

# COVID-19 Death Risk Assessment in Iran using Artificial Neural Network

Adebayo Ifeoluwapo. R<sup>1</sup>, Eko Supriyanto<sup>1,2\*</sup>, and Sahar Taheri<sup>1</sup>

<sup>1</sup>School of Biomedical Engineering and Health Sciences, Faculty of Engineering, Universiti Teknologi Malaysia

<sup>2</sup>IJN-UTM Cardiovascular Engineering Centre, Universiti Teknologi Malaysia  
Email: \*eko@utm.my

**Abstract.** Since the pandemic spread of COVID-19, it has posed a unique public health concern worldwide due to its increased death rate all around the world. The pandemic disease is caused by the SARS-CoV-2, which is the main cause of Middle East Respiratory Syndrome (MERS) and severe acute respiratory syndrome (SARS). Risk assessment is a vital action toward disease risk reduction as it increases the understanding of the risk factors associated with the disease and allows existing data to decide on adequate preventive and mitigation measures. Machine learning techniques have gained strength since 2000, as it has crucial role in data analysis and is really helpful to develop standard mortality models. This study aims to find the best model for data analysis using the Artificial Neural Network (ANN) and other risk factors, which contribute to the high mortality and morbidity associated with COVID-19 in Iran, to predict the risk of death for the people with different situation. A systematic review and meta-analysis were examined by using patient risk factor data from studies done by researchers to estimate COVID-19 death risk. Risk factors for the disease were extracted from an existing study. Using ANN, the best risk prediction for the disease is calculated. Assessment of a different number of hidden neurons with a different training function using the Bayesian Regularization algorithm, the best training function for the ANN model with 5 hidden neurons is found to have the most satisfying results. The coefficient of determination (R) and Root Mean Square Error (RMSE) was 9.99999e-1 and 4.54201e-19 respectively.

## 1. Introduction

Since the emergence of the known coronavirus disease, it has been a great threat to both global health and a socio-economic in almost all countries around the world. A continuous occurrence of the virus was reported first in Wuhan City, Hubei Province, China, since December 12<sup>th</sup>, 2019 and it was linked to have originated from the South China Seafood Market [1]. On the December 29<sup>th</sup>, 2019, the first 4 cases of COVID-19 were first reported all linking to the seafood wholesale market in Huanan (Southern China). Due to this occurrence, the World Health Organization (WHO) China Country Office was informed of a pneumonia of unknown cause, detected in the city of Wuhan in the Hubei province, China on the December 31<sup>st</sup>, 2019 [2]. The COVID-19 disease has been associated to a great number of morbidity and mortality among both young and old people worldwide but, the answer to the origin of this disease remains unknown. Human- to- human transition is one of the biggest challenges in curtailing this epidemic leading to the WHO on January 30<sup>th</sup>, 2020 to issue a worldwide health emergency warning notice. (COVID-19) is zoonotic virus belonging to a large family of positive-sense; single-stranded RNA viruses that can cause both mild and moderate upper respiratory tract illnesses with symptoms similar to that of the Severe Acute Respiratory Syndrome (SARS) and Middle East Respiratory Syndrome (MERS) infection such as common cold [3,4].



COVID-19 has spread greatly to other countries including South Korea, Japan, Thailand, Singapore, Italy, US, UK, Spain and Iran, and on the March 12<sup>th</sup>, 2020, it was classified by the WHO as a pandemic [1]. As of October 22<sup>nd</sup>, 2020, based on the WHO reports, we had globally 41,104,946 confirmed cases with 1,128,325 deaths [5]. Iran has reported an increased number of cases of people infected with the virus in the Middle East and the highest cumulative number of deaths outside China. As of October 23<sup>rd</sup>, 2020, there were 550,757 COVID-19 cases, including 31,650 associated deaths in Iran. About six neighbouring countries (Afghanistan, Bahrain, Kuwait, Iraq, Oman, and Pakistan) had reported imported cases from Iran [6]. According to recent statistics, Iran is ranked 13<sup>th</sup> in the number of deaths due to COVID-19 as reported as at October 23<sup>rd</sup>, 2020. Multiple reports have shown that pre-existing illnesses, such as cardiovascular disorders, respiratory diseases, cancer, infectious diseases and substance abuse, can lead to increased morbidity and mortality of COVID-19. By the aid of existing data from China, cardiovascular disease, diabetes and hypertension are highly predominant among the patients. The disease can act as a factor to make the existing conditions worse and accordingly leading to death. To reduce the risk of this disease, risk assessment is important due to its help in understanding the risk factors associated with the disease and allows the opportunity to decide on adequate preventive and mitigation measures [7,8].

