IASJ المجلات الاكاديمية العراقية

Tikrit Journal of Pure Science

ISSN: 1813 – 1662 (Print) --- E-ISSN: 2415 – 1726 (Online)

Journal Homepage: http://tjps.tu.edu.iq/index.php/j

Prediction of Corona-Virus Using Deep Learning

Laith Al-Ali

Computer Engineering, Okan University, Istanbul, Turkey

https://doi.org/10.25130/tjps.v27i1.89

ARTICLE INFO.

Article history:

-Received: 5 / 9 / 2021

-Accepted: 15 / 12 / 2021

-Available online: / / 2022

Keywords: COVID-19, artificial neural network, artificial intelligence, random forest, decision tree, deep learning.

Corresponding Author:

Name: Laith Al-Ali

E-mail: Laith.ibrahim2@gmail.com

Tel:

ABSTRACT

With the rapid spread of the Corona virus in most parts of the worldwide, it has become necessary to find solutions to contain and treat this epidemic. This research presents a method to predict the occurrence of COVID-19 based on different symptoms of the disease, using non-clinical methods such as artificial intelligence, to help medical staff, save the cost of testing (PCR), and get results in a short time. Artificial intelligence provides many tools for data analysis, statistical analysis, and intelligent research. In this paper, we focus on predicting COVID-19 infection, using Artificial Neural Networks (ANN), random forests and decision trees, to effectively analyze medical datasets, based on the most common and acute symptoms, such as cough, fever, headache, diarrhea, living in infected areas Pain and shortness of breath. Breathing, chills, nasal congestion and some other symptoms of the disease. A data set consisting of (1495) patients is used to determine whether or not a person has this disease, after determining the symptoms that appear on it. The data set is divided into 75% of the training data and 25% of the test data after applying deep learning algorithms. Python libraries such as pandas, NumPy, and matplotlib are also used in addition to sklearn and Keras. The search results show very high accuracy indicated by 91% of Random Forest with estimators = 200 and 91% of the decision tree. the accuracy of an artificial neural network is 85%. Thus, this research provides an important indicator for the possible prediction of COVID-19 infection.

Introduction

Coronavirus was initially detected in the Chinese city of Wuhan in December 2019. This virus has caused unknown cases of pneumonia[1]. also known as SARS-CoV-2[2]. The World Health Organization (WHO) declared on August 5, 2021, that COVID-19 is a global pandemic[3] due to the great health tragedy in all countries of the world when the number of infections reached more than 200,586.376 million cases and 426,2114 million deaths. Serious damage occurred to the health systems due to its rapid spread among humans. According to the World Health Organization, our reliable source in our study, some symptoms of COVID-19 belong to the category of cold, fever, cough, shortness of breath, loss of taste, and acute respiratory syndrome[4]. A person with COVID-9 can recover within two weeks or can last longer. This pandemic poses a constant threat to all societies due to the lack of fully effective treatment or vaccine. Thus, researchers around the world are taking projects to help health physicians/practitioners to reveal and stop the spread of this disease by relying on artificial intelligence, for example, to build prediction models that can optimally allocate the medical resources to assist the health care systems against this pandemic[5]. In this regard, mathematical, dynamic and statistical methods have been used to predict the outbreak of COVID-19[6-10]. Early detection of COVID-19 cases is important to isolate the patients with this disease in order to reduce any possible transmission of the virus to others or have an appropriate treatment before worsening their health condition. It is well known that the primary method for diagnosing COVID-19 infection is RT-PCR, to obtain the result which takes about 4-6 hours or it probably takes several days in some health centers. However, there is a shortage of

RT-PCR test kits, and the costs of this test are high. To overcome this inadequacy and reduce the cost of testing and speed-up obtaining the results, computational models have been performed in deep learning that includes an artificial neural network and Random Forest and Decision Tree. These models are used to estimate/explore if a patient has this virus or not.

Objectives of the study

Artificial intelligence can greatly contribute to humanitarian assistance, especially in the medical field. Therefore, this research aims to use artificial intelligence find simple and low-cost solutions to detect COVID-19 infection and invent the best prediction model which delivers correct and accurate results, diagnoses the disease, and make it easy to use.

