*Revised Manuscript (unmarked)

A fusion of data science and feed-forward neural network-based modelling of COVID-19 outbreak forecasting in IRAQ

Ahmed J. Aljaaf[a,c,*], Thakir M. Mohsin[b], Dhiya Al-Jumeily[c] and Mohamed Alloghani[d]

[a] Centre of Computer, University of Anbar, Ramadi, IRAQ
[b] College of Medicine, University of Anbar, Ramadi, IRAQ
[c] Faculty of Engineering and Technology, Liverpool John Moores University, Liverpool, UK
[d] Artificial Intelligence Office, Dubai, UAE

ARTICLE INFO

Keywords:
COVID-19 outbreak
Forecasting
Neural networks
IRAQ

ABSTRACT

Background: Iraq is among the countries affected by the COVID-19 pandemic. As of 2 August 2020, 129,151 COVID-19 cases were confirmed, including 91,949 recovered cases and 4,867 deaths. After the announcement of lockdown in early April 2020, situation in Iraq was getting steady until late May 2020, when daily COVID-19 infections have raised suddenly due to gradual easing of lockdown restrictions. In this context, it is important to develop a forecasting model to evaluate the COVID-19 outbreak in Iraq and so to guide future health policy.
Methods: COVID-19 lag data were made available by the University of Anbar through their on-line analytical platform (https://www.uoanbar.edu.iq/covid/), engaged with the day-to-day figures form the Iraqi health authorities. 154 days of patient data were provided covering the period from 2 March 2020 to 2 August 2020. An ensemble of feed-forward neural networks has been adopted to forecast COVID-19 outbreak in Iraq. Also, this study highlight some key questions about this pandemic using data analytics.
Results: Forecasting were achieved with accuracy of 87.6% for daily infections, 82.4% for daily recovered cases, and 84.3% for daily deaths. It is anticipated that COVID-19 infections in Iraq will reach about 308,996 cases by the end of September 2020, including 228,551 to recover and 9,477 deaths.
Conclusion: The applications of artificial neural networks supported by advanced data analytics represent a promising solution through which to realise intelligent solutions, enabling the space of analytical operations to drive a national health policy to contain COVID-19 pandemic.

1. Introduction

Since the emerging of novel coronavirus, i.e. COVID-19, in the Wuhan city of China late December 2019, it has rapidly spread to all over the world and become a global pandemic with about 18 million confirmed cases worldwide, while the global death toll reaching 692,000 cases by end of July 2020. More than 190 countries have been affected with COVID-19 pandemic, with major outbreaks in Italy, Spain, China, France, United States of America, and regionally Iran, and Saudi Arabia.

The first COVID-19 case has been confirmed in Iraq was for a foreign student who enrolled in the religious school of Al-Najaf holy city in 24 February 2020. By the next day, another four COVID-19 cases were confirmed for a family returned from Iran to their homeland in Kirkuk province, northern Iraq. Thereafter, the first COVID-19 cases were confirmed in the capital city of Baghdad and precisely in Al-Rusafa region.

Current evidence has shown that COVID-19 can transmit just like other respiratory infections through droplets and contact routes [1, 2]. The global community has started to adopt various measures to handle this pandemic in line with their capacity and healthcare systems. Iraqi medical authorities have started to carry out COVID-19 tests since the detection of the earliest case on 24 February. According to the world health organisation (WHO)[1], the mean period between exposing to COVID-19 and symptom onset is 5-6 days, although some cases can take up to 14 days. The basic reproduction number has been widely used as contagiousness indicator of the COVID-19 virus. It is representing the mean number of newly infected individuals who got the virus from an infectious person in an entirely naive population. The number of infected individuals is highly likely to increase when $R_0 > 1$, while transmission is probably to fade away when $R_0 < 1$. A study by Ying and his colleagues has revealed that the mean R_0 of COVID-19 was around 3.28 [3], which exceeds the WHO estimates that stated the R_0 for COVID-19 to be in between 2.0 and 2.5 [4].

There are still many vaguest about COVID-19. We cannot precisely and definitively predict the outcomes of this pandemic in a general term. People are in fear and worry about their own health while this pandemic is getting more stressful. The economic risks of this pandemic are also not insignificant. Most countries are in some sort of lockdown while production lines and supply chains have also stalled both nationally and internationally. It is been anticipated that small and medium-sized enterprises with partial operating flexibility would deplete its holdings cash in about two years [5]. In addition to political and security crises, Iraqi government is experiencing great challenges by losing half of its financial revenues, with oil prices

*Corresponding author
A.J.Aljaaf@uoanbar.edu.iq (A.J. Aljaaf)
ORCID(s):

[1] https://covid19.who.int/

dropping below $40 per barrel due to COVID-19 outbreak. This might aggravate the risk of public desperation and renewed social unrest if the government has failed to demonstrate its commitments to manage this pandemic with its socio-economic consequences.

The lack of evidence complicates the design of appropriate national response policies. Although we do not know what is going to occur, with artificial intelligence we have a better opportunity to forecast the near future of this pandemic with unprecedented accuracy. Artificial intelligence methods is required not only to develop forecasting models, but also to determine which pathway we shall follow. While data analytics can help visualising data and place raw numbers into perspective, neural networks can be fitted to COVID-19 lag data to forecast the future of this pandemic. Results from these methods will inform, along with big data integration and analysis, decision makers how to combat the spread of this pandemic [6].

Since we only count those with a lab-confirmed infection, which maybe raises the hypothesis that our counts relay on the number of actual performed test. A total of 1,027,620 diagnostic tests were done until 2 August 2020 for a population of over 40 million, which corresponds to the testing of almost 25,690 in a million. As presented in figure 1, medical authorities in Iraq have dramatically increased COVID-19 testing to reach an average of 14,689 tests per day in July 2020 covering all four corners of the country. When comparing lab-confirmed infections to the performed COVID-19 testings, a strong positive correlation has been defined with 95% confidence interval (R = 0.91, p-value < 0.001). This clearly suggests that increasing daily COVID-19 testings would results in more COVID-19 cases in Iraq.

