Iran Journal of Computer Science (2022) 5:99–107
https://doi.org/10.1007/s42044-021-00092-4

REVIEW ARTICLE

# Analysis and prediction of worldwide novel coronavirus (COVID-19) infections, using neural network-based techniques

**Sachin Kamley**[1] · **R. S. Thakur**[2]

Received: 30 May 2020 / Accepted: 15 October 2021 / Published online: 16 November 2021


**Abstract**

The novel coronavirus (COVID-19) outbreak has recently become a major public health concern around the world. It is commonly known that some of the world's most powerful countries, such as Iran and the United States, are suffering more than others from the effects of this horrific pandemic. It has spread throughout communities and has endangered the health of many people. Governments must take the necessary steps to stop the virus from spreading globally. The three most widely used backpropagation neural network (BPNN) techniques, i.e., Levenberg–Marquardt, Bayesian regularization (BR), and scaled conjugate gradient (SCG), are used to either predict the future or evaluate the current status of COVID-19 in this research. This study uses a real-time COVID-19 dataset from the Worldometer website, which contains 204 samples from 30 January to 15 April 2020. The 12 most important parameters are selected for study purposes, including country, total cases (TC), new cases (NC), total deaths (TD), new deaths (ND), total recoveries (TREV), active cases (AC), serious cases (SC), total tests (TT), death rate (DR), recovery rate (RR), and case rate (CR). Finally, countries are classified into three risk levels, i.e., high, medium, and low, based on the above parameters. In addition, some new countries are discovered at these levels.

**Keywords** Prediction · Coronavirus · COVID-19 · Neural network · MATLAB

## 1 Introduction

The novel coronavirus that causes coronavirus disease 2019 (COVID-19) is one of the most infectious viruses of the family Coronaviridae, which is widely distributed among humans and other mammals. The first reports of COVID-19 occurred in Wuhan, China, on 31 December 2019, causing 259 deaths. Outside China, Thailand reported their first case on 13 January 2020 [1]. At present, more than 50 countries have been infected by this outbreak [2].

The World Health Organization (WHO) declared the COVID-19 outbreak a public health emergency of international concern (PHEIC) on 30 January 2020 [2, 3]. Over 76,000 cases of COVID-19 had been confirmed globally as of 20 February 2020 [4, 5].

As of 21 March 2020, 186 countries worldwide had reported COVID-19 infections, with more than 2,80,000 confirmed cases and 11,842 deaths [6]. Countries such as the United States, China, Spain, Italy, Germany, and the United Kingdom have suffered the most as a result of neglecting to take adequate measures against the outbreak and failure to prevent the spread of this virus in humans. The primary symptoms of the infection include cough, fever, breathing difficulty, and pneumonia in both lungs, leading to death in many. Most importantly, individuals with COVID-19 require a quarantine period of at least 2 weeks as a precaution. If the patient fails to quarantine during this period, the virus can spread to the community through close contact and respiratory droplets [7]. Figure 1 shows the worldwide growth in COVID-19 cases.

Figure 1 clearly shows that since December 2019, the virus has spread rapidly to every contingent, and the number of cases is continuing to rise. COVID-19 cases globally surpassed 1 million on April 2nd, and 2 weeks later, they surpassed 2 million. On April 24th, the number of cases had

✉ Sachin Kamley
skamley@gmail.com

R. S. Thakur
ramthakur2000@yahoo.com

[1] Department of Computer Applications, S.A.T.I., Vidisha, MP, India

[2] Department of Computer Applications, M.A.N.I.T., Bhopal, MP, India

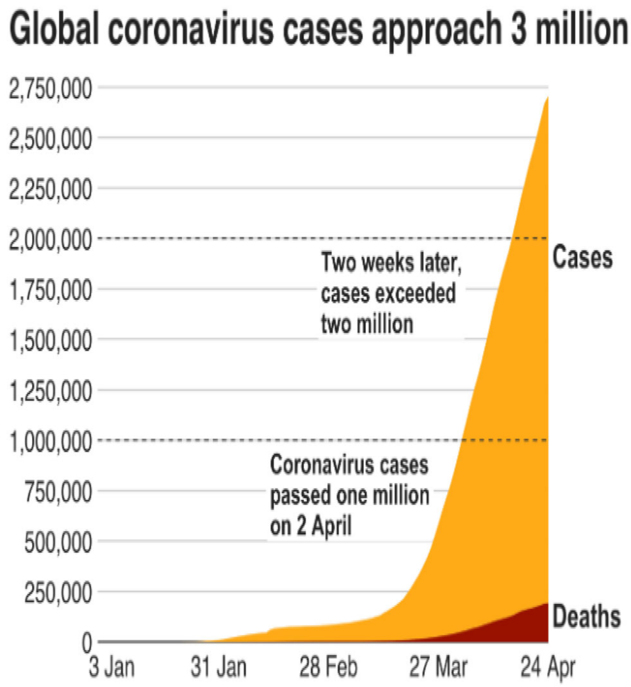


**Fig. 1** Worldwide growth in COVID-19 cases [6, 7]

surpassed 3 million, with 194,456 deaths. As a result, there is a dangerous situation all across the planet.

Artificial intelligence (AI) approaches can imitate human intelligence. They might be able to track patients' geographic locations, diagnose them, and speed up the process of discovering a cure for COVID-19 [5, 8]. AI approaches have been widely used in a variety of fields over the years, including water quality prediction, stock market prediction, weather prediction, and health care diagnosis.

