ORIGINAL PAPER

INFECTIOUS DISEASES

THE INTERNATIONAL JOURNAL OF
CLINICAL PRACTICE WILEY

A study of the possible factors affecting COVID-19 spread, severity and mortality and the effect of social distancing on these factors: Machine learning forecasting model

Hossam M. Zawbaa[1] | Ahmed El-Gendy[2] | Haitham Saeed[3] | Hasnaa Osama[3] | Ahmed M. A. Ali[4] | Dina Gomaa[5,6] | Mona Abdelrahman[3] | Hadeer S. Harb[3] | Yasmin M. Madney[3] | Mohamed E. A. Abdelrahim[3]

[1]Faculty of Computers and Artificial Intelligence, Beni-Suef University, Beni-Suef, Egypt

[2]Microbiology and Immunology Department, Faculty of Pharmacy, Beni-Suef University, Beni-Suef, Egypt

[3]Clinical Pharmacy Department, Faculty of Pharmacy, Beni-Suef University, Beni-Suef, Egypt

[4]Department of Pharmaceutics and Industrial Pharmacy, College of Pharmacy, Taif University, Taif, Saudi Arabia

[5]Egyptian Drug Authority, Cairo, Egypt

[6]RAMEDA Pharmaceuticals, Cairo, Egypt

Correspondence

Mohamed E. A. Abdelrahim, Department of Clinical Pharmacy, Faculty of Pharmacy, Beni-Suef University, Beni-Suef, Egypt.
Email: mohamedemam9@yahoo.com

[Correction added on 22 March 2021, after first online publication: First author's name has been corrected from 'Hossam Zawbaa' to 'Hossam M. Zawbaa' in this current version.]

Abstract

Backgrounds: SARS-CoV-2 is affecting different countries all over the world, with significant variation in infection-rate and death-ratio. We have previously shown a presence of a possible relationship between different variables including the Bacillus Calmette–Guérin (BCG) vaccine, average age, gender, and malaria treatment, and the rate of spread, severity and mortality of COVID-19 disease. This paper focuses on developing machine learning models for this relationship.

Methods: We have used real-datasets collected from the Johns Hopkins University Center for Systems Science and Engineering and the European Centre for Disease Prevention and Control to develop a model from China data as the baseline country. From this model, we predicted and forecasted different countries' daily confirmed-cases and daily death-cases and examined if there was any possible effect of the variables mentioned above.

Results: The model was trained based on China data as a baseline model for daily confirmed-cases and daily death-cases. This machine learning application succeeded in modelling and forecasting daily confirmed-cases and daily death-cases. The modelling and forecasting of viral spread resulted in four different regions; these regions were dependent on the malarial treatments, BCG vaccination, weather conditions, and average age. However, the lack of social distancing resulted in variation in the effect of these factors, for example, double-humped spread and mortality cases curves and sudden increases in the spread and mortality cases in different countries.

The process of machine learning for time-series prediction and forecasting, especially in the pandemic COVID-19 domain, proved usefulness in modelling and forecasting the end status of the virus spreading based on specific regional and health support variables.

Conclusion: From the experimental results, we confirm that COVID-19 has a very low spread in the African countries with all the four variables (average young age, hot weather, BCG vaccine and malaria treatment); a very high spread in European countries and the USA with no variable (old people, cold weather, no BCG vaccine and no malaria). The effect of the variables could be on the spread or the severity to

the extent that the infected subject might not have symptoms or the case is mild and can be missed as a confirmed-case. Social distancing decreases the effect of these factors.

What's known

SARS-CoV-2 is affecting different countries all over the world, with significant variation in infection-rate and death-ratio. We have previously shown a presence of a possible relationship between different variables, including the Bacillus Calmette–Guérin (BCG) vaccine, average age, gender, and malaria treatment, and the rate of spread, severity and mortality of COVID-19 disease. This paper focuses on developing machine learning models for this relationship.

What's new

From the experimental results, we confirm that COVID-19 has a very low spread in the African countries with all the four variables (average young age, hot weather, BCG vaccine and malaria treatment); a very high spread in European countries and the USA with no variable (old people, cold weather, no BCG vaccine and no malaria). The effect of the variables could be on the spread or the severity to the extent that the subject would not have symptoms or the case is mild and can be missed as a confirmed case. Social distancing decreases the effect of these factors.

1 | INTRODUCTION

The COVID-19 disease was first recognised in December 2019 in Wuhan, the capital of China's Hubei province. As of the 14th December 2020 (according to the world health organization report[1]), over 70 million confirmed cases have been reported in more than 200 countries, and over one and a half million confirmed deaths and progressing. This pandemic is odd when compared with the magnitude of previous infectious outbreaks, including influenza, Middle East respiratory syndrome (MERS), and severe acute respiratory syndrome (SARS); however, the influenza pandemic in 1918 was the closest in terms of transmission and fatality. With the increased threats of the COVID-19 outbreak, based on the previous infectious outbreaks control plans, several countries responded by applying the mitigation model and focused on the delay of the virus arrival to flatten the pandemic curve.[2,3]

The SARS-CoV-2 virus is mainly spread after close contact with infected people during coughing or sneezing. Moreover people may also become infected by touching a contaminated surface or object, then touching their eyes, nose or mouth.[4]

China's death rate was around 4%, while it was greater than 10% in Italy and the USA. SARS-CoV-2 spread from China to most countries around the world. The spread of the disease became higher in Europe and the USA compared with China. The number of confirmed COVID-19 cases detected in countries neighbouring China, for example, Kazakhstan, India and Korea, was lower than that detected in the USA and Europe, for example, Italy, Spain, France and the UK. To the extent that a close contact area to COVID-19 source, Hong Kong has fewer cases with a very low mortality rate of 0.71%. The USA now became the highest country with confirmed COVID-19 cases and progressing. Something else other than social distancing could

be the cause of this variation in infection severity. One confusing debate raised right now is the possible relation between the vaccination schedule in different countries and the prevention of the spread and lower severity of SARS-CoV-2 infection.[5] Also, during the current COVID-19 outbreak, it was reported that infants are less susceptible to such a violent virus or mostly presented with mild cases.[6,7]

A model was developed by Anderson et al 2020 to predict the spread of COVID-19 within each country, taking into account only the social distancing as a possible limitation that could decrease the infection spread.[8] However, this model failed to predict why the spread in Europe and the USA were higher than China and Iran and why Africa and the Arab countries have lower spread. Possibly, they have a less severe infection because of other factors to the extent that most of the cases in Africa and the Arab countries are undetected. As previously reported, other possible factors can be included as limitations of the model, for example, Bacillus Calmette–Guérin (BCG) vaccine, antimalarial administration, country average age and country average weather temperatures.[5] Nowadays, most of the developed applications and systems for different businesses are based on artificial intelligence and machine learning. Deep learning is a part of machine learning that produces proper performance and completely surpasses classical machine learning methods, especially when the scale of data increases.

Unlike linear regression, in artificial intelligence and machine learning, the model is first trained using a baseline model, then fitting the sequence samples by framing the sequence samples to have certain inputs and output. In the second step, retraining the model is established using the data that need to be modelled and the variables used.

Sequence or time-series prediction has been around for a long time. It is considered one of the most challenging problems to solve in machine learning and data science fields. The main objective of this study was to develop a comprehensive analysis of COVID-19 time-series data and create a machine learning model to forecast the number of confirmed and death cases for such a pandemic based on the suggested variables. The machine learning model was based on the four possible variables (BCG vaccine, Malaria treatment, weather temperatures and age) that could affect the scenarios of the COVID-19. Even though not many pieces of evidence are shown for those factors, but we tried to compare and predict the confirmed cases and death cases and determine if these variables have an effect, as previously reported, and suggest an explanation; Taking into account that social distancing varies in the degree of accomplishment and we do not have any certain data stating how the social distancing in each country is and could not measure it with confidence like the variables we are using here.

2 | METHODS

2.1 | Multilayer perceptron (MLP)

Each perceptron recognises linearity during the linear classification of data by dividing the input data into two groups by a straight line.[9] The input data x is a feature vector multiplied by w weights and summed to the b bias as in the Equation (1):

$$y = w * x + b \tag{1}$$

The multilayer perceptron (MLP) is a deep artificial neural network, which is formed of more than one perceptron. MLP is comprised of three layers:

1. The input layer to get the input signal
2. The output layer gives a decision or prediction about the input data
3. In between those two, the number of hidden layers that are the accurate computational engine of the MLP

MLP with one hidden layer is capable of approximating any continuous function.[9]

MLPs are made up of several neurons that have learnable weights and biases. Every perceptron provides a single output based on different real-valued inputs by performing a dot product (linear combination) and applying its input weights (optionally follows it with a non-linearity) as in the Equation (2):

$$y = \varphi \left(\sum_{i=1}^{n} w_i x_i + b \right) = \varphi \left(W^T X + b \right) \tag{2}$$

where x denotes the vector of inputs, w indicates the vector of weights, b expresses the bias, and φ is the nonlinear activation function.