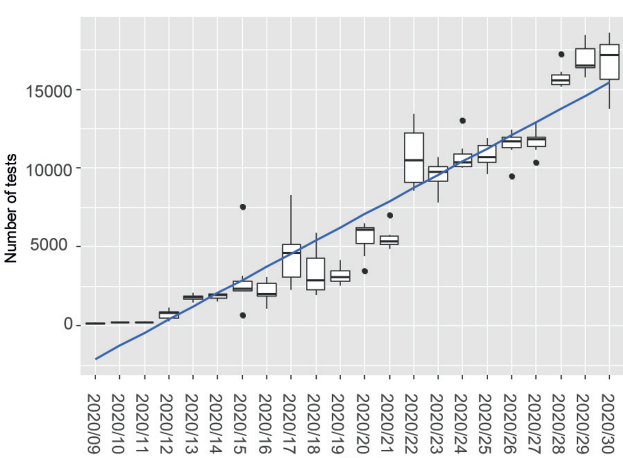

Figure 1: Number of performed COVID-19 tests on a weekly basis

As COVID-19 pandemic continues to propagate worldwide, this study presents the results of COVID-19 outbreak forecasting in Iraq using an ensemble of feed-forward neural networks. A novel combination of linear and nonlinear activation functions has been employed to process COVID-19 lagged values through networks' layers in a homogeneous ensemble of neural networks. Also, a bench-marking analysis was achieved to compare the relative forecasting of COVID-19 daily infections with two well-known time series forecasting methods; namely the Auto-regressive Integrated Moving Average (ARIMA) and the Exponential Smoothing State Space model (ETS).

2. Neural networks as a forecasting model

2.1. Background

The interest in feed-forward neural network has dramatically increased to became central to much of the advanced works on time series forecasting and beyond [7–11]. Feed-forward neural network consists of a several interconnected neurons with linear or nonlinear transfer functions that can be adopted for univariate time series forecasting or nonlinear signals such as speech and images. Feed-forward neural network can approximate reasonable functions to any required degree of accuracy using only two layers of neurons [12]. Further to this, neural networks can learn by adaptively adjusting their synaptic weights within a supervised and unsupervised learning settings. Ideally, the learning process will adjust the weights of the network to map a given input pattern into desired output.

COVID-19 outbreak forecasting, including infectious, deaths, and recovered, has been achieved in this study using an ensemble of feed-forward neural networks with random starting values. Each of the networks was a 10-6-1 architecture with a linear output unit and 73 weights options. COVID-19 lagged values of infectious, deaths, and recovered have separately fed to the network in an ensemble learning paradigm. A feed-forward network is a non-recurrent network which consists of inputs, hidden, and outputs layers as presented in figure 2. In this type of neural network, the data can only pass in one direction and onto a layer of processing units [11, 13, 14]. Each processing unit perform its computation based on a weighted sum of its inputs. The output of a processing unit then become the new input values that passed to the next layer. This process continues through all the layers until the network determines the output.

Consider a feed-forward neural network of three layers. An index k will be used to reference the input units, index i represents the output units, and the hidden neurons will be referred by the index j. We have fed the lagged values of COVID-19 daily infections in Iraq to our feed-forward neural networks as inputs and a single hidden layer with 6 nodes. A total of 20 networks were fitted with random starting weights, and then averaged to forecast the upcoming COVID-19 spread in Iraq, including new infections, survivors and deaths.

Let M refer to the number of inputs, N to the number of outputs, and S to the number of hidden units. Let x refers to the M-tuple inputs to the network, and y denotes the N-

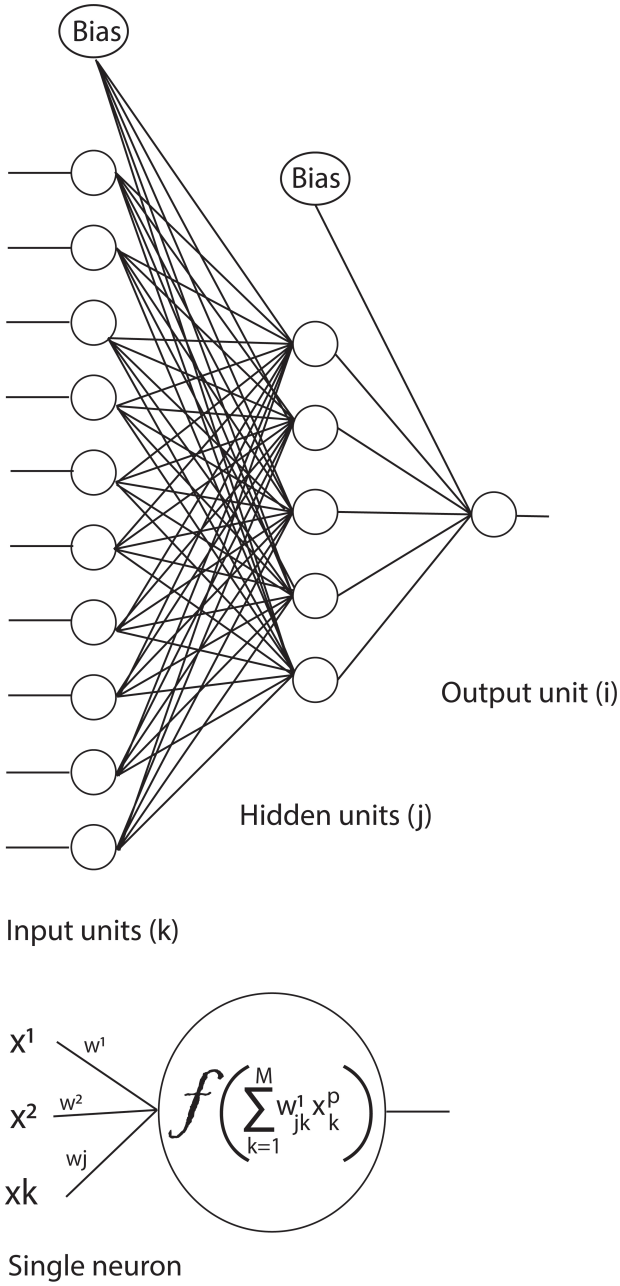

Figure 2: Neural network architecture

tuple outputs of the output layer. There were two sets of weights matrices in our feed-forward neural network. The first weight matrix has $S * M$ elements and represented as W^1_{jk}, while the weight second matrix represented as W^2_{ij} and contains $N * S$ elements. The biases of the feed-forward neural network can either be represented independently or via an extra input line of value for each layer in the network [8, 13]. When the lagged values are passed to the network, the hidden neuron j receives a net input nj denoted as shown in equation 1, and the output of this hidden neuron will be represented as shown in equation 2, where f is a nonlinear transfer function.

$$n_j^p = \sum_{k=1}^{M} w_{jk}^1 x_k^p \tag{1}$$

$$V_j^p = f\left(n_j^p\right) = f\left(\sum_{k=1}^{M} w_{jk}^1 x_k^p\right) \tag{2}$$