Literature survey

In general, ANNs have been used in many research areas of medicine, such as electroencephalogram[11], A decision tree has been used to help identify people with Covid-19, through the use of CT scans and chest X-rays [12], (ANN) was developed to detect lung cancer based on symptoms such as fatigue, cough, chest pain, etc., using a data set (Lung Cancer Survey), which showed an efficiency of the model 96.67% in terms of accuracy [13], An AI tool that uses deep learning is designed to quickly predict the spread of saliva droplets produced by a sneezing passenger on a bus[14], The AI is able to extract specific graphic features of COVID-19 and provide a clinical diagnosis prior to pathogen testing, relying on CT images.[15]. Important deep learning research has been done on cardiac diagnostic systems [16-17]. The experimental results proved that deep learning algorithms in neural networks gave high accuracy and reduced the error rate in training algorithms[18]. The backpropagation neural network was used to identify patients undergoing keratectomy, and the backpropagation neural network was trained with 100% accuracy[19]. The neural network was used in the data of lymph node-positive patients and got correct results about the potential tumor recurrence rate in 960 out of 1008 cases[20]. (ANN) have been applied to predict the outbreak of COVID-19. Also, ANN was used to predict the recoveries and deaths using mist clinical characteristics[21]. The (ANN) has shown that it is possible to predict the future cases of an outbreak very efficiently in a country, and it is possible to predict the growth in cases of COVID-19 virus in different countries with the expectation of improving cases in one week, and predicting the future cases in any country as well[22]. AI-based statistical approaches have been proposed to predict the spread of COVID-19. Autoregressive moving average (ARIMA) and autoregressive nonlinear artificial neural networks (NARANN) were used. The results exposed good agreement with the officially announced cases[23]. A system for predicting heart disease was developed using the J-48 decision tree algorithm and used in different ways to predict, implement and conclude the cause of heart attacks and strokes is due to heart disease, in addition, due to neglect of personal health and wrong diet, not exercising sports, the accuracy of the model was 68% [24]. Maji et al. (2019) has suggested a hybrid technique in which an artificial neural network is used with a decision tree to achieve better performance in predicting heart disease using WEKA program[25]. The model was tested on a heart patient dataset from the UCI repository to analyze the accuracy and sensitivity of the model singles and Hybrid Technology ten times[25]. A model of machine learning algorithms is designed to calculate and evaluate the performance of each algorithm, to understand the associations between different features of the data set, and to understand the impact of COVID-19 on people and recovery times. The results showed that the Random Forest Regressor and Random Forest Classifier algorithms outclassed other models such as KNN + NCA, SVM, Decision Tree Classifier, Multilinear Regressor, Gaussian Naïve Bayesian Classifier, XGBoost Classifier and Logistic Regressor in terms of accuracy[26].

1. Neural networks

An artificial neural network is a model similar to the human nervous system, in human neurons, they are connected to each other by synapses [(27)]. Neurons are computational units that receive inputs from other neurons, The input is calculated and fed to other neurons. Sometimes the behavior of each neuron changes depending on the stimuli. This stimulus is external to learn these weights through the training data set. When the predictions in the current weights are incorrect by changing these weights appropriately it is gradually changed to achieve the desired results [(28)]. The basic structure of a neural network is simple single-layer or complex and multi-layer [27]. The efficiency of an ANN depends on the architecture used to arrange connections between nodes[28].

1.1 Backpropagation algorithm

A neural network mostly consists of three various layers with a feed-forward structure. The input layer of the network is a set of input units, which are completely linked to the hidden layer by hidden units. These, in turn, are related to the output layer by the units of the output layer. These layers provide a neural network with a response to the activation pattern that is applied to the input layer. The information provided to the neural network is processed from one layer to another, that is, from the input layer to the output layer, across one or more hidden layers. The specification of the number of hidden layers and the number of units in the hidden layers are crucial matters in the Multilayer Perceptron (MLP) design. The number of inputs and outputs is determined by the problem since the number of hidden units and layers are far from clarity. However, the definition of one hidden layer is enough to resolve most problems, as using more than one hidden layer

does not help most of the time because it is slow. It is important to choose an appropriate number of hidden nodes to avoid the occurrence of overprocessing. A common method is usually used by starting with a higher number of nodes, thereby using network segmentation algorithms for optimization[29].

