PREDICTING THE COVID-19 SPREAD, RECOVERIES AND MORTALITIES RATES IN SAUDI ARABIA USING ANN

Article *in* Journal of Theoretical and Applied Information Technology · December 2020

1 author:

Mohammed Alsuwaiket
University of Hafr Al Batin

Journal of Theoretical and Applied Information Technology
15th December 2020. Vol.98. No 23

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

PREDICTING THE COVID-19 SPREAD, RECOVERIES AND MORTALITIES RATES IN SAUDI ARABIA USING ANN

MOHAMMED A. ALSUWAIKET

Assistant Professor, University of Hafr Al Batin, Computer Science and Engineering Technology Department, Saudi Arabia

E-mail: malsuwaiket@uhb.edu.sa

ABSTRACT

The worldwide pandemic of the COVID-19 has become the main national security issue for almost all nations. The advancement of accurate prediction models provides insights into the spread of this infectious disease. In fact, the high uncertainty and low size of data have caused some epidemiological models that show low accuracy for long-term prediction. Although the related works include many attempts to deal with this issue, the robustness abilities of current models need to be enhanced. In this paper, to achieve the main contribution, a prediction model using Artificial Neural Networks (ANNs) approach is developed based on the COVID-19 data from March 2, 2020 to August 5, 2020 to predict COVID-19 spread rate, recoveries rate, and mortalities rate in Saudi Arabia using Python programming language for the implementation stage and code has been developed to achieve the final results. The evaluation in this paper has conducted through calculating the values for Correlation Coefficient (CC), Mean Absolute Error (MAE), and Mean Square Error (MSE). However, the results are promising by achieving low MAE with average value 0.05 and MSE with average value 0.02, and high Correlation Coefficient for all targets' rates with average value 0.97. Paper further recommends that real novelty in spread prediction can be realized through using other machine learning models with different types of COVID-19 data.

Keywords: *COVID-19, Neural Networks, Artificial Intelligence, Coronavirus, Pandemic.*

1. INTRODUCTION

Coronavirus disease (COVID-19) as named by the World Health Organization (WHO) [1] is a respiratory disease originating from coronavirus firstly occurred in Wuhan City of China in 2019. It was first derived from bats and seafood, transmitted to humans through intermediate hosts maybe the raccoon dog (*Nyctereutes procyonoides*) and palm civet (*Paguma larvata*) [2, 3, 4].In March 2020, COVID-19 was defined as a pandemic by the WHO [5] due to the high spread in the globe.

The main symptoms of COVID-19 include cough, high fever, and shortness of breath which is similar to flu [6]. COVID-19 has considered as pandemic which claimed the lives of a lot of people across the globe and human-to-human transmission of the disease from infected individuals with mild symptoms have been reported [6, 7].

As of August 12, 2020, COVID-19 has affected most the countries around the globe Globally, there have been 20,162,474 confirmed cases of COVID-19, including 737,417 deaths [8]. According to WHO situation reports by the August 12, 2020 [9], the number of people infected with COVID-2019 in European Region is 3,641,603, 10,799,062 (the highest) in Region of the Americas, 383,739 in Western Pacific Region, 1,669,933 in Eastern Mediterranean Region, 2,757,822 in South-East Asia Region, and 909,574 in African Region.

As of August 8, 2020, Saudi Arabia has reported more than 19,342 confirm cases including 18,367 active cases, 4,111 deaths, 16,108 recoveries, and 118 critical cases [10] as shown in Figure 1. At the beginning of COVID-19 spread as of April 2020 in Saudi Arabia, the main cases appear to be in returning travelers and their immediate contacts [11].

Journal of Theoretical and Applied Information Technology
15th December 2020. Vol.98. No 23

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

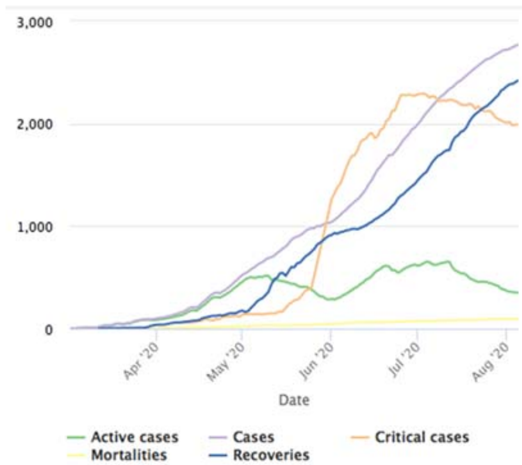

Figure 1: COVID-2019 reported cases in Saudi Arabia [10]

In fact, Saudi Arabia is considered the second largest country in the Arab world with more than thirty-four million in which non-Saudis people represent approximately 37% [12]. Furthermore, most of the population in KSA is in the middle age between 15 and 64 years, while teenagers and elderly make up 32.4% and 2.8%, respectively [12].

During the global pandemic, Saudi Arabia could have a serious impact due to the huge number of visitors during all the year. Muslims from all the world visit KSA for Umrah, which is an Islamic ritual performed by thousands of Muslims every day in the city of Makkah, also more than two million Muslims come for Haj, which is an annual pilgrimage in Makkah that lasts for five days. The pilgrims are visiting Saudi Arabia from all the world, which makes the country hotspot for spreading Coronavirus disease. Furthermore, the international investments, growing tourism, and businesses with many countries make Saudi Arabia an important travel destination.