The output layer receives its inputs form the hidden layer in our case. Therefore, the net input to the output unit i can be represented in equation 3, while and the output unit i produces an output value as indicated in equation 4.

$$n_i^p = \sum_{j=1}^{S} w_{ij}^2 V_j^p = \sum_{j=1}^{S} w_{ij}^2 f\left(\sum_{k=1}^{M} w_{jk}^1 x_k^p\right) \tag{3}$$

$$y_j^p = f\left(\sum_{j=1}^{S} w_{ij}^2 V_j^p\right) = f\left(\sum_{j=1}^{S} w_{ij}^2 f\left(\sum_{k=1}^{M} w_{jk}^1 x_k^p\right)\right) \tag{4}$$

It is imperative to note that the transfer function at the out layer is a linear transfer function, which is different from what it is in the hidden layer. The cost function or the overall network error can be depicted by the following equation, where $e_i^p = d_i^p - y_i^p$ and d_i^p refers to the target value of the output unit i while the lagged values (i.e. COVID-19 daily infections) are represented to the input layer of the network. Further to this, when substituting the output y, the cost function becomes as presented in equation 6 [7, 15].

$$J = \frac{1}{2} \sum_{i=1}^{N} \left[e_i^p\right]^2 = \frac{1}{2} \sum_{i=1}^{N} \left[d_i^p - y_i^p\right]^2 \tag{5}$$

$$J = \frac{1}{2} \sum_{i=1}^{N} \left[d_i^p - f\left(\sum_{j=1}^{S} w_{ij}^2 f\left(\sum_{k=1}^{M} w_{jk}^1 x_k^p\right)\right)\right]^2 \tag{6}$$

The change in the weights that link a hidden unite to an output unite can be calculated using equation 7, where η is the learning rate and $\frac{\partial J}{\partial w_{ij}^2} = -\left(d_i^p - y_i^p\right) f'\left(n_i^p\right) V_j^p$.

$$\Delta \mathrm{w}_{ij}^2 = -\eta \frac{\partial J}{\partial w_{ij}^2} \tag{7}$$

The algorithm starts by feeding the lagged values of COVID-19 daily infections to all input nodes and initialising the weights and biases of the network to small random values. Then, the network calculates sum of hidden node values and using a nonlinear activation function, it calculates hidden node activation values. These outputs are then used as inputs to the output unites, which in our case uses a liner activation function to calculate activation values of the output unite. This is the batch training in which the weights are only updated when all the lagged values of COVID-19 daily infections are presented to the network. A back-propagation training algorithm could also be used to propagate the error signals backwards to update the weights matrix as reported in our previous work [12].

2.2. COVID-19 time series

The data were made available by University of Anbar (https://www.uoanbar.edu.iq/covid/), through their online analytical platform, engaged with the day-to-day COVID-19 figures form the Iraqi medical authorities, to raise public awareness of COVID-19 outbreak on a national level. Past 154 days of patients data were provided in a csv file format. Data has been imported to R and converted into time series object as a daily number of lab-confirmed COVID-19 cases in Iraq starting from 24 February 2020 to 2 August 2020, as presented in appendix A.

A total of 129,151 COVID-19 cases over 154 days were considered to train our forecasting model using using ensemble learning approach. As presented in figure 8, the number of COVID-19 infections has slightly dropped in mid-April 2020 due to closing main airports and banning the entry of travellers' form neighbour countries. After that in late May 2020, the daily registered COVID-19 infections have increased dramatically following the mitigate of lockdown restrictions in some parts of Iraq including Kurdish region. Figure 3 reveals that the trend of COVID-19 infections registered at Iraqi hospitals was going upwards while the exponential growth curve of COVID-19 infections has appeared to originate in the first two weeks of July, which means the virus has started to spread through people, especially in incubation period. In this context, we have conducted forecasting analysis of COVID-19 cases using an ensemble of feed-forward neural network, assuming the same trend and course of action has continued for the next two months, i.e. August and September 2020.

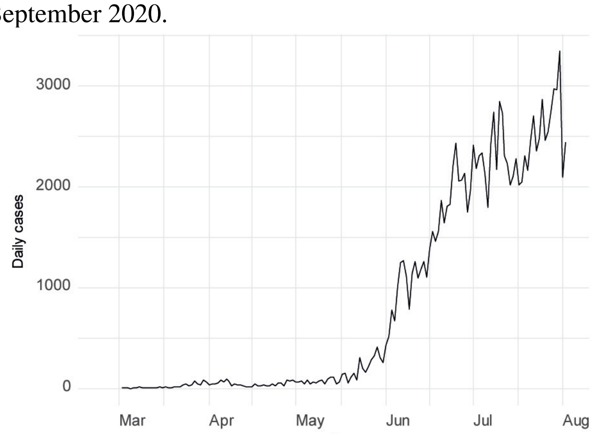

Figure 3: Daily lab-confirmed COVID-19 cases in Iraq

2.3. Model development using R

R is a language and environment for statistical computing with a pertinent importance for epidemiologists. R language provides a rapid access to many libraries which dedicated to outbreak management and analysis. To develop a forecasting model for the COVID-19 outbreak in Iraq, we have imported "forecast" packages in R [16]. This package provides tools and methods for analysing univariate time series data including exponential smoothing, automatic ARIMA modelling, and feed-forward neural networks. Using forecast package, we have fitted 20 neural networks and have averaged their outcomes in the ensemble learning setting. A lagged version of COVID-19 time series has been normalised to [0,1] to fit in for the training phase.

The neural network model was set to forecast two-months ahead of COVID-19 infections in Iraq along with survivors and the likely deaths led by this pandemic. The number of non-seasonal lags has been specified to 10, which means that the neural network function would include the $lag_1, lag_2, ..., lag_{10}$ of the series as inputs. Since we have fitted a neural network with a single hidden layer, we have defined the number of nodes in this layer to be half of the number of input nodes plus one. A logit activation function has been employed in our neural network forecasting model to map the lags from input nodes to the hidden nodes. Using linear activation function, neural network maps the forecasting values from the output unite. To this point, we end up with a 10-6-1 network architecture and 73 connection weights. The complete R script for developing a neural network forecasting model is presented in appendix A.

2.4. COVID-19 outbreak forecasting

To fit a neural network forecasting model for COVID-19 daily infections, recovered and deaths, we have performed the R script shown in Appendix A. The probability of new COVID-19 positive cases, recovered cases, and daily deaths on a nationwide scale for the next 60 days, i.e. August and September 2020, based on the lag data were forecasted and presented in figure 4.