The backpropagation algorithm consists of two main stages that are applied in the process of updating the weight for each training condition:

1.1.1 Forward propagation stage

The forward propagation stage begins with presenting what the network input looks like, where each processing element from the input layer is assigned to a packet component that represents the input. The values of the input vector components cause the units of the input layer to activate, followed by the forward propagation of that excitation through the rest of the network layers, i.e., the network is fed by a front-end feed system and no modulation occurs weigh the correlations during this phase.

1.1.2 Reverse propagation stage:

The reverse propagation stage is an inverse method of updating the weights of all nodes, the weights adjustment phase where the network outputs are compared during training for the set of integer shapes fed from the outside and the difference between the two is calculated. Then the relative weights of the connection nodes are adjusted in order to process the elements, where the signal allows. By propagating from the output layer to the input layer in the reverse direction during the weight adjustment phase, the process is repeated until the network output matches the specified integer shapes.

1.2 Selection of weights

Weights are selected, on all connections, in small random numbers[30], (untrained). The weights of each neuron are randomly chosen and range from +1 to -1. In ANN, the weight training is usually created by decreasing the error function [X_{in}], the mean squared error, between the desired target and the actual output by repeatedly adjusting the weights. To calculate the grid by changing the weights, the weights are corrected to decrease the rate of error. The likelihood of a successful rappr0chement depends on the weight configuration scheme[30].

1.3 Hidden layer:

It is a set of neurons between the input and output layer[31]. It calculates the weighted sum of the inputs and weights, adds bias and performs the activation function[32]. Usually, one hidden layer is enough for the vast common problems, but many unnecessary neurons in the network, "Overfitting"[31] may occur. The optimal number of neurons in a hidden layer can be determined using equation(1):

$$\text{NUMBER OF NEURONS}=\frac{TRADING\ DATA\ SAMLES}{FACTOR*(INPUT\ NEURONS+OUTPUT\ NEURONS} \quad \ldots(1)$$

The factor is used to prevent overfitting and a number is between 1-10[32].

Epoch

Epoch refers to a single cycle through the full training data set. Sometimes it takes training tens of times to thousands of ages to get reasonable accuracy in the test data[33].

Loss

It is a method for evaluating the quality of the performance of an algorithm model for a data set. The lower the loss value indicates that the model is working efficiently, and the loss is related to the accuracy of the model[34].

Recall

The fraction of the True Positive (TP) samples divided by the total summation of False Negative (FN), and True Positive (TP) samples is called Recall. Equation (2) can be used to calculate the Recall[35]:

$$Recall=\frac{TP}{TP+FN} \quad \ldots.(2)$$

Precision

The fraction of the True Positive (TP) samples divided by the total summation of True Positive (TP), and False Positive (FP) samples is called precision. Equation (3) can be used to calculate the Precision[35]:

$$Precision=\frac{TP}{TP+FP} \ldots..(3)$$

F1 Score

F1 Score or F score can be defined as the harmonized mean of Precision and Recall values. The F1 Score reaches the ideal balance between Recall and Precision giving a correct estimation of the model's execution in categorizing the patients of COVID-19. F1 Score can be determined using Equation (4)[35]:

$$F1\ Score=2\ x\frac{Precision\ x\ Recall}{Precision+Recall} \ldots\ldots(4)$$

Advantages of Classifying Neural Networks[36]:

1. Neural networks are stronger because of the weights.
2. Neural networks improve their performance through learning.
3. The error rate is low and therefore has high accuracy with the only proper training procedure.
4. Neural networks are more powerful in a noisy environment.

2. Random Forests Algorithm

Random forests are defined as a set of decision trees, where the model is generated randomly for each decision tree[37], and then combined to obtain more accurate and stable predictions, and several specific decision trees are produced applying a trait selection index like information gain ratio or genetic index for each trait. every single tree is based on a random independent sample. Random forest is based on a divide and conquer approach[38].

Main advantages:

1. Random forests are a very accurate and efficient method because of the number of decision trees involved in the process.
2. There is no overfitting problem.
3. The algorithm can be utilized for regression and categorization problems.