In data analysis, since 2000, machine learning techniques have gained strength as they played a vital role in epidemiological data analysis by developing standard mortality models [3]. Artificial Neural Networks (ANNs) are statistical and relatively unique approaches for modelling complex non-linear relationships in spatial epidemiology. The technique is considered as one of the most efficient methods in processing large sets of data that can be analysed computationally to reveal trends, patterns, prediction forecasting etc.

These techniques have been applied in a number of areas, including artificial intelligence, public health, epidemiology, economics, environmental science and agriculture [8,9]. ANN is a collection of a large number of units called neurons. Some of these neurons receives its information and sends the results to other neurons. The ANN structure is computer-based and consists of several basic parallel-operated processing elements. The most common ANN consists of a three-layer network, an input layer, one or more hidden layers, and an output layer. The input layer includes independent variables linked to the hidden layer for processing. The hidden layer consists of activation functions that helps calculate the weights of variables to examine the influence of predictors upon the dependent variables. The prediction process is completed in the output layer and the results are given with a small estimation error. Generally, a feedforward network is trained in a backpropagation algorithm. The back propagation algorithm learns relation between a defined set of input-output pairs in the training phase [10].

Unfortunately, to date, there is no global and standard response to the COVID-19 pandemic and each country is facing the crisis based on their own possibilities, expertise and hypotheses. Thus, there are different criteria for testing, hospitalisation and estimating of cases making it difficult to estimate the number of people affected by the epidemic. Early screening and risk estimation of the disease can help reduce the death chances and improve survival rate [11]. This study therefore aimed to evaluate the ability of ANN-based method in accessing the risk factors associated with the high morbidity and mortality of COVID-19 and to also estimate the risk of a patient dying when infected. To this end, we compared results from the conventional method and ANN modelling.

## 2. Methodology

### 2.1. Conventional Method

Data extracted from a journal represented a retrospective, epidemiological study that has been performed on hospitalized cases of COVID-19 in Baqiyatallah hospital in Tehran, Iran, for patients admitted from February 19<sup>th</sup>, 2020, to April 15<sup>th</sup>, 2020 [12]. The study variables were as follows: age, gender, final outcome (including death or survival), and type of comorbidities. The number of people infected and death cases varied per gender alongside the comorbidities. Descriptive analyses of the variables were

expressed with representation of the risk factors as LEL=100, MEL=100. LEL (lower energy level) included data for death cases of COVID-19 across their age and gender as stated in the journal (Table 1 and Table 2). It was represented as:

$$LEL = (Age + Gender) = 100 \tag{1}$$

**Table 1.** Risk factors of the model [12].

S/N	Low Energy Level (LEL)	High Energy Level (HEL)
1.	Age	Pre-existing medical conditions (Cancer, Hypertension, Chronic Respiratory disease, Diabetes, Chronic Kidney disease, Cardiovascular disorder).
2.	Gender	-

MEL included data for death cases of COVID-19 with pre-existing medical condition as also stated in the journal (Table 1 and Table 2). It was represented as:

$$\begin{aligned}
 MEL &= \text{Diabetes} + \text{Chronic Respiratory Disease} + \text{Cardiovascular Disease} \\
 &\quad + \text{Kidney Disease} + \text{Cancer} + \text{Hypertension Other Diseases)} \\
 &= 100
 \end{aligned}
 \tag{2}$$

**Table 2.** Cases with COVID-19 based on their gender, age, comorbidities and death rate in Iran [12].

	Cases	Died
<b>Gender</b>		
<i>Male</i>	66	69.9
<i>Female</i>	34	30.1
<b>Age Group</b>		
<i>0-10</i>	0.07	0
<i>11-20</i>	0.27	0
<i>21-30</i>	3.27	0.42
<i>31-40</i>	14.03	5.9
<i>41-50</i>	15.25	5.4
<i>51-60</i>	27.83	22.6
<i>61-70</i>	20.98	29.7
<i>71-80</i>	11.8	21.75
<i>&gt; 80</i>	6.5	15.5
<b>Comorbidities</b>		
<i>Cancer</i>	5.2	0.4
<i>Hypertension</i>	18.3	3.4
<i>Diabetes</i>	35	4.6
<i>Cardiovascular Disorder</i>	11.5	1.7