After the sudden decrease in registered daily infections late July, the forecasting model shows that the number of people who are anticipated to get COVID-19 infection in Iraq will be gradually increases over the next two months. It is anticipated that daily infections to reach its peak late August with a possibility of registering 4200 people with COVID-19 per day. The pattern of daily infections appear to recurring over the next 8 weeks but with higher numbers. Over this period, there will be a total of 179,845 ($mean = 2996, SD = 438$) people who are expected to test positive for COVID-19 in Iraq. This would raise the total number of COVID-19 cases to 308,996 by end of September 2020, assuming the same trend and course of action has continued by the Iraqi health authorities.

Part B of figure 4 represents an ensemble of neural networks to forecast the daily deaths attributed to

COVID-19 in Iraq over the next two months, i.e. Augus and September 2020. The forecasting model has als anticipated that number of deaths due to COVID-19 wil drop in early August and will dramatically increase again t reach its peak in early September. The number of death due to COVID-19 will then slightly declining in mi September 2020. An ensemble of feed forward neura networks has predicted that the mean number of dail deaths over the coming 60 days is about 78 cases per da on a nationwide level (min=51, max=115). By the end c September 2020, it is expected that a total of 4610 case t die due to COVID-19 in Iraq, which in turn increas COVID-19 victims to about 9,477 on a nationwide scale.

Part C of figure 4 neither follow the forecasting patter of daily infections nor daily deaths. The pattern c anticipated daily recovered seems to fluctuating up an down in an iterative way. This is highly likely attributed t the fact that different regional healthcare centres wer trying various therapeutic protocols as there is no approve treatment for COVID-19 up to the time of writing thi paper. Also such pattern may reflects the variation i patients recovery time, which ranges somewhat between weeks for mild cases and 3-6 weeks for severe or critica cases. It has anticipated that, according to an ensemble c feed forward neural networks, a total of 191,15 $(mean = 2640, SD = 523)$ patients to recover fror COVID-19 in Iraq by the end of September 2020. Thi would boost the total number of survivors to 228,55 assuming the same trend and course of action ha continued by the Iraqi health authorities.

2.5. Residuals analysis and evaluation

ACF plot is a result of an auto-correlation function t explain trend, seasonality and residuals in a time serie data. ACF plots our COVID-19 daily cases along with th confidence band, i.e. the dashed blue lines in the AC figure, to show any possible auto-correlation of COVID-1 time series data with its lagged values. In other words ACF plot describes how well the current value of th COVID-19 series is correlated with its past values.

The ACF of the residuals from the feed forward neura network shows that the produced forecast appear to accoun for all available data of COVID-19 daily cases in Iraq. As we can observe in figure 5, the mean residuals of our forecasting model are close to zero and the residuals series displays with no significant correlation. Apart of few outliers, the time plot of the residuals from the feed forward neural network shows that the variation of the residuals is almost ± 60 for two third of the COVID-19 lag values, while residuals were slightly skewed up or down across the rest of the data. This can also be demonstrated on the histogram of the residuals. When ignoring these outliers, i.e. the right and left tails, the histogram proposes that the residuals are almost normal. Therefore, the 60-days forecasting of COVID-19 outbreaks in Iraq using an ensemble of feed forward neural networks is quite reasonable.

Along with ACF, we have also assessed our forecasting model using prediction errors. Based on the feed forward neural network model accuracy evolution of COVID-19 pandemic outbreaks in Iraq, and using the lag data, we have considered the mean absolute prediction error (MAPE) parameter using the following equation.

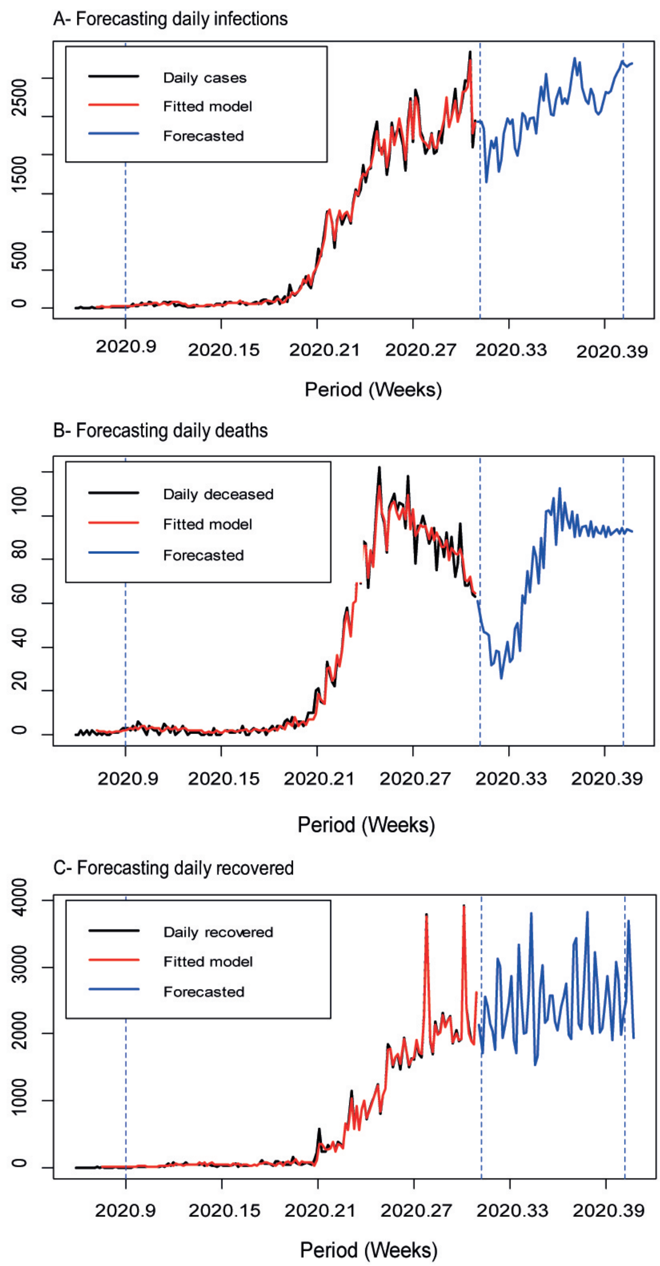

Figure 4: COVID-19 outbreak forecasting using an ensemble of neural networks

$$Accuracy = (100 - MAPE) * 100 \tag{8}$$

Our forecasting model using an ensemble of 10-6-1 feed forward neural networks has been validated with a prediction accuracy of 87.6% for daily infections, 82.4% for

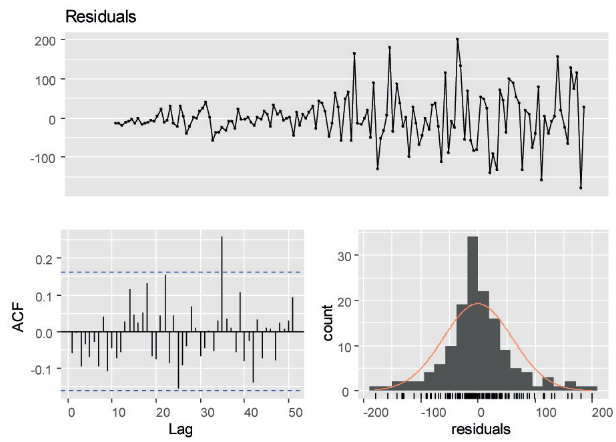

Figure 5: ACF and residuals of the neural networks

daily recovered, and 84.3% for daily deaths.

