How to Cite:
Sujatha, D. C., Aruna, P., & Jeyachidra, J. (2022). Building predictive model for COVID-19 using artificial neural network (ANN) algorithm. *International Journal of Health Sciences*, *6*(S2), 4522–4532. https://doi.org/10.53730/ijhs.v6nS2.6077

Building predictive model for COVID-19 using artificial neural network (ANN) algorithm

D. Christy Sujatha
Assistant Professor, Department. of Software Engineering, Periyar Maniammai Institute of Science and Technology, Vallam, Thanjavur
Email: christy_se@pmu.edu

P. Aruna
Associate professor, Department of Software Engineering, Periyar Maniammai Institute of Science and Technology, Vallam, Thanjavur
Email: aruna_se@pmu.edu

J. Jeyachidra
Professor, Department of Computer Science and Applications, Periyar Maniammai Institute of Science and Technology, Vallam, Thanjavur
Email: chithu_raj@pmu.edu

***Abstract*---**Machine learning plays an important role in addressing the pandemic crisis to analyse, identify and to forecast the infection and the spread of any contagious virus. Nowadays, most of the organizations and researchers are moving towards machine learning algorithms to develop predictive models, trying to reduce the death rate and to identify the patients who are at the increased risk of mortality. The major challenge of Covid-19 is, its identification and classification, due to the fact that the symptoms of Covid -19 are similar to other infectious diseases such as viral fever, typhoid, dengue, pneumonia and other lung infectious diseases. The objective of this paper is to build a predictive model for covid-19 using the Artificial Neural Network (ANN), a supervised machine learning Algorithm. In this study, the data set from *Kaggle Sírio-Libanês data for AI and Analytics by the Data Intelligence Team* has been used to build the predictive model. It is observed that there is 73% of accuracy in predicting the number of corona infected cases.

***Keywords*---**COVID-19, machine learning algorithm, artificial neural network.

Introduction

In this pandemic world, the disease caused by the SARS-COV-2 (Severe Acute Respiratory Syndrome Corona-Virus 2) virus named COVID-19 has been considered as the highly contagious viral infection disease worldwide by the World Health Organization [1]. COVID-19 has distracted the entire world into an unpredicted situation, causing various health and financial issues, not only in each and every individual but also in the entire society of the world [2]. Everyone in this globe is getting scared to go out and to blend in with others, especially in working environments like Educational Institutions, Industries, Government and Private sectors [3]. Since the period when it was first confirmed as COVID in 2019, the infected cases in all over the world were rapidly increasing and reaching 86.7 million cases, including 1.87 million deaths by January 2021[4].

In general, COVID-19 starts with mild symptoms, such as fever and cough, and improves gradually, sometimes leading to organ failure and death. The major challenges of COVID-19 is its identification and classification, due to the fact that the symptoms of COVID -19 is similar to other infectious diseases such as viral fever, typhoid, dengue, pneumonia and other lung infectious diseases. Moreover, it spreads at a very rapid speed from human to human and hence reducing the severity of the outbreak is also considered as another significant challenge [5].

It is very essential to early detect any infectious disease to reduce its spread and to save more lives. Machine Learning is one of the widely used states of art technologies to detect, track and to predict the growth rate of the coronavirus infected patients. Hence, to lessen the burden on the healthcare industries and to provide the best care for patients, it is necessary to predict the impact of the disease. Most of the organizations and researchers are moving towards Machine Learning algorithms to develop predictive models, trying to reduce the death rate to identify patients who are at the increased risk of mortality and providing support to them [6].

The objective of this paper is to build a predictive model using Artificial Neural Network (ANN) Algorithm for the corona infected persons. In this study the data set from *Kaggle Sírio-Libanês data for AI and Analytics by the Data Intelligence Team* has been used to build the predictive model. It is observed that there is 73 % of accuracy in predicting the number of corona infected cases. Organisation of this paper : Following Section II shows the outline of the survey which focus on various data set used to build the model, ML algorithms applied for prediction and the performance metrics used to find the quality of the proposed work. Section III elaborated the ANN algorithm to build the predictive model. Section IV shows the Implementation of the ANN model. Section V ends with conclusion and future work.

Observations from the survey

Based on the following questions towards research a survey has been performed and analysed

(i) What are the Covid -19 data set used by several researchers?

(ii) What are the machine learning algorithms used by the researchers for the classification and prediction of Covid -19 infections?
(iii) What are the performance metrics used by the researchers to find the quality of the proposed predictive model?
(iv) What are the parameters or characteristics used by the researchers in order to identify corona symptoms?