One successful AI technology is the neural network (NN), which can translate input data patterns into output data patterns [8]. The three most popular NN approaches, Levenberg–Marquardt (LM), Bayesian regularization (BR), and scaled conjugate gradient (SCG), are used in this study to anticipate and analyze the worldwide COVID-19 dataset from 30 January to 16 April 2020. The LM approach is ideally suited for small and medium-sized problems and has a consistent convergence rate. The method's major flaw is that many weight and squared errors might appear.

The BR technique adds additional term regularization, enhances network performance in terms of minimum mean squared error (MSE), and decreases the enormous penalty of weight by selecting the ideal combination of input patterns.

SCG is the third most powerful approach in the NN family. Using the conjugate direction, the approach can calculate step size in each iteration and yield fastest convergence. The major advantage of the method is that it has a lower MSE than the other approaches.

We divided the COVID-19 dataset into three risk levels, i.e., low, medium, and high, and as a result, some additional countries were discovered in each of these categories. The anatomy of a neural network is depicted in Fig. 2, which shows the mapping of input data patterns to output data patterns.

Figure 2 depicts the mapping of input data patterns to output data patterns, with input neurons transferring information to hidden layers and hidden layers propagating information to the output layer.

Section 2 provides a brief a literature review describing some significant research studies. Section 3 describes a data preprocessing process. Section 4 describes the proposed methodologies in detail. Sections 5 and 6 describe the experimental results and discussion, respectively. Lastly, the conclusions and future scope of the study are described in Sect. 7.

## 2 Literature review

This section briefly highlights some significant research work in the health care field.

Venkatalakshmi and Shivshankar [10] presented a comparative study of the decision tree (DT) and naive Bayes (NB) algorithms. The dataset contained 294 samples, including 13 attributes. The experimental results showed that NB outperformed the DT algorithm in terms of accuracy.

Bellachia [11] suggested various data mining techniques including the backpropagation neural network (BPNN), NB, and the C4.5 DT for predicting breast cancer survivability. The real dataset was collected from the Surveillance, Epidemiology, and End Results (SEER) Program data source. Their experimental results showed that C4.5 DT provided more accurate performance than the other techniques.

Sandhu et al. [12] designed a prediction system based on Bayesian belief networks (BBN) for the Middle East respiratory syndrome coronavirus (MERS-CoV) patient classification. Google Maps were used to track the geographic position of infected patients using their mobile phones. Based on the experimental results, the authors concluded that the proposed system would help individuals avoid infected areas, and the model demonstrated 83.1% accuracy, which is better than other existing approaches.

Turaiki et al. [13] built a prediction model for MERS-CoV infection based on data mining techniques. The dataset consisted of 1082 records from 2013 to 2015. The NB and J48 DT algorithms were used to build the prediction model. Three accuracy measures, i.e., precision, recall, and accuracy, were used to evaluate the performance of the model. The proposed model recorded accuracy of between 53.6 and 71.58%.

Ferrira et al. [14] reported a comparative study of various data mining techniques including J48, classification and

