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.
An innovative ensemble model based on multiple neural networks and a novel heuristic optimization algorithm for COVID-19 forecasting

Zongxi Qu a,b, Yutong Li a,b, Xia Jiang c, Chunhua Niu a,b,*

a School of Management, Lanzhou University, Lanzhou 730000, China
b Research Center for Emergency Management, Lanzhou University, Lanzhou 730000, China
c Affiliated Hospital of Northwest Minzu University/Second Provincial People’s Hospital of Gansu, Lanzhou 730009, China

ARTICLE INFO

Keywords:
COVID-19 forecasting
Ensemble forecasting
Neural network
Optimization algorithm

ABSTRACT

During the global fight against the novel coronavirus pneumonia (COVID-19) epidemic, accurate outbreak trend forecasting has become vital for outbreak prevention and control. Effective COVID-19 outbreak trend prediction remains a complex and challenging issue owing to the significant fluctuations in the COVID-19 data series. Most previous studies have limitations only using individual forecasting methods for outbreak modeling, ignoring the combination of the advantages of different prediction methods, which may lead to insufficient results. Therefore, this paper develops a novel ensemble paradigm based on multiple neural networks and a novel heuristic optimization algorithm. First, a new hybrid sine cosine algorithm-whale optimization algorithm (SCWOA) is exercised on 15 benchmark tests. Second, four neural networks are used as predictors for the COVID-19 outbreak forecasting. Each predictor is given a weight, and the proposed SCWOA is used to optimize the best matching weights of the ensemble model. The daily COVID-19 series collected from three of the most-affected countries were taken as the test cases. The experimental results demonstrate that different neural network models have different performances in various complex epidemic prediction scenarios. The SCWOA-based ensemble model can outperform all comparable models with its high accuracy and robustness.

1. Introduction

The global outbreak of COVID-19 caused by novel coronavirus pneumonia began in December 2019. By October 2021, the epidemic had spread to more than 200 countries worldwide, with nearly 230 million confirmed cases and 4.8 million deaths. Its strong infectivity, rapid spread, broad epidemic scope, and unpredictability pose a serious threat to global human health, social stability, and public security (Chakraborty & Maity, 2020). To slow the spread of COVID-19, governments around the world have implemented different degrees of prevention and control measures. However, strict measures such as lockdowns and quarantines may cause severe socio-economic consequences, and the variation of the epidemic is highly uncertain, making it difficult for governments to make the best decisions (Haug et al., 2020). When the number of new cases increases by thousands every day, even the health systems of developed countries have become overwhelmed and unable to deal with this large number of patients in such a short time. Effective prediction and indoctrination of prediction models can help governments estimate healthcare requirements and provide advice and information to the public (Dehning et al., 2020). For example, from the identification of exclusion zones and the organization of economic activities to the management of medical resources and the planning of emergency hospitals, effective forecasting is of strategic importance to decision-makers, which helps governments decide whether to impose or relax a restriction, thus minimizing the economic and political effects of the pandemic (Petrooulos, Makridakis, & Stylianou, 2020). Hence, forecasting the outcome of outbreaks as accurately as possible is crucial for decision-making and policy implementations.

Recently, many significant studies have been devoted to forecasting the upcoming number of cases and the spread of COVID-19 in the near future. Traditional epidemiological models have been widely adopted in predicting COVID-19 cases. The time-dependent susceptible, infectious,
and/or recovered (SIR) model is frequently used to model the growth of COVID-19 and to predict the future condition of infection and recovery rates (Alenezi, Al-Anzi, & Alabdalrazzaq, 2021; Cumsille, Rojas-Díaz, de Espanés, & Verdugo-Hernández, 2022; Masuhara & Hosoya, 2021). In addition, many studies also used the susceptible, exposed, infectious, and/or recovered (SEIR) model for COVID-19 epidemic prediction (Annas, Isbar Pratama, Rifandi, Sanusi, & Side, 2020; Das, Dhar, Goyal, Kundu, & Pandey, 2021; Paul, Mahata, Ghosh, & Roy, 2021; Piovela, 2020). The epidemiological approach attempts to model disease states, considering biological and disease processes, which requires preliminary assumptions, thus making the calculation process more complex. Another method of epidemic prediction is the statistical forecasting model. Ceylan (Ceylan, 2020) have applied autoregressive integrated moving average (ARIMA) model to forecast the epidemiological trend in Italy, Spain, and France. Ghosal et al. (Ghosal, Sengupta, Majumder, & Sinha, 2020) used linear and multiple linear regression methods to predict the number of deaths in India over a short period of 6 weeks. Moftakhar & Seif (Moftakhar & Seif, 2020) used the ARIMA model to forecast the patients of COVID-19 in Iran in the next 30 days. A further noteworthy forecasting method is to use machine learning models such as artificial neural networks (NNs) and support vector algorithms, which have recently become more prevalent in predicting infectious diseases. For example, Ly (Ly, 2020) employed an adaptive neuro-fuzzy inference system (ANFIS) to predict COVID-19 cases in the UK. The result shows that data from Spain and Italy can increase the forecasting ability of COVID-19 cases in the UK, Parbat & Chakraborty (Parbat & Chakraborty, 2020) used support vector regression (SVR) for a 60-day forecast of 2019 coronavirus cases in India based on time-series data reported from March 1, 2020, to April 30, 2020. Tomar & Gupta (Tomar & Gupta, 2020) assessed the number of confirmed cases of COVID-19 in India in the next 30 days and tested the effectiveness of quarantine measures using a long short-term memory (LSTM) model and curve fitting. Chimmula & Zhang (Chimmula & Zhang, 2020) used the LSTM model to predict the time series of COVID-19 transmission in Canada and compared the transmission rates in Canada, Italy, and the United States. A LSTM model based on a recurrent neural network (RNN) was also used to predict the future mutation rate of SARS-COV-2. Behnoood et al. (Behnoood, Mohammadi Golafshani, & Hosseini, 2020) used ANFIS and LSTM models to predict new infection cases in Bangladesh and compared the results of the two experiments, believing that LSTM results were more satisfactory.

Most previous studies are based on NNs for forecasting outbreaks in a data-driven manner. Nevertheless, different forecasting techniques have their strengths and weaknesses, implying that there is no single best model that can be applied all the various complex forecasting scenarios. To deal with sampling and modeling uncertainties, NNs are typically used as an ensemble of several network models. The ensemble combines the results of the different models that compose them to improve the accuracy and robustness of the predictions (Kourentzes, Barrow, & Crone, 2014). Although the utilization of ensembles is accepted nowadays as the standard for NN prediction (Crone, Hibon, & Nikolopoulos, 2011), their performance is a function of combining the individual forecasts (Stock & Watson, 2004). Improving the operation mode of ensemble prediction has a direct effect on prediction accuracy and the decision-making of prediction support, including some prediction applications in different fields, such as economic modeling and policy-making (McAdam & McNelis, 2005; Stock & Watson, 2004), temperature forecast, weather (Angelella, Basile, Bonfante, & Greco, 2010), wind energy forecasting (Xiao, Wang, Dong, & Wu, 2015; W. Zhang et al., 2017), fault detection (X. Zhang et al., 2021), impulsive reaction-diffusion NNs (Wei, Li, & Stojanovic, 2021), and climate modeling (Filides & Kourentzes, 2011).