The MLP can be observed as a logistic regression classification where the input is first reconstructed using a well-informed nonlinear transformation. Moreover this transformation is converting the input data into a different space where the data becomes linearly separable. This intermediate layer is identified as the hidden layer. A single hidden layer is enough to produce MLPs. Nevertheless, there are extraordinary advantages to use many hidden layers, ie, the very assumption for deep learning. The MLP deep learning expresses a single differentiable score function from the raw image pixels as an input and outputs the corresponding class scores. Also, the MLP deep learning still has a loss function on the last (fully connected) layer.[10]

Multilayer Perceptron (MLP) has been applied to model and forecast the univariate time series COVID-19 data. Univariate COVID-19 time series is a dataset composed of a single series of observations (the total number of confirmed cases and the total number of death cases) with a temporal ordering (date). Therefore, the MLP model is used to learn from the COVID-19 series of past observations (inputs) and predict the next value in the sequence (output). We have done two main phases to achieve this work as follows:

2.2 | Data preparation

A python machine learning application based on a multilayer perceptron (MLP) is used to model and forecast the COVID-19 virus spreading pattern.

Data preparation was managed using the data example in Table 1 at which COVID-19 time series real data were prepared and transformed into suitable interpretation and coding to fit into the MLP model for the forecasting problem. The MLP model learns a function that maps a sequence of time series data as input and converts these observations as in the following steps using the data example in Table 1.

The COVID-19 time series data for nine countries was downloaded, as shown in Table 2. For Chinese data, we have downloaded from the European Centre for Disease Prevention and Control

TABLE 1 Snapshot of Egypt's COVID-19 confirmed cases from JHU CSSE[1]

Date	COVID-19 confirmed cases
04/03/2020	0
05/03/2020	1
06/03/2020	12
07/03/2020	0
08/03/2020	34
09/03/2020	6
10/03/2020	4
11/03/2020	1
12/03/2020	7
13/03/2020	13

(ECDC) for the period 29 December 2019 till 13 December 2020 (we removed the sample of the data for 17 April 2020 as an outlier—because the number of death cases was 1290 accidentally)[11,12] For the other eight countries, data were downloaded for the period

TABLE 2 Selected countries and their corresponding factors for the experiment

Regions	Countries	Factors
Europa/USA	Italy, USA	Old people, Cold weather
Asia	China, Japan, Iran	BCG
Arab countries	Egypt, Alegria	Young people, Hot weather, BCG
African countries	Kenya, Cote d'Ivoire	Young people, Hot weather, BCG, Malaria

TABLE 3 Input/output patterns for the Table (1) time-series data

Inputs	Output
[0, 1, 12, 0, 34, 6, 4]	1
[1, 12, 0, 34, 6, 4, 1]	7
[12, 0, 34, 6, 4, 1, 7]	13

22 January 2020 till 13 December 2020 from the Johns Hopkins University Center for Systems Science and Engineering (JHU CSSE). Also, we removed the following observations of the time series data (06 June 2020 from Japan death cases), (24 June 2020 from Italy death cases), and (19 June 2020 from Italy confirmed cases) as outliers—because these numbers were negative.

After that, we have deleted the date column from the time series data. The sequence of observations was as follows: [0, 1, 12, 0, 34, 6, 4, 1, 7, 13] as in Table 2. The sequence of observations must be transformed into multiple samples from the MLP can learn. We divided the sequence into multiple input/output patterns called samples. In this paper, the sequence of observations was transformed using a window size equal to one week (because the spread of COVID-19 can be significant from week to week). We have considered seven-time steps (one-week of the daily reported COVID-19 cases) as the input and one-time step as the corresponding output for the prediction model that was being learned, as shown in Table 3.

Adding four more inputs to represent the different factors with 0 and 1 encoding as follows: 1- Average age (Youth:1, Old:0), 2- Average weather temperature (Hot:1, Cold: 0), 3- BCG vaccination (Yes:1, No:0), 4- Malaria treatment (Yes:1, No:0) as shown in Figure 1. Figure 1 shows a map displaying BCG vaccination policy by country[13]; the estimate of world malaria burden[14]; average yearly weather temperature by country[15]; and average age by country.[16]

A: The country currently has universal BCG vaccination program. B: The country used to recommend BCG vaccination for everyone, but currently does not. C: The country never had universal BCG vaccination program.
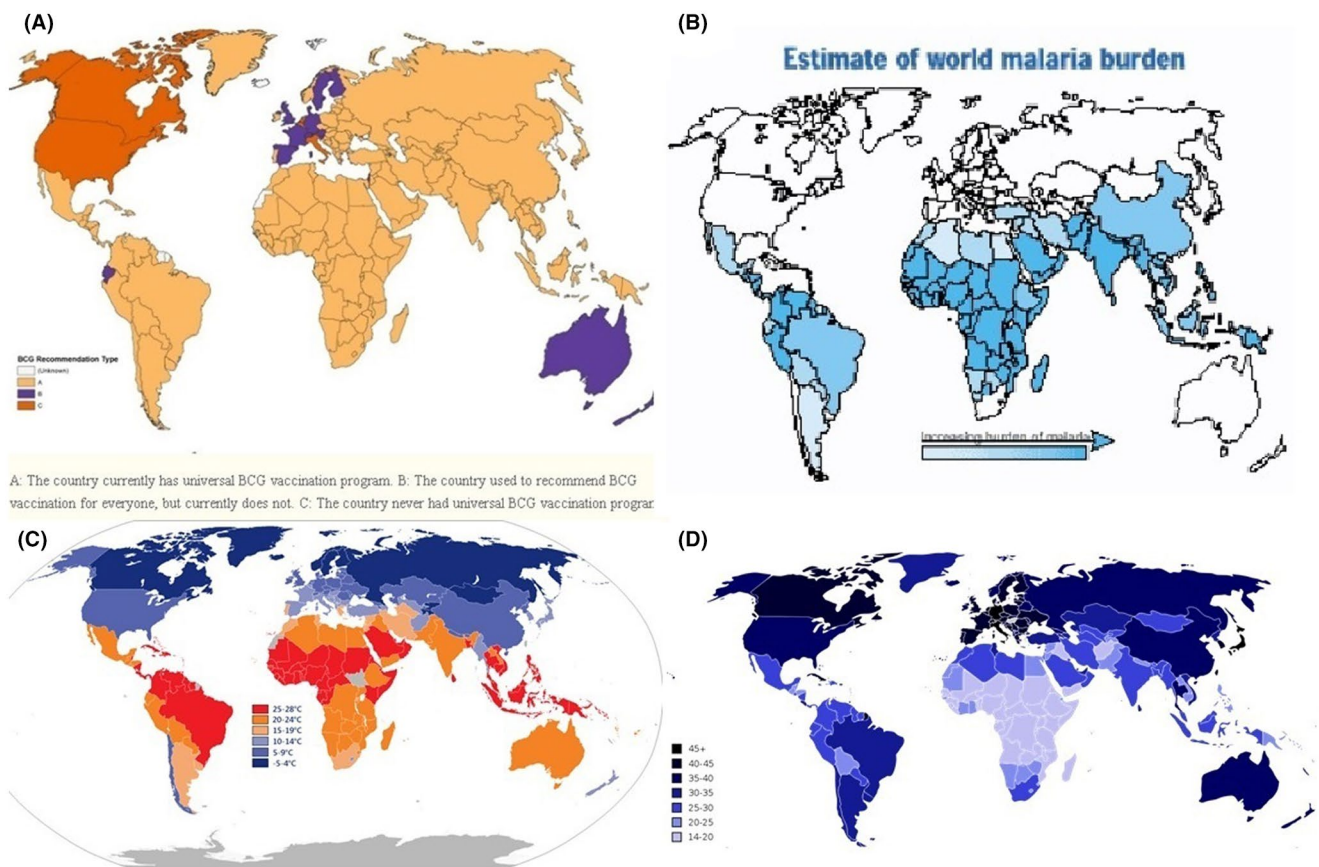
FIGURE 1 A, Map displaying BCG vaccination policy by country (BCG vaccinated countries in yellow)[13] B, The estimate of world malaria burden[14] C, Average yearly weather temperature by country[15] D, Average age by country[16]

TABLE 4 Input/output patterns for the first sample (row) in Table (3) time-series data after adding the four factors

Countries	Inputs	Output
Europa/USA	[0, 0, 0, 0, 0, 1, 12, 0, 34, 6, 4]	1
China/Japan/Iran	[1, 0, 1, 0, 0, 1, 12, 0, 34, 6, 4]	1
Arab countries	[1, 1, 1, 0, 0, 1, 12, 0, 34, 6, 4]	1
African countries	[1, 1, 1, 1, 0, 1, 12, 0, 34, 6, 4]	1

We have used the cutoffs for the binary categories as the following: 1- Average age > 35; considered Old, otherwise, Youth, 2- Average weather temperature > 25; considered Hot, otherwise, Cold, 3- BCG vaccination exists in the country; considered Yes, otherwise, No, 4- Malaria and its treatment is common in the country; considered Yes, otherwise, No; as shown in Figure 1 and Table 4.