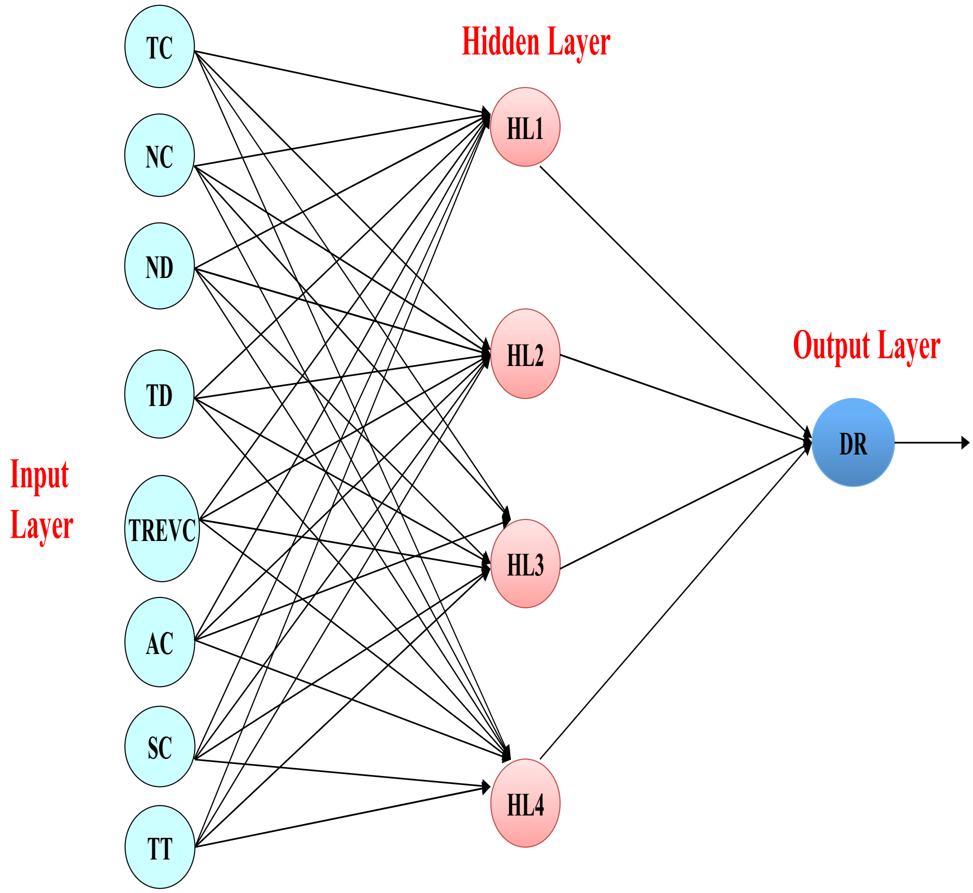


**Fig. 2** Anatomy of NN [9]

regression tree (CART), NB, multilayer perceptron (MLP), simple logistic regression (SLR), and sequential minimal optimization (SMO) for diagnosing neonatal jaundice. A dataset of 227 healthy newborns, including 70 variables, was collected. The experimental results showed that NB, MLP, and SLR outperformed other techniques.

Pal et al. [15] presented the long short-term memory (LSTM) and Bayesian-optimization-based neural network methods to predict the performance of a proposed model. In this study, trend and weather data for 170 countries were collected and analyzed. The authors concluded that the proposed model would enable earlier preventive actions to be taken.

Al-Najjar and AL-Rouson [16] presented a classification model to predict COVID-19 status in South Korea. The Korea Centers for Disease Control and Prevention (KCDC) used an actual dataset of 7869 COVID-19 patients from 20 January to 9 March 2020 for this study. The feature selection method was used to select only relevant attributes. As a result, out of 15 attributes, only seven attributes, i.e., sex, birth year, country, region, group, infection reason, and confirmed data, were considered for study purposes. Their experimental results showed that the proposed model successfully predicted cases of death and enabled early diagnosis.

Petropoulos and Makridakis [17] performed statistical time series analysis to forecast COVID-19 occurrence. They obtained a dataset from the Center for Systems Science and Engineering (CSSE), Johns Hopkins University, Maryland, from 22 January 2020 to 11 March 2020. The parameters considered for study purposes were confirmed cases, deaths, and recoveries. They included both lab-confirmed and clinically diagnosed cases for better prediction results. They found a significant increase in the trend of COVID-19 cases globally coupled with an increase in associated uncertainty.

Finally, this research presents a risk classification of COVID-19 countries based on various parameters and BPNN techniques including LM, BR, and SCG for COVID-19 dataset prediction and analysis.

**Table 1** Sample of the COVID-19 dataset in abbreviated form [18]

| S. No | Attributes | Abbreviation | Min–Max value |
|---|---|---|---|
| 1 | Country | Country | – |
| 2 | Total cases | TC | [1, 678,210] |
| 3 | New cases | NC | [0, 3258] |
| 4 | Total deaths | TD | [0, 34,641] |
| 5 | New deaths | ND | [0, 1290] |
| 6 | Total recoveries | TREV | [0, 81,800] |
| 7 | Active cases | AC | [0,585,725] |
| 8 | Serious cases | SC | [0, 13,369] |
| 9 | Total tests | TT | [10, 3,411,394] |
| 10 | Death rate | DR | [1.5, 19.47] |
| 11 | Case rate | CR | [1.06, 22.22] |
| 12 | Recovery rate | RR | [1.21, 34.54] |

## 3 Data preprocessing

The Worldometer website was used to acquire global data on COVID-19 from 30 January 2020 to 16 April 2020 for this investigation [18]. The dataset consisted of 204 samples including nine attributes. However, the dataset contained redundant data. Therefore, data prepossessing procedures were used to eliminate corrupt or incorrect information or tuples [19]. We included the three most essential attributes, i.e., death rate (DR), recovery rate (RR), and case rate (CR), to help forecast more accurate results. The formulas for calculating DR, RR, and CR are as follows [18].

$$\text{Death Rate (DR)} = \frac{\text{Total No. of Cases (TC)}}{\text{Total No. of Deaths (TD)}} \times 100 \tag{1}$$

$$\text{Recovery Rate (RR)} = \frac{\text{Total No. of Cases (TC)}}{\text{Total No. of Recovered Cases (TREV)}} \times 100 \tag{2}$$

$$\text{Case Rate (CR)} = \frac{\text{Total No. of Deaths (TD)}}{\text{Total No. of Cases (TC)}} \times 100 \tag{3}$$

Due to the large number of attribute values, we are only showing a sample of attributes. Table 1 describes the sample of the COVID-19 dataset in abbreviated form.

## 4 Proposed methodology: backpropagation neural network (BPNN)

BPNN is a well-known and widely used approach for training multilayer perceptron (MLP) networks. It belongs to the category of supervised training algorithms. However, in MLP networks, errors are generated at an output layer and always propagate in the "backward" direction at hidden layers, where the activation function conducts a calculation, giving rise to the term "backpropagation" [19, 20]. To train the feed-forward networks in this study, the three most prominent backpropagation techniques, i.e., LM, BR, and SCG, are utilized, which are each presented below.

### 4.1 Levenberg–Marquardt (LM)

The LM method is used to provide a numerical solution to the problem of minimizing a nonlinear function. The method can provide fast and stable convergence and is suited for training small and medium-sized problems [20, 21].