The ensemble of NNs is the basis for achieving precise forecasts for these various applications; hence, it is essential to enhance the construction of the ensembles. Currently, a commonly accepted way to combine individual models is to assign a weighting coefficient to each individual model, i.e., the weighting-based combination method (Xiao et al., 2015). These methods assign a weighting coefficient to each component model based on its performance in the determination procedure (by checking the measured and predicted values). Compared with ensemble averaging, the weighted ensemble can combine the strengths of various models to obtain more stable forecasting results. However, the determination of optimal weights is vital to ensemble prediction. Several previous studies have attempted to utilize heuristic algorithms for weight optimization. For example, Yang et al. (Yang, Chen, Wang, Li, & Li, 2016) used a differential evolution (DE) heuristic algorithm to construct ensemble models to determine the optimal weights for electricity demand ensemble forecasting. Xiao et al. (Xiao et al., 2015) used the cuckoo search optimization (CSO) algorithm to optimize the weight coefficients of the combined model. The search ability of the heuristic algorithm affects the prediction performance of the weight-based ensemble model. The whale optimization algorithm (WOA), proposed by Mirjalili & Lewis (Mirjalili & Lewis, 2016b), is a novel heuristic algorithm that imitates the hunting mechanism of humpback whales. Despite its reasonable convergence rate, it is not suitable for highly complicated functions and may still encounter problems such as being trapped in an optimal local solution and slow convergence (Deepa & Venkataraman, 2021; Vijaya Lakshmi & Mohanaiah, 2021). To overcome these weaknesses and improve its search capability, we propose a new hybrid heuristic algorithm, which is a combination of WOA and the sine cosine algorithm (SCA). SCA is based on sine and cosine functions and can be applied for exploitation and exploration phases in global optimization functions. Therefore, we use SCA to search for the initial random position of WOA to improve the number of iterations and the convergence rate. The proposed hybrid variant is called SCWOA.

Through this analysis, a novel ensemble paradigm based on multiple NNs and a new type of heuristic optimization algorithm (SCWOA) is introduced in this paper. First, four individual NN methods, back-propagation neural network (BPNN), Elman neural network (ENN), ANFIS, and LSTM, are selected to predict the COVID-19 outbreak. Then, an ensemble forecasting method, which integrates the characteristics of each forecasting method, is developed to improve the fitting accuracy and forecasting capability of the model. Finally, the best weight coefficients for each prediction member are automatically obtained by a new optimization algorithm SCWOA. To verify and validate the proposed model, the daily COVID-19 series collected from three of the most-affected countries were taken as test cases to conduct the empirical study. The main contributions in this paper are as follows. (1) An innovative ensemble prediction framework is proposed for COVID-19 epidemic prediction. In this architecture, four NN models are combined, and the weights of the ensemble model are optimized by SCWOA to achieve better performance than obtained with a single prediction model (2) A new type of heuristic optimization algorithm (SCWOA) is introduced. To improve both the exploration and exploitation capacities, the hybrid SCWOA is proposed, and to evaluate the improved algorithm, 15 benchmark functions are used. (3) The proposed ensemble model can effectively integrate the advantages of multiple NN models to achieve stable and accurate prediction results.

The remainder of this paper is organized as follows. Section 2 introduces four NN methods and the proposed heuristic optimization algorithm SCWOA. Section 3 describes the main procedure of the proposed ensemble forecasting model. Section 4 describes data collection and evaluation criteria. Section 5 discusses the forecasting results of the proposed model and provides a comparison of results with other models. Finally, Section 6 concludes with the essential results of this paper.
2. Methodology

2.1. BPNN

BPNN is a multilayer feedforward NN with a wide variety of applications. The diagram of BPNN includes an input layer, one or more hidden layers, and an output layer, as shown in Fig. 1.

BPNN has two learning processes, forward and reverse propagation. Assuming an L-layer, n-node NN, each layer of neurons only accepts the output information of the previous layer and then propagates to the next layer of neurons. Suppose the network outputs $y_i$, $N$ samples $(x_k, y_k)$ $(k = 1, 2, ..., N)$, and the output of node $i$ is $O_k$. The network input is $x_k$, the output is $y_k$, then the output of node $i$ is $O_k$. For node $i$ in the first layer, when the sample $k$ is input, the input of node $i$ is $E_k = \frac{1}{2}(y_k - y_p)^2$.

where $y_p$ is the actual output of the NN and $y_i$ is the ideal output. 

Assume $O_k = f(\text{net}_k)$ then,

$$\frac{\partial E}{\partial O_k} = \frac{\partial E_k}{\partial \text{net}_k} \frac{\partial \text{net}_k}{\partial O_k} = \frac{\partial E_k}{\partial \text{net}_k} O^{-1}$$

If the $i$th node is the output unit, then $O_k = y_p$,

$$\delta_i = \frac{\partial E}{\partial O_i} = \frac{\partial E_k}{\partial y_p} = - (y_k - y_p) f'(\text{net}_k)$$

If the $i$th node is not the output unit, then,

$$\delta_i = \frac{\partial E}{\partial O_i} = \frac{\partial E_k}{\partial O_k} \frac{\partial O_k}{\partial \text{net}_k} \frac{\partial \text{net}_k}{\partial \text{net}_i} f'(\text{net}_i)$$

where $\delta_i$ is the input of the next layer of $L$ layers. To calculate $\frac{\partial E}{\partial O_i}$, we calculate it from the next layer of $L$ layers:

$$E = \frac{1}{2N} \sum_{k=1}^{N} E_k$$

where $m$ represents the $m$th unit in the next layer of the $L$ layers.

Finally, from the above two formulas, we can obtain.

$$\delta_i = \sum_{m} \delta_{i+1} w_{im} f'(\text{net}_m)$$

2.2. ENN

Elman first proposed the ENN in 1990 (Elman, n.d.). ENN is a common example of a dynamic recurrent network, and its structure consists of an input layer with a specific context node, a hidden layer, and an output layer. The major benefit of ENN is that the context nodes may be utilized to remember previous hidden node activations, which makes it suitable in the fields of dynamic system identification and predictive control. The mathematics of the ENN can be described as follows.

Let the external input of the network be $u$. The output is $y$, and the output of the hidden layer is $x$. Then we have.

$$x(k) = f(w_1^1 x(k) + w_2^1 u(k - 1))$$

$$y(k) = g(w_3^1 x(k))$$

where $w_1^1$, $w_2^1$, and $w_3^1$ denote the connection weighting matrix from the context layer to the implicit layer, the input layer to the hidden layer, and the hidden layer to the output layer, respectively. Here $f$ and $g$ are the transfer functions of the implicit layer and the output layer.

From Eqs. (8)-(10),

$$x(k) = x(k-1)$$

Then

$$x(k-1) = x(k-2)$$

where $x(k)$ relies on $w_{1,2}$ at different moments, so $x(k)$ is a dynamic recursive process. Accordingly, the backpropagation algorithm used for Elman regression NN training is the dynamic backpropagation learning algorithm.

2.3. ANFIS

ANFIS, proposed by Jang (Jang, 1993), combines the advantages of
The fuzzy membership grade of the inputs, which are given by.

cally extract input and sample data rules, thereby forming an adaptive

\( \mu \) denotes the sample size.

O

first-order polynomial (for a first-order Sugeno model). Thus, the out

this layer is simply the product of the normalized firing strength and a

indicating that they perform as a simple multiplier. The outputs of this layer

the previous layer. The outputs of this layer can be represented as.

where \( \mu_k(x) \) and \( \mu_{k+1}(y) \) can adopt any fuzzy membership function.

Layer 2: The nodes are fixed nodes. They are labeled with \( M \), indicating that they perform as a simple multiplier. The outputs of this layer can be represented as.

\[
O^1_i = \mu_k(x) i = 1, 2
\]

\[
O^1_i = \mu_{k+1}(y) i = 3, 4
\]

Layer 3: The nodes are also fixed nodes. They are labeled with \( N \), indicating that they play a normalization role to the firing strengths from the previous layer. The outputs of this layer can be represented as.

\[
O^2_i = \omega_i = \mu_k(x) \mu_k(y) i = 1, 2
\]

Layer 4: The nodes are adaptive nodes. The output of each node in this layer is simply the product of the normalized firing strength and a first-order polynomial (for a first-order Sugeno model). Thus, the outputs of this layer are given by.

\[
O^3_i = \omega_i = \frac{\omega_0}{\omega_1 + \omega_2} i = 1, 2
\]

Layer 5: There is only one single fixed node labeled with \( S \). This node

performs the summation of all incoming signals. Hence, the overall output of the model is given by.

\[
O^4_i = \sum_{i=1}^{\omega_1} \frac{a_0}{\omega_1 + \omega_2} + \frac{a_2}{\omega_1 + \omega_2} \]

The final output of ANFIS is.

\[
f_{out} = \bar{m}_1 \bar{f}_1 + \bar{m}_2 \bar{f}_2 = \frac{a_0}{\omega_1 + \omega_2} + \frac{a_2}{\omega_1 + \omega_2} f_1 + f_2
\]

\[
= (\bar{m}_1 x) p_1 + (\bar{m}_2 x) q_1 + (\bar{m}_1 x) r_1 + (\bar{m}_2 x) p_2 + (\bar{m}_1 x) q_2 + (\bar{m}_2 x) r_2
\]

2.4. LSTM

LSTM was introduced by Hochreiter and Schmidhuber (Hochreiter & Schmidhuber, 1997) and refined and popularized by many people in the following work. LSTM was designed explicitly to avoid the long-term dependency problem. Like RNNs, LSTMs also have a chain-like structure, but the repeating module has a different structure.