The government of Saudi Arabia undertakes hard efforts to ensure the safety and healthiness of pilgrims, but unfortunately Umrah and Hajj still represent a high risk for spreading COVID-19 in Saudi Arabia [13, 14].

Artificial Neural Networks (ANNs) are an advanced artificial intelligence (AI) technique that is used for discovering new knowledge, and valid hidden patterns from dataset [15, 16]. this technique finds relationships and patterns among the dataset [17, 18]. In this paper is proposing an intelligent model using Artificial Neural Networks (ANNs) for predicting COVID-19 patients spread rate, recovered patients rate and mortifies rates in Saudi Arabia. the proposed system has been trained, validated and tested using Python programming language on datasets which collected from health ministry of kingdom Saudi Arabia.

The rest of this paper is organized as follows. Section 2 presents the related works on recent COVID-19 detection and prediction methods using Artificial Neural Networks and other machine learning algorithms. Section 3 presents the methodology with the detailed model design, dataset description, data preprocessing and performance-evaluation metrics. Sections 4 present the results and discussions. Section 6 concludes the paper and provides an outlook to future research.

2. RELATED WORK

Recently, COVID-19 has become a global pandemic due to its rapid spread. It is very hard to detect infected persons because they have no disease symptoms at the first couple of days of infection. So, it is necessary to find a method that might help to predict and estimate the number of infected persons to adopt the appropriate measures. The prediction and analysis of COVID-19 have been extensively investigated in the last few months, Artificial Neural Networks technique (ANNs) has increasingly been seen as a potential technology.

In general, Machine learning has been used widely as a computing technique with high potential in diseases outbreak prediction. Following some examples of the popular machine learning algorithms have used for some global diseases. Artificial Neural Networks (ANNs) for H1N1 flu, dengue fever, and Oyster norovirus [19, 20, 21], Genetic Algorithm (GA) programming for Oyster norovirus [22], Random Forest (RF) for swine fever [23, 24], Bayesian Network and LogitBoost for Dengue [25, 26], Decision Tree (DT) for Dengue [27], multi-regression for Dengue outbreak prediction [28]. furthermore, many studies have examined different aspects of the disease such as, identifying the source of the virus and analyzing its gene sequences [29, 30], analyzing patient information [31], presenting several methods of virus detection [32, 33], analyzing the epidemiological spread [34, 35]. In term of COVID-19, Several recent literature works point out the potential of machine learning for prediction of Corona virus disease [36–43].

AI can also be used for predicting possibilities in the near future which can help to adopt the necessary measures [44]. Authors in their paper [45] focused on two main tasks. The first task is studying the related research to the diagnosis of COVID-19, and the second concerns about studies that related to the prediction of the number of people who might be infected in the coming days. In their study, they maintained that most of the existing poor and biased models. The authors recommended that research based COVID-19 data must be publicly available to adopt more specifically designed prediction models.

According to [41], authors presented the exponential curve for forecasting the increases of new cases for the nest two-week, based on the world health organization reports and the heuristic method. In their proposed model, they tested the 58th situation report based on old reports. They assumed that the current trend can continue for the next seventeen days, also they predicted one million new cases outside of China by March 30.

In [35], the Corona Tracker team proposed a model called Susceptible Exposed Infectious Recovered (SEIR) based on the data from their website, they made 240-day prediction of COVID-19 confirmed cases inside and outside of China. They predicted that the spread of the disease is reached its peak on May 23, 2020 and predicted the maximum number of infected people will be 425.066 million all over the world.

In [42], the authors in this paper examined several models to predict the next five and ten day of cumulative cases in Guangdong and Zhejiang by February 23, 2020. Authors used Richards growth, generalized logistic growth, and a sub-epidemic wave models to predict the previous infectious outbreaks.

According to [46], authors proposed two scenarios for sampling their data. Scenario one considered sampling the odd days, and Scenario two used even days for training the data, they used two machine learning models, ANFIS and MLP-ICA. They concluded that using different scenarios for data sampling has a minimum effect on the model performance. Both models showed promising results in terms of predicting the time series without the assumptions that epidemiological models require. Both machine learning models, as an alternative to epidemiological models, showed potential in predicting COVID-19 outbreak as well as estimating total mortality

Authors in [47] presented several Prediction models such as the PA, ARIMA, and LSTM algorithms were used to predict the number of COVID-19 confirmations, recoveries, and deaths over the next seven days. They found that PA delivered the best performance comparing with other models. Also, they built A diagnosis model using VGG16 to detect COVID-19 using chest X-ray images. Their model allowed rapid and reliable detection of COVID-19. Authors concluded that the most highly infected areas have similar characteristics, and the spread of the disease in coastal areas is significantly higher than that in other non-coastal areas.

According to [48] developed data mining models for the prediction of COVID-19 infected patients' recovery using epidemiological dataset of COVID-19 patients of South Korea. They applied some machine learning algorithms such as DT, SVM, NB, LR, RF, and K-NN algorithms using python programming language. They concluded that the model developed with DT was found to be the most efficient with the highest accuracy of 99.85%, followed by RF with 99.60% accuracy, then SVM with 98.85% accuracy, then K-NN with 98.06% accuracy, then NB with 97.52% accuracy and LR with 97.49% accuracy.