	Cases	Died
<i>Chronic Respiratory Disease</i>	18.6	3.8
<i>Chronic Kidney Disease</i>	5.6	1.3
<i>Other Diseases</i>	5.9	0.8
<i>No comorbidities</i>	89.1	84.1

2.2. Artificial Neural Network (ANN)

ANN was trained in MATLAB using the outputs of the conventional model predictions. To develop a machine learning model, a two-layer feed forward neural network was designed with input data of the age, gender, and comorbidities of covid-19 death cases in Iran as seen in (Table 2) and the target data (output of the conventional model). The network type used is the Feed-forward with a function of ‘TRAINUM’ to train data in the MATLAB. A total of 35 data samples were used as the input 25 of which were used for training, 5 for validation and 5 for testing. The neural network structure is shown in (Figure 1) with a total of 9 input, 5 hidden neuron and 1 output. The algorithm Levenberg-Marquardt (LM) was applied to train the network. Transfer function of ‘TRANSIG’ for the hidden layer and linear were considered. Number of hidden layers, neurons, and training algorithms are all influential parameters in defining the best prediction model using ANN.

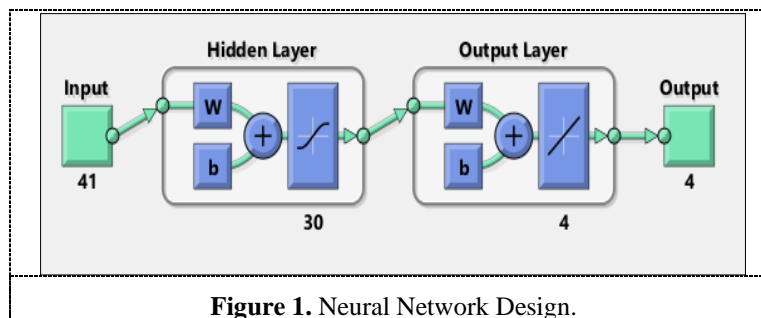


Figure 1. Neural Network Design.

3. Results

3.1. Conventional Modelling Result

According to Table 2, the probability that a patient would die if confirmed to have COVID-19 was generated. Probability of people dying of COVID-19 based on their age and gender is calculated as follows:

$$LEL = (Age + Gender) \times 1.005 \tag{3}$$

Probability of people dying of COVID-19 based on their pre-existing medical condition is calculated as follows:

$$MEL = (Cancer + Chronic Respiratory Disease + Diabetes + Cardiovascular Disease + Hypertension + Chronic Kidney Disease + Others) * 6.25 \tag{4}$$

Total risk of dying when infected with COVID-19 is calculated as follows:

$$\begin{aligned}
 \text{Death Risk\%} = & ((\text{Age} + \text{Gender}) * 1.005 + (\text{Cancer} \\
 & + \text{Chronic Kidney Disease} + \text{Diabetes} + \text{Hypertension} \\
 & + \text{Chronic Respiratory Disease} + \text{Cardiovascular Disease} \\
 & + \text{Others}) * 6.25
 \end{aligned}
 \tag{5}$$

In conventional method by the aid of all information extracted from previous studies (Table 2), an algorithm was written to have probability of each disease based on prediction of the various risk factors. The algorithm dedicated one possible answer for the probability that a patient will die of COVID-19 depending on their age, gender and pre-existing medical conditions repeating the method 35 times (see Appendix A). With prediction made, the probability for 35 samples was calculated using the MATLAB software.