Disadvantages:

1. Random forests generate predictions slowly as they contain several decision trees.
2. The model is hard to understand compared to a decision tree because in the later it can easily see how the decision is made by following the path in the tree.

3. Decision Trees Algorithm

The decision tree is a tree-like schema consisting of nodes and paths, each node represents a test attribute, each branch represents a test result, and each attribute (node) corresponds to the path it represents from root to tip. The decision tree consists of two stages: (the growth stage and the pruning stage). Decision trees are machine learning model, achieves high accuracy in many tasks and are easy to understand[39].

Basic steps in medical diagnosis based on artificial intelligence

The most important and basic steps to be taken for an accurate medical diagnosis and to predict a specific disease are two steps: the first step - patient data, and the second step - identification of features that are important and influential in determining the disease (symptoms, mechanistic data, etc.) which provides the necessary information to know the patient's condition. Later, when a database, that impacts the disease, is created, it must be carefully selected, validated and cleaned "of outliers and avoiding over-information". The data then is divided into (Training and Testing) and consequently, the prediction algorithms are used to diagnose the patient, and hence the expected diagnosis is evaluated by the clinical specialist.

The main steps can be summarized as follows[40]: In order to predict and reach an accurate diagnosis of the disease:

- Features selection.
- A database building.
-Data cleaning and validation.
-Practice perfect data processing.
-Training and validating databases using the Training algorithm.

Methodology

The data set used in this research was collected from Kaggle[41], which contains 1495 cases of patients with Corona Virus. The data included the characteristics and symptoms of COVID-19 namely: age, gender, fever, cough, fatigue, pains, nasal congestion, shortness of breath, runny nose, sore throat, diarrhea, chills, headache, vomiting, and lives in the affected area. Several algorithms were utilized, such as ANN in deep learning, and Random Forest, and decision tree in machine learning. These algorithms were applied to all symptoms mentioned above to determine whether the patient is infected with COVID-19 or not.

ANN algorithm structure was used, where the number of entries was 15 characteristics or symptoms of COVID-19 disease obtained from our search data set and used with one hidden layer containing 20 neurons with 150 batch size and 1500 epochs. In machine learning, two algorithms, Decision Tree and Random Forest were used as, 200 n_estimators.

Results and Discussion

The Effect of Selected Features

Among the fifteenth selected features, Figure (1) shows the correlation matrix of the most important ten features when ANN is applied to detect the COVID-19 infection. Applying ANN algorithm explored that the optimal influential features among the ten symptoms in determining whether a patient has COVID-19 or not are as follows: the fever with 53%, the (lives_in_affected_area) with 30% followed by cough with 20%. While the rest ranges from 8% with nasal congestion to 2% for the shortness of breath and chillness, as shown in Figure (1).

Figure (2) shows that 77.18% of the people with fever has been identified as infected people with COVID-19 (≈ 574 people out of 745, Figure 3), while only 24.26% of the people without fever has been identified as affected people with COVID-19 (≈ 186 people out of 750 people, Figure 3).

Therefore, ANN explores that the fever is the most important symptom which aids to discover the infected people and it can predict the number of infection cases depending on the fever symptom with high precision without the necessity to make the PCR test, Figures (1).