2.6. Benchmarking analysis and evaluation

A benchmarking analysis was carried out for further evaluation of the ensemble of neural networks by comparing the relative forecasting of COVID-19 daily infections with two popular time series forecasting approaches; namely Auto-regressive Integrated Moving Average (ARIMA) and Exponential Smoothing State Space model (ETS). A country wide probability of new COVID-19 positive cases in the coming 60 days using ARIMA and ETS models were forecasted and presented in figure 6.

ARIMA model is typically specified by three structured parameters, namely (p, d, q); where p refers to the lag values, d refers to the degree of difference, and q denotes the model error [16]. When combining these parameters in a non-seasonal settings, ARIMA model can be expressed linearly in equation 9, where $\phi(Y)(1 - \phi_1 Y - \cdots - \phi_p Y^p)$ and $\theta(Y) = \left(1 + \theta_1 Y + \cdots + \theta_q Y^q\right)$ are the polynomial order that respectively expressing p^{th} and q^{th} in Y.

$$\phi(Y)(1 - Y)^d y_t = c + \theta(Y)\varepsilon_t \quad (9)$$

While considering differences in observations along with residual errors at prior timestamp, ARIMA forecasts the upcoming COVID-19 infections as a sequence of linear function. It is commonly known that ARIMA models are more predictive than regressive models especially in univariate time series without seasonal components. A total of 13 ARIMA models were fitted using approximations to figure out the best performing model, which was ARIMA (2,1,1) with drift of 2029.235 and 31 lagged values. It is evident from the 60 days forecasting using ARIMA model that COVID-19 infections are anticipated to increase between 174,465 and 421,555 at 95% forecast interval.

On the other hand, ETS model forecasts 60-days of COVID-19 infections using the iteration of equation 10.

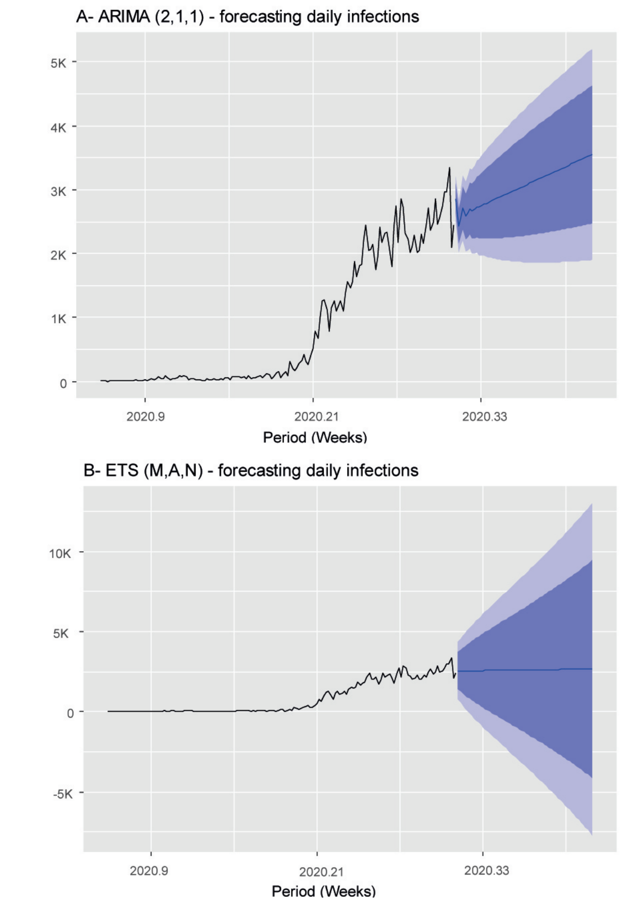

Figure 6: Forecasting of COVID-19 daily infections using ARIMA and ETS models

For $ETS(M, A, N)$, $\hat{y}_{T+1|T} = \iota_T + b_T$ and $y_{T+2} = \left(\iota_{T+1} + b_{T+1}\right)\left(1 + \varepsilon_{T+2}\right)$, and so on.

$$ETS(M, A, N), y_T + 1 = \left(\iota_T + b_T\right)\left(1 + \varepsilon_{T+1}\right) \quad (10)$$

ETS model forecasts upcoming COVID-19 infections as a weighted averages of lagged observations while recent observations were gained extra weight. As presented in figure 6-B, ETS model appears to achieve a bad forecast as the trend has moved opposite to that in the COVID-19 time series. At 95% forecast interval, ETS model has anticipated that COVID-19 infections over the next 60-days would range between -206,132 and 519,856. This was due to the relatively high smoothing parameters and the existence of outliers.

Table 1
Comparison of neural networks, ARIMA and ETS models.

	RMSE	MAE	MPE	MAPE
Neural networks	32.23	22.21	-9.86	12.4
ARIMA (2,1,1)	176.7	100.8	-67.1	75.06
ETS (M,A,N)	186.6	105.9	-7.4	27.48

Table 1 provides a head to head comparison of our three forecasting models. Although ARIMA model fits the univariate time series of COVID-19 daily infections slightly better than the ETS model, but the ensemble of neural networks provides more accurate forecasts with lower overall errors. It is clear that the ensemble of neural networks have registered considerably lower root mean squared error (RMSE), mean absolute error (MAE), and the mean absolute percentage error (MAPE).