The LM method utilizes the following formula for the weight updating process, which is shown by Eq. (4) [21, 22].

$$\Delta W = \left(\mathrm{JM}^{\mathrm{T}}(W)\,J(W) + \mu I\right)^{-1} \mathrm{JM}^{\mathrm{T}}(W)\,e(W) \tag{4}$$

where JM denotes the Jacobian matrix, $W$ denotes the weight, and $\mu$ is a regularization parameter automatically adjusted by the algorithm.

When the error function increases by step results $\mu$, then the LM method is adjusted by a multiplication factor $\beta$, and when the error function decreases by step results, it is adjusted by a division factor. The step-by-step procedure of the LM algorithm is shown by Algorithm 1 [19, 22].

**Algorithm 1.** Levenberg Marquardt (LM).

1) Start
2) Initialize the network weights (W) and biases (b) and other training parameters.
3) Start network training and calculate the error value for all input.
4) Calculate Jacobian Matrix JM (W).
5) Calculate $\Delta W$ by using eq. no. (4).
6) Calculate new error value by using weight updating formula: $\overline{W}=W+\Delta W$
7) If Error (Er) < $Er_{Min}$
   7.1) Stop the network training and calculate output performance.
   Otherwise
   7.2) Check if Epoch ≥ $Epoch_{Max}$ if yes stop network training.
   Otherwise
   Go to step 5 (update weights)
8) Increment epoch by 1:
   Epoch=Epoch+1
9) Go to step 2 and continue the same procedure.
10) Stop

### 4.2 Bayesian regularization (BR)

The BR approach is another prominent backpropagation method for updating weights and biases based on the LM optimum. To increase network performance, it minimizes a squared error and then assesses whether the generalized network performance can be improved by a correct combination of input. The BR algorithm offers a substantial benefit over the LM approach in that it adds term regularization to penalize large weight values [21–23]. The BR method's Algorithm 2 is shown below [23].

**Algorithm 2.** Bayesian Regularization (BR).

1) Start with the LM optimization method and find the weights ($W_i$) and biases (b) values to minimize the objective function. The eq. (5) denotes the objective function.

$$F(Z)=xW_i+yE_r \quad (5)$$

Where x and y are determined by Bays theorem and $E_r$ denotes the error term.

1. a) Calculate the Jacobian Matrix (JM) i.e. containing the first derivatives of network errors w.r.t. to all network parameters like weights and biases, Levenberg damping factor ($\mu$) and parameter update vector ($\partial$).

Where ($\partial$) denotes how much magnitude of weights need to be changed to achieve better prediction accuracy.

1. b) Calculate error gradient i.e. $EG = JM^T.E_r$

1.c) Approximate the Hessian matrix i.e. $HM=JM^TJ$

1. d) Solve $(JM^TJ+\mu I)\partial = JM^TE_r$ to calculate objective function.

1. e) Update the weights by using equation no. (4).

2) To minimize the error, calculate the effective no. of parameters by using Newton's method to Hessian Matrix (generated from LM algorithm).

$$\rho=n-2x^{MAP}\text{train}(HM)^{-1}$$

3) Calculate updated values of x and y.

$$x^{new}=\frac{\rho}{2E_r(W^{new})}$$

$$y^{new}=\frac{\rho}{2W_i(W^{new})}$$

4) Repeat steps 2-4 until convergence.
5) Stop.

### 4.3 Scaled conjugate gradient (SCG)

The major drawback of the LM algorithm is that it does not produce a faster convergence. On the other hand, conjugate gradient algorithms can provide faster convergence against the steepest descent direction [21, 23]. The major advantage of the SCG algorithm is that it minimizes error through conjugate gradient directions and it is used to determine the step size. Algorithm 3 describes the step-by-step procedure of the SCG method [24].

**Algorithm 3.** Scaled Conjugate Gradient (SCG).

1) Start with the initial vector $W_1$ and set $q1=s1= -E'(W_1)$, K=1.

2) Find second-order derivatives

$$P_k=-E''(W_k)q_k$$

$$\partial_K=q^T{}_kP_K$$

3) Calculate step size

$$\mu_k=q^T{}_kS_K$$

$$\alpha_K=\frac{\mu_k}{\partial_K}$$

4) Weight vector updated by

$$W_{K+1}=W_k+\alpha_Kq_k$$

$$S_{K+1}=-E^t(W_{K+1})$$

5) If mod N=0 then restarts again

$$q_{k+1}=S_{K+1}$$

Otherwise

Go and search new conjugate direction

$$\beta_k=\frac{|S_{K+1}|^2-S_{K+1}S_K}{\mu_k}$$

$$q_{k+1}=S_{K+1}+\beta_kq_k$$

6) If Steepest descent direction ($S_K \neq 0$) then increment the value of k by 1(k++) and go to step 2.

7) Stop the algorithm and return the weight vector $W_{K+1}$ as the desired minimum.

## 5 Experimental results

In this study, a MLP model is adopted to predict the DR, CR, and RR in different countries. An 11*10*1 architecture is used, and the training functions trainlm, trainscg, and trainbr are used to train the MLP network. Here, 11*10*1 denotes 11 input variables, i.e., total cases (TC), new cases (NC), new deaths (ND), total deaths (TD), total recoveries (TREV), actual cases (AC), serious cases (SC), total tests (TT), CR, and RR, and 1 denotes the output parameter. Here, the DR is considered as an output parameter. Similarly, CR and RR are also considered output parameters. The number of hidden neurons at the hidden layer is denoted by 10. The real dataset from the Worldometer website is considered for study purposes. The dataset consists of 204 samples, and data are divided into three categories: training (70%), validation (15%), and testing (15%). The training, validation, and testing data comprise 142, 31, and 31 samples, respectively. For experimental generation purposes, the MATLAB R2017a tool is used. Table 2 shows the training status of backpropagation (BP) techniques.

