposition and Hilbert transform. Advances in Meteorology 2019, 2019: 1715673(1-8).

[44] Yang H, Gao LP, and Li GH. Underwater acoustic signal prediction based on correlation variational mode decomposition and error compensation. IEEE Access 2020, 8: 103941-103955.

[45] Zheng JD, Pan HY, Cheng JS. Rolling bearing fault detection and diagnosis based on composite multiscale fuzzy entropy and ensemble support vector machines. Mechanical Systems and Signal Processing 2017, 85: 746-759.

[46] Li GH, Chang WN, Yang H. A novel combined prediction model for monthly mean precipitation with
error correction strategy. IEEE Access 2020, 8: 141432-141445.

[47] Huang G, Huang GB, Song SJ, You, KY. Trends in extreme learning machines: A review. Neural Networks 2015, 61: 32-48.

[48] Ding SF, Zhao H, Zhang YN, Xu XZ, Nie R. Extreme learning machine: Algorithm, theory and applications. Artificial Intelligence Review 2015, 44(1): 103-115.

Biographies

Guohui Li received the bachelor degree in electronic information engineering from Chongqing University of Technology, Chongqing, China, in 2001. He received the master in circuit and system degree from University of Electronic Science and Technology of China, Chengdu, China, in 2004. He received the doctor degree in acoustics from Northwestern Polytechnical University, Xi’an, China, in 2015. He is an associate professor in school of electronic engineering, Xi’an University of Posts and Telecommunications, Shaanxi, China. His research interests include chaotic signal processing and neural network prediction.

Kang Chen received the bachelor degree in optoelectronic Information Science and Engineering from Jiangxi Science and Technology Normal University, Jiangxi, China, in 2019. He is currently pursuing the master degree in electronics and communication engineering at Xi’an University of Posts and Telecommunications, Xi’an, China. His research interests include chaotic signal processing and neural network prediction.

Hong Yang received the bachelor and master degree in mechanical & electronic engineering from University of Electronic Science and Technology of China, Chengdu, China, in 2003 and 2006 respectively. She received the doctor degree in acoustics from Northwestern Polytechnical University, Xi’an, China, in 2015. She is an associate professor in school of electronic engineering, Xi’an University of Posts and Telecommunications, Shaanxi, China. Her research interests include chaotic signal processing and neural network prediction.

Declaration of interests

☐ √ 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.

Graphical abstract

Figure 3. The flow chart of GVMD-ELM-ARIMA