We have used the multilayer perceptron (MLP) as a regression model to predict and forecast the COVID-19 spread for the coming months based on the number of confirmed cases and the number of death cases.

2.3 | MLP model

We have implemented a simple MLP model with an input layer, a single hidden layer, and an output layer, which is used to make a prediction. The input layer has several neurons equal to the number of sequence steps (11 inputs; 4 for factors and 7 for one-week COVID-19 time series data). The hidden layer is a dense layer with 100 neurons and a rectified linear unit (ReLU) as an activation function. The output layer had a dense layer with 1 unit for predicting the corresponding output. Moreover we have used 500 as the number of epochs, Adam as the optimiser, and mean squared error (MSE) as the loss function. Finally, we fit the model with prepared data to make a prediction. The obtained results may vary given the stochastic nature of the MLP algorithm; therefore, we have run the model several times. Finally, we enter the last observation sequence and its corresponding output to forecast the next value in the sequence.

2.4 | Datasets

We have selected nine countries (the highest in their region) and their corresponding factors (Average age, Average weather temperature, BCG vaccination, Malaria Treatment) as shown in Table 4.

2.5 | Experimental setup

We have developed the multilayer perceptron (MLP) for time series forecasting with the following steps: the univariate problem proposed in the COVID-19 time-series after the data preparation phase was considered a sequence of integers. In the first step, training the model was carried out using all the COVID-19 sequence samples for both the total number of confirmed cases and the total number of death cases of China as the baseline model because of China's success of almost control the spread of COVID-19. Fitting the COVID-19 sequence samples of China into the model was undertaken by framing the sequence samples to have 11 inputs (4 factors and 7 for one-week COVID-19 recorded numbers from CSSE data) and 1 output (for example, the 8th daily COVID-19 recorded number). Two models were developed for China, one for confirmed cases and the other for the death cases. The Chinese model's settings included 500 as the number of epochs, Adam as the optimiser, and mean squared error (MSE) as the loss function.

In the second step, retraining the model was established using all COVID-19 sequence samples for both the total number of confirmed cases and the total number of death cases of all the eight countries (two models for every country, we have 16 different models). Fitting the eight countries' COVID-19 sequence samples into the model was also performed by framing the sequence samples to have 11 inputs.

For the prediction model, the model predicts the previously trained data (actual data vs. model prediction). For the forecasting model, the model forecasts the following values in the sequence (the forecast for the next two months based on the last record's predicted output). We have split our results into four groups, as shown in Figures 2–4:

1. The model training dataset (in purple colour)
2. The actual data of the testing dataset (in blue colour)
3. The model prediction data for the testing dataset (in red colour)
4. The model forecasting for the confirmed cases and the death cases (in green colour).

3 | RESULTS AND DISCUSSION

The confirmed cases and death cases models prediction and forecasting for China, as the baseline country, are shown in Figure 2. Based on these four possible factors, when looking at the COVID-19 data in different countries, we can divide countries into four categories; Countries with a very hot climate, with most of their people taking antimalarial and BCG vaccine, and with an average young age (four factors); those are most of the African countries with their low spread, Figures 3A,B, and mortality Figures 4A,B, of COVID-19; Countries with all the above limiting criteria except the antimalarial (three factors) which are the Arab countries with their higher data compared with the African countries as shown in Figures 3C,D and 4C,D; Countries with low climate temperature but take BCG vaccine and with younger average age than Europe and the USA (2 factors), for example, Iran, and Japan with their high data as shown in Figures 3E,F and 4E,F. Lastly, countries that do not have any of the above limitation (0 factors), for example, most of Europe and the USA with their very high COVID-19 disease spread and mortality compared with the other three categories as shown in Figures 3G,H and 4G,H.
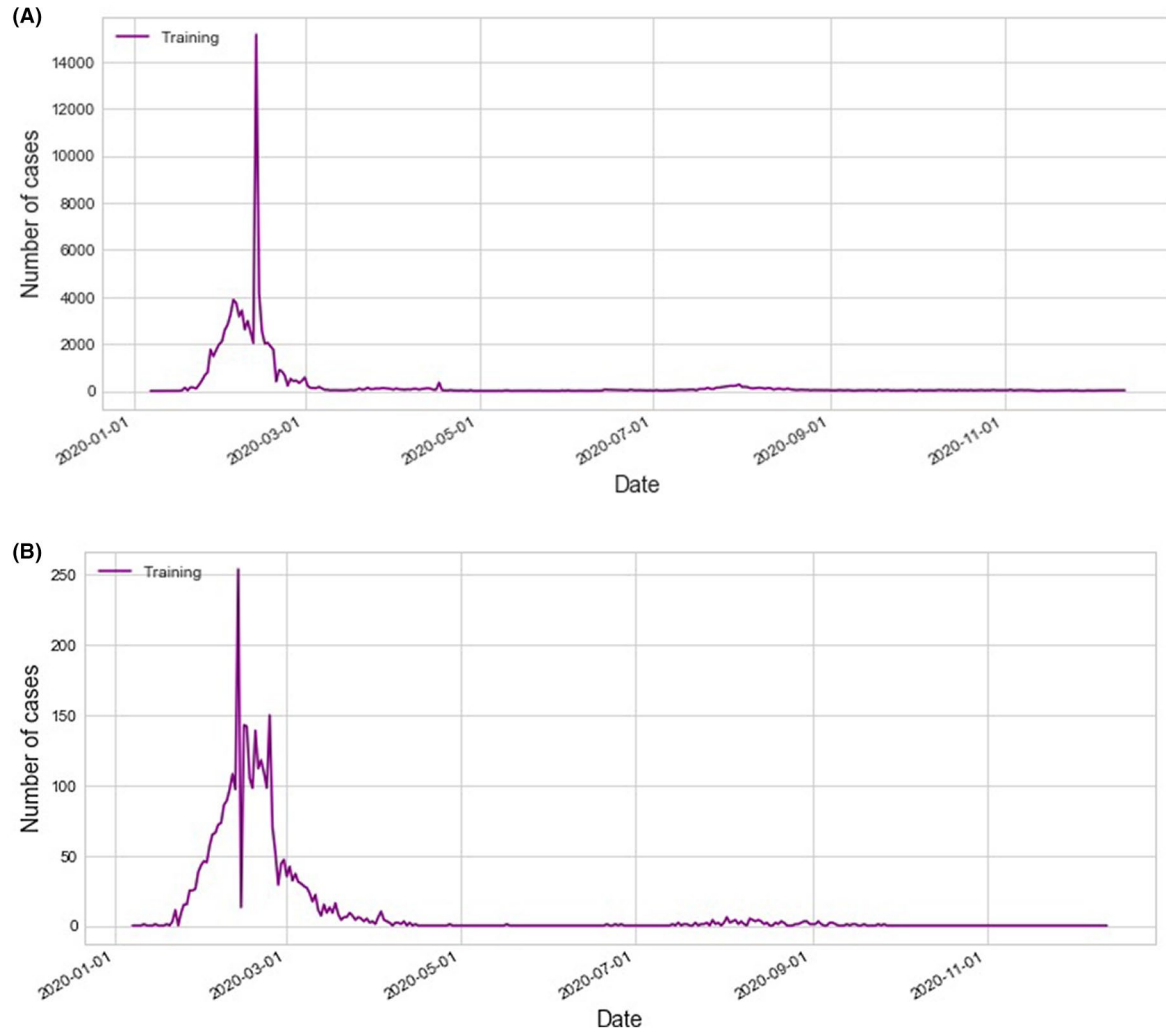
FIGURE 2 Model training data (Purple) for the confirmed cases (a) and death cases (b) for China as the baseline country