The first step of LSTM is to decide what information should be discarded from the cell state. This decision is made by a sigmoid layer called the “forget gate layer.” Inputs are \( h_{t-1} \) and \( x_t \), and the output is a number between 0 and 1 for each number in the cell state \( C_t \). A “1” means “keep this completely,” whereas a “0” means “forget this completely.”

\[
f_t = \sigma(W_f h_{t-1} + x_t)
\]

The next step is to decide what information to store in the cell state, including two parts. First, the “input gate layer” determines which values will be updated, and then the tanh layer creates a vector of new candidate values \( \tilde{C}_t \). Next, these two layers are combined to create an update to the state:

\[
i_t = \sigma(W_i h_{t-1} + x_t)
\]

\[
C_t = \tanh(W_C h_{t-1} + x_t) + C_{t-1}
\]

The next step is to multiply the old state by \( f_t \), forgetting the things we decided to earlier. Then, add \( \tilde{C}_t \) to it.

\[
C_t = f_t \cdot C_{t-1} + i_t \cdot \tilde{C}_t + C_{t-1}
\]

The final step is to run a sigmoid layer which decides what parts of the cell state to output. Then, put the cell state through tanh (to push the values to be between \(-1 \) and \( 1 \)) and multiply it by the output of the sigmoid gate:

\[
O_t = \sigma(W_o h_{t-1} + x_t)
\]

\[
h_t = o_t \cdot \tanh(C_t)
\]
2.5. A novel heuristic optimization algorithm: SCWOA

To improve the performance of the ensemble model, an improved heuristic algorithm SCWOA is proposed to determine the optimal weight coefficient of the ensemble model. This section introduces the basic WOA, SCA, and the novel SCWOA.

Table 2

Unimodal benchmark functions.

| Function type | Function | Dim | Domain | Optimum value |
|---------------|----------|-----|--------|---------------|
| Unimodal      | \(F_1(x) = \sum_{i=1}^{n} x_i^2\) | 30  | [-100,100] | 0             |
|               | \(F_2(x) = \sum_{i=1}^{n} |x_i| + \sum_{i=1}^{n} |x_i|\) | 30  | [-10,10] | 0             |
|               | \(F_3(x) = \sum_{i=1}^{n} (x_i^2)^2\) | 30  | [-100,100] | 0             |
|               | \(F_4(x) = \max(\{|x_i, 1\} ; |x_i|\})\) | 30  | [-100,100] | 0             |
|               | \(F_5(x) = \sum_{i=1}^{n} \left[100(x_{i+1} - x_i^2)^2 + (x_i - 1)^2\right]\) | 30  | [-30,30] | 0             |

Multimodal

\(F_6(x) = \sum_{i=1}^{n} [x_i^2 - 10 \cos(2\pi x_i)] + 10\)
\(F_7(x) = -20 \exp(-0.2 \sqrt{\frac{1}{n} \sum_{i=1}^{n} x_i^2} + \exp(\frac{1}{n} \sum_{i=1}^{n} \cos(2\pi x_i))) + 20 + e\)
\(F_8(x) = \frac{1}{8000} \sum_{i=1}^{n} x_i^2 + \sum_{i=1}^{n} \cos(x_i) + 1\)
\(F_9(x) = \frac{x}{n} \left[10 \sin(x_1) + \sum_{i=1}^{n-1} (y_{i+1} - 1)^2 \left[1 + 10 \sin^2(y_{i+1}) + (y_{i+1})^2\right]\right] + \sum_{i=1}^{n} u(x_i, 10, 100, 4) + \frac{x_{i+1}}{4} u(x_i, a, k, m)\)
\(F_{10}(x) = 0.1 \left[\sin^2(3\pi x_1) + \sum_{i=1}^{n-1} (x_{i+1} - 1)^2 \left[1 + \sin^2(3\pi x_{i+1} + 1)\right] + (x_1 - 1)^2 \left[1 + \sin^2(2\pi x_1)\right]\right] + \sum_{i=1}^{n} u(x_i, 5, 100, 4)\)

Fixed-dimension multimodal

\(F_{11}(x) = \left(\frac{1}{500} + \frac{1}{\sum_{i=1}^{n} (x_i - a_i)^2}\right)^{-1}\)
\(F_{12}(x) = \left[\frac{x_i \left(\frac{b_i}{2} + b_i x_i\right)}{b_i + b_i x_i + x_i}\right]^2\)
\(F_{13}(x) = \left(x_2 - \frac{5}{4} \left(\frac{1}{x_1^2} + \frac{5}{\pi^2} - 6\right) + 10(1 - \frac{1}{8\pi} \cos x_1) + 10\right)\)
\(F_{14}(x) = \left[1 + (x_1 + x_2 - 1)^2 \left[19 - 14x_3 + 3x_3^2\right] + (x_1 - 1)^2 \left[1 + \sin^2(2\pi x_1)\right] + (x_2 - 1)^2 \left[1 + \sin^2(2\pi x_2)\right] + 30 + (2x_1 - 3x_2)^2 x_1 + 12x_2^2 + 4x_2 - 36x_1x_2 + 27x_2^2\right]\)
\(F_{15}(x) = -\sum_{i=1}^{n} \exp\left[ -\sum_{i=1}^{n} (x_i - p_k)^2\right]\)

Fig. 3. The graphs of each benchmark functions.

2.5.1. Overview of WOA

Miraljili and Lewis proposed WOA in 2016 to simulate the feeding mechanism of humpback whales (Miraljili & Lewis, 2016a). Humpback whales hunt near the surface by creating distinctive bubbles along a “9-shaped” or circular path as they circle their prey. First, the humpback whale dives about 10–15 m, then releases many bubbles creating a spiral
encirclement of the prey, and then swims toward the prey at the surface. The humpback whale’s fins flash and use the light from the fins to surround and immobilize the prey and prevent it from escaping. The mathematical models for encircling prey, spiral bubble net foraging maneuvers, and searching for prey are now described.

(a) Encircling prey

The humpback whale circles its prey, increasing the number of iterations to update its position to the best search agent. This process can be expressed mathematically as:

\[
\vec{D} = \left[ \vec{C} \vec{X}^* - \vec{X} \right] (26)
\]

\[
\vec{X}(t + 1) = \vec{X}^* - \vec{A} \cdot \vec{D} (27)
\]

where \( \vec{X}^* \) indicates the position vector of the best solution achieved so far, \( t \) is the current iteration, \( \vec{X} \) denotes the position vector, and \( \vec{C} \) and \( \vec{A} \) denote the coefficient vectors.

(b) Bubble-net attacking method

To mathematically model the humpback whale’s bubble net behavior, we designed the following two methods.

Shrinking encircling mechanism: By setting a random value \( \vec{A} \) in \([-1, 1]\), the new position of the search agent can be defined as any position between the position of the current best agent and the initial position of the search agent.

Spiral update position: The helical equation between the humpback whale and prey positions, mimicking the spiral motion of a humpback whale, is as follows:

\[
\vec{X}(t + 1) = \vec{D} \cdot e^{-p \cdot \cos(2\pi t)} + \vec{X}(t) (28)
\]

The probability \( p \) is a random number in \([0,1]\) and is assumed to be chosen between the spiral-shaped path and the shrinking encircling during the optimization process.