Existing Machine Learning Based Approaches on COVID-19

From the survey it is observed that most of the researchers towards COVID -19 used Machine learning algorithms such as Decision tree, Naïve Bayes, Random forest, Regression model and Support Vector as and evaluated the model using Accuracy Precision, Recall, F1 Score. Table 2.1 shows the comparison on the above said arguments.

Problem Description	Data Set	Algorithm	Performance Metrics	Reference
To monitor the risk of COVID-19 in public areas.	Multidisciplinary University Research Initiatives (MURI) dataset	Decision tree, Naïve Bayes, Random forest, K Nearest Cluster	TP rate, FP rate, precision, Recall, F-measure, MCC, ROC area, PRC area, Kappa, MAE, RMSE, RAE, and RRSE	[7]
To identify the symptoms of the coronavirus	—--	Rule-based approaches	Accuracy Precision , Recall, F1 Score	[8]
To determine COVID-19 in the patients	Israeli government website	---	---	[9]
To classify the disease into four different categories	Open-source data repository GitHub	Support vector machine (SVM), Naı¨ve Bayes (MNB), Logistic regression, Decision tree, Random forest, Bagging, Adaboost,Stochastic gradient boosting	Precision, Recall, F1 score, Accuracy	[10]
Proposed a predictive model to identity Covid -19	Collected from participants personal device data, self-reported symptoms and diagnostic test results during the	Decision Tree	Accuracy	[11]

	data collection period.			
Developed a predictive model for covid -19	GitHub registry, supplied by Johns Hopkins University,	Regression model (LR), the lowest absolute and selective shrinking operator (LASSO), vector supports (SMS) and Exponential smoking (ES)	R-squared value (R2 value), R-squared modified score (R2), MSD, MSD (MEA), and RMSE (Root Medium Square Error)	[12]
To predict mortality in COVID-19 hospitalized patients	The registry of Ayatollah Taleghani Hospital, Abadan city, Iran	J48 decision tree, random forest (RF), k-nearest neighborhood (k-NN), multi-layer perceptron (MLP), Naïve Bayes (NB), eXtreme gradient boosting (XGBoost), and logistic regression (LR)	Sensitivity, Specificity, Accuracy, Precision, and ROC	[13]

Table 2.1: Data set, ML Algorithm and Performance Metrics used for Covid -19 Detection & Prediction

Predictors used by the researchers to identify corona symptoms

Elbasi et al., [7] initially detected distance between the crowd objects in the first phase, because when a sick person coughs, sneezes, or talks, COVID-19 is transmitted through respiratory droplets expelled from the nose or mouth, and hence it is so important to stay at least one meter away from others. Then the author extracted features of the corona in the second phase using high-level processing. In his proposed work he has used 800 crowd data vectors {sw V1, V2, ..., V800} for a total of 7200 data and used VL531x time of flight (ToF) sensor to measure distance between human objects, LM34/TMP34, LM35/TMP35, or TMP36 temperature sensors to detect temperature and the computer vision-based Raspberry Pi single-board computer equipped with a camera module for mask detection.

Sample UNSAFE Class: CONDITION: (Distance {A, B} ≤ 0.7) && (Distance {A, C} > 0.8) && (Mask {A, B} = 0) where A, B, and C are three different people in the same crowd.
Sample SAFE Class CONDITION: (Distance {A, B} ≤ 0.7) && (Distance {A, C} > 0.8) && (Mask {A, B} = 1) SET: Class = {Safe}
Hamid Mukthar et al., [8] proposed a classification model using rule based algorithm where he classified corona symptoms in to the following four classes.

Class	SpO2	Cough rate	Heartbeat Rate	Temperature
Class - “0” Non-symptomatic	≥ 95%	NIL	≤ 100 bpm;	≤ 37.2 ◦C;
Class 1: Mild symptoms	≥ 95%;	≤ 5/min;	≤ 100 bpm;	36 ◦C to 38 ◦C;
Class 2: Moderate clinical symptoms	93% to 94%;	5/min to 30/min;	> 100 bpm;	e > 100 bpm;
Class 3: Serious clinical symptoms	≤ 92%	≥ 30/min;	> 120 bpm;	> 38 ◦C.