The created series of prediction and forecasting models, based on Chinese data, in the eight countries (USA, Italy, Iran, Japan, Algeria, Egypt, Cote d'Ivoire and Kenya) was very simple and easy. The MLP model has the lowest RMSE, for the confirmed cases, is 105.94 and 0.91 for death cases; both are obtained from Cote d'Ivoire. Also, the model has the highest RMSE, for the confirmed cases, which are 77822.38 and 792.07 for death cases; both are obtained from the USA as shown in Table 5. From the MLP models, we confirm that the USA and Europe, as shown from Italy results, would take the longest period to recover from the SARS-CoV-2 virus. The lack of social distancing was very obvious in the sudden increase in the number of confirmed cases and death cases on day 30.06.2020 in the USA Figures 3H and 4H. This could be because of the demonstrations and the presidential campaigns in the USA around this time.[17,18]

Moreover countries with all the four factors (average young age, hot weather, BCG vaccine and malaria treatment), for example, African countries, and countries with three factors (young people, hot weather and BCG vaccine), for example, in Arab countries have a low severity and spread of the virus. However, because of the lack of social distance in Iran, Egypt and Algeria COVID-19 spread will last longer; that can be seen from the obvious double-humped confirmed cases and death cases curves of Iran, Algeria Figures 3D,F and 4D,F and the confirmed cases and death cases curves of Egypt that started low and suddenly increased (2.5 time) as shown in Figures 3C and 4C. That occurs around the mid of May 2020 which was around the mid of Ramadan the holy month in these three countries which has a special social gathering that could cause a lack of social distancing resulting in these abrupt increases.[19,20] Also, it is very obvious when you look at the results of Cote d'Ivoire and Kenya Figure 3A,B. Cote d'Ivoire confirmed cases Figure 3A with its bell-shaped curve which is controlled now and Kenya confirmed cases Figure 3B with its double-humped curve that started low but still going. This difference could be because of either social distancing or lake of tests performed to detect SARS-CoV-2 infection since both countries should have similar curves as both have the 4 factors. On the other hand, if we look at the data of China (Figure 2 with its proper application of social distancing, we can see that the effect of the above-mentioned factors was overcome by proper social distancing.

The effect of lack of social distancing on death cases curves was higher than in the curves of the confirmed case. That could be

because a huge number of confirmed cases can be missed, especially the mild and moderate cases, and different in countries number of tests performed to detect SARS-CoV-2 infection which varies very much. For example, the USA performed around 200 million tests and Kenya performed around one million tests. That could significantly affect the number of confirmed cases. Moreover from Figures 3 and 4, most of the countries have started their second cycle of the SARS-CoV-2 virus spread but it seems controllable because of
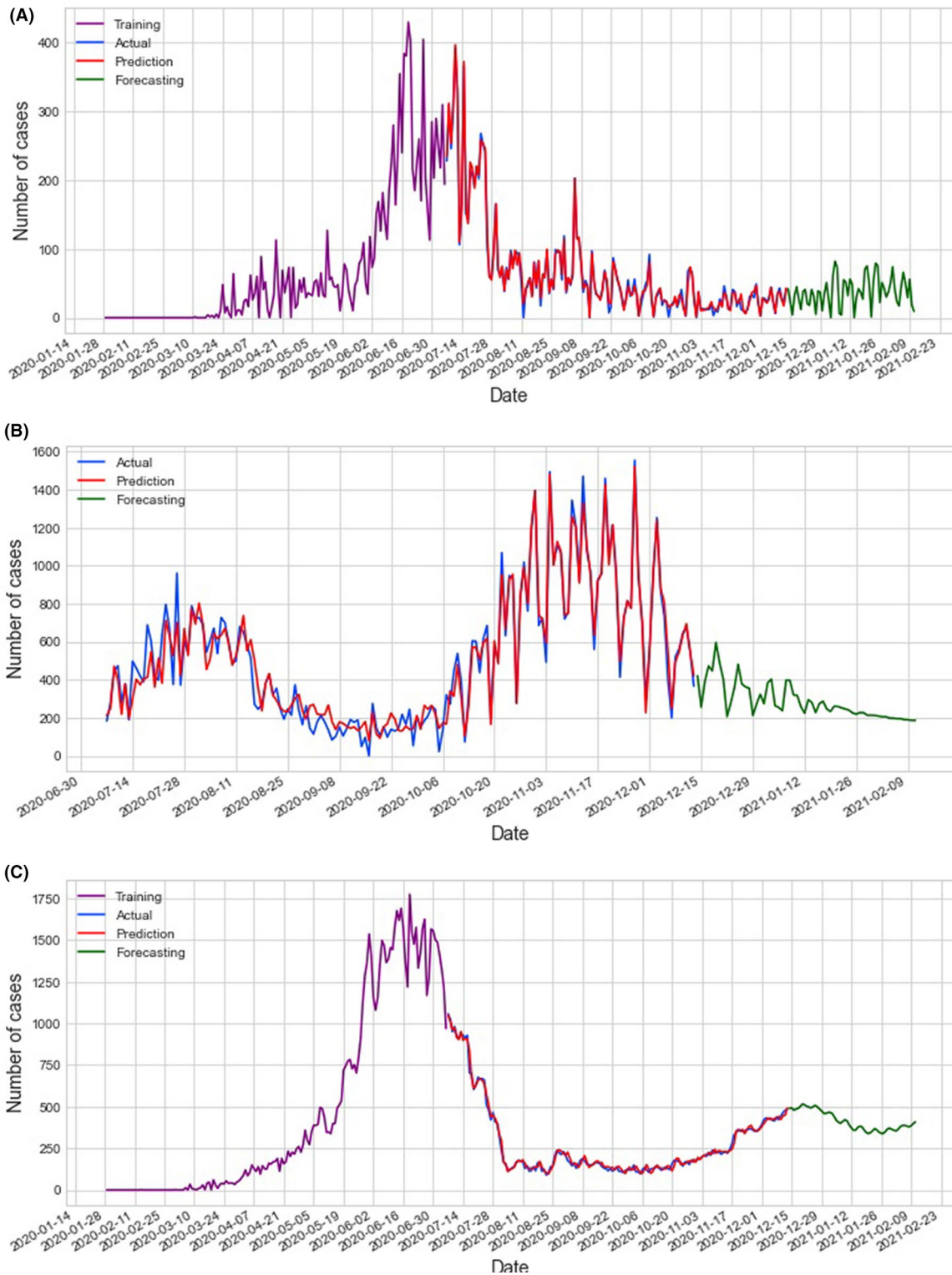
FIGURE 3 Model training data (Purple), prediction (Red), and forecasting (Green) vs. actual data (Blue) for the confirmed cases; Confirmed cases model prediction and forecasting for eight countries a. Cote d'Ivoire, b. Kenya, c. Egypt, d. Algeria, e. Japan f. Iran, g. Italy, and h. the USA

the increases in the awareness of the disease except in Europe, as shown from Italy and the USA.[4,21] Therefore, the social distancing as a variable could only affect the model of spread, severity and mortality of any pandemic when the social distancing process ends earlier than it should be or not properly achieved, for example,