In [49], authors used COVID-19 data from January 22 to March 29 to develop a prevalence prediction model with an error of 17%, which can predict the prevalence of COVID-19 in all infected areas by April 12, with an accuracy of 83%. In their proposed model, they found that the Europe and America would have the highest prevalence of the disease. Finally, they recommended that their model should be continuously implemented with daily data to determine when the prevalence of COVID-19 will begin to decrease.

Authors in [50] developed a prediction model using machine learning algorithms to fight COVID-19 in China and other infected countries in the globe. Their proposed model estimated the number of reported confirmed cases and mortalities. The data used to build the models were collected between 20th of January and the first March. They concluded that the estimated number of COVID-19 cases would be approximately 89,000 in China and 403,000 worldwide during the outbreak. It is clear that their prediction was closely similar to the actual situation in China [51].

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

In [52], the authors predicted the COVID-19 outbreak in the fifteen most infected countries in the globe using the autoregressive integrated moving average (ARIMA) model. They found in their prediction model that the situation would worsen in Europe and Iran. On the other side, they found that the number of cases in South Korea and China would be more stable and indicated that COVID-19 would spread exponentially in the United States. Authors in [53] applied a CNN to a small dataset to estimate and evaluate the number of reported cases in China. The authors used the mean absolute and RMSE to compare their model with other deep learning models. Finally, the authors concluded in their paper that the achieved results promise a high predictive efficiency.

All previous studies are building several prediction models to predict new coronavirus cases (outbreak ratio), or the number of death cases may be caused by the virus. Some of other related work were used image processing methods to decide if the person is infected by Coronavirus or not. However, very limited papers have considered all factors that related to COVID-19.

In this paper seven factors which include, number of cases, accumulative daily recoveries, accumulative daily mortalities, accumulative active cases, daily spread rate, daily recoveries rate, daily mortalities rate have been considered to build the prediction model.

In this study, Artificial Neural Networks (ANNs) model will be built and applied on Coronavirus dataset for Kingdom of Saudi Arabia to predict COVID-19 spread, recoveries, and mortalities rates considering seven important factors which discussed previously.

In the next section the methodology including the data collection, data preparation, methods and experiment work will be presented in detail.

3. METHODOLGY

In this section, the methodology of this work will be presented including the data source and description, data preparation and preprocessing, experiment environment, model description and design, and the evaluation criteria will be covered as well.

3.1 Data Collection and Description

The dataset was obtained from Saudi Arabia ministry of health which was made available on King Abdullah Petroleum Studies and Research Center (KAPSARC) Website [10]. Dataset of COVID-19 cases of Saudi Arabia has been used in this paper. The dataset of 23 weeks starting on March 2, 2020 to August 5, 2020 with seven attributes which include, number of cases, accumulative daily recoveries, accumulative daily mortalities, accumulative active cases, daily spread rate, daily recoveries rate, daily mortalities rate.

3.2 Data Preparation and Preprocessing

The dataset was prepared and cleaned where only relevant attributes were extracted from the original dataset. The extracted dataset has four inputs with three outputs attributes. For the preprocessing phase, Python programing language has used for data normalization to convert numbers to be between 0 and 1. Data transformation has done through the alternative standardization which is scaling features to lie between a given minimum and maximum value, This can be achieved using MinMaxScaler or MaxAbsScaler in Python. Figure 2. Shows the total statistics for the number of weekly spread rate, recoveries rate, and mortalities rate for all the period (23 weeks).

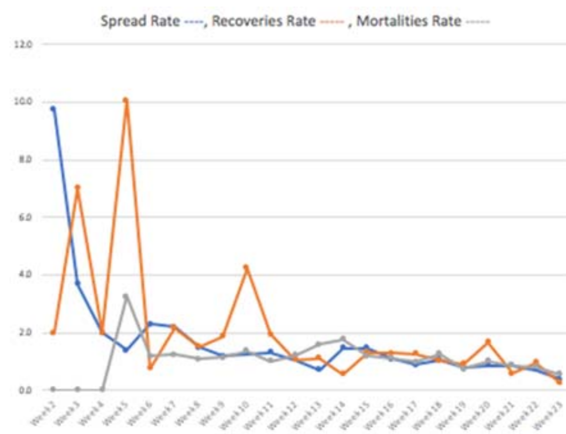

Figure 2: Total statistics for the number of weekly spread rate, recoveries rate, and mortalities rate

Table1 1. shows a sample of preprocessed data, where the target (outputs) are representing the spread rate, recoveries rate, and mortalities rate for COVID-19, while the other four attributes are the inputs which are the number of cases, accumulative daily recoveries, accumulative daily mortalities, accumulative active cases of the proposed model.

3.3 Environment

A computer with macOS was used for the experiment. It has the following specifications: Intel Core i7, 1.80-GHz processor, 16 GB of DDR4 RAM, and 1 TB of hard disk. Python programming language is considered in this paper because it is the most suitable and designed to be

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

extendable with compiler code for proficiency. Many tools are available to facilitate Python integration and software code. The following libraries and software have been installed:

- Matplotlib: https://matplotlib.org/
- NumPy: https://numpy.org/
- Pandas: https://pandas.pydata.org/
- Python: https://www.python.org/
- Scikit: https://scikit-learn.org/
- SciPy: https://www.scipy.org/
- TensorFlow: https://www.tensorflow.org/
- Keras: https://keras.io

3.4 Artificial Neural Networks Model

Artificial Neural Networks (ANNs) are computational paradigms based on mathematical models that different from the traditional computing. ANNs are also called connectionist systems or parallel distributed systems because they are composed by a series of interconnected processing nodes that operate in parallel. Furthermore, the interconnected processing nodes adapt simultaneously with the information flow and adaptive rules. The main objective of ANNs is to understand the functional characteristics and the computational properties of the human brain when it performs cognitive processes such as concept association, concept categorization, sensorial perception, and learning.