While the ensemble of neural networks was superior to ARIMA and ETS models for predicting 60-days of COVID-19 infections, the resulting model is often a black box and hard to interpret. ARIMA and ETS models are analogous to the typical linear regression, but without considering exogenous variables. Such forecasting models allow a straightforward calculation of their coefficients, which is intractable for neural networks because of their large connection weights. An ensemble of 20 neural networks with 73 connection weights each result in a weights matrix of 1,460 parameters. While estimating a huge number of connection weights is computationally expensive compared to ARIMA and ETS, the neural networks are yet widely adopted to solve challenging problems such as analysing genomics data [17], diagnosing of skin cancer [18], inspecting of radio astronomy [19], analysing video footage for self-driving cars [20], and other endless fields.

The wide adoption of artificial neural networks to solve challenging real-world problems is attributed to the fact that they possess lower bound of generalisation error when compared to other forecasting models including ARIMA and ETS. Although neural networks can model nonlinear complex relationships using their hidden nodes, grasp how they operate is somewhat unsatisfying especially to validate medical applications. While artificial neural networks has already conveyed revolutionary impacts to countless real-world applications, they represent promising solution through which to develop forecasting model rooted in epidemiology and predict infections, recovery, and deaths from COVID-19 pandemic and beyond.

3. Conclusion and recommendations

COVID-19 belongs to the family of coronavirus that cause a wide range of illnesses starting from the common cold to severe pneumonia such as the severe acute respiratory syndrome (SARS) and the Middle East respiratory syndrome (MERS) [21, 22]. Although epidemiological and clinical features of patients with COVID-19 have been reported, many aspects of this pandemic are not completely understood. Forecasting the outbreak of this pandemic at a national scale using an ensemble of feed-forward neural networks were conducted and compared to other forecasting models in a competitive benchmarking analysis. As of 2 August, 129,151 COVID-19 infections were confirmed in Iraq (4867 deaths). Daily infection data shows a spike in the curve from late May onward.

COVID-19 mortality rate in Iraq has jumped from 2.56% to 3.97% in just 20 days, i.e. between 10 and 30 June 2020. This substantial increase comes in line with the increase of daily infections (R=0.95, p-value<0.001). While current COVID-19 testing in Iraq involves IgM/IgG Rapid test, real-time reverse transcription polymerase chain reaction (rRT-PCR), and computed tomography scan (CT) for the highly suspected, asymptomatic cases, these screening tests must be conducted by a qualified laboratory scientist and quality control protocols should be followed to ensure a precise result. An ensemble of neural networks can effectively help defining COVID-19 form CT scans, which in turn minimise the pressure on medical doctors while increase the capacity of current health service model.

While the feed-forward neural network approach is straightforward to implement and typically no special assumptions are required to start, we have applied a novel combination of linear and nonlinear activation functions to process COVID-19 lagged values through networks' layers in a homogeneous ensemble of neural networks. Such ensemble of neural networks can learn and generalise to produce reasonable outputs for inputs that are not encountered in the learning phase. Such information processing ability allows ensemble of neural networks to produce valid solutions to complex problems that would take a very long time for a human to do. Therefore, neural networks, especially in the ensemble learning paradigm, are widely adopted as a research tool to address complicated problems such as for understanding neurological phenomena's, forecasting financial time series, natural language processing, and the most recent of which is the deep learning approach for the classification of medical images and self-driven cars assistant. Further to this, neural networks have been taken to newer levels of challenges, especially the hardware design techniques, as having a high chance of being fault tolerance due to their reliability when it comes to losing a significant portion of a network.

While quantifying uncertainty is a crucial aspect for practical forecasting applications like the one in this study, neural networks tend to be poor at this aspect and typically overconfident predictions when lagged inputs are noisy or uncertain. Several methods such as Monte-Carlo (MC) sampling method and Bayesian inference to reformatting neural networks parameters were commonly used to estimate uncertainty. The MC sampling maintains dropout at testing time to prevent over-fitting by ignoring random subset of neurons in a network layer with every batch evaluation. Our ensemble of neural networks is approximating the MC sampling methods since every network iteration starts with a random subset of lagged inputs. Thus, such approximating behaviour can be leveraged while training multiple forecasting models with random starting points for each of the iterations to improve generalisation process. Further to this, sufficient data points of 154 days were provided to the network for an adequate precision to balance random noise component that may, possibly, accompanying the signal. Many other factors can also influence generalisation power of neural networks including the complexity of mapping function and the number of hidden units. Too- small or -large number of hidden units is highly likely to produce poor generalisation for data sample outside the training set, in which the network may fail to thoroughly determine the signal in complex data set and cause over-fitting or under-fitting.

According to our analysis, increasing the number of daily screening being carried out in Iraq would highly likely to rise the number of COVID-19 infections. In other words, the number of daily COVID-19 infections is not only influenced by how quickly the virus spreads, but also the ability of the Iraqi health authorities to provide the test to everyone who presents with COVID-19 concerns. It is also anticipated that the total number of COVID-19 infections in Iraq to reach approximately 308,996 cases by the end of September 2020, assuming the same trend and course of action has continued by the Iraqi health authorities. At present, the Iraqi health authorities begins easing lockdown measures and it is expected that life will return to normal while maintaining social distancing and closures of schools, colleges, and universities. It is imperative to mention that another outbreak is inescapable once lockdown restrictions are removed and the anticipated number of COVID-19 infections are most likely to double by the end of September 2020.

The number of recovered patients from COVID-19 are expected to reach a total of 228,551 by the end of September 2020, while the average daily number of deaths due to COVID-19 is expected to be 78 cases nationwide. Not to mention that, numbers of daily survivors and deaths due to COVID-19 will also influenced by governments' lockdown restrictions, and current predictions have been derived while assuming the same trend and course of action has continued over the coming 60 days, i.e. August and September 2020.

Although Iraqi health authorities are issuing regular official statements to update public and diminish their fears, however this information is currently inadequate to allow for a rapid assessment of the country's emergency preparedness and evidence-based medicine. Currently, there is no system to track or follow up patients on self-quarantine and accessing all the potential points of transmission along the chain seems currently impossible. While the cross-border control was essential to prevent this pandemic, contact history of visitors or patient are currently not accessible. The lack of national health system to cover all Iraqi citizens has resulted in missing national big-data integration and analysis. And therefore, misinformation and spreading panic about the COVID-19 pandemic are being widely shared on social media [23]. To sum up, main hospitals should work in a one collaborative emergency network and exploit innovative applications of artificial intelligence to help contain COVID-19 pandemic.

Acknowledgements

This work received no funding. The authors would like to appreciate `www.en.siha.info` to provide lag data of COVID-19 daily infections, recovered, and deceased cases in Iraq. The authors want to publicly thank our healthcare workers and local doctors for working in the frontline of this global pandemic.

Conflict of interest

The authors have no conflicts of interest.