Denver mortality double-humped curve of the Spanish flu pandemic in 1918[22]; Where the outbreak started to flatten and progress to end when social distancing was properly achieved but, when they ended social distancing process and resumed normal life early, another stronger pandemic with higher mortality occurs in Denver, the
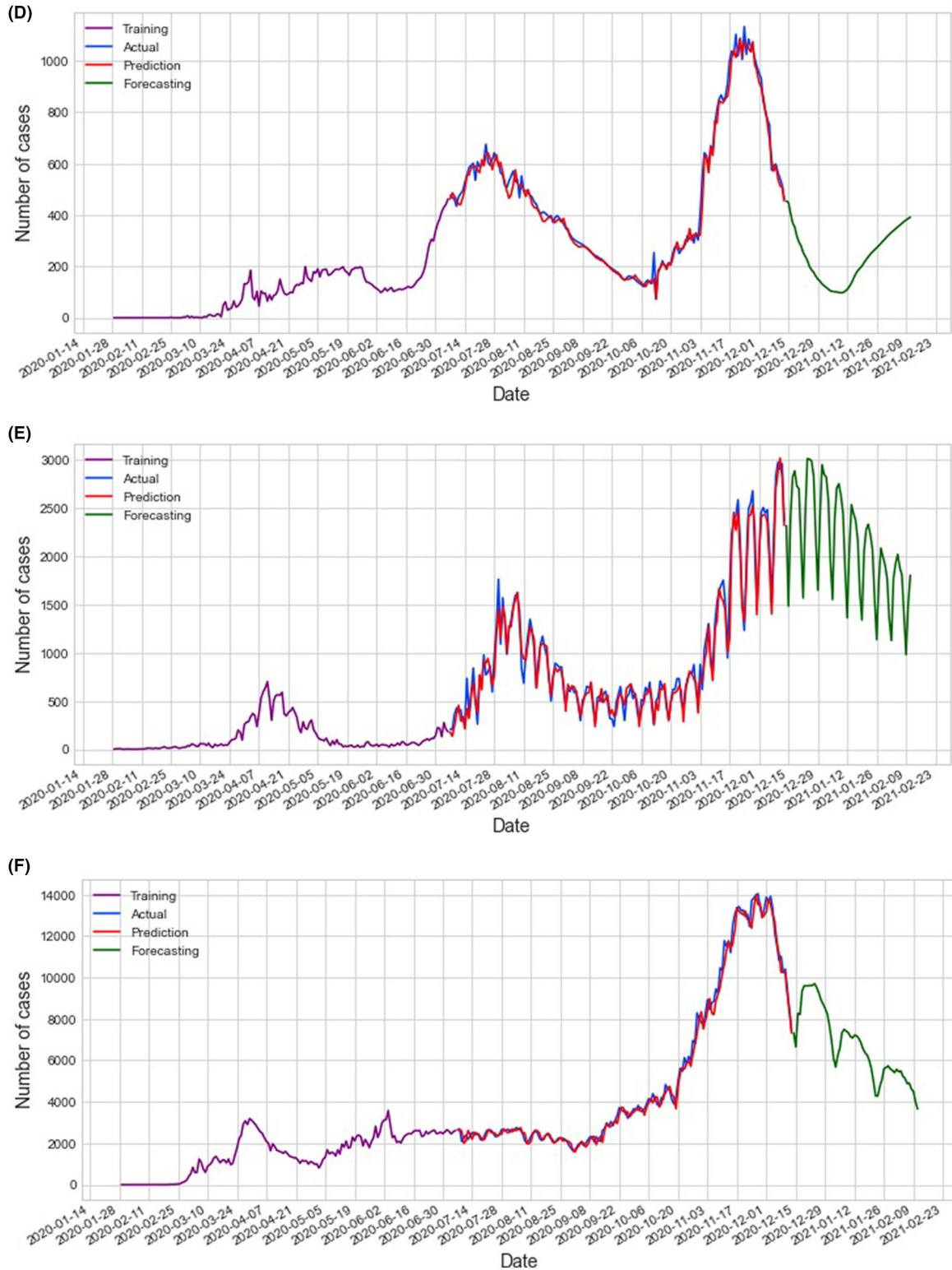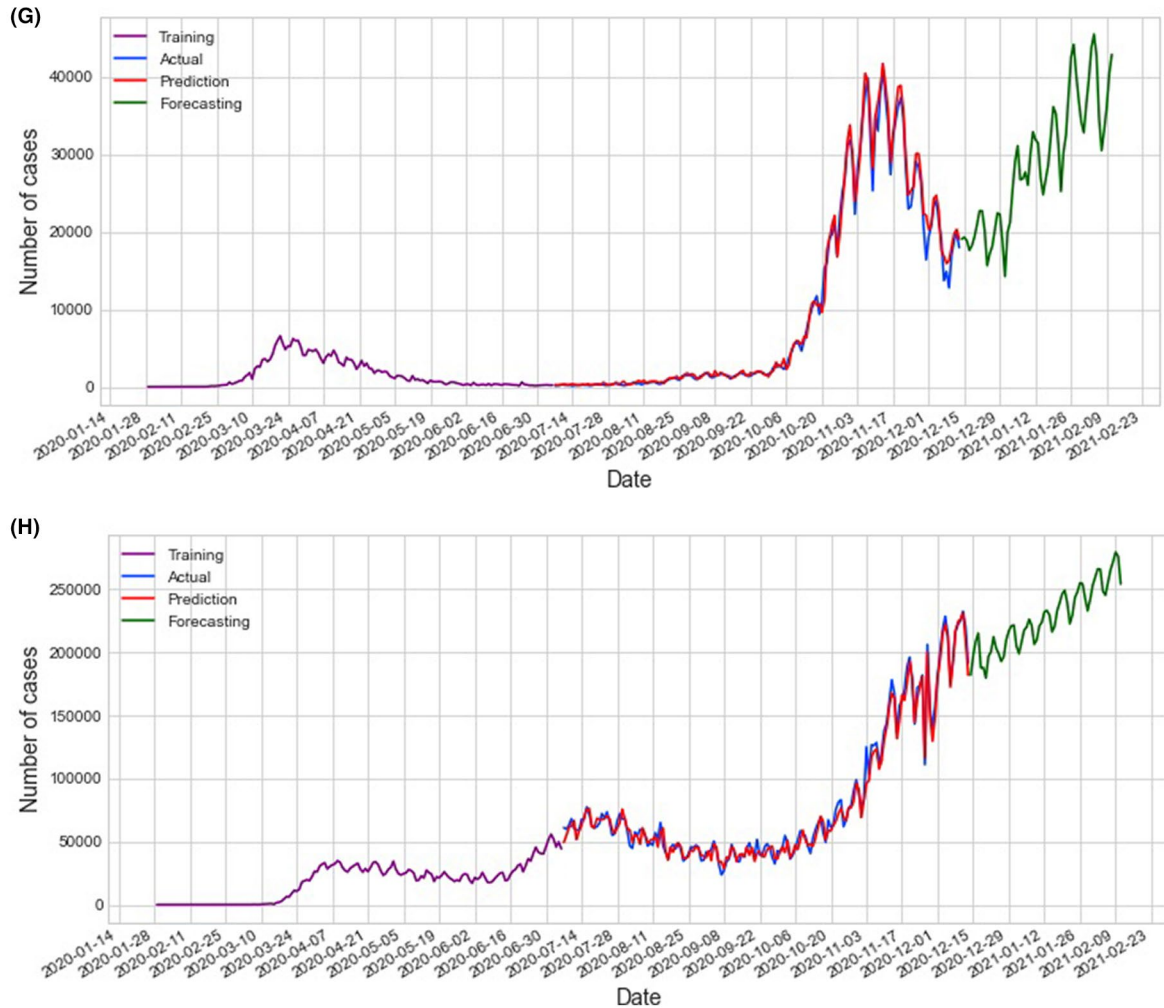

FIGURE 3 (Continued)

FIGURE 3 (Continued)

USA in 1918.[22] Moreover, proper application of social distancing and isolation policies during previous similar outbreaks such as H5N1 in 2001, SARS in 2003, H1N1 in 2009 and MERS in 2012 is successful in controlling and limiting viral transmission among the population. So, it is urgently advised and requested that preventive measures and quarantine should be strictly followed.[23] Also, there is an urgent requirement to have educational science and technology to fight against any such pandemic in the future.[4]

Our advice to reallocate the BCG vaccine for protection against COVID-19 is based on the capability of the BCG vaccine to induce non-specific innate immunity against various unrelated pathogens including viruses and parasites through an epigenetic reprogramming mechanism that recalls the innate immunity. The recalled immune cells, for example, monocytes, macrophages and natural killer T cells modify histones by methylation and acetylation and consequently stimulate the expression of interleukin-6 (IL-6), tumour necrosis factor-α (TNF-α) and IL-1β after vaccination using BCG vaccine.[24]

Hence, the innate immunity tends to increase the production of the pro-inflammatory cytokines, which in turn decreases the infectivity and severity of COVID-19 disease through the antiviral inhibitory cytokines, activated natural killer cells, and the activated cytotoxic T-cells.[5] Another assumption is based on the continuous cytokine production in pulmonary cells after vaccination with the BCG vaccine. This will promote pulmonary cells to be more trained and adapted to cytokines for a long time and will be less responsive to side effects of cytokine storm induced by COVID-19; hence, decreasing severity since the cytokine storm is the main cause of respiratory failure and death of COVID-19.[5]

The induction of cytokine is attributed to what is known as heterologous non-specific effects of the BCG vaccine which is also caused by other live-attenuated vaccines (eg, mumps-measles-rubella vaccine "MMR," oral polio vaccine "OPV" and smallpox vaccine).[25] In the case of the BCG vaccine, heterologous effects can protect against other non-mycobacterial infections to generally lower respiratory tract infections through activation of innate immunity memory-based myeloid cells, a process termed trained immunity.[25,26]