Table 2 .2 : Predictor Variables used for Covid -19 Detection & Prediction (Hamid Mukthar et al.,)

Matteo Gadaleta, et t al., [11] considered the following features to identify Covid 19 and classify in to 3 different categories

Highly discriminative symptoms	Cough and decrease in taste and smell, with ≥10% relative contribution
Medium discriminative symptoms	Runny nose, fever, chills or sweating and congestion or runny nose with <10% and ≥5% relative contribution
Low discriminative Symptoms	Body aches, headache, fatigue, with < 5% relative contribution

Table 2 .3 : Predictor Variables used for Covid -19 Detection & Prediction (Matteo Gadaleta, et al.,)

Building Predictive model using ANN Algorithm

Artificial Neural Networks (ANN) method with historical and present day data to estimate destiny values via associating complex correlations hidden inside the data just like the human brain. Figure 3.1 suggests the structure of a neural-network algorithm which has 3 layers and the black circles constitute nodes of the neural network [14].

Artificial Neural community has the subsequent 3 layers.

- The input layer feeds past or preceding data values into the next (hidden) layer.
- The hidden layer encapsulates several complicated features that create predictors and those features are hidden from the user. The nodes inside the hidden layer constitute mathematical features that alter the input records.

- The output layer collects the predictions made inside the hidden layer and produces the very last result is the prediction given through the ANN version.