ANNs are typically organized in layers structure. Each layer in the network is an array of processing neurons. Information flows through each element in an input-output manner. A two-layer feed-forward network with sigmoid hidden layer and output neurons can classify vectors arbitrarily well. The network is trained with scaled conjugate gradient back-propagation.

Multi Layered Perceptron (MLP) is popular technique that used ANNs method for prediction and modeling purposes. This technique provides acceptable accuracy for prediction tasks in the simple and semi-complex datasets. However, in the case of applying modeling phases in complex datasets, there is a need for more powerful techniques [54, 55].

ANNs will be applied in this paper to the dataset that contains four features (inputs) and three targets (outputs) as mentions previously in section 3.1. However. The dataset in the ANNs has divided into three samples, first sample is the training set (70% of the dataset) which is presented to the network during training, second is the validation set (15% of the dataset) which is used to measure network generalization, and to cut off training when generalization stops improving, and final sample is the testing set (15% of the dataset) which doesn't effect on training and provides an independent measure of network performance during and after training. In addition to two hidden layers have been considered for model.

The following Figure 3 shows the neural networks diagram for the three outputs (daily spread rate, daily recoveries rate, daily mortalities rate). The inputs are four features (number of cases, accumulative daily recoveries, accumulative daily mortalities, accumulative active cases)

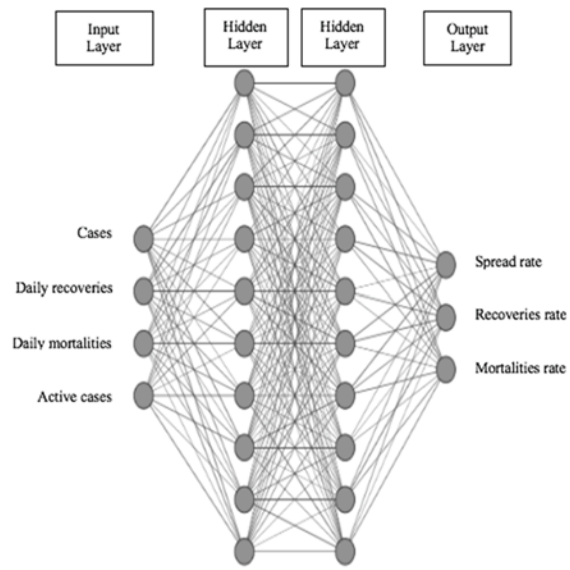

Figure 3: Neural networks diagram for the proposed model

3.5 Evaluation Criteria

Evaluation is conducted through calculating the values for Correlation Coefficient, Mean Absolute Error (MAE), and Mean Square Error (MSE). Those factors are evaluating the target values and for further estimation of the model performance through calculating an index score. However, indexes are used to estimate the model accuracy [56, 57]. Follows present the evaluation criteria equations used in this paper.

Correlation coefficient: a statistical measure of the strength of the relationship between the relative movements of two variables. It is defined as follows:

$$\text{Correlation Coefficient} = \frac{N\Sigma\,(XY) - \Sigma(Y)\,\Sigma(Y)}{\sqrt{[N\,\Sigma\,x^2 - (\Sigma X)^2][N\,\Sigma\,Y^2 - (\Sigma XY)^2]}} \quad (1)$$

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

Mean Absolute Error (MAE): measures of errors between paired observations expressing the same phenomenon. It is defined as follows:

$$\text{MAE} = \frac{1}{N} \left| \frac{XY}{X} \right| \quad (2)$$

In the above evaluation metrics, N represents the number of data. Also, X and Y are the predicted and desired values, respectively.

Mean Square Error (MSE): Measures the differences between the actual (A) and the predicted (P) numbers of COVID-19 spread, recoveries, and mortalities (*N*). MSE is penalizing large prediction errors. It is defined as follows:

$$\text{MSE}=\frac{1}{N}\Sigma(A-P)^2 \quad (3)$$

4. EXPERIMENTAL RESULTS AND DISCUSSIONS

In this section, the experiment work and results will be presented in detail. However, in the proposed model, the number of COVID-19 spread, recoveries, and mortalities rates in Saudi Arabia has predicted using the ANNs technique. The model was trained to make predictions for the next two weeks (weak 24 and 25) to make the prediction from datasets that were collected from a statistics website [10].

To assess the performance of the implemented predicting model, the numbers of COVID-19 spread rate, recoveries rate, and mortalities rate were collected. Figures 4 shows the actual COVID-19 spread rates, recoveries rates, and mortalities rates from 2 March 2020 to 5 August 2020, and Figure 5 shows the predicted COVID-19 spread rates, recoveries rates, and mortalities rates in Saudi Arabia from 2 March 2020 to 20 August 2020.