Viruses as mumps, measles, rubella and polio do not primarily attack lung cells while both Mycobacterium tuberculosis (TB) and SARS-CoV-2 attack the lungs and interfere with host immunity.[27] This could interpret that, although the aforementioned vaccines can induce cytokine production, only the BCG vaccine was found to be linked to decrease infectivity, severity and mortality of COVID-19.[28]

Cytokines as IL-1beta, TNF-alpha, IFN-gamma, IL-4 and IL-13 are known to induce cellular antiviral state.[29] Cytokines-production after BCG vaccination is represented in increasing IFNγ and pro-inflammatory cytokines Interleukins IL-1β, IL-2, IL-6, IL-8, IL-10, IL-12, IL-17 and TNF.[30-32] Such cytokine induction, especially IL-2, IL-12, TNF and IFN-γ, can lead to the activation of cytotoxic T-lymphocytes with their antiviral behaviours. Also, IFN-γ has many roles as antiviral, strong regulatory of the-immune
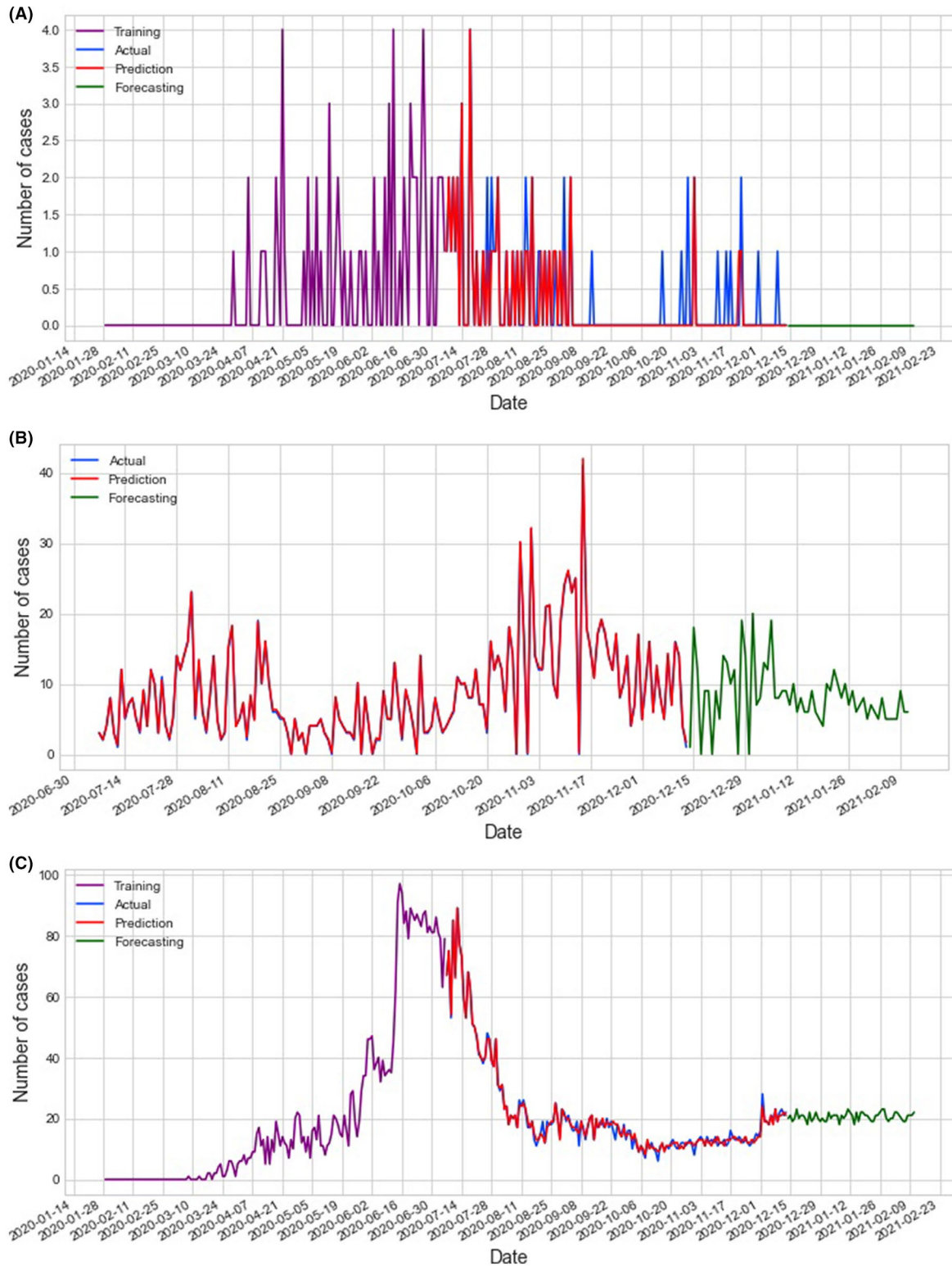
FIGURE 4 Model training data (Purple), prediction (Red), and forecasting (Green) vs actual data (Blue) for the death cases; Death cases model prediction and forecasting for seven countries A, Cote d'Ivoire, B, Kenya, C, Egypt, D, Algeria, E, Japan F, Iran, G, Italy, and H, the USA