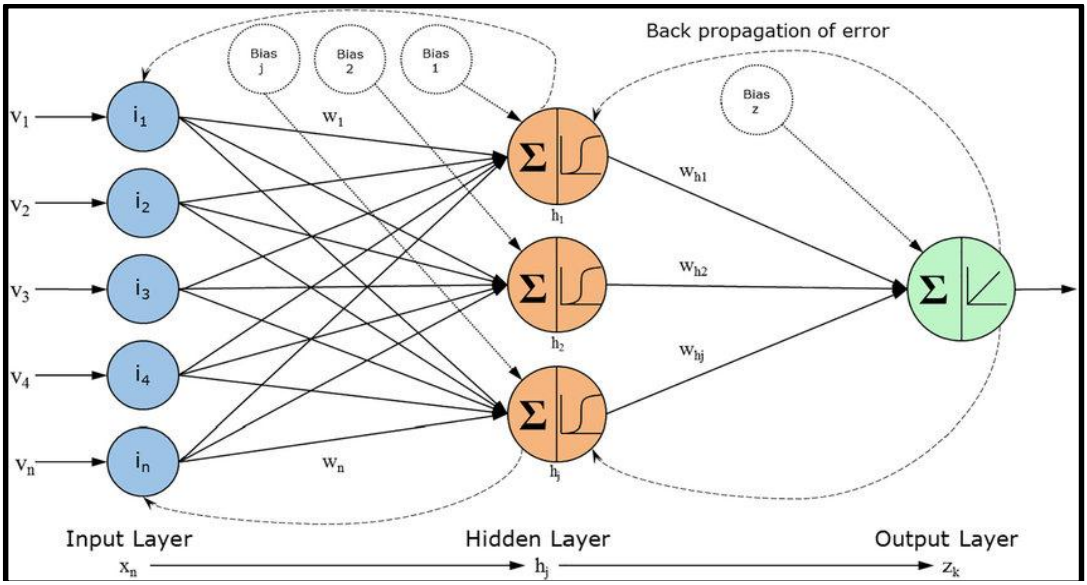

Figure 3.1 . The Structure of Artificial Neural Network (ANN) Algorithm

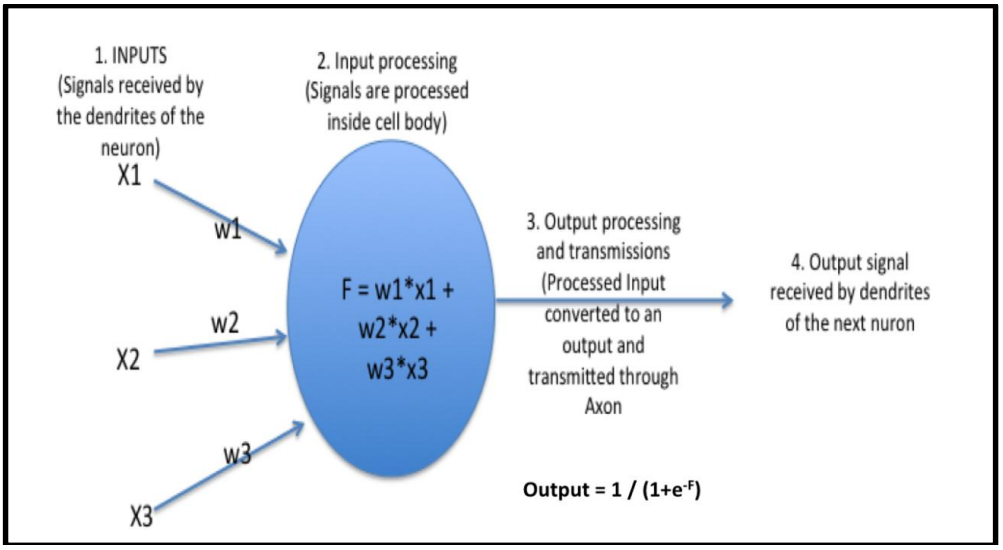

Figure 3.2 . Calculation of activation function in (ANN) Algorithm

Every neuron takes a set of input values which has been related to a *weight* and a numerical cost known as *bias*. The output of every neuron is a characteristic of the output of the weighted sum of every input plus the bias. Maximum of the neural networks use mathematical features of the neurons. A *function* is a relation between a set of inputs and a set of outputs, with the rule of thumb that every input corresponds to an output. The weight for each neuron is a numerical

value that can be derived using either supervised learning or unsupervised learning along with records clustering.

- In supervised learning, weights are derived with the aid of feeding pattern inputs and outputs to the set of rules until the weights are tuned till there's a near-perfect match between inputs and outputs.
- Inside the case of unsupervised learning, the neural network is only presented with inputs and the set of rules generates their corresponding outputs.

Each neuron has some weights (in above picture w1, w2, w3) and biases and based in this computations are done as, *combination = bias + weights * input (F = w1*x1 + w2*x2 + w3*x3)* and eventually activation function is implemented *output = activation(combination)* in above picture activation is sigmoid represented by 1/(1 + e-F) [15].

Advantages

- ANN has the capability to analyze and model non-linear and complex relationships as many relationships between input and output are non-linear.
- In the case of supervised training, ANN can infer unseen relationships from unseen statistics, and therefore it is generalized.
- In the case of unsupervised training ANN does not have restrictions on datasets like facts have to be Gaussian distributed or every other distribution.

Artificial Neural Network (ANN) Implementation

System Configuration

The following configuration & tool has been used to build predictive model.

- Core i5 processor
- Memory with 8GB RAM
- 1TB Hard Disc
- Python has been used to build the predictive model,
- Jupyter Notebook (Anaconda 3) platform has been used to build the model

Data Set

Covid -19 data set has been obtained from Kaggle Sírio-Libanês data for AI and Analytics by the Data Intelligence Team [16]. Data were obtained and grouped by windows in chronological order. Data set contains *Patient_Visit_Identifier,Bloodpressure_Diastolic_Mean,Bloodpressure_Sistolic_Mean, Heart_Rate_Mean, Respiratory_Rate_Mean, Temperature_Mean, Oxygen_Saturation_Mean, Bloodpressure_ Diastolic_ Mean, Window, ICU.* Totally the data set has 1925 records and 8 predictors or columns. ICU is target output and Window is not considered as predictor variable.

Window	Description
0 -2	From 0 to 2 hours of the admission
2 -4	From 2 to 4 hours of the admission
4 - 6	From 4 to 6 hours of the admission
6 - 12	From 6 to 12 hours of the admission
Above 12	above 12 hours of the admission

Table 4.1. Patient observation

Data Pre-processing

It is possible to get quality output only if we process with quality input data and hence it is necessary to pre-process the input data. Hence the collected data is cleaned and the null values are removed using the following steps.

Data Cleaning

All the missing data have been filled with the mean value of the attribute and the noisy which may be generated due to faulty data collection, data entry errors has been removed by clustering method. This approach groups the similar data in a cluster. The outliers may be undetected or it will fall outside the clusters.

Data Transformation

In order to transform the data in appropriate forms suitable for mining process. Normalisation has been done in order to scale the data values in a specified range (-1.0 to 1.0 or 0.0 to 1.0). Hence Data has been cleaned and scaled by column according to Min Max Scaler to fit between -1 and 1.