(c) Search for prey

In the exploration phase, the change of vectors \( \vec{A} \) can be searched for prey. Thus, \( \vec{A} \) can use random values greater than 1 or less than -1 to move away from the reference whale. At this stage, the mathematical model is as follows:

\[
\vec{X}(t + 1) = \vec{X}_{\text{rand}} - \vec{A} \cdot \vec{D} \quad (29)
\]

where \( \vec{X}_{\text{rand}} \) denotes a vector of random positions (random whales) chosen from the current population.

2.5.2. Overview of SCA

Mirjalili (Mirjalili, 2016) proposed SCA based on sine and cosine functions to explore different regions of the search space. In SCA, the search space dimension is determined by the number of parameters required for optimization. The SCA creates different initial random positions chosen from the current population.

To mathematically model the humpback whale (t) = \( \vec{X} \) – \( \vec{A} \cdot \vec{D} \) (27)

where \( \vec{X}^* \) indicates the position vector of the best solution achieved so far, \( t \) is the current iteration, \( \vec{X} \) denotes the position vector, and \( \vec{C} \) and \( \vec{A} \) denote the coefficient vectors.

2.5.2. Overview of SCA

Mirjalili (Mirjalili, 2016) proposed SCA based on sine and cosine functions to explore different regions of the search space. In SCA, the search space dimension is determined by the number of parameters required for optimization. The SCA creates different initial random positions chosen from the current population.

To mathematically model the humpback whale (t) = \( \vec{X} \) – \( \vec{A} \cdot \vec{D} \) (27)

where \( \vec{X}^* \) indicates the position vector of the best solution achieved so far, \( t \) is the current iteration, \( \vec{X} \) denotes the position vector, and \( \vec{C} \) and \( \vec{A} \) denote the coefficient vectors.

2.5.2. Overview of SCA

Mirjalili (Mirjalili, 2016) proposed SCA based on sine and cosine functions to explore different regions of the search space. In SCA, the search space dimension is determined by the number of parameters required for optimization. The SCA creates different initial random positions chosen from the current population.

To mathematically model the humpback whale (t) = \( \vec{X} \) – \( \vec{A} \cdot \vec{D} \) (27)

where \( \vec{X}^* \) indicates the position vector of the best solution achieved so far, \( t \) is the current iteration, \( \vec{X} \) denotes the position vector, and \( \vec{C} \) and \( \vec{A} \) denote the coefficient vectors.

2.5.2. Overview of SCA

Mirjalili (Mirjalili, 2016) proposed SCA based on sine and cosine functions to explore different regions of the search space. In SCA, the search space dimension is determined by the number of parameters required for optimization. The SCA creates different initial random positions chosen from the current population.

To mathematically model the humpback whale *X*(t) = \*X* - \*A* ⋅ \*D* (27)

where \*X* indicates the position vector of the best solution achieved so far, \( t \) is the current iteration, \*X denotes the position vector, and \*C and \*A denote the coefficient vectors.

2.5.2. Overview of SCA

Mirjalili (Mirjalili, 2016) proposed SCA based on sine and cosine functions to explore different regions of the search space. In SCA, the search space dimension is determined by the number of parameters required for optimization. The SCA creates different initial random positions chosen from the current population.

To mathematically model the humpback whale *X*(t) = \*X* - \*A* ⋅ \*D* (27)

where \*X* indicates the position vector of the best solution achieved so far, \( t \) is the current iteration, \*X denotes the position vector, and \*C and \*A denote the coefficient vectors.

2.5.2. Overview of SCA

Mirjalili (Mirjalili, 2016) proposed SCA based on sine and cosine functions to explore different regions of the search space. In SCA, the search space dimension is determined by the number of parameters required for optimization. The SCA creates different initial random positions chosen from the current population.

To mathematically model the humpback whale *X*(t) = \*X* - \*A* ⋅ \*D* (27)

where \*X* indicates the position vector of the best solution achieved so far, \( t \) is the current iteration, \*X denotes the position vector, and \*C and \*A denote the coefficient vectors.

2.5.2. Overview of SCA

Mirjalili (Mirjalili, 2016) proposed SCA based on sine and cosine functions to explore different regions of the search space. In SCA, the search space dimension is determined by the number of parameters required for optimization. The SCA creates different initial random positions chosen from the current population.

To mathematically model the humpback whale *X*(t) = \*X* - \*A* ⋅ \*D* (27)

where \*X* indicates the position vector of the best solution achieved so far, \( t \) is the current iteration, \*X denotes the position vector, and \*C and \*A denote the coefficient vectors.

2.5.2. Overview of SCA

Mirjalili (Mirjalili, 2016) proposed SCA based on sine and cosine functions to explore different regions of the search space. In SCA, the search space dimension is determined by the number of parameters required for optimization. The SCA creates different initial random positions chosen from the current population.

To mathematically model the humpback whale *X*(t) = \*X* - \*A* ⋅ \*D* (27)

where \*X* indicates the position vector of the best solution achieved so far, \( t \) is the current iteration, \*X denotes the position vector, and \*C and \*A denote the coefficient vectors.
conducted by WOA. When the SCA finds an optimal solution, the acceptance of these new solutions determined by the WOA obtain a better solution near the optimum. The purpose is to overcome the blindness of the random initialization, enrich the searching behavior, and accelerate the local convergence of the WOA.

The pseudo-code of the proposed SCWOA algorithm is outlined as follows.

**Algorithm.** Pseudocodes of SCWOA.

**Objective:**
Minimize and maximize the objective function $f(x_i) = \{x_1, x_2, ..., x_i\}$

**Parameters:**
- $\text{iter}$ - iteration number.
- $M$ - the maximum number of iteration.
- $I$ - a population pop.
- $p$ - the switch probability
- $X_i(i = 1, 2, 3, ..., I)$

(continued)

(continued on next column)
Update the position of the current agent by $X(t+1) = \overrightarrow{X_{\text{rand}}} \cdot \overrightarrow{D}$

end if
end for

Calculate the fitness of each search agent
Update $X^*$ if there is a better solution

iter = iter + 1
end while
return $X^*$

3. Framework of the ensemble forecasting model

This section describes the details of the ensemble forecasting model framework, which is shown in Fig. 2. The framework includes the ensemble forecasting theory and determining the combination weights by SCWOA.

3.1. Theory of the ensemble forecasting theory

It is increasingly recognized that the combination of models has advantages over the choice of an individual model, not only in terms of accuracy and error variability but also in terms of simplifying model building and selection. How to combine existing forecasting technologies to achieve perfect prediction results is a widely discussed topic. The ensemble prediction theory suggests that if there exist $n$ prediction techniques for addressing a certain prediction problem with appropriately assigned weighting coefficients, the prediction results of several techniques can be summed. A weighted ensemble is a methodology that allows multiple models to contribute to a forecast in proportion to their predictive or estimated performance.

Then the predicted value of the ensemble method can be formulated as:

Fig. 5. Convergence Curve of PSO, SCA, WOA, GWO and SCWOA variants on F6-F10 function.
\[ \hat{y}_t = \sum_{i=1}^{n} \hat{w}_i f_{it}, t = 1, 2, \ldots, m \]  
\[ \sum_{i=1}^{n} w_i = 1 \]  

where \( y_t \) is the actual time series data, \( w_i \) is the weight coefficient for the \( i \)th predictor, and \( f_{it} \) is the forecasting value of \( i \)th predictor at time \( t \).