A. Appendix

```
# Importing the data
data <- read.csv("C:/Covid19/
daily-infection.csv",
na.strings=c(".", "NA", "", "?"),
strip.white=TRUE, encoding="UTF-8")
# Descriptive statistics
library(plyr)
Des <- ddply(data, c("Gender"),
summarise, N = length(Age),
min = min(Age), max = max(Age),
mean = round(mean(Age),2),
sd = round(sd(Age),2),
median = median(Age))
# Plotting example
ggplot(data, aes(weeks, tests))+
geom_boxplot()+
geom_smooth(method = "lm",
se=FALSE, aes(group=1))+
xlab("Weeks")+
ylab("Number␣of␣tests")+
theme(axis.text.x = element_text
(angle = -90, vjust = 0.5))+
theme(axis.text=element_text(size=12))
# Pearson's correlation
cor.test(data$tests,
data$cases, method="pearson")
library("ggpubr")
ggscatter(data, x = "cases",
y = "tests", add = "reg.line",
conf.int = TRUE, cor.coef = TRUE,
cor.method = "pearson",
xlab = "Daily␣confirmed␣cases",
ylab = "Daily␣test")
#Installing forecast package
install.packages('forecast',
dependencies = TRUE)
library(forecast)
# Time series formatting
data$Date <- as.Date(data$Date,
format = "%d-%m-%Y")
daily_cases <- ts(data[,2],
start = c(2020,55), frequency = 366)
daily_recovered <- ts(data[,5],
start = c(2020,55), frequency = 366)
daily_deaths <- ts(data[,4],
start = c(2020,55), frequency = 366)
# Fitting a neural network
```

```r
set.seed(1985)
covid.nnetar <- nnetar(daily_cases,
repeats = 20, p=10,
P = 1, size=6,
lambda="auto", scale.inputs=TRUE)
# Print model summary
summary(covid.nnetar$model[[1]])
# Forecasting upcoming 60 days
covid.fcast <- forecast(covid.nnetar,
h=60, level = c(80, 95))
# Print forecasts summary
print(summary(covid.fcast))
# Plotting of figure 9-A
plot(covid.fcast, col="black",
ylab="COVID-19␣daily␣cases␣in␣IRAQ",
xlab="Period", lwd=2, lty=1)+
lines(covid.fcast$fitted,
lwd=2, col="red")
abline(v=2020.2, col="blue", lty=2)
abline(v=2020.67, col="blue", lty=2)
abline(v=2020.82, col="blue", lty=2)
legend(x= "topleft", inset=.02,
legend=c("Daily␣cases",
"Fitted␣model", "Forecasted"),
col=c("black", "red", "blue"),
lty = c(1, 1, 1), lwd = 2, cex=0.8)
# Measure performance
checkresiduals(covid.fcast)
plot(covid.fcast$residuals,
ylab="Residuals", xlab="Period")
abline(h=30, col="blue", lty=2)
abline(h=-30, col="blue", lty=2)
# Fitting ARIMA model
fit_arima <- auto.arima(daily_cases,
ic="aic", trace = TRUE)
fcast_arima <- forecast(fit_arima, h=60,
level = c(80, 95))
autoplot(fcast_arima)
print(summary(fcast_arima))
# Fitting ETS model
fit_ets <- ets(daily_cases)
fcast_ets <- forecast(fit_ets,
h=60, level = c(80, 95))
autoplot(fcast_ets)
print(summary(fcast_ets))
```

CRediT authorship contribution statement

Ahmed J. Aljaaf: Conceptualization, Methodology, Software, Formal analysis, Visualization, Writing - Original Draft. **Thakir M. Mohsin:** Conceptualization, Investigation, Project administration. **Dhiya Al-Jumeily:** Conceptualization, Investigation, Supervision. **Mohamed Alloghani:** Investigation, Data Curation, Writing - Review Editing.

References

[1] Liu Jiaye et al. Community transmission of severe acute respiratory syndrome coronavirus 2, shenzhen, china, 2020. *Emerging Infectious Disease journal*, 26(6):1320, 2020. ISSN 1080-6059. doi: 10.3201/eid2606.200239. URL https://wwwnc.cdc.gov/eid/article/26/6/20-0239_article.

[2] Jasper Fuk-Woo Chan et al. A familial cluster of pneumonia associated with the 2019 novel coronavirus indicating person-to-person transmission: a study of a family cluster. *The Lancet*, 395(10223):514–523, 2020. ISSN 0140-6736. doi: 10.1016/S0140-6736(20)30154-9. URL https://doi.org/10.1016/S0140-6736(20)30154-9.

[3] Ying Liu et al. The reproductive number of covid-19 is higher compared to sars coronavirus. *Journal of Travel Medicine*, 27(2), 2020. ISSN 1708-8305. doi: 10.1093/jtm/taaa021. URL https://doi.org/10.1093/jtm/taaa021.

[4] Nalini Chintalapudi, Gopi Battineni, and Francesco Amenta. Covid-19 virus outbreak forecasting of registered and recovered cases after sixty day lockdown in italy: A data driven model approach. *Journal of Microbiology, Immunology and Infection*, 53(3):396–403, 2020. ISSN 1684-1182. doi: https://doi.org/10.1016/j.jmii.2020.04.004. URL http://www.sciencedirect.com/science/article/pii/S1684118220300980.

[5] Antonio De Vito and Juan-Pedro Gómez. Estimating the covid-19 cash crunch: Global evidence and policy. *Journal of Accounting and Public Policy*, 39(2):106741, 2020. ISSN 0278-4254. doi: https://doi.org/10.1016/j.jaccpubpol.2020.106741. URL http://www.sciencedirect.com/science/article/pii/S0278425420300144.

[6] Fang-Ming Chen et al. Big data integration and analytics to prevent a potential hospital outbreak of covid-19 in taiwan. *Journal of Microbiology, Immunology and Infection*, 2020. ISSN 1684-1182. doi: https://doi.org/10.1016/j.jmii.2020.04.010. URL http://www.sciencedirect.com/science/article/pii/S1684118220301043.

[7] Salah L. Zubaidi et al. A method for predicting long-term municipal water demands under climate change. *Water Resources Management*, 34(3):1265–1279, 2020. ISSN 1573-1650. doi: 10.1007/s11269-020-02500-z. URL https://doi.org/10.1007/s11269-020-02500-z.

[8] Ricardo de A. Araújo et al. A deep increasing–decreasing-linear neural network for financial time series prediction. *Neurocomputing*, 347:59–81, 2019. ISSN 0925-2312. doi: https://doi.org/10.1016/j.neucom.2019.03.017. URL http://www.sciencedirect.com/science/article/pii/S0925231219303194.