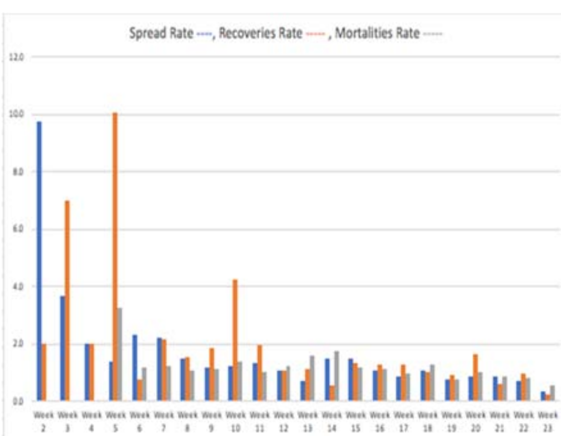

Figure 4: COVID-19 actual cases, recoveries, and mortalities rates in Saudi Arabia

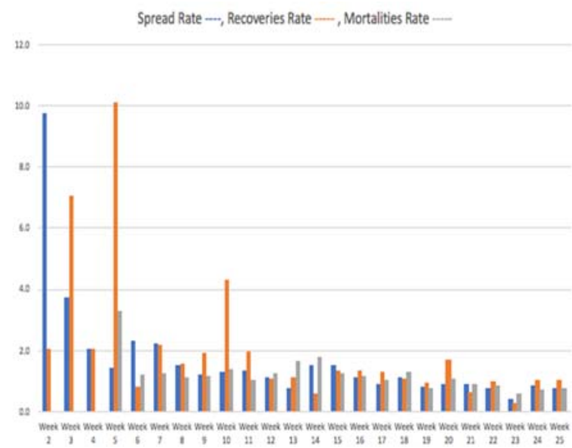

Figure 5: COVID-19 predicted spread, recoveries, and mortalities rates in Saudi Arabia

The ANNs model was compared in terms of MSE, MAE, and Correlation Coefficient, as detailed in Tables 2 below. results show 0.008 MSE value for spread rate, 0.046 for recoveries rate, and 0.032 for mortalities rate, MAE value for spread rate is 0.025, 0.067 for recoveries rate, and 0.053 for mortalities rate. For the MSE and MAE measures, the lowest value is the better. The correlation coefficient shows high values for all targets, the spread rate is the highest with 0.99, then the mortalities rate with 0.96, and the recoveries rate with 0.95.

Table 2: Results of the ANNs model in Saudi Arabia

ANNs (MLP)	Correlation Coefficient	MAE	MSE
Spread rate	0.99	0.025	0.008
Recoveries rate	0.95	0.067	0.046
Mortalities rate	0.96	0.053	0.032

The following Figure 6 shows the MSE of proposed model for the training and validation data using 50 epochs.

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

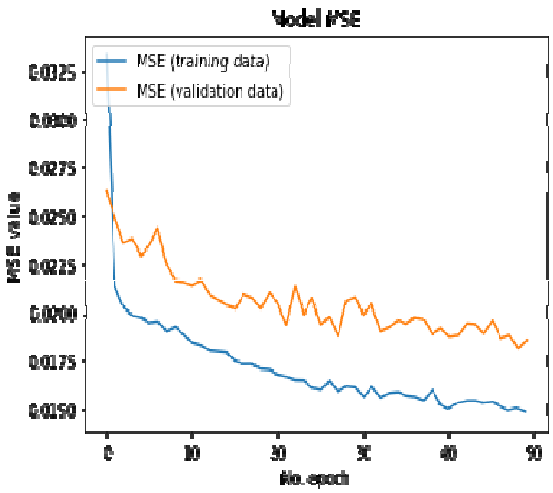

Figure 6: MSE of proposed model for the training and validation data

This paper evaluated the applicability the proposed ANNs model for predicting the COVID-19 spread rates, recoveries rates, and mortalities rates in Saudi Arabia. The model showed promising results in terms of predicting the targets of the proposed approach. ANNs model, as an alternative to epidemiological models, showed potential in predicting COVID-19. Considering the availability of only a small amount of training data, it is expected that the proposed model will be developed further as the basis for future prediction models. Finally, using ANNs model in this paper delivered the best performance for COVID-19 prediction over two weeks.

Comparing with all previous studies which are building several prediction models to predict new coronavirus cases (outbreak ratio), or the number of death cases may be caused by the virus. Some of other related work were used image processing methods to decide if the person is infected by Coronavirus or not. However, very limited papers have considered all factors that related to COVID-19. This paper proposed new factors that may help to achieve more accurate results. However, the limitation of data related to COVID-19 are still the main challenge for all researchers due to newly appearance of this disease.

In the next section of this paper, the conclusions and future work will be presented.

5. CONCLUSIONS AND FUTURE WORK

The worldwide pandemic of the COVID-19 has become the main national security issue for almost all nations. Advancement of accurate prediction models provide insights into the spread of this infectious disease. In fact, the high uncertainty and low size of data have caused some epidemiological models that show low accuracy for long-term forcast.

In this paper, based on the COVID-19 data from March 2 to August 5, a prediction model is developed to predict COVID-19 spread rate, recoveries rate, and mortalities rate in Saudi Arabia with low error rates. the results are promising by achieving low MAE with average value 0.05 and MSE with average value 0.02, and high Correlation Coefficient for all targets' rates with average value 0.97. By having these results, it can be concluded that the main contribution has been achieved successfully. Therefore, it has not yet reached the peak of this disease, so it is recommended that this model be continuously implemented with daily data of this disease to determine when the spread of this disease will begin to decrease.