The prediction error of the ensemble method is expressed as follows:

\[ e_t = y_t - \hat{y}_t = \sum_{i=1}^{n} w_i y_t - \sum_{i=1}^{n} \hat{w}_i f_{it} = \sum_{i=1}^{n} w_i (y_t - f_{it}) = \sum_{i=1}^{n} w_i e_t \]  

The ensemble method used in this paper combines four commonly used predictors: BPNN, ENN, ANFIS, and LSTM. Then the predicted value in Eq. (36) can be written as:

\[ \hat{Y}_{\text{Ensemble}}(t) = w_1 \hat{Y}_{\text{BP}}(t) + w_2 \hat{Y}_{\text{ELMAN}}(t) + w_3 \hat{Y}_{\text{ANFIS}}(t) + w_4 \hat{Y}_{\text{LSTM}}(t), t = 1, 2, \ldots, m \]  

where \( \hat{Y}_{\text{Ensemble}}(t) \), \( \hat{Y}_{\text{BP}}(t) \), \( \hat{Y}_{\text{ELMAN}}(t) \), \( \hat{Y}_{\text{ANFIS}}(t) \), and \( \hat{Y}_{\text{LSTM}}(t) \) are the predicted value by the ensemble method BPNN, ENN, ANFIS, and LSTM at time \( t \), respectively, \( w_i \) is the weight coefficient assigned to BPNN, ENN, ANFIS, and LSTM, with \( \sum_{i=1}^{n} w_i = 1 \), \( -1 \leq w_i \leq 1 \).

3.2. Determining the weights of the ensemble method by SCWOA

To forecast a certain problem, determining the weight coefficients of
each forecasting method is a vital issue for the ensemble method. A modeling averaging ensemble is one of the most commonly used combination strategies, which combines the prediction from each model equally and often results in better performance on average than a given single model. Clemen (Clemen, 1989) demonstrated that the simple averaging ensemble (i.e., a combination of equal weights) performs almost the best among the combination methods. de Menezes et al. (de Menezes, Bunn, & Taylor, 2000) reviewed several research works and concluded that the simple averaging ensemble performed best when the performance of individual predictions was comparable. Nevertheless, there are some well-performing prediction methods that we expect to contribute more to an ensemble prediction, and perhaps some less-skilled methods that may be useful but should have a smaller contribution to an ensemble prediction.

To obtain the weight coefficients of each predictor automatically, the proposed optimization algorithm SCWOA in this paper is adopted to search for optimal matching ensemble model weights. Moreover, the mean absolute percentage error function (MAPE) between the predicted and actual values of the models is chosen as the fitness function of SCWOA. Finding the optimal weights is the process of the heuristic search problem that minimizes MAPE by iteratively finding the optimal global solution. The fitness function is as follows:

$$\min_{\text{MAPE}} = \frac{1}{N} \sum_{n=1}^{N} \left| \frac{y_n - \hat{y}_n}{y_n} \right| \times 100\%$$

(37)

where $y_n$ and $\hat{y}_n$ denote actual and predicted values and $N$ is the total sample size.

Finally, the weights computed by SCWOA are multiplied with each single prediction result separately, and the results of all ensemble members are added to obtain the final prediction value.

4. Data description and evaluation criteria

4.1. Data description

In this paper, the model’s validity is verified by using the new crown

| Countries | Methods | Total cases | New cases | Total death cases | New death cases |
|-----------|---------|-------------|-----------|------------------|----------------|
| USA       | BPNN    | -0.6120     | 0.3320    | 0.0976           | 262.4478       |
|           | ENN     | 0.7603      | 0.1106    | 0.1443           | 75.8460        |
|           | ANFIS   | 1.0000      | 0.1374    | 1.0000           | 74.6520        |
|           | LSTM    | -0.1517     | 0.3508    | 0.0115           | 85.4976        |
| India     | BPNN    | 0.0110      | 0.3327    | -0.6712          | 30.1301        |
|           | ENN     | 1.0000      | 0.0230    | 0.1037           | 1.0000         |
|           | ANFIS   | -0.1110     | 0.1445    | 1.0000           | 17.8443        |
|           | LSTM    | 0.0003      | 0.0561    | 0.3926           | 21.2914        |
| Brazil    | BPNN    | 0.0015      | 0.5151    | 0.0014           | 28.5270        |
|           | ENN     | 0.2245      | 0.6607    | 0.2260           | 28.9823        |
|           | ANFIS   | 0.0000      | 0.4839    | 0.1412           | 40.9626        |
|           | LSTM    | 0.7740      | 0.2397    | 0.4581           | 31.3894        |

| Countries | Methods | RMSE | MAPE | $R^2$ | IA | TIC |
|-----------|---------|------|------|-------|----|-----|
| USA       | ARIMA   | 286215.7956 | 0.5443 | 0.5376 | 0.7710 | 0.0042 |
|           | LSSVM   | 115039.0835  | 0.2375 | 0.9253 | 0.9776 | 0.0017 |
|           | BPNN    | 172032.0221  | 0.3320 | 0.8329 | 0.9390 | 0.0025 |
|           | ENN     | 52578.6236   | 0.1169 | 0.9844 | 0.9956 | 0.0008 |
|           | ANFIS   | 79062.9332   | 0.1374 | 0.9723 | 0.9923 | 0.0010 |
|           | LSTM    | 163120.2984  | 0.3508 | 0.8498 | 0.9520 | 0.0024 |
|           | Averaging-Ensemble | 109190.0711 | 0.2091 | 0.9327 | 0.9792 | 0.0016 |
|           | SCWOA-Ensemble | 34113.6566  | 0.0815 | 0.9881 | 0.9973 | 0.0005 |
| India     | ARIMA   | 610954.9038  | 1.4814 | -0.5265 | 0.8222 | 0.0098 |
|           | LSSVM   | 547312.1701  | 18.0088 | 121.5077 | 0.0339 | 0.0951 |
|           | BPNN    | 115347.4440  | 0.3327 | 0.9456 | 0.9847 | 0.0019 |
|           | ENN     | 10048.1000   | 0.0230 | 0.9996 | 0.9999 | 0.0002 |
|           | ANFIS   | 53861.3340   | 0.1445 | 0.9881 | 0.9969 | 0.0009 |
|           | LSTM    | 194170.7690  | 0.5561 | 0.8458 | 0.9552 | 0.0032 |
|           | Averaging-Ensemble | 55846.8377 | 0.1368 | 0.9872 | 0.9965 | 0.0009 |
|           | SCWOA-Ensemble | 9025.2129   | 0.0185 | 0.9997 | 0.9999 | 0.0001 |
| Brazil    | ARIMA   | 453183.3331  | 1.7983 | 0.3322 | 0.9013 | 0.0118 |
|           | LSSVM   | 92408.7638   | 0.4386 | 0.9722 | 0.9933 | 0.0024 |
|           | BPNN    | 1113125.7660 | 0.5151 | 0.9584 | 0.9895 | 0.0030 |
|           | ENN     | 145190.2240  | 0.6607 | 0.9315 | 0.9812 | 0.0038 |
|           | ANFIS   | 10280.3420   | 0.4839 | 0.9656 | 0.9920 | 0.0027 |
|           | LSTM    | 51934.1284   | 0.2397 | 0.9912 | 0.9979 | 0.0014 |
|           | Averaging-Ensemble | 41681.8038 | 0.1889 | 0.9944 | 0.9986 | 0.0011 |
|           | SCWOA-Ensemble | 29145.0877  | 0.1141 | 0.9972 | 0.9993 | 0.0008 |
outbreak prediction as a case study. The accuracy of the prediction mainly depends on the quality of the data and requires sufficient historical data. The experimental data were collected from the “COVID-19 data repository”, which updates the global daily epidemic data by the Center of Systems Science and Engineering (CSSE) of Johns Hopkins University at Github (Dong, Du, & Gardner, 2020). We select three countries significantly affected by COVID-19: the USA, India, and Brazil. To validate the model’s performance, this paper collects daily data of total cases, new cases, total deaths, and new death cases in the three countries from January 1, 2021 to August 1, 2021. The obtained data are divided into two parts, with 80 % as the training set and the remaining 20 % as the test set.

### 4.2. Evaluation criteria

This paper considers five evaluation criteria to effectively evaluate the model’s performance, as shown in Table 1. Specifically, the root-mean-squared error (RMSE), MAPE, $R^2$, IA, and TIC are chosen as error criteria to reflect the prediction performance of the forecasting models.