Figure 3 shows the performance graph of the LM algorithm against MSE and epochs.

Figure 3 clearly shows that the best validation performance is achieved at epoch 19, i.e., 1148.8717.

Table 3 clearly indicates that the performance of the SCG method is better than the best validation performance of the two other approaches. Thus, various performance indicators

**Table 2** Training status of backpropagation (BP) techniques

| S. no | BP techniques | Parameters | | | | | |
|---|---|---|---|---|---|---|---|
| | | Gradient | Mutation | Validation checks | Epoch no. | No. parameters | Sum squared parameters |
| 1 | Levenberg–Marquardt (LM) | 93.6565 | 1 | 6 | 25 | – | – |
| 2 | Bayesian regularization (BR) | 12.2841 | $5 \times 10^{10}$ | 0 | 402 | 19.4972 | 7.1896 |
| 3 | Scaled conjugate gradient (SCG) | 43.7373 | - | 6 | 26 | – | – |

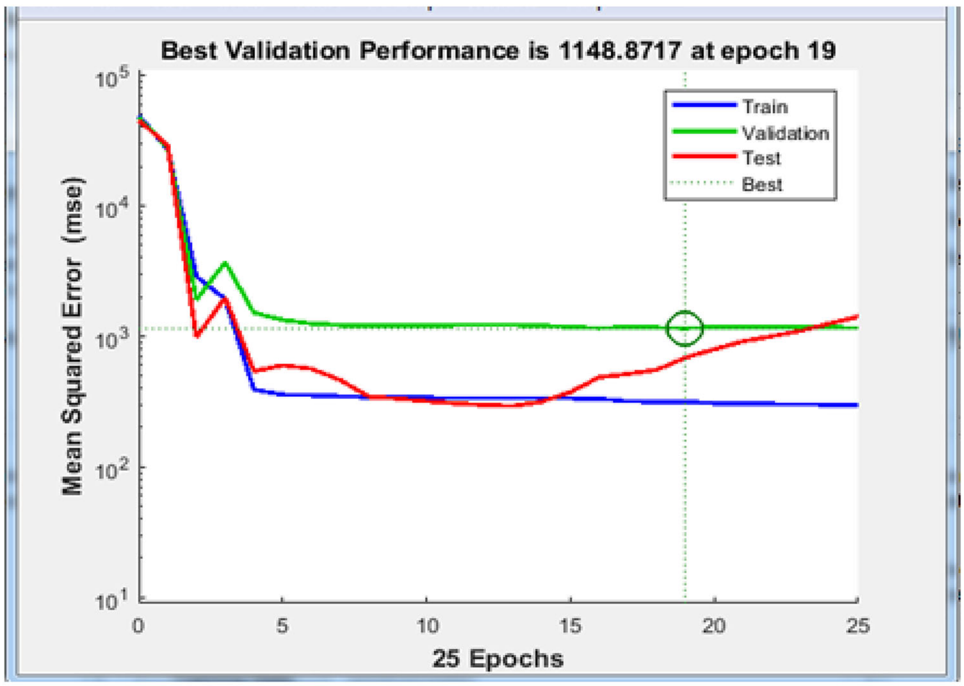


**Fig. 3** Performance graph of LM BP algorithm mean squared error (MSE) vs. epochs

**Table 3** Performance statistics for BP techniques against MSE and epochs

| S. no | BP techniques | Best validation/training performance | Epochs |
|---|---|---|---|
| 1 | Levenberg–Marquardt (LM) | 1148.8717 | 19 |
| 2 | Bayesian regularization (BR) | 301.9782 | 251 |
| 3 | Scaled conjugate gradient (SCG) | 325.8208 | 20 |

are calculated to predict the performance of the proposed model, which is shown in Table 4.

Table 5 denotes the performance of the proposed model.

Table 5 clearly shows that the SCG method achieves the lowest error among the techniques. Table 6 shows the country-wise performance comparison between the actual value and predicted value.

**Table 4** Performance indicators of proposed model

| S. no | Performance indicators | Description | Formula |
|---|---|---|---|
| 1 | Mean absolute error (MAE) | Measuring the performance w.r.t. mean of squared errors | $\sum \frac{(AD-PD)^2}{N}$ |
| 2 | Root mean squared error (RMSE) | Square root of mean squared error | $\sqrt{\sum \frac{(AD-PD)^2}{N}}$ |
| 3 | Mean absolute percentage error (MAPE) | Simple average of absolute percentage errors | $\frac{\sum \frac{\lvert AD-PD \rvert}{AD} \times 100}{N}$ |
| 4 | Root relative squared error (RRSE) | The ratio of the square root of the difference between the actual value and the predicted value | $\sqrt{\sum \frac{(AD-\overline{PD})^2}{(\overline{AD}-AD)^2}}$ |

*AD* actual data, *PD* predicted data, *N* no. of sample

**Table 5** Performance of proposed model

| Techniques | Error measures | | | |
|---|---|---|---|---|
| | MAE | RMSE | MAPE | RRSE |
| Levenberg-Marquardt (LM) | 433.37 | 20.81 | 141.89 | 0.026 |
| Bayesian regularization (BR) | 270.84 | 16.19 | 2.46 | 0.035 |
| Scaled conjugate gradient (SCG) | 239.80 | 15.48 | 0.209 | 0.026 |

Based on the results in Table 6, we have identified high-risk, medium-risk, and low-risk countries. A lower DR, higher RR, and lower CR are considered superior metrics for predicting a country's risk level. For the objectives of this study, we investigated 204 countries; therefore, all of the data for each country cannot be displayed here. In Table 7, we have only included a few well-known counties. Table 7 gives a sample of country risk classifications. Green indicates low risk, red indicates high risk, and black indicates medium risk.