response, stimulation of phagocytic bactericidal activity, induction of antigen presentation through major histocompatibility-complex (MHC) molecules, the arrangement of leukocyte-endothelium interactions, and effects on cell proliferation and apoptosis.[33] The inducting of cytokine by BCG vaccine is mediated by monocytes and natural killer (NK) cells with a memory-like effect via a mechanism of epigenetic reprogramming and chromatin remodelling through histone modifications, leading to enhanced gene transcription and upregulation of IL-1ß which is considered as an antiviral cytokine.[29]
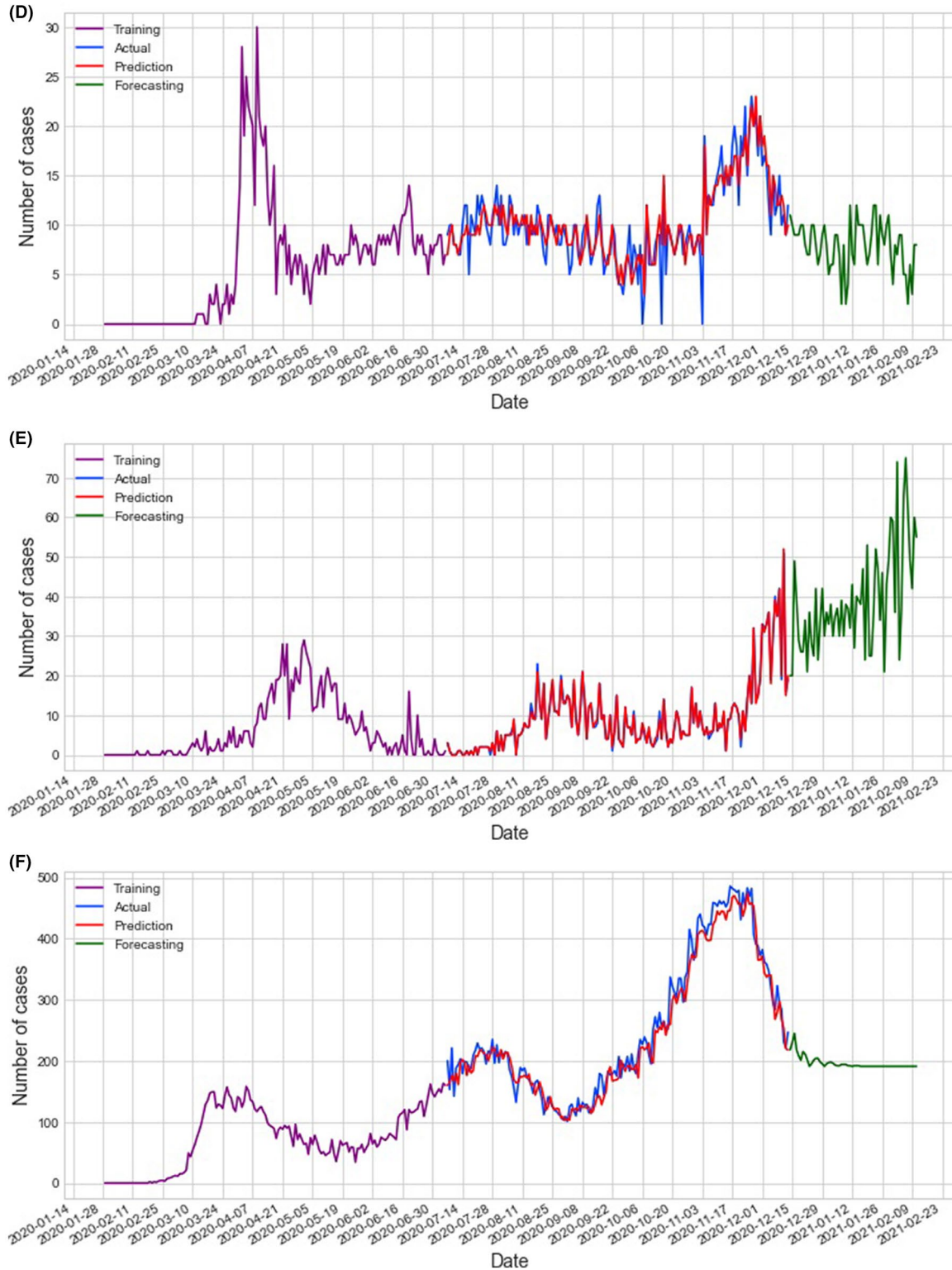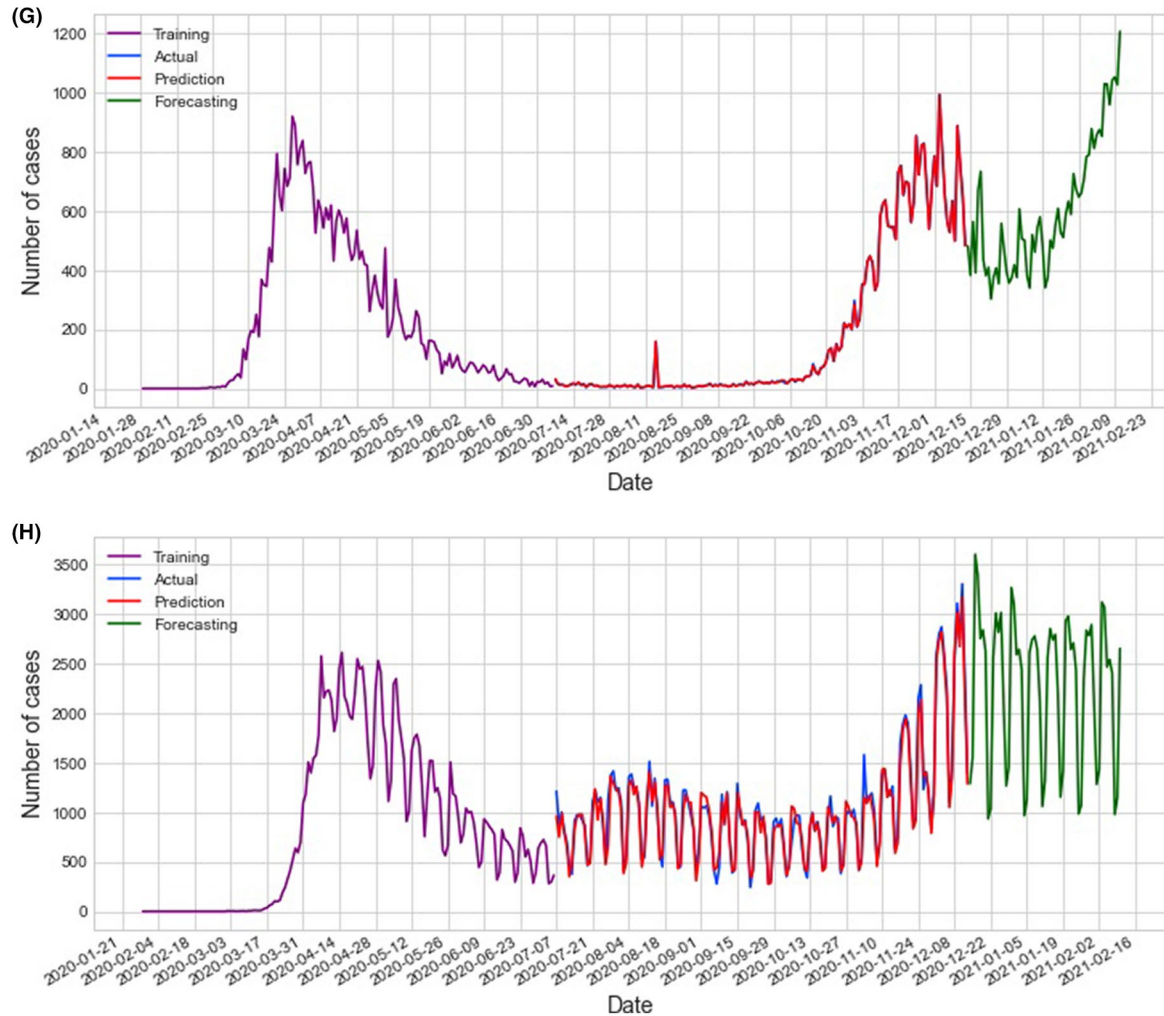

FIGURE 4 (Continued)

FIGURE 4 (Continued)

TABLE 5 Root mean square error (RMSE) between the actual and predicted values of each country for confirmed and death cases

Country	RMSE for confirmed cases	RMSE for death cases
China	1460.57	38.98
Cote d'Ivoire	105.94	0.91
Kenya	486.13	9.79
Egypt	318.29	22.97
Algeria	371.40	5.59
Japan	952.29	12.67
Iran	5632.84	160.60
Italy	17922.58	384.54
USA	77822.38	792.07

In African countries, the number of confirmed cases is considered very low compared with Europe or even China. The highest country with death cases is Egypt with over 120,000 cases now and progressing which is even with its low numbers is in the category of Arab countries with three factors not four like the rest of the African countries. Malaria disease is considered a common disease in those countries, as shown in Figure 1B.[14] The finding in countries encountered malaria could be related to the malaria treatment, for example, chloroquine, which has around 50 days half-life,[34] and hydroxychloroquine, which have around 40 days half-life,[34] that was proven effective against SARS-CoV-2 virus in in-vitro studies, while clinical trials had a controversial finding, some trial showed a beneficial effect while other showed no difference between subjects used the drug compared with control group.[35,36] This long biological half-life could be the reason for such a low number of confirmed cases and worth further investigation. However, the side effect of chloroquine and hydroxychloroquine, and any possible interaction should be taken into account.[37,38] However, No evidence-based information is known yet for such effects of chloroquine and hydroxychloroquine, and their side effect and any possible interaction should be taken into account.

3.1 | Limitation

We could not divide weather temperature and age groups other than two categories, which was to simplify the model.

Not much evidence supports these factors' effects on the SARS-CoV-2 virus. However, we are trying here to compare the confirmed cases and death cases and examine if these variables have any effect, and suggest further studies to support our findings.

Based on our model, these factors showed some possible effect that requires further studies. The low numbers of cases in Africa and the Arab countries might suffer from reporting bias and the small number of SARS-CoV-2 tests comparing to the country population. The factors that we used are fixed; either present or absent in each country but social distancing is variable and we could not be sure who it was achieved in a different country. Besides, it can be archived, partially with a different percentage or not archived at all. Also, no evidence-based data can be used to relay on like the four factors presented in Figure 1 support socially distancing other than the Denver mortality double-humped curve of the Spanish flu pandemic in 1918 and the global recommendation of social distancing from which we stated that all our results can happen but certain deviation can occur if the social distancing was not archived properly within the country.

4 | CONCLUSIONS

The process of machine learning for time-series prediction and forecasting, especially in the pandemic COVID-19 domain, proved useful in modelling and forecasting the status of the virus spreading based on specific factors affect and health support variables.

In this study, the obtained results from the MLP models have confirmed that the SARS-CoV-2 virus has a very low spread in the African countries because of the existence of the four factors (average young age, hot weather, BCG vaccine and malaria treatment) and has a very high spread in European countries and the USA because of the absence of these factors.

BCG vaccine has a very important role in stimulating the immune system but requires time to do so and help in COVID-19 prevention. Antimalarial-like chloroquine and hydroxychloroquine has a possible role in prophylactic against SARS-CoV-2 infection and transmission and worth further investigation. Age and weather temperature have also a possible effect on the spread, severity and mortality of COVID-19 disease. Lack of social distancing has a very powerful effect on those factors. Further investigations and future researches are required.