[9] Anastasia Borovykh, Cornelis W. Oosterlee, and Sander M. Bohté. Generalization in fully-connected neural networks for time series forecasting. *Journal of Computational Science*, 36:101020, 2019. ISSN 1877-7503. doi: https://doi.org/10.1016/j.jocs.2019.07.007. URL http://www.sciencedirect.com/science/article/pii/S1877750319301838.

[10] Chao Luo et al. An evolving recurrent interval type-2 intuitionistic fuzzy neural network for online learning and time series prediction. *Applied Soft Computing*, 78:150–163, 2019. ISSN 1568-4946. doi: https://doi.org/10.1016/j.asoc.2019.02.032. URL http://www.sciencedirect.com/science/article/pii/S1568494619300961.

[11] Leandro Anghinoni et al. Time series trend detection and forecasting using complex network topology analysis. *Neural Networks*, 117:295–306, 2019. ISSN 0893-6080. doi: https://doi.org/10.1016/j.neunet.2019.05.018. URL http://www.sciencedirect.com/science/article/pii/S0893608019301583.

[12] A. Hussain et al. Backpropagation approach supported by image compression algorithm for the classification of chronic condition diseases. *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, pages 1–1, 2018.

[13] Lin Wang et al. Optimal forecast combination based on neural networks for time series forecasting. *Applied Soft Computing*, 66:1–17, 2018. ISSN 1568-4946. doi: https://doi.org/10.1016/j.asoc.2018.02.004. URL http://www.sciencedirect.com/science/article/pii/S156849461830053X.

[14] Qiegen Liu and Henry Leung. Variable augmented neural network for decolorization and multi-exposure fusion. *Information Fusion*, 46:114–127, 2019. ISSN 1566-2535. doi: https://doi.org/10.1016/j.inffus.2018.05.007. URL http://www.sciencedirect.com/science/article/pii/S1566253517305298.

[15] Adrian Moldovan, Angel Caţaron, and Răzvan Andonie. Learning in

feedforward neural networks accelerated by transfer entropy. *Entropy*, 22(1):102, 2020. doi: https://doi.org/10.3390/e22010102.

[16] Rob Hyndman and Yeasmin Khandakar. Automatic time series forecasting: The forecast package for r. *Journal of Statistical Software, Articles*, 27(3):1–22, 2008. ISSN 1548-7660. doi: 10.18637/jss.v027.i03. URL https://www.jstatsoft.org/v027/i03.

[17] Santos Kumar Baliarsingh, Swati Vipsita, Khan Muhammad, Bodhisattva Dash, and Sambit Bakshi. Analysis of high-dimensional genomic data employing a novel bio-inspired algorithm. *Applied Soft Computing*, 77:520–532, 2019.

[18] Philipp Tschandl, Cliff Rosendahl, Bengu Nisa Akay, Giuseppe Argenziano, Andreas Blum, Ralph P Braun, Horacio Cabo, Jean-Yves Gourhant, Jürgen Kreusch, Aimilios Lallas, et al. Expert-level diagnosis of nonpigmented skin cancer by combined convolutional neural networks. *JAMA dermatology*, 155(1):58–65, 2019.

[19] Michael Mesarcik, Albert-Jan Boonstra, Christiaan Meijer, Walter Jansen, Elena Ranguelova, and Rob V van Nieuwpoort. Deep learning assisted data inspection for radio astronomy. *Monthly Notices of the Royal Astronomical Society*, 2020.

[20] Wael Farag. Cloning safe driving behavior for self-driving cars using convolutional neural networks. *Recent Patents on Computer Science*, 12(2):120–127, 2019.

[21] N. Petrosillo et al. Covid-19, sars and mers: are they closely related? *Clinical Microbiology and Infection*, 26(6):729–734, 2020. ISSN 1198-743X. doi: 10.1016/j.cmi.2020.03.026. URL https://doi.org/10.1016/j.cmi.2020.03.026.

[22] Daolin Tang, Paul Comish, and Rui Kang. The hallmarks of covid-19 disease. *PLOS Pathogens*, 16(5):e1008536, 2020. doi: 10.1371/journal.ppat.1008536. URL https://doi.org/10.1371/journal.ppat.1008536.

[23] Araz Ramazan Ahmad and Hersh Rasool Murad. The impact of social media on panic during the covid-19 pandemic in iraqi kurdistan: Online questionnaire study. *J Med Internet Res*, 22(5):e19556, 2020. ISSN 1438-8871. doi: 10.2196/19556. URL http://www.jmir.org/2020/5/e19556/https://doi.org/10.2196/19556http://www.ncbi.nlm.nih.gov/pubmed/32369026.

Ahmed J. ALjaaf is a senior lecturer of Computational Intelligence in the Centre of Computer at the University of Anbar, Iraq. He received his PhD in data science and machine learning from Liverpool John Moores University, UK, at which He continued as a post-doctoral researcher working on a wide variety interdisciplinary application of machine learning, data science, knowledge discovery, and applied artificial intelligence in health and medicine. Ahmed is a member of the Institute of Electrical and Electronics Engineers (IEEE) and the Applied Computing Research Group, LJMU. He is also associated follow of the UK Higher Education Academy.

Thakir M. Mohsin is an assistant professor of ophthalmology and also the dean of the college of medicine at the University of Anbar, Iraq. He is a member of the Iraqi Ophthalmological Society, the European Society of Cataract and Refractive Surgery, and an international member of the American Academy of Ophthalmology.

Dhiya Al-Jumeily is a professor of Artificial Intelligence at Liverpool John Moores University, UK. He has extensive research interests covering a wide variety of interdisciplinary perspectives concerning the theory and practice of Applied Computing in medicine, human biology, and health care. Dhiya has extensive leadership experience including the Development and Management of the Professional Doctorate programme in Engineering and Technology, a founder and Chair of the IEEE International Conference Series on Developments in eSystems Engineering DeSE (www.dese.org.uk) since 2007. He has one patent and coordinated over 10 projects at national and international level. Prof. Al-Jumeily is a Senior Member of the IEEE and a Chartered IT Professional. He is also a fellow of the UK Higher Education Academy.

Mohamed Alloghani is an advisor of research and capacity development at the UAE's Minister of State for Artificial Intelligence. He has 16 years of working experience in diverse organisations such as public, private and semi-government organisations. He received his PhD in Artificial Intelligence from Liverpool John Moores University. His research interest includes data science, machine learning and mathematical modelling.