Figure 4 shows a classification of countries based on the risk level.

As shown in Fig. 4, countries are categorized from higher to lower risk based on the DR, RR, and CR. Black indicates that countries are well known and that we have heard about the COVID-19 pandemic in daily newspapers and on television. However, we also identified some new countries from higher to lower risk levels, including Belgium, Mexico, Sweden, Indonesia, Zimbabwe, the Netherlands, Argentina, Poland, Romania, South Africa, Afghanistan,

**Table 6** Country-wise performance comparison between actual value and predicted value

| Country | Actual Value (Death Rate) | Predicted Value (Death Rate) | Actual Value (Recovery Rate) | Predicted Value (Recovery Rate) | Actual Value (Case Rate) | Predicted Value (Case Rate) |
|---|---|---|---|---|---|---|
| U.S.A | 5.3 | 5.1 | 11.72 | 11.56 | 5.11 | 5.14 |
| Spain | 10.4 | 10.2 | 2.47 | 2.98 | 10.44 | 11.21 |
| Italy | 13.2 | 12.12 | 4.21 | 3.92 | 13.12 | 13.56 |
| France | 12.9 | 10.87 | 5.03 | 5.01 | 10.86 | 10.74 |
| Germany | 3.2 | 2.95 | 1.68 | 1.54 | 2.94 | 2.68 |
| U.K | 13.4 | 12.94 | 3.46 | 2.53 | 13.32 | 12.98 |
| China | 5.4 | 6.83 | 1.06 | 1.23 | 5.60 | 5.62 |
| Iran | 6.1 | 6.9 | 1.49 | 1.49 | 6.24 | 6.56 |
| Turkey | 2.3 | 1.98 | 10.47 | 10.83 | 2.21 | 2.21 |
| Brazil | 6.2 | 5.81 | 2.20 | 2.22 | 6.32 | 6.34 |
| Canada | 4.98 | 5.14 | 3.09 | 3.65 | 3.97 | 3.76 |
| Russia | 0.7 | 0.87 | 12.13 | 11.84 | 0.83 | 0.56 |
| Switzerland | 5.1 | 5.9 | 1.68 | 1.74 | 4.79 | 4.57 |
| S. Korea | 2.24 | 2.75 | 1.36 | 1.31 | 2.16 | 2.21 |
| India | 3.2 | 3.6 | 7.59 | 7.11 | 3.32 | 3.41 |
| Japan | 2.65 | 2.56 | 9.87 | 9.41 | 2.06 | 2.11 |
| Pakistan | 2.12 | 2.79 | 3.98 | 3.93 | 1.92 | 1.89 |
| Australia | 1.13 | 1.22 | 1.73 | 1.72 | 0.97 | 1.12 |
| Mexico | 9.19 | 11.65 | 2.96 | 2.91 | 7.72 | 6.92 |
| Malaysia | 1.7 | 1.5 | 1.87 | 1.79 | 1.62 | 1.63 |
| Singapore | 0.11 | 1.2 | 6.48 | 7.21 | 0.23 | 0.34 |
| Thailand | 1.75 | 1.69 | 1.60 | 1.49 | 1.74 | 1.71 |
| South Africa | 1.9 | 2.12 | 2.88 | 2.95 | 1.84 | 1.69 |
| Bangladesh | 3.03 | 3.61 | 7.96 | 7.32 | 3.82 | 3.79 |
| Iraq | 4.95 | 5.56 | 1.68 | 1.72 | 5.58 | 4.94 |
| Afghanistan | 3.28 | 3.56 | 15.56 | 14.98 | 3.57 | 3.54 |
| Sri Lanka | 3.7 | 3.92 | 11.61 | 11.81 | 2.94 | 2.96 |
| Belgium | 14.2 | 13.98 | 4.60 | 3.94 | 13.95 | 13.71 |
| Israel | 1.3 | 1.3 | 5.33 | 5.21 | 1.15 | 1.27 |
| U.A.E | 0.64 | 0.64 | 5.32 | 5.27 | 0.60 | 0.53 |