AUTHOR CONTRIBUTION
Mohamed E. A. Abdelrahim contributed to conception and design. All authors contributed to administrative support, provision of study materials or patients, collection and assembly of data, data analysis and interpretation, manuscript writing and final approval of the manuscript. Final approval of manuscript: All authors.

DISCLOSURE
The authors declared no conflict of interest.

DATA AVAILABILITY STATEMENT
The datasets analysed during the current study are available from the corresponding author on reasonable request.

ORCID
Haitham Saeed https://orcid.org/0000-0002-2040-9466
Mona Abdelrahman https://orcid.org/0000-0002-8748-8090
Hadeer S. Harb https://orcid.org/0000-0003-2259-9686
Yasmin M. Madney https://orcid.org/0000-0003-4105-7874
Mohamed E. A. Abdelrahim https://orcid.org/0000-0003-0227-8404

REFERENCES
1. WHO, W.h.o. COVID-19 situation-reports. [2020, 6 April 2020]; Available from: https://www.who.int/docs/default-source/coronaviruse/situation-reports/20200416-sitrep-87-covid-19.pdf?sfvrsn=9523115a_2.
2. Baker MG, Kvalsvig A, Verrall AJ, Telfar-Barnard L, Wilson N. New Zealand's elimination strategy for the COVID-19 pandemic and what is required to make it work. *New Zealand Med J (Online).* 2020;133:10-14.
3. Hsu LY, Chia PY, Vasoo S. A midpoint perspective on the COVID-19 pandemic. *Singapore Med J.* 2020;61:381-383.
4. Elgendy MO, El-Gendy AO, Abdelrahim MEA. Public awareness in Egypt about COVID-19 spread in the early phase of the pandemic. *Patient Educ Couns.* 2020;103:2598-2601.
5. El-Gendy AO, Saeed H, Ali AM, et al. Bacillus Calmette-Guérin vaccine, age and gender relation to COVID-19 spread and mortality. *Vaccine.* 2020;38:5564-5568.
6. Xu X-W, Wu XX, Jiang XG, et al. Clinical findings in a group of patients infected with the 2019 novel coronavirus (SARS-Cov-2) outside of Wuhan, China: retrospective case series. *BMJ.* 2020;368:1-7.
7. Huang C, Wang Y, Li X, et al. Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China. *Lancet.* 2020;395:497-506.
8. Anderson RM, Heesterbeek H, Klinkenberg D, Hollingsworth TD. How will country-based mitigation measures influence the course of the COVID-19 epidemic? *Lancet.* 2020;395:931-934.
9. Taravat A, Proud S, Peronaci S, et al. Multilayer perceptron neural networks model for meteosat second generation SEVIRI daytime cloud masking. *Remote Sens.* 2015;7:1529-1539.
10. Popescu M-C, Balas VE, Perescu-Popescu L, Mastorakis N. Multilayer perceptron and neural networks. *WSEAS Trans Circuits Syst.* 2009;8:579-588.
11. CSSE. Johns Hopkins University Center for Systems Science and Engineering (JHU CSSE). 2020 16 April 2020]; Available from: https://raw.githubusercontent.com/CSSEGISandData/COVID-19/master/csse_covid_19_data/csse_covid_19_time_series/time_series_covid19_confirmed_global.csv.
12. ECDC. European Centre for Disease Prevention and Control. 2020 April 16. Available from: https://www.ecdc.europa.eu/en/publications-data/download-todays-data-geographic-distribution-covid-19-cases-worldwide.

13. Zwerling A, Behr MA, Verma A, et al. The BCG World Atlas: a database of global BCG vaccination policies and practices. *PLoS Med*. 2011;8:e1001012.

14. Ricci F. Social implications of malaria and their relationships with poverty. *Mediterr J Hematol Infect Dis*. 2012;4:e2012048.

15. Wikipedia. List of countries by average yearly temperature. 2020, April 20. Available from: https://en.wikipedia.org/wiki/List_of_countries_by_average_yearly_temperature.

16. Wikipedia. List of countries by median age. 2020, April 20. Available from: https://en.wikipedia.org/wiki/List_of_countries_by_median_age.

17. Hardy LJ. Connection, contagion, and COVID-19. *Med Anthropol*. 2020;39:655-659.

18. Dyer O. *Covid-19: Trump Stokes Protests Against Social Distancing Measures*. London: British Medical Journal Publishing Group; 2020;369:m1596.

19. Asfahan S, Chawla G, Dutt N. Ramadan and COVID-19: a challenge amongst challenges. *Turk Thorac J*. 2020;21:285.

20. Andrikopoulos P, Cui Y, Gad S, Kallinterakis V. Feedback trading and the ramadan effect in frontier markets. *Res Int Business Finance*. 2020;51:101085.

21. Elgendy MO, Abd Elmawla MN, Abdel Hamied AM, El Gendy SO, Abdelrahim ME. COVID-19 patients and contacted person awareness about home quarantine instructions. *Int J Clin Pract*. 2020;in press.

22. Markel H, Navarro JA. *To Save Lives, Social Distancing Must Continue Longer Than We Expect*. Washington, DC: The Washington Post; 2020.

23. Ali I, Alharbi OM. COVID-19: Disease, management, treatment, and social impact. *Sci Total Environ*. 2020;728:138861.

24. Johnson BS, Laloraya M. A cytokine super cyclone in COVID-19 patients with risk factors: the therapeutic potential of BCG immunization. *Cytokine Growth Factor Rev*. 2020;54:32-42.

25. Shann F. Nonspecific effects of vaccines and the reduction of mortality in children. *Clin Ther*. 2013;35:109-114.

26. Netea MG, Quintin J, Van Der Meer JW. Trained immunity: a memory for innate host defense. *Cell Host Microbe*. 2011;9:355-361.

27. Bassat Q, Moncunill G, Dobaño C. Making sense of emerging evidence on the non-specific effects of the BCG vaccine on malaria risk and neonatal mortality. *BMJ Specialist J*. 2020;5:e002301.

28. Osama El-Gendy A, Saeed H, Ali AM, et al. Bacillus Calmette-Guérin vaccine, antimalarial, age and gender relation to COVID-19 spread and mortality. *Vaccine*. 2020;38:5564-5568.

29. Boehm U, Klamp T, Groot M, Howard JC. Cellular responses to interferon-γ. *Annu Rev Immunol*. 1997;15:749-795.

30. Jensen KJ, Larsen N, Biering-Sørensen S, et al. Heterologous immunological effects of early BCG vaccination in low-birth-weight infants in Guinea-Bissau: a randomized-controlled trial. *J Infect Dis*. 2015;211:956-967.

31. Smith SG, Kleinnijenhuis J, Netea MG, Dockrell HM. Whole blood profiling of bacillus Calmette–Guérin-induced trained innate immunity in infants identifies epidermal growth factor, IL-6, platelet-derived growth factor-AB/BB, and natural killer cell activation. *Front Immunol*. 2017;8:644.

32. Kleinnijenhuis J, Quintin J, Preijers F, et al. Bacille Calmette-Guerin induces NOD2-dependent nonspecific protection from reinfection via epigenetic reprogramming of monocytes. *Proc Natl Acad Sci*. 2012;109:17537-17542.

33. Arts RJ, Moorlag SJ, Novakovic B, et al. BCG vaccination protects against experimental viral infection in humans through the induction of cytokines associated with trained immunity. *Cell Host Microbe*. 2018;23:89-100.e5.

34. Tett S, Cutler DJ, Day RO, Brown KF. Bioavailability of hydroxychloroquine tablets in healthy volunteers. *Br J Clin Pharmacol*. 1989;27:771-779.

35. Wang M, Cao R, Zhang L, et al. Remdesivir and chloroquine effectively inhibit the recently emerged novel coronavirus (2019-nCoV) in vitro. *Cell Res*. 2020;30:269-271.

36. Chowdhury MS, Rathod J, Gernsheimer J. A rapid systematic review of clinical trials utilizing chloroquine and hydroxychloroquine as a treatment for COVID-19. *Acad Emerg Med*. 2020 Jun;27:493-504.

37. Lopez-Izquierdo A, Ponce-Balbuena D, Ferrer T, et al. Chloroquine blocks a mutant Kir2. 1 channel responsible for short QT syndrome and normalizes repolarization properties in silico. *Cell Physiol Biochem*. 2009;24:153-160.

38. Vereckei A, Fazakas Á, Baló T, et al. Chloroquine cardiotoxicity mimicking connective tissue disease heart involvement. *Immunopharmacol Immunotoxicol*. 2013;35:304-306.