3.2. Artificial Neural Network (ANN) Result

The training has been considered for a total of 35 data sample. 9 inputs have been in the data with 5 hidden neuron and 1 output neural network structure. The ANN was optimized using the Bayesian Regularization algorithm. For the training function 25 samples were taken. A Mean Squared Error (MSE) of 4.54201e-19 with a regression value of 9.99999e-1 were obtained. For the validation function, 5 samples were gotten, MSE of 0.00000e-0 with regression value 0.00000e-0 and for the testing function, 5 samples were gotten with MSE of 8.89662e-0 and regression value of 9.94067e-1. The MSE is the average squared difference between outputs and targets. Lower values are better and zero means no error. Regression R values measure the correlation between outputs and targets. An R value of 1 means a close relationship, 0 a random relationship. The result of the training is showed in Figure 2. The algorithm typically requires more time, but can result in good generalization for difficult, small or noisy datasets. The training stopped according to adaptive weight minimization (regularization).

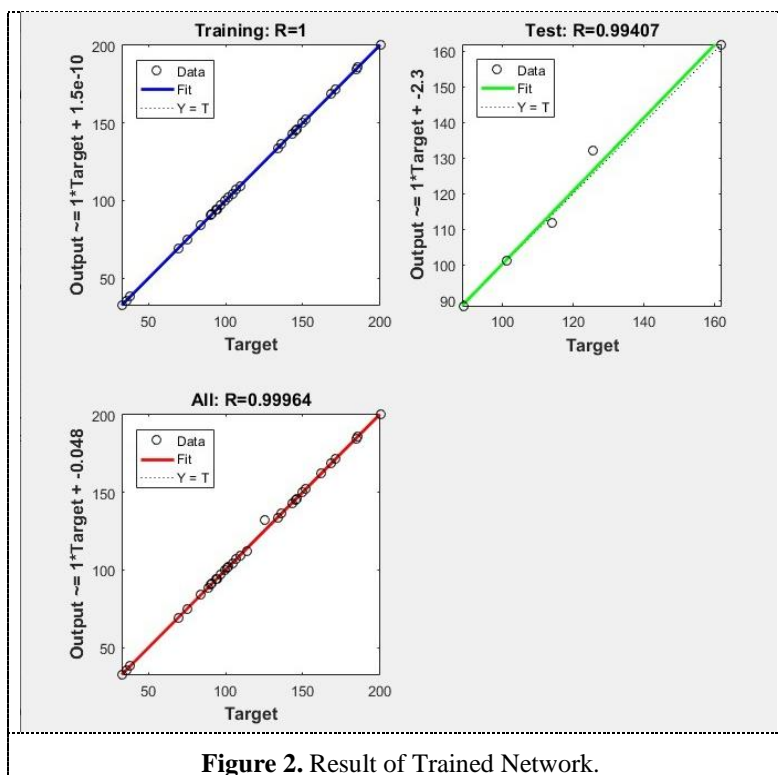


Figure 2. Result of Trained Network.

#### 4. Discussion

Comparing this study to other similar ones, Assaf et al. [13] used three different methods such as Neural Network, Random Forest, and Classification and Regression Decision Tree (CART) to predict patient deterioration. They reached 88.0% sensitivity, 92.7% specificity and 92.0% accuracy in predicting critical COVID-19. A study by Dhamodharavadhani et al. [3] employed Statistical Neural Network (SNN) models such as Probabilistic Neural Network (PNN), Radial Basis Function Neural Network (RBFNN), and Generalized Regression Neural Network (GRNN). They showed that the PNN and RBFNN-based MRP model had the best performance between all models for COVID-19. Another study by Clift et al. [14] to derive and validate a risk prediction algorithm to estimate hospital admission and mortality outcomes from COVID-19 in adults. The models were fitted in the derivation cohort to derive risk equations using a range of predictor variables. Although, the risk prediction model QCOVID population-based risk algorithm performed well, showing very high levels of discrimination for deaths due to covid-19; the algorithm will be regularly updated as the understanding of COVID-19 changes, as more data are provided. This study could expand to other diseases to help the healthcare system respond more effectively during an outbreak or a pandemic especially with most countries currently having a second wave of COVID-19.

#### 5. Conclusion

In this study, we aimed to assess the utility of machine-learning models for predicting a high mortality rate for patients infected with COVID-19 based on their gender, age group and pre-existing medical condition. Machine Learning (ML) models enabled assessment of the relations between input and output of complex processes. ML models generated better performance than traditional prediction models, owing to their ability to reveal nonlinear associations, superior optimization of multi-factorial. Data analysed was just a few numbers with case study only in Iran therefore, extended studies with larger sample size of the variables and their relationship with the disease could provide more accurate information for COVID-19 monitoring. With the huge number of mortalities associated with COVID-19 especially for people with pre-existing medical conditions. The proposed method can help medical facilities, hospitals, and caregivers in deciding who needs to get attention first before other patients, especially when there is a lot of patients to be attended to. It can also help eliminate delays in providing the necessary care in order to reduce the risk of a patient with COVID-19 dying.