The objectives of this paper have achieved by developing a prediction model using Artificial Neural Networks (ANNs) that predict COVID-19 spread rate, recoveries rate, and mortalities rate in Saudi Arabia with low error rates and high coefficient correlation using Python Programming language for the implementation stage and code has been developed to achieve the final results.

For future work, for the advancement of higher performance models for long-term prediction, future research should be devoted to comparative studies on various machine learning algorithms such as Decision Tree (DT) and k-Nearest Neighbors k-NN [58] for bigger data size with different types of data, or using fuzzy logic approach [59].

REFERENCES

[1] Guo Y, Cao Q, Hong Z, Tan Y, Chen S, Jin H, et al. The origin, transmission and clinical therapies on coronavirus disease 2019 (COVID-19) outbreak - an update on the status. Mil Med Res 2020 Mar 13;7(1):11 [doi: 10.1186/s40779-020-00240-0]

[2] Li Y et al. A machine learning-based model for survival prediction in patients with severe COVID-19 infection medRxiv 2020.02.27.20028027. 2020. https ://doi.org/10.1101/2020.02.27.20028027 15th May, 2020

[3] Raphael Dolin MD, Stanley Perlman MD. Novel coronavirus from Wuhan China, 2019–20, Chapter 155, Mandell, Douglas, and Bennett's principles and practice of infectious diseases. 9th edition. Elsevier. 2020.

[4] Shang J, Ye G, Shi K, et al. Structural basis of receptor recognition by SARS-CoV-2. Nature. 2020. https://doi.org/10.1038/s4158 6-020-2179-y.

[5] Lu R, Zhao X, Li J, Niu P, Yang B, Wu H, et al. Genomic characterization and epidemiology of 2019 novel coronavirus: implications for virus origins and receptor binding. The Lancet 2020 Feb;395(10224):565-574. [doi: 10.1016/S0140-6736(20)30251-8]

[6] Leonenko, V.N.; Ivanov, S.V. Fitting the SEIR model of seasonal influenza outbreak to the incidence data for Russian cities. *Russ J Numer Anal Math Modell* 2016, *31*, 267-279, doi:10.1515/rnam-2016-0026.

[7] Zhao, S.; Musa, S.S.; Fu, H.; He, D.; Qin, J. Simple framework for real-time forecast in a data-limited situation: The Zika virus (ZIKV) outbreaks in Brazil from 2015 to 2016 as an example. Parasites Vectors 2019, 12, doi:10.1186/s13071-019-3602-9.

[8] Prathap L, Jagadeesan V, Suganthirababu P, Ganesan D. Online Journal of Health and Allied Sciences. 2017. Association of quantitative and qualitative dermatoglyphic variable and DNA polymorphism in female breast cancer population URL: https://www.ojhas.org/issue62/2017-2-2.pdf [accessed 2020-04-07]

[9] World Health Organization, Coronavirus disease 2019 (COVID-19) Situation Report august 2020.

[10] Kingdom of Saudi Arabia ministry of health. https://datasource.kapsarc.org/explore/dataset/saudi-arabia-coronavirus-disease-covid-19-situation/information. (accessed August 16, 2020).

[11] Saudi center for disease prevention and control (SCDC). https://covid19.cdc.gov.sa/daily-updates/ (accessed August 12, 2020).

[12] General authority of statistics, Kingdom of Saudi Arabia. https://www.stats.gov.sa/en/indicators/1 (accessed August 12, 2020).

[13] Memish ZA, Zumla A, Alhakeem RF, et al. Hajj: infectious disease surveillance and control. Lancet 2014;383:2073–2082. doi:10.1016/S0140-6736(14)60381-0.

[14] Hashem AM, Al-Subhi TL, Badroon NA, Hassan AM, Bajrai LHM, Banassir TM, et al. MERS-CoV, influenza and other respiratory viruses among symptomatic pilgrims during 2014 Hajj season. J Med Virol. 2019; 91(6):911–917. doi:10.1002/jmv.25424.

[15] Hussain S, et al. Performance evaluation of various data mining algorithms on road traffic accident dataset. In: Satapathy S, Joshi A, editors. Information and communication technology for intelligent systems. Smart innovation, systems and technologies. Singapore: Springer; 2019. p. 106.

[16] Muhammad LJ et al. Performance evaluation of classification data mining algorithms on coronary artery disease dataset: IEEE 9th international conference on computer and knowledge engineering (ICCKE 2019), Ferdowsi University of Mashhad 978-1-7281-5075-8/19/$31.00 ©2019 IEEE.

[17] Muhammad LJ, et al. Using decision tree data mining algorithm to predict causes of road traffic accidents, its prone locations and time along Kano Wudil highway. Int J Database Theory Appl. 2017;10(11):197–208.

[18] Muhammad LJ, Usman SS. Power of artificial intelligence to diagnose and prevent further COVID-19 Outbreak: a short com-munication.2020. ArXiv: 2004.12463 [cs.CY]. Accessed 15 May 2020.

[19] Koike, F.; Morimoto, N. Supervised forecasting of the range expansion of novel non-indigenous organisms: Alien pest organisms and the 2009 H1N1 flu pandemic. Glob. Ecol. Biogeogr. 2018, 27, 991–1000.