### Table 6

The comparative forecasting error of different models for COVID-19 new cases.

| Countries | Methods       | RMSE    | MAPE  | $R^2$ | IA    | TIC   |
|-----------|---------------|---------|-------|-------|-------|-------|
| USA       | ARIMA         | 39625.9112 | 53.8695 | -0.1961 | 0.4116 | 0.6022 |
|           | LSSVM         | 50495.5089 | 59.1437 | -0.9422 | 0.5093 | 0.4420 |
|           | BPNN          | 40631.3412 | 262.4478 | -0.2575 | 0.5226 | 0.4341 |
|           | ENN           | 36895.8196 | 75.8460 | -0.0369 | 0.5564 | 0.3983 |
|           | ANFIS         | 57308.4210 | 74.6520 | -1.3017 | 0.5437 | 0.4893 |
|           | LSTM          | 37168.2175 | 85.4976 | 0.2313 | 0.6985 | 0.3263 |
|           | Averaging-Ensemble | 33129.8781 | 77.3119 | 0.1639 | 0.6768 | 0.3576 |
|           | SCWOA-Ensemble | 10882.6212 | 30.3627 | 0.8450 | 0.9607 | 0.1250 |

| India     | ARIMA         | 7915.9498 | 11.7499 | -0.0429 | 0.4069 | 0.0891 |
|           | LSSVM         | 13613.2906 | 28.6263 | 0.9422 | 0.5093 | 0.4420 |
|           | BPNN          | 13501.7673 | 30.1301 | 0.7942 | 0.3572 | 0.1633 |
|           | ENN           | 13049.4068 | 23.5619 | 0.6759 | 0.2745 | 0.1520 |
|           | ANFIS         | 12346.3386 | 17.8443 | 0.5564 | 0.3983 | 0.3263 |
|           | Averaging-Ensemble | 12437.0265 | 21.2914 | 0.5223 | 0.2854 | 0.1482 |
|           | SCWOA-Ensemble | 10256.7920 | 16.1490 | 0.7509 | 0.3714 | 0.1193 |

| Brazil    | ARIMA         | 30803.8363 | 73.1414 | 0.8711 | 0.6110 | 0.2389 |
|           | LSSVM         | 21460.9495 | 29.8514 | 0.8239 | 0.2878 | 0.1571 |
|           | BPNN          | 20824.7056 | 28.5270 | 0.7942 | 0.3572 | 0.1633 |
|           | ENN           | 22265.4856 | 28.9823 | 0.6759 | 0.2745 | 0.1520 |
|           | ANFIS         | 25101.7439 | 40.9626 | 0.5564 | 0.3983 | 0.3263 |
|           | Averaging-Ensemble | 20149.1588 | 27.0733 | 0.1994 | 0.6900 | 0.1798 |
|           | SCWOA-Ensemble | 17425.6981 | 21.2914 | 0.5223 | 0.2854 | 0.1482 |

### Table 7

The comparative forecasting error of different models for COVID-19 total death cases.

| Countries | Methods       | RMSE    | MAPE  | $R^2$ | IA    | TIC   |
|-----------|---------------|---------|-------|-------|-------|-------|
| USA       | ARIMA         | 2585.9083 | 0.3327 | 0.3285 | 0.7658 | 0.0021 |
|           | LSSVM         | 238.3001 | 0.0332 | 0.9943 | 0.9986 | 0.0002 |
|           | BPNN          | 869.4171 | 0.1122 | 0.9241 | 0.9752 | 0.0007 |
|           | ENN           | 860.8173 | 0.1345 | 0.9256 | 0.9818 | 0.0007 |
|           | ANFIS         | 537.0568 | 0.0742 | 0.9710 | 0.9922 | 0.0004 |
|           | LSSVM         | 209.3502 | 0.0268 | 0.9956 | 0.9989 | 0.0002 |
|           | BPNN          | 344.8095 | 0.0457 | 0.9981 | 0.9967 | 0.0003 |
|           | ENN           | 197.5213 | 0.0259 | 0.9960 | 0.9990 | 0.0002 |

| India     | ARIMA         | 18396.8831 | 3.4737 | -2.0467 | 0.7240 | 0.0221 |
|           | LSSVM         | 18826.4698 | 3.9881 | -2.1906 | 0.4678 | 0.0235 |
|           | BPNN          | 3880.1558 | 0.8763 | 0.8645 | 0.9629 | 0.0048 |
|           | ENN           | 3664.1177 | 0.8201 | 0.8791 | 0.9669 | 0.0045 |
|           | ANFIS         | 6497.4214 | 1.4032 | 0.6200 | 0.8848 | 0.0080 |
|           | LSTM          | 2768.9240 | 0.6363 | 0.9310 | 0.9623 | 0.0034 |
|           | Averaging-Ensemble | 1860.6057 | 0.3636 | 0.9688 | 0.9910 | 0.0023 |
|           | SCWOA-Ensemble | 935.3028 | 0.1955 | 0.9921 | 0.9980 | 0.0011 |

| Brazil    | ARIMA         | 14900.6884 | 2.0865 | 0.1096 | 0.8754 | 0.0138 |
|           | LSSVM         | 3069.7632 | 0.5192 | 0.9622 | 0.9913 | 0.0029 |
|           | BPNN          | 2525.0654 | 0.4007 | 0.9744 | 0.9935 | 0.0024 |
|           | ENN           | 879.7529 | 0.1422 | 0.9969 | 0.9992 | 0.0008 |
|           | ANFIS         | 857.7813 | 0.1425 | 0.9970 | 0.9993 | 0.0008 |
|           | LSTM          | 4519.5498 | 0.7265 | 0.9181 | 0.9823 | 0.0042 |
|           | Averaging-Ensemble | 1817.6157 | 0.3041 | 0.9868 | 0.9968 | 0.0017 |
|           | SCWOA-Ensemble | 360.9762 | 0.0569 | 0.9995 | 0.9999 | 0.0003 |
5. Experimental results and analysis

In this section, we have established two experiments: Experiment I provides the evaluation of the SCWOA by benchmark functions; Experiment II investigates the performance of the ensemble model for forecasting the COVID-19 epidemic. Details are presented in the following sections.

5.1. Experiment I: The evaluation of the SCWOA by benchmark functions.

To measure and evaluate the performance of the novel optimization algorithm, the algorithms need to process some well-defined test functions. In this experiment, various optimization benchmark problems are utilized as benchmark algorithms. Set the number of search agents for all algorithms to 30 and the maximum number of iterations to 500. In the experiments, each algorithm runs 50 times independently on all benchmark functions and stops when the maximum number of iterations is reached.

5.1.2. The computational results of the proposed algorithm

To statistically evaluate the proposed hybrid variant in comparison with other algorithms, the average and standard deviation of optimal values are calculated. The results of the proposed SCWOA and PSO, GWO, SCA, and WOA are reported in Table 3. In addition, the convergence curves for all benchmark functions through the proposed SCWOA and the standard PSO, GWO, SCA, and WOA are as shown in Figs. 4–6.

For the unimodal benchmark functions, the experimental results prove that the new proposed algorithm outperforms the other algorithms in all cases tested: The average optimal values of F1–F5 reach 3.19E-161, 1.49E-147, 3.68E-109, 1.05E-27, and 2.62E+01, respectively, and the best results are written in bold font. The convergence performance of the algorithm is shown in Fig. 4, which similarly reveals that the SCWOA converges faster and shows a strong optimization capability.

For the multimodal benchmark functions, the experimental results also prove that the new algorithm has the best performance in all test cases. The average optimal values of F6–F10 reach 1.6752, 0.6252, 0.2370, 0.5798, 0.3400, 0.3375, 1.2109, 0.3202, 0.3850, and 0.5767, 0.3591, 0.3482, 0.7278, 0.3304, 0.3645, 1.3577, 0.3511, and 0.3197, respectively, and the best results are written in bold font. The convergence performance of all algorithms has been compared using Fig. 5.