#### Acknowledgments

The authors are grateful to Universiti Teknologi Malaysia on the support throughout this study and for financial support through High Impact Research Grant (Vot no. 04G29).

#### References

- [1] Elengoe A 2020 COVID-19 outbreak in Malaysia *Osong Public Health and Research Perspectives* **11**(3) 93–100.
- [2] Harapan H, Itoh N, Yufika A, Winardi W, Keam S, Te H, Megawati D, Hayati Z, Wagner A L and Mudatsir M 2020 Coronavirus disease 2019 (COVID-19): A literature review *Journal of Infection and Public Health*.
- [3] Dhamodharavadhani S, Rathipriya R and Chatterjee J M 2020 COVID-19 Mortality Rate Prediction for India Using Statistical Neural Network Models *Frontiers in Public Health* **8**.
- [4] Zehra Z, Luthra M, Siddiqui S M, Shamsi A, Gaur N and Islam A 2020 Corona virus versus existence of human on the earth: A computational and biophysical approach *International Journal of Biological Macromolecules* **161** 271-81.
- [5] World Health Organization 2020 Coronavirus disease 2633.
- [6] Zhuang Z, Zhao S, Lin Q, Cao P, Lou Y, Yang L and He D 2020 Preliminary estimation of the novel coronavirus disease (COVID-19) cases in Iran: A modelling analysis based on overseas cases and air travel data *International Journal of Infectious Diseases* **94** 29-31.
- [7] Guan W J, Ni Z Y, Hu Y, Liang W H, Ou C Q, He J X, Liu L, Shan H, Lei C L, Hui D S and Du B 2020 Clinical characteristics of coronavirus disease 2019 in China *New England Journal of Medicine* **382**(18) 1708-20.

- [8] Mollalo A, Rivera K M, Vahedi B 2020 Artificial neural network modeling of novel coronavirus (COVID-19) incidence rates across the continental United States *International Journal of Environmental Research and Public Health* **17(12)** 4204.
- [9] Tamang S K, Singh P D, Datta B 2020 Forecasting of Covid-19 cases based on prediction using artificial neural network curve fitting technique *Global Journal of Environmental Science and Management* **6 Special Issue (Covid-19)** 53-64.
- [10] Kayri M 2016 Predictive abilities of Bayesian regularization and Levenberg-Marquardt algorithms in artificial neural networks: A comparative empirical study on social data *Mathematical and Computational Applications* **21(2)** 20.
- [11] Di Gennaro F, Pizzol D, Marotta C, Antunes M, Racalbutto V, Veronese N and Smith L 2020 Coronavirus diseases (COVID-19) current status and future perspectives: A narrative review *International Journal of Environmental Research and Public Health* **17(8)** 2690.
- [12] Nikpouraghdam M, Farahani A J, Alishiri G, Heydari S, Ebrahimnia M, Samadinia H, Sepandi M, Jafari N J, Izadi M, Qazvini A, Dorostkar R 2020 Epidemiological characteristics of coronavirus disease 2019 (COVID-19) patients in IRAN: A single center study *Journal of Clinical Virology* **127** 104378.
- [13] Assaf D, Gutman Y, Neuman Y, Segal G, Amit S, Gefen-Halevi S, Shilo N, Epstein A, Mor-Cohen R, Biber A, Rahav G, Levy I and Tirosh A 2020 Utilization of machine-learning models to accurately predict the risk for critical COVID-19 *Internal and Emergency Medicine* **15(8)** 1435–43.
- [14] Clift A K, Coupland C A C, Keogh R H, Diaz-Ordaz K, Williamson E, Harrison E M, Hayward A, Hemingway H, Horby P, Mehta N, Bengler J, Khunti K, Spiegelhalter D, Sheikh A, Valabhji J, Lyons R A, Robson J, Semple M G, Kee F, Hippisley-Cox J 2020 Living risk prediction algorithm (QCOVID) for risk of hospital admission and mortality from coronavirus 19 in adults: national derivation and validation cohort study *BMJ* **371**.