[20] Anno, S.; Hara, T.; Kai, H.; Lee, M.A.; Chang, Y.; Oyoshi, K.; Mizukami, Y.; Tadono, T. Spatiotemporal dengue fever hotspots associated with climatic factors in taiwan including outbreak predictions based on machine-learning. Geospat. Health 2019, 14, 183–194.

[21] Chenar, S.S.; Deng, Z. Development of artificial intelligence approach to forecasting oyster norovirus outbreaks along Gulf of Mexico coast. Environ. Int. 2018, 111, 212–223.

[22] Chenar, S.S.; Deng, Z. Development of genetic programming-based model for predicting oyster norovirus outbreak risks. Water Res. 2018, 128, 20–37.

[23] Liang, R.; Lu, Y.; Qu, X.; Su, Q.; Li, C.; Xia, S.; Liu, Y.; Zhang, Q.; Cao, X.; Chen, Q.; et al. Prediction for global African swine fever outbreaks based on a combination of random forest algorithms and meteorological data. Transbound. Emer. Dis. 2020, 67, 935–946.

[24] Tapak, L.; Hamidi, O.; Fathian, M.; Karami, M. Comparative evaluation of time series models for predicting influenza outbreaks: Application of influenza-like illness data from

sentinel sites of healthcare centers in Iran. BMC Res. Notes 2019, 12.
[25] Raja, D.B.; Mallol, R.; Ting, C.Y.; Kamaludin, F.; Ahmad, R.; Ismail, S.; Jayaraj, V.J.; Sundram, B.M. Artificial intelligence model as predictor for dengue outbreaks. Malays. J. Public Health Med. 2019, 19, 103–108.
[26] Iqbal, N.; Islam, M. Machine learning for dengue outbreak prediction: A performance evaluation of different prominent classifiers. Informatica 2019, 43, 363–371.
[27] Titus Muurlink, O.; Stephenson, P.; Islam, M.Z.; Taylor-Robinson, A.W. Long-term predictors of dengue outbreaks in Bangladesh: A data mining approach. Infect. Dis. Model. 2018, 3, 322–330.
[28] Agarwal, N.; Koti, S.R.; Saran, S.; Senthil Kumar, A. Data mining techniques for predicting dengue outbreak in geospatial domain using weather parameters for New Delhi, India. Curr. Sci. 2018, 114, 2281–2291.
[29] Ji, W., et al., Cross-species transmission of the newly identified coronavirus 2019-nCoV. Journal of Medical Virology, 2020. 92(4): p. 433-440.
[30] Paraskevis, D., et al., Full-genome evolutionary analysis of the novel corona virus (2019-nCoV) rejects the hypothesis of emergence as a result of a recent recombination event. Infect Genet Evol, 2020. 79: p. 104212.
[31] Huang, C., Y. Wang, and X. Li, Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China (vol 395, pg 497, 2020). Lancet, 2020. 395(10223): p. 496-496.
[32] Corman, V.M., et al., Detection of 2019 novel coronavirus (2019-nCoV) by real-time RT-PCR. Eurosurveillance, 2020. 25(3): p. 23-30.
[33] Zhang, N.R., et al., Recent advances in the detection of respiratory virus infection in humans. Journal of Medical Virology, 2020. 92(4): p. 408-417.
[34] Dey, S.K., et al., Analyzing the epidemiological outbreak of COVID-19: A visual exploratory data analysis approach. Journal of Medical Virology.
[35] Binti Hamzah, F.A., et al., CoronaTracker: World-wide COVID-19 Outbreak Data Analysis and Prediction. 2020.
[36] Alimadadi, A.; Aryal, S.; Manandhar, I.; Munroe, P.B.; Joe, B.; Cheng, X. Artificial Intelligence and Machine Learning to Fight COVID-19; American Physiological Society: Bethesda, MD, USA, 2020.
[37] Alimadadi, A.; Aryal, S.; Manandhar, I.; Munroe, P.; Joe, B.; Cheng, X. Artificial Intelligence and Machine Learning to Fight COVID-19. Physiol. Genom. 2020.
[38] Ardabili, S.F.; Mosavi, A.; Ghamisi, P.; Ferdinand, F.; Varkonyi-Koczy, A.R.; Reuter, U.; Rabczuk, T.; Atkinson, P.M. COVID-19 Outbreak Prediction with Machine Learning. medRxiv 2020.
[39] Miralles-Pechuán, L.; Jiménez, F.; Ponce, H.; Martínez-Villaseñor, L. A Deep Q-learning/genetic Algorithms Based Novel Methodology for Optimizing Covid-19 Pandemic Government Actions. arXiv 2020, arXiv:2005.07656.
[40] Roosa, K., et al., Real-time forecasts of the COVID-19 epidemic in China from February 5th to February 24th, 2020. Infect Dis Model, 2020. 5: p. 256-263.
[41] Koczkodaj, W.W., et al., 1,000,000 cases of COVID-19 outside of China: The date predicted by a simple heuristic. Global Epidemiology, 2020: p. 100023.
[42] Roosa, K., et al., Short-term Forecasts of the COVID-19 Epidemic in Guangdong and Zhejiang, China: February 13-23, 2020. J Clin Med, 2020. 9 (2).
[43] Nishiura, H., et al., The Extent of Transmission of Novel Coronavirus in Wuhan, China, 2020. Journal of Clinical Medicine, 2020. 9(2).
[44] K. Santosh, "AI-Driven Tools for Coronavirus Outbreak: Need of Active Learning and Cross-Population Train/Test Models on Multitudinal/Multimodal Data," Journal of Medical Systems, vol. 44, pp. 1-5, 2020.
[45] L. Wynants, B. Van Calster, M. M. Bonten, G. S. Collins, T. P. Debray, M. De Vos, et al., "Systematic review and critical appraisal of prediction models for diagnosis and prognosis of COVID-19 infection," medRxiv, 2020.
[46] Gergo Pinter, Imre Felde , Amir Mosavi, Pedram Ghamisi and Richard Gloaguen, "COVID-19 Pandemic Prediction for Hungary; A Hybrid Machine Learning Approach". Mathematics 2020, 8, 890; doi:10.3390/math8060890
[47] Moutaz Alazab, Albara Awajan, Abdelwadood Mesleh, Ajith Abraham, Vansh Jatana, and Salah Alhyari. "COVID-19 Prediction and Detection Using Deep Learning". International Journal of Computer Information Systems and Industrial Management Applications. ISSN 2150-7988 Volume 12 (2020) pp. 168-181
[48] L. J. Muhammad, Md. Milon Islam, Sani Sharif Usman, and Safial Islam Ayon. "Predictive Data Mining Models for Novel Coronavirus (COVID-19) Infected Patients'