Out[1]:

	PATIENT_VISIT_IDENTIFIER	BLOODPRESSURE_DIASTOLIC_MEAN	BLOODPRESSURE_SISTOLIC_MEAN	HEART_RATE_MEAN	RESPIRATORY_RATE_MEAN
0	0	0.086420	-0.230769	-0.283019	-0.593220
1	0	0.333333	-0.230769	-0.132075	-0.593220
2	0	NaN	NaN	NaN	NaN
3	0	NaN	NaN	NaN	NaN
4	0	-0.243021	-0.338537	-0.213031	-0.317859

Figure 4.1. Transformed Data set

Building Predictive Model using Artificial Neural Algorithm

Initially the records set is fed into the input layer which then transfers it to the hidden layers, and interconnection among those two layers assign weights to each

input randomly at the initial factor. Then bias is brought to each input neuron and after this, the weighted sum which is a aggregate of weights and bias is surpassed thru the activation function. Activation function has the duty of which node to remove for characteristic extraction and finally the output is calculated. This entire procedure is known as forward Propagation. . After getting the output model to compare it with the original output the error is known and finally, weights are updated in backward propagation to reduce the error and this process continues for a certain number of epochs (iteration). Finally, model weights get updated and prediction is done.

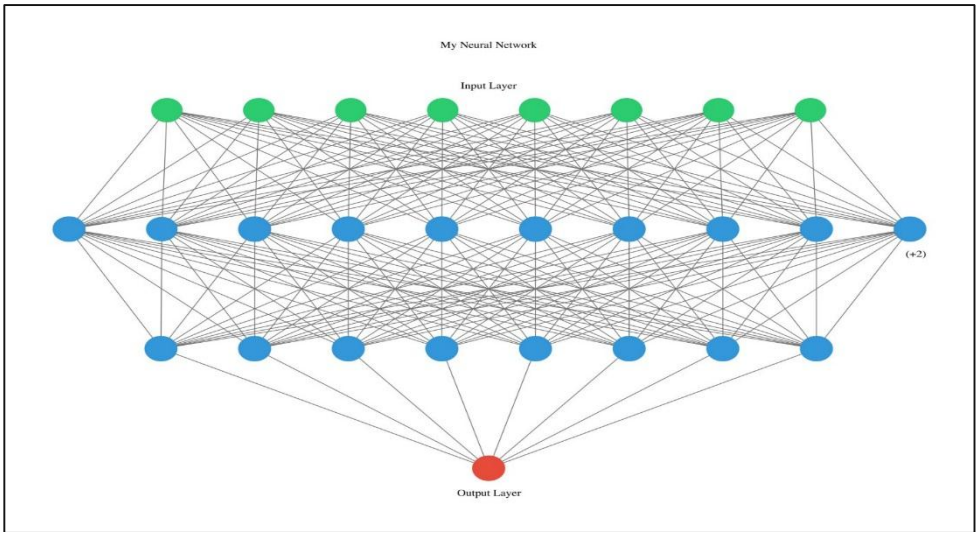

Figure 4.2. ANN Predictive Model

```
Epoch 5/500
49/49 [==============================] - 0s 1ms/step - loss: nan - accuracy: 0.7364
Epoch 6/500
49/49 [==============================] - 0s 1ms/step - loss: nan - accuracy: 0.7364
Epoch 7/500
49/49 [==============================] - 0s 1ms/step - loss: nan - accuracy: 0.7364
Epoch 8/500
49/49 [==============================] - 0s 1ms/step - loss: nan - accuracy: 0.7364
Epoch 9/500
49/49 [==============================] - 0s 1ms/step - loss: nan - accuracy: 0.7364
Epoch 10/500
49/49 [==============================] - 0s 1ms/step - loss: nan - accuracy: 0.7364
Epoch 11/500
49/49 [==============================] - 0s 1ms/step - loss: nan - accuracy: 0.7364
Epoch 12/500
49/49 [==============================] - 0s 1ms/step - loss: nan - accuracy: 0.7364
Epoch 13/500
49/49 [==============================] - 0s 1ms/step - loss: nan - accuracy: 0.7364
Epoch 14/500
49/49 [==============================] - 0s 3ms/step - loss: nan - accuracy: 0.7364
```

Figure 4.3. Accuracy of the ANN Predictive Model

Conclusion

In this paper the observations from the survey of Machine Learning Algorithm in a covid-19 context from different perspectives such as (i) covid-19 data set used by several researchers (ii) the parameters or characteristics used to identify corona symptoms (iii) various machine learning algorithms used by several researchers for classification and prediction (iv) various performance metrics used by several researchers to find the quality of the predictive model have been performed. In addition to that, a predictive model to identify COVID -19 cases has been built using Artificial Neural Network (ANN) Algorithm and the prediction accuracy of the predictive model has been evaluated as 73%. In future it is planned to compare ANN model with Decision Tree, Support Vector Machine and Regression model and the best model will be selected to detect the infected persons earlier and it is also planned to send an alert message to the infected persons as well as to the management of their working place.