Fixed-dimensional multimodal problems have multiple local optima, and their number grows exponentially with increasing dimensionality, making them a benchmark for measuring the ability of a technology to explore. According to the results of F11–F15 in Table 3, the new hybrid algorithm finds superior results on these problems, performing much better than PSO, GWO, SCA, and WOA and indicating that SCWOA has a robust exploratory capability. The convergence performance of the algorithm in multimodal problems with fixed dimensions is shown in Fig. 6. It can be seen that SCWOA decreases rapidly and can effectively avoid getting into local minima.

5.2. Experiment II: The performance of the ensemble model for forecasting the COVID-19 epidemic

As individual NNs cannot be applied to various complex epidemic

### Table 8

The comparative forecasting error of different models for COVID-19 new death cases.

| Countries | Methods         | RMSE    | MAPE    | $R^2$    | IA     | TIC     |
|-----------|-----------------|---------|---------|----------|--------|---------|
| USA       | ARIMA           | 167.918 | 57.480  | -0.0067  | 0.3893 | 0.3092  |
|           | LSSVM           | 218.509 | 54.3070 | -0.7119  | 0.3511 | 0.3197  |
|           | BPNN            | 206.867 | 45.2748 | -0.5344  | 0.2856 | 0.3096  |
|           | ENN             | 234.954 | 46.5709 | -0.9793  | 0.3990 | 0.3141  |
|           | ANFIS           | 207.359 | 48.8013 | -0.5417  | 0.4520 | 0.2979  |
|           | LSTM            | 233.273 | 204.470 | -0.9511  | 0.4271 | 0.5115  |
|           | Averaging-Ensemble | 163.5745 | 41.3553 | 0.0406  | 0.4540 | 0.2577  |
|           | SCWOA-Ensemble  | 146.1758 | 41.6379 | 0.2339  | 0.5764 | 0.2418  |
| India     | ARIMA           | 767.0861 | 96.6555 | -0.6908  | 0.3845 | 0.3104  |
|           | LSSVM           | 768.9617 | 49.4638 | -0.6167  | 0.3158 | 0.3439  |
|           | BPNN            | 759.4036 | 59.8135 | -0.5767  | 0.3591 | 0.3482  |
|           | ENN             | 794.9469 | 64.6156 | -0.7278  | 0.3304 | 0.3645  |
|           | ANFIS           | 928.6265 | 75.1196 | -1.3577  | 0.4030 | 0.4028  |
|           | LSTM            | 899.2495 | 100.5203 | -1.2109  | 0.3202 | 0.3850  |
|           | Averaging-Ensemble | 741.4895 | 52.0475 | -0.5798  | 0.3400 | 0.3375  |
|           | SCWOA-Ensemble  | 684.0992 | 142.5934 | -0.2975  | 0.3909 | 0.3657  |
| Brazil    | ARIMA           | 847.1567 | 68.0459 | -1.6752  | 0.6252 | 0.2370  |
|           | LSSVM           | 309.4662 | 18.3241 | 0.6430   | 0.8821 | 0.1073  |
|           | BPNN            | 282.0152 | 17.5701 | 0.7035   | 0.9175 | 0.0967  |
|           | ENN             | 291.6718 | 17.3553 | 0.6829   | 0.9128 | 0.0989  |
|           | ANFIS           | 397.2349 | 20.1853 | 0.4118   | 0.8591 | 0.1334  |
|           | LSTM            | 454.5381 | 22.9903 | 0.2299   | 0.8401 | 0.1471  |
|           | Averaging-Ensemble | 256.7650 | 15.9107 | 0.7543   | 0.9333 | 0.0866  |
|           | SCWOA-Ensemble  | 197.3885 | 14.6789 | 0.8548   | 0.9553 | 0.0704  |
prediction scenarios, it is necessary to consider combining multiple NNs for ensemble prediction to overcome the one-sidedness of the individual model prediction. Therefore, this paper selects four commonly used NNs, BPNN, ENN, ANFIS, and LSTM (see Section 2 for model details), and the prediction performance of the four methods is compared through four epidemic data sets of three countries. However, the limitations of single models in prediction mean that each method can only provide corresponding information from a particular perspective, making it difficult to fully characterize the COVID-19 epidemic trend. Therefore, to eliminate the weaknesses of every individual model, suitable combinations of the four models are needed to constitute the ensemble prediction. The newly proposed SCWOA is applied to optimize the ensemble prediction model in this paper. The SCWOA searches for the optimal combination weights of the four single models. The MAPE error indicator is used as the objective function of SCWOA to obtain the optimal solution of the weight coefficients by iterative search. The optimal weights and MAPE results are obtained as listed in Table 4.

As indicated in Table 4, the four methods of BPNN, ENN, ANFIS, and LSTM have their advantages and disadvantages in predicting different epidemic COVID-19 data in the three countries. Among them, BPNN obtains the smallest MAPE value in the total death cases prediction of the USA and India. ENN obtains the smallest MAPE value in the total cases prediction of the USA and India and Brazil’s total death cases prediction. ANFIS obtained the best results in the new cases prediction of the USA and India. LSTM obtained the minimum MAPE in the total death cases prediction of the USA and India and the total cases prediction of Brazil. The values marked in bold are used to indicate the best values of the model, and the optimal prediction model is selected accordingly. In summation, the situation of the COVID-19 epidemic is complex and variable, each NN model has its advantages.
and applicability, and it is difficult to determine the only NN model suitable for all epidemic situations.

Further, ensemble weights obtained from the SCWOA search were weighted into each of the four models to obtain the ensemble prediction results, and the effectiveness of the model was verified by four different epidemic data sets in three countries. To fully assess the predictive performance of the proposed model, BPNN, ENN, ANFIS, LSTM, ARIMA, least squares support vector machine (LSSVM), and Averaging-Ensemble are selected as benchmark models to compare with the proposed model. In addition, five evaluation criteria, such as RMSE, MAPE, $R^2$, IA, and TIC, are used to reflect the prediction performance of the models, and the results are listed in Tables 5–8. Figs. 7–9 shows the predicted values and observed values between the proposed model and other models. The further discussion of the experimental results is as follows.

Tables 5–8 indicate that the proposed SCWOA-based ensemble model (SCWOA-Ensemble) shows more robust predictive performance compared with ARIMA, LSSVM, BPNN, ENN, ANFIS, and LSTM, and Averaging-Ensemble has the best performance among all COVID-19 data sets of three countries. For the total cases, compared with ARIMA, LSSVM, BPNN, ENN, ANFIS, and LSTM, SCWOA-Ensemble leads to 93.39 %, 79.55 %, 82.19 %, 41.74 %, 68.73 %, and 72.77 % average reductions in RMSE, 92.48 %, 79.86 %, 82.58 %, 44.14 %, 68.09 %, and 75.28 % average reductions in MAPE, 93.36 %, 79.50 %, 82.15 %, 41.69 %, 68.67 %, and 72.75 % average reductions in TIC of three countries. For the new cases, compared with ARIMA, LSSVM, BPNN, ENN, ANFIS, and LSTM, SCWOA-Ensemble leads to 52.33 %, 40.95 %, 38.17 %, 38.21 %, 43.18 %, and 35.90 % average reductions in RMSE, 58.77 %, 38.19 %, 50.91 %, 37.66 %, 39.88 %, and 39.37 % average reductions in MAPE, 49.19 %, 35.60 %, 35.90 %, 34.72 %, 36.81 %, and 32.03 % average reductions in TIC of three countries. For the total death cases, compared with ARIMA, LSSVM, BPNN, ENN, ANFIS, and LSTM, SCWOA-Ensemble leads to 94.95 %, 66.80 %, 79.63 %, 70.17 %, 68.91 %, and 72.77 % average reductions in RMSE, 94.95 %, 66.80 %, 79.63 %, 70.17 %, 68.91 %, and 72.77 % average reductions in MAPE, 94.95 %, 66.80 %, 79.63 %, 70.17 %, 68.91 %, and 72.77 % average reductions in TIC of three countries.
%, and 54.63 % average reductions in RMSE, 94.62 %, 68.71 %, 80.13 %, 72.29 %, 70.40 %, and 54.86 % average reductions in MAPE, 94.91 %, 66.82 %, 79.65 %, 70.13 %, 68.95 %, and 54.59 % average reductions in TIC of three countries. For the new death cases, compared with ARIMA, LSSVM, BPNN, ENN, ANFIS, and LSTM, SCWOA-Ensemble leads to 33.49 %, 26.79 %, 23.09 %, 28.02 %, 35.38 %, and 39.28 % average reductions in RMSE, 44.39 %, 0.31 %, 2.28 %, 5.70 %, 16.09 %, and 48.59 % average reductions in MAPE, 24.33 %, 17.07 %, 14.30 %, 16.79 %, 26.06 %, and 36.28 % average reductions in TIC of three countries. In addition, compared with Averaging-Ensemble, SCWOA-Ensemble can obtain better prediction results than the simple averaging ensemble approach, indicating that SCWOA can search to obtain the optimal ensemble model weight coefficients, thus effectively combining the advantages of the four single models.