**Table 7** Sample of risk-wise classification of countries

| S.No. | Country | High risk | Medium risk | Low risk |
|---|---|---|---|---|
| 1 | USA | Yes | – | – |
| 2 | Germany | – | Yes | – |
| 3 | Russia | – | – | Yes |
| 4 | S. Korea | – | Yes | – |
| 5 | Spain | Yes | – | – |
| 6 | Italy | Yes | – | – |
| 7 | France | Yes | – | – |
| 8 | UK | Yes | – | – |
| 9 | China | Yes | – | – |
| 10 | Iran | Yes | – | – |
| 11 | Turkey | – | – | Yes |
| 12 | Belgium | Yes | – | – |
| 13 | Brazil | Yes | – | – |
| 14 | Canada | – | Yes | – |
| 15 | Japan | – | – | Yes |
| 16 | India | – | – | Yes |
| 17 | Pakistan | – | Yes | – |
| 18 | Australia | – | Yes | – |
| 19 | Mexico | – | Yes | – |
| 20 | Malaysia | – | Yes | – |
| 21 | Singapore | – | – | Yes |
| 22 | Thailand | – | Yes | – |
| 23 | South Africa | – | Yes | – |
| 24 | Bangladesh | – | – | Yes |
| 25 | Iraq | Yes | – | – |
| 26 | Afghanistan | – | – | Yes |
| 27 | Sri Lanka | – | – | Yes |
| 28 | Switzerland | – | Yes | – |
| 29 | Israel | – | – | Yes |
| 30 | U.A.E | – | – | Yes |

Australia, Israel, Austria, and the United Arab Emirates. To prevent COVID-19 from spreading from a higher-risk to a lower-risk country, the government should undertake a lockdown strategy that includes halting all vital economic activity for 5–6 weeks.

## 6 Discussion

The applicability of the LM, BR, and SCG NN techniques for analysis and prediction of worldwide COVID-19 consisting of 204 samples from 30 January 2020 to 16 April 2020 was evaluated in this study was evaluated in this

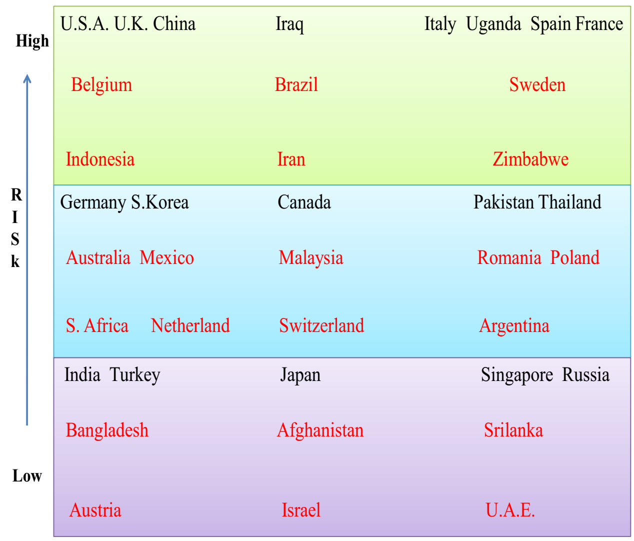


**Fig. 4** Classification of countries based on risk level

study. The MATLAB R2017a tool was used to run various experiments. The proposed model was assessed using a variety of input and training parameters including gradient, mutation, and validation checks, among others. The best validation training performance achieved with the LM technique was 1148.8717 at epoch 19, with 93.6565 gradients. Similarly, the best validation performance observed with the BR method was 301.9782 at epoch no. 251, with 12.2841 gradients. Finally, the best validation performance with the SCG approach was 325.8208 at epoch no. 20 and 43.7373 gradients. The LM technique had a mean absolute error (MAE) of 433.37, RMSE of 20.81, mean absolute percentage error (MAPE) of 141.89, and root relative squared error (RRSE) of 0.026. The BR approach yielded MAE of 270.84, RMSE of 16.19, MAPE of 2.46, and RRSE of 0.035. Finally, the SCG approach achieved the lowest error rates among the methods, with MAE of 239.80, RMSE of 15.48, MAPE of 0.209, and RRSE of 0.026.

Data from 204 countries were taken into account for analysis and classification, and countries were categorized as low-, medium-, or high-risk based on the output parameters DR, RR, and CR. Table 7 shows that high-risk countries include the United States, Spain, Italy, France, the United Kingdom, China, Iran, Belgium, Brazil, and Iraq, while medium-risk countries include Germany, South Korea, Canada, Pakistan, Australia, Mexico, Malaysia, Thailand, South Africa, and Switzerland. Russia, Turkey, Japan, India, Singapore, Bangladesh, Afghanistan, Sri Lanka, Israel, and the United Arab Emirates are all considered low-risk countries.

If the pandemic is not contained soon, these countries' risk ratings may shift, i.e., countries listed as high-risk may shift to medium- and low-risk, and vice versa.

## 7 Conclusion and future perspectives

The three most popular BPNN approaches are discussed in this study: LM, BR, and SCG. To estimate the performance of the suggested model, the RRSE, RMSE, MAPE, and MAE were calculated. The SCG method outperformed both of the other approaches, with the lowest errors of 239.80 (MAE), 0.209 (MAPE), 15.48 (RMSE), and 0.026 (RRSE).

We also classified countries based on their DR, RR, and CR, and discovered that COVID-19 has spread to a number of countries, including the United States, the United Kingdom, China, Uganda, Spain, India, Russia, France, Italy, Switzerland, Japan, Iran, Sri Lanka, and South Africa, among others.

We further classified some new countries as high-, medium-, or low-risk, including Belgium, Mexico, the Netherlands, Poland, Argentina, Zimbabwe, Romania, Israel, Austria, Malaysia, and the United Arab Emirates, among others. This could result in a global pandemic, which would have an impact on the global economy.

Due to this major epidemic, the world's large economies will experience a recession in the next years. To save the world's economy and people's lives, governments should take precautionary measures such as executing a lockdown strategy, maintaining social distancing, and temporarily shutting down all economic activity.