References

1. Mohammed A Soghaier, Khwaja MI Saeed, and Khushhal K Zaman, "Public health emergency of international concern (PHEIC) has declared twice in 2014; polio and ebola at the top" AIMS Public Health, 2015 Jun 5;2(2):218-222. doi: 10.3934/publichealth.2015.2.2.
2. Nora EI Rishidy, Sameer Abdelrazik, Tamue Abumad,"Comprerehensive Survey of Using Machine Learning in the COVID-19 Pandemic" ,Diagnostics 2021, 11(7), 1155; https://doi.org/10.3390/diagnostics11071155
3. Wang, N. Fu, Y.; Zhang, H. Shi, H, " An evaluation of mathematical models for the outbreak of COVID-19" . Precis. Clin. Med. 2020, 3, 85–93
4. Chae, S.Y. Lee, Mafwele," Estimation of Infection Rate and Predictions of Disease Spreading Based on Initial Individuals Infected With COVID-19" ,Front. Phys. 2020, 8, 1–6.
5. Mahdi Mahdavi , Hadi Choubdar ,Erfan Zabeh , Michael Rieder, "A machine learning based exploration of COVID-19 mortality risk" PLOS ONES, July 2, 2021https://doi.org/10.1371/journal.pone.0252384
6. Yazeed Zoabi, Shira Deri-Rozov & Noam Shomron ,"Machine learning-based prediction of COVID-19 diagnosis based on symptoms" NPJ Digital Medicine volume 4, Article number: 3 (2021)
7. Elbasi, E.; Topcu, A.E.; Mathew, S " Prediction of COVID-19 Risk in Public Areas Using IoT and Machine Learning" Electronics 2021, 10, 1677. https://doi.org/10.3390/ electronics10141677
8. Hamid Mukhtar, Saeed Rubaiee,Moez KrichenAn IoT Framework for Screening of COVID-19 Using Real-Time Data from Wearable Sensors,Int. J. Environ. Res. Public Health 2021, 18(8), 4022; https://doi.org/10.3390/ijerph18084022
9. Marium Malik1, Muhammad Waseem Iqbal1,*, Syed Khuram Shahzad2, Muhammad Tahir Mushtaq2, Muhammad Raza Naqvi3,4, Maira Kamran1, Babar Ayub Khan4, Muhammad Usman Tahir4,Determination of COVID-19 Patients Using Machine Learning Algorithms,Tech Science Press,Vol.31, No.1, 2022, pp.207-222, doi:10.32604/iasc.2022.018753
10. Akib Mohi Ud Din Khanday, Syed Tanzeel Rabani, Qamar Rayees Khan, Machine learning based approaches for detecting COVID-19 using clinical

text data,International Journal of Information Technology ,volume 12, pages 731–739 (2020)
11. Matteo Gadaleta, Jennifer M. Radin, Katie Baca-Motes, Edward Ramos, " Passive detection of COVID-19 with wearable sensors and explainable machine learning algorithms" , NPJ Digital Medicine , volume 4, Article number: 166 (2021)
12. Ramesh Kumar Mojjada a,, Arvind Yadav a , A.V. Prabhu b , Yuvaraj Natarajan c, "Machine learning models for covid-19 future forecasting " Materials Today: Proceeding , 214-7853/ 2021 Elsevier Ltd. s, https://doi.org/10.1016/j.matpr.2020.10.962
13. Khadijeh Moulaei, Mostafa Shanbehzadeh, Zahra Mohammadi‑Taghiabad and Hadi Kazemi‑Arpanahi,"Comparing machine learning algorithms for predicting COVID-19 mortality" BMC Medical Informatics and Decision Making (2022) 22:2 https://doi.org/10.1186/s12911-021-01742-0
14. Hamid Reza Niazkar, Majid NiazkarApplication of artificial neural networks to predict the COVID-19 outbreak , in Global Health Research and Policy (2020) 5:50, Policy (2020) 5:50, https://doi.org/10.1186/s41256-020-00175-y
15. Nawaf N. Hamadneh, Waqar A. Khan, Waqar AshrafArtificial Neural Networks for Prediction of COVID-19 in Saudi Arabia, Vol.66, No.3, 2021, pp.2787-2796, doi:10.32604/cmc.2021.013228
16. https://www.kaggle.com/code/divendraseeram/covid-19-icu-prediction#Building-a-Predictive-Model-for-COVID-19-ICU-Admission