Recovery”. SN Computer Science (2020) 1:206, June 2020.

[49] Fatemeh Ahouz, amin golabpour. “Predicting the COVID-19 Prevalence Rate Using Data Mining”. Preprint on ResearchGate. Under review on research square. DOI: 10.21203/rs.3.rs-21247/v1. July 2020.

[50] M. Li, Z. Zhang, S. Jiang, Q. Liu, C. Chen, Y. Zhang, et al., "Predicting the epidemic trend of COVID-19 in China and across the world using the machine learning approach," medRxiv, 2020.

[51] Worldometers. (2020, April. 6). Coronavirus Cases. Available: https://www.worldometers.info/coronavirus/

[52] P. Kumar, H. Kalita, S. Patairiya, Y. D. Sharma, C. Nanda, M. Rani, et al., "Forecasting the dynamics of COVID-19 Pandemic in Top 15 countries in April 2020 through ARIMA Model with Machine Learning Approach," medRxiv, 2020.

[53] C.-J. Huang, Y.-H. Chen, Y. Ma, and P.-H. Kuo, "Multiple-Input Deep Convolutional Neural Network Model for COVID-19 Forecasting in China," medRxiv, 2020.

[54] Rawabi A. Aroud, Anas H. Blasi, and Mohammed A. Alsuwaiket. “Intelligent Risk Alarm for Asthma Patients using Artificial Neural Networks”. International Journal of Advanced Computer Science and Applications 11(6):95-100. June 2020. DOI: 10.14569/IJACSA.2020.0110612

[55] Anas H. Blasi. “Performance Increment of High School Students using ANN model and SA algorithm”. Journal of Theoretical and Applied Information Technology 95(11):2417-2425. January 2017.

[56] Ardabili, S.F.; Mahmoudi, A.; Gundoshmian, T.M. Modeling and simulation controlling system of HVAC using fuzzy and predictive (radial basis function, RBF) controllers. J. Build. Eng. 2016, 6, 301–308.

[57] Ardabili, S.; Mosavi, A.; Mahmoudi, A.; Gundoshmian, T.M.; Nosratabadi, S.; Varkonyi-Koczy, A.R. Modelling Temperature Variation of Mushroom Growing Hall Using Artificial Neural Networks. Engineering for Sustainable Future, Lecture Notes in Networks and Systems; Springer: Cham, Switzerland, 2019.

[58] Mohammad A. Lababede, Anas H. Blasi, and Mohammed A. Alsuwaiket. “Mosques Smart Domes System using Machine Learning Algorithms”. International Journal of Advanced Computer Science and Applications 11(3):373-378. March 2020. DOI:10.14569/IJACSA.2020.0110347

[59] Anas H. Blasi. “Scheduling Food Industry System using Fuzzy Logic”. Journal of Theoretical and Applied Information Technology 96(19):6463-6473. October 2018.

Table 1: Sample of preprocessed data

Accumulative Cases	Accumulative Daily Recoveries	Accumulative Daily Mortalities	Accumulative Daily active	Spread rate	Recoveries rate	Mortalities rate
67	2	0	230	0.44	0.00	0.00
36	0	0	266	0.13	0.00	0.00
70	0	0	336	0.48	0.00	0.00
48	16	0	368	0.17	0.00	0.00
119	0	0	487	0.62	0.00	0.00
51	0	0	538	0.11	0.00	0.00
205	4	1	738	1.00	0.00	0.00
133	1	1	869	0.16	0.06	0.25
112	4	1	976	0.21	1.00	0.25
92	2	0	1066	0.20	0.12	0.00
99	2	1	1162	0.27	0.25	0.00
96	29	4	1225	0.24	0.61	1.00
154	49	0	1330	0.40	0.42	0.00
110	50	2	1388	0.18	0.25	0.00
157	99	6	1440	0.36	0.49	0.75
165	64	5	1536	0.26	0.16	0.21