The most significant benefit of this research is that it will be useful for academics and scientists to gain a better understanding of the COVID-19 country statistics to improve their research. It would also be beneficial for the government and social workers to make timely strategic decisions to prevent COVID-19 situations and save the lives of millions of individuals around the world.

Only a few data samples (30 January to 15 April) are taken into account in this investigation. Large data samples and a fuzzy NN technique will be used in the future to improve performance.

## References


1. National Health Commission of the People's Republic of China. Update on pneumonia of new Corona virus infection as of 21:00 on 31st January available online: http://www.nhc.gov.cnxcs/yqtb/202002/84faf71e096446fdb1ae44939ba5c528.shtm. (2020). Accessed 12 May 2020
2. WHO. Novel coronavirus – Thailand (ex-China) (14th January, 2020) http://www.who.int/csr/don/14-january-2020-novel-coronavirusthailand/en/. Accessed 19 Jan 2020
3. Koh, G., F1000: Prime recommendation of report of the WHO-China joint mission on Corona Virus Disease 2019 (COVID-19). F1000 Post-pub. Peer Rev. Bio-med. Lit. (2020). https://doi.org/10.3410/f737509210.793572110
4. Read, J.M., Bridgen, J.R., Cummings, D.A., Ho, A., Jewell, C.P.: Novel coronavirus 2019-nCoV: early estimation of epidemiological parameters and epidemic. Predictions (2020). https://doi.org/10.1101/2020.01.23.20018549
5. Maier, B.F., Brockmann, D.: Effective containment explains subexponential growth in confirmed cases of recent COVID-19 outbreak in Mainland China. Science (2020). https://doi.org/10.1101/2020.02.18.20024414
6. Topcuoglu, N.: Public health emergency of International concern: Corona virus Disease 2019 (COVID-19). Open Dent. J. **14**(1), 71–72 (2020). https://doi.org/10.2174/1874210602014010071
7. BBC: Corona virus (2020) Sharp increase in deaths and cases in Hubei. https://www.bbc.co.uk/news/worldasiachina-51482994, Accessed 18 Feb 2020
8. Russell, S., Norvig, P.: Artificial Intelligence: A Modern Approach, 3rd edn. Prentice Hall (2009)
9. Witten, I.H., Frank, E.: Data Mining: Practical Machine Learning Tools and Techniques, 2nd edn. Morgan Kaufmann (2005)
10. Venkatalakshmi, B., Shivsankar, M.: Heart disease diagnosis using predictive data mining. Int. J. Innov. Res. Sci. Eng. Technol. **3**, 1873–1880 (2014)
11. Bellaachia, A., Guven, E.: Predicting breast cancer survivability using data mining techniques. In: Ninth Workshop on Mining Scientific and Engineering Data Sets in Conjunction with the Sixth SIAM International Conference on Data Mining. Bethesda, MD, USA, 20-22 April 2006
12. Sandhu, R., Sood, S.K., Kaur, G.: An intelligent system for predicting and preventing MERS-COV infection outbreak. J. Super Comput. **72**, 1–24 (2015)
13. Al, T.I., Mona, A., Tahani, A.: Building predictive models for MERS-COV infections using data mining techniques. J. Infect. Public Health **9**, 744–748 (2016)
14. Ferreira, D., Oliveira, A., Freitas, A.: Applying data mining techniques to improve diagnosis in neonatal jaundice. BMC Med. Inf. Decis. Mak. **12**, 1–6 (2012)
15. Pal, R., Sekh, A.A., Kar, S., Prasad, D.K.: Neural network based country wise risk prediction of COVID-19. IEEE Trans. J. **4**, 1–9 (2016)
16. AL-Najjar H, AL-Rousan N,: A classifier prediction model to predict the status of Corona Virus (COVID-19) patients in South Korea. Eur. Rev. Med. Pharmacol. Sci. **24**, 3400–3403 (2020)
17. Petropoulos, F., Makridakis, S.: Forecasting the novel coronavirus COVID-19. PLoS ONE **15**(3), 1–8 (2020)
18. Data Set Downloaded from Worldometer Website "http://worldometers.info" Accessed 16 Apr 2020
19. Han, J., Kamber, M.: Data Mining: Concepts and Techniques, 2nd edn. Morgan Kaufmann (2006)
20. Mezzatesta, S., Torino, C., De Meo, P., Fiumara, G., Vilassi, A.: A machine learning-based approach for predicting the outbreak of cardiovascular diseases in patients on dialysis. Comput. Methods Programs Biomed. **177**, 9–15 (2019)
21. Feng, N., Wang, F., Qiu, Y.: Novel approach for promoting the generalization ability of neural networks. Int. J. Signal Process. **2**, 131–135 (2006)
22. Ping, G., Michael, R.L., Chen, C.L.P.: Regularization parameter estimation for Feed forward neural networks. IEEE Trans. Syst. Man Cybern.etics PART B Cybern. **33**, 35–44 (2003)
23. Foresee, F.D., Hagan, M.T.: Gauss-Newton approximation to Bayesian learning. In: Proceedings: IEEE International Conference on Neural Networks, Vol. 3, pp. 1930–1935 (1997)
24. Moller, M.F.: A scaled conjugate gradient algorithm for fast supervised learning. Neural Netw. **6**, 525–533 (1993)


**Publisher's Note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.