## Appendix A

NO	AGE GROUP	NO DIED PER AGE GROUP	NO DIED PER GENDER	NO DIED WITH THEIR COMORBITIES							
				CANCER	HYPERTENSION	DIABETES	CARDIOVASCULAR DISEASE	CHRONIC RESPIRATORY DISEASE	CHRONIC KIDNEY DISEASE	OTHERS	TOTAL
1.	0-10	0	69.9	0.4	0	0	1.7	3.8	0	0	107.1245
2.	11-20	0	30.1	0.4	0	0	1.7	3.8	1.3	0	75.2505
3.	21-30	0.42	30.1	0	0	4.6	0	3.8	1.3	0	91.2976
4.	31-40	5.9	69.9	0	3.4	4.6	1.7	0	1.3	0.8	149.929
5.	41-50	5.4	69.9	0	3.4	4.6	1.7	0	0	0	136.3015
6.	61-70	29.7	30.1	0.4	3.4	4.6	1.7	3.8	0	0.8	151.974
7.	71-80	21.75	69.9	0	3.4	4.6	1.7	3.8	1.3	0	184.60825
8.	31-40	5.4	30.1	0	0	0	0	0	0	0	35.6775
9.	71-80	21.75	30.1	0.4	3.4	4.6	1.7	3.8	1.3	0.8	152.10925
10.	31-40	5.9	69.9	0.4	3.4	4.6	1.7	3.8	1.3	0	171.179
11.	>80	15.5	30.1	0.4	3.4	4.6	1.7	3.8	1.3	0.8	145.828
12.	61-70	29.7	301	0	0	4.6	0	0	0	0	88.849
13.	71-80	21.75	69.9	0.4	0	0	0	0	0	0	94.60825
14.	31-40	5.4	30.1	0.4	0	4.6	1.7	0	1.3	0.8	90.6775
15.	51-60	22.6	69.9	0	3.4	4.6	1.7	3.8	1.3	0	185.4625
16.	0-10	0	30.1	0.4	0	0	0	0	0	0	32.7505
17.	61-70	29.7	69.9	0.4	3.4	4.6	1.7	3.8	1.3	0.8	200.098
18.	11-20	0	69.9	0	0	0	1.7	3.8	0	0.8	109.6245



1 9 .	21- 30	0.4 2	69.9	0	0	4.6	1.7	3.8	0	0	133.7 966
2 0 .	21- 30	0.4 2	69.9	0.4	0	4.6	0	3.8	0	0	125.6 716
2 1 .	21- 30	0.4 2	30.1	0.4	0	0	0	0	0	0.8	38.17 26
2 2 .	31- 40	5.9	30.1	0.4	0	4.6	1.7	3.8	0	0	101.8 05
2 3 .	31- 40	5.9	69.9	0.4	3.4	0	0	0	0	0	99.92 9
2 4 .	41- 50	5.4	69.9	0	3.4	0	0	0	0	0	96.92 65
2 5 .	51- 60	22. 6	69.9	0	3.4	4.6	1.7	0	1.3	0	161.7 125
2 6 .	41- 50	5.4	30.1	0.4	3.4	4.6	0	0	1.3	0.8	101.3 025
2 7 .	51- 60	22. 6	30.1	0.4	0	4.6	0	0	0	0	84.21 35
2 8 .	61- 70	29. 7	69.9	0.4	0	0	1.7	3.8	1.3	0	145.0 98
2 9 .	61- 70	29. 7	30.1	0.4	3.4	4.6	0	3.8	0	0	136.3 49
3 0 .	71- 80	21. 75	69.9	0	0	4.6	1.7	3.8	1.3	0.8	168.3 5825
3 1 .	71- 80	21. 75	30.1	0.4	3.4	4.6	0	0	0	0	104.6 0925
3 2 .	>80	15. 5	69.9	0	0	0	0	0	1.3	0	93.95 2
3 3 .	>80	15. 5	30.1	0.4	3.4	0	0	0	0	0	69.57 8
3 4 .	>80	15. 5	30.1	0.4	3.4	4.6	1.7	0	0	0.8	113.9 53
3 5 .	61- 70	22. 6	69.9	0	3.4	4.6	0	0	0	0	142.9 625