To visually compare each model’s time series prediction ability, the line graphs of the prediction results of different models compared with the actual values are shown in Figs. 7–9. From Figs. 7–9, it can be seen that the four single models have different prediction performances in total cases, new cases, total death cases, and new death cases. In contrast, by combining the advantages of each single model, the ensemble model can simulate the overall trend of epidemic changes well

![Fig. 9. Comparison of predicted and actual values of different models in Brazil.](image-url)
and utilize the sudden change information of epidemic prediction and obtain better prediction performance in all kinds of situations. The pooled model can obtain better prediction performance in all scenarios.

**Remark.** Each single NN model has its advantages in predicting different epidemic situations, and no model can obtain the best results in various situations. The performance of each single NNs in different epidemic prediction scenarios varies, and no single model can obtain optimal results in all situations. Moreover, the proposed ensemble model based on the SCWOA has strong prediction ability and can effectively address the complexity and nonlinearity of total cases, new cases, total death cases, new death cases. The proposed optimization method SCWOA plays an essential role in improving the prediction accuracy of the ensemble model.

### 5.3. Discussion

From the analysis, it can be found that each NN model has its advantages in different epidemic prediction scenarios. In addition, the ensemble prediction model optimized by SCWOA achieves the best prediction performance in all prediction scenarios, which shows that the swarm intelligence algorithm SCWOA can effectively obtain the optimal weight coefficient so that the ensemble model can synthesize the advantages of each single model to achieve better prediction performance. Specifically, the prediction methods with good performance should make a more significant contribution to the ensemble model, whereas the methods with poor performance should contribute less. For a given set of data sets, if a prediction method has a better prediction performance, it makes a more significant contribution to the ensemble model and has a greater weight coefficient. Thus, in various complicated forecast scenarios, through the SCWOA to intelligently search for the optimal solution and assign the optimal weight to the set model, we can successfully combine the advantages of every single model and obtain more accurate prediction results.

We have also analyzed the computational complexity and burden of our method. The computation experience of the proposed method is the two stages of parameter estimation. The ensemble model contains a large number of parameters and the ideal parameters form a complex hypersurface, so model optimization via SCWOA results in over-computing loads. The optimal method fixes some of them in the pre-trained model while the single model is constructed in the first stages, and then the remaining parameters, such as the weights of single models, are estimated by SCWOA. As the proposed method integrates each separately trained model, its complexity is about the sum of the complexity of five single models, which means that our ensemble method will not cause more computational burden than other single models. The code is run on a machine with an Intel® Core™i7-11800H, 2.30 GHz CPU and 32 GB of RAM. The model time cost is listed in Table 9. The results of Table 9 indicate that the SCWOA optimization process does not take the longest running time compared with the training of every single model, so it will not cause a computational burden. Although the time cost of the final ensemble model is about 1 min, the ensemble model significantly improves the accuracy of prediction, and this time cost is acceptable in public health practice.

### 6. Conclusions and Future Prospects

Reliable and precise COVID-19 epidemic forecasting is vital for outbreak prevention and control. It is increasingly recognized that the combination of models has advantages over the choice of an individual model. How to combine existing forecasting technologies to achieve perfect prediction results is a widely discussed topic. This paper has proposed a novel ensemble forecasting paradigm based on multiple NNs and a new heuristic intelligence algorithm SCWOA. First, four individual NNs were selected to predict the COVID-19 outbreak. To eliminate the weaknesses of each model, an ensemble forecasting method, which integrates the characteristics of each forecasting method, has been developed. The proposed SCWOA algorithm was applied to optimize the ensemble prediction model. The SCWOA algorithm was used to search for the best matching weights of the ensemble model. The performance of the SCWOA has been evaluated by 15 classical benchmark functions, and the results show that the SCWOA performs much better than PSO, GWO, SCA, and WOA, which indicates that SCWOA has a robust exploratory capability and can effectively avoid getting into local minima. The daily COVID-19 series collected from three of the most-affected countries were taken as the test cases to conduct the empirical study to verify the proposed ensemble model. The comparison results obtained in this study demonstrate that different NN models have different prediction performances in various complex epidemic prediction scenarios. The SCWOA-based ensemble prediction model significantly outperforms all other comparable models with its high prediction accuracy, which implies that the proposed model can effectively integrate the advantages of multiple NN models to achieve stable and accurate prediction results.

For future work, there are several potential research paths: (a) the selection of the base predictor in the current ensemble model tends to be more subjective; optimizing the model selection strategy is a promising direction in the next step; (b) building a big-data-driven prediction model, more epidemic-related variables, such as epidemic prevention policy measures, search indices, and population mobility, need to be incorporated and how these variables affect changes in epidemic trends should be analyzed.

### Declaration of Competing Interest

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

### Data availability

Data will be made available on request.

### Acknowledgment

This work was supported by the National Natural Science Foundation project of China (72004086).

### References

Alenzi, M. N., Al-Anzi, F. S., & Alabdirfrazzah, H. (2021). Building a sensible SIR estimation model for COVID-19 outbreak in Kuwait. Alexandria Engineering Journal, 60(3), 3161–3175. https://doi.org/10.1016/j.aeje.2021.01.025

Annas, S., Ishar Pratama, M., Rifandi, M., Sanusi, W., & Side, S. (2020). Stability analysis and numerical simulation of SEIR model for pandemic COVID-19 spread in Indonesia. Chaos, Solitons & Fractals, 139, Article 110057. https://doi.org/10.1016/j.chaos.2020.110057

Behnood, A., Mohammad Gofarshani, E., & Hosseini, S. M. (2020). Determinants of the infection rate of the COVID-19 in the U.S. using ANFIS and virus optimization algorithm (VOA). Chaos, Solitons & Fractals, 139, Article 110051. https://doi.org/10.1016/j.chaos.2020.110051

Ceylan, Z. (2020). Estimation of COVID-19 prevalence in Italy, Spain, and France. Science of the Total Environment, 729. https://doi.org/10.1016/j.scitotenv.2020.138817

Chakraborty, I., & Maity, P. (2020). COVID-19 outbreak: Migration, effects on society, global environment and prevention. Science of the Total Environment, 728. https://doi.org/10.1016/j.scitotenv.2020.138882

Chhimula, V. K. R., & Zhang, L. (2020). Time series forecasting of COVID-19 transmission in Canada using LSTM networks. Chaos, Solitons & Fractals, 135, Article 109864. https://doi.org/10.1016/j.chaos.2020.109864

Clemen, R. T. (1989). Combining forecasts: A review and annotated bibliography. International Journal of Forecasting, 5(4), 559–583. https://doi.org/10.1016/0169-2070(89)90012-5

Cone, S. F., Hibon, M., & Nikolopoulos, K. (2011). Advances in forecasting with neural networks? Empirical evidence from the NN3 competition on time series prediction. International Journal of Forecasting, 27(3), 635–660. https://doi.org/10.1016/j.ijforecast.2011.04.001

Cunillera, P., Rojas-Díaz, O., de España, P. M., & Verdugo-Hernández, P. (2022). Forecasting COVID-19 Chile’ second outbreak by a generalized SIR model with constant time delays and a fitted positivity rate. Mathematics and Computers in Simulation, 193, 1–18. https://doi.org/10.1016/j.mcs.2021.09.016