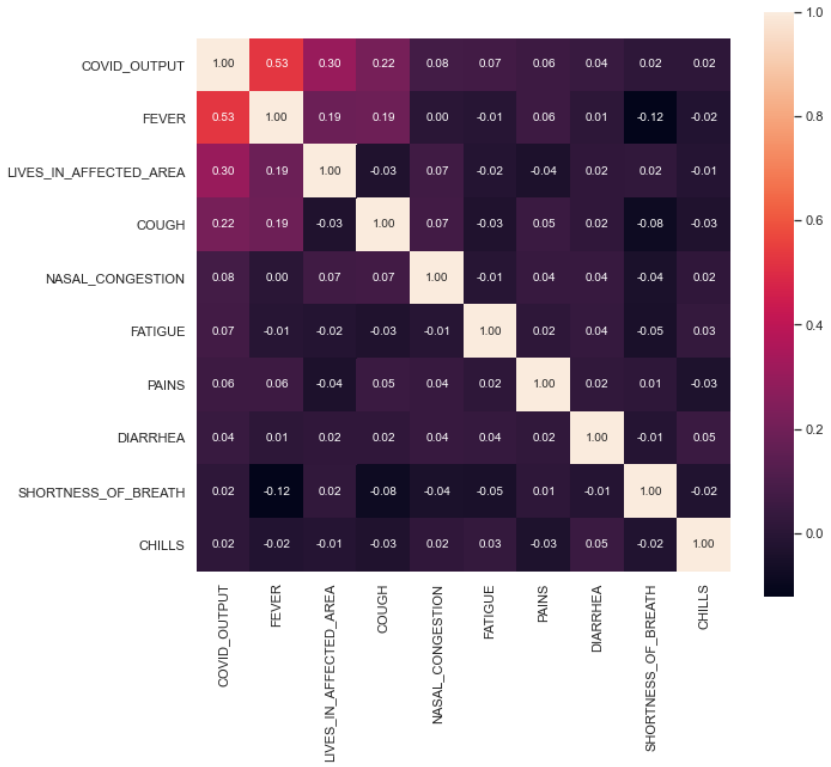


Fig. 1: The correlation matrix of the ten important attributes that affect COVID-19

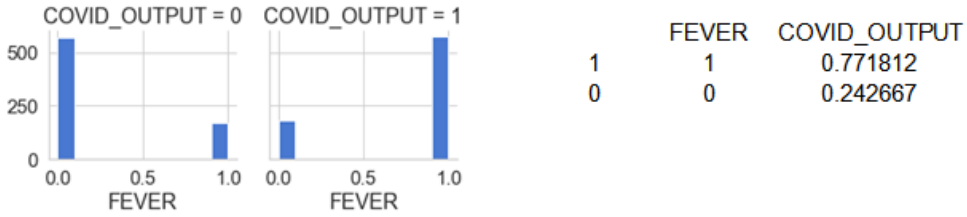


Fig. 2: Prediction ratio of COVID-19 infection based on fever symptoms using ANN.

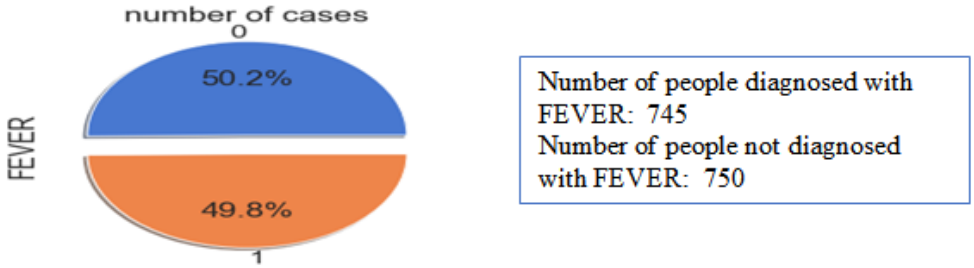
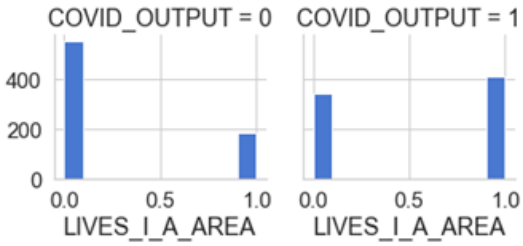


Fig. 3: The percentages and number of patients with fever and without fever symptoms.

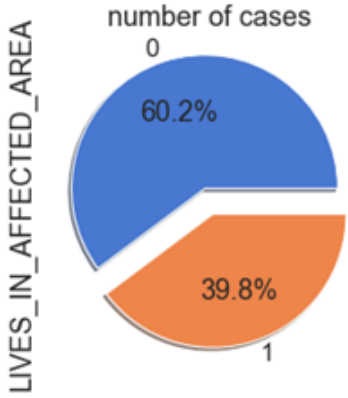
Similarly, Figures (4 and 5) display the prediction ratio of COVID-19 infection based on lives-in-affected areas and the percentages and number of the lives-in-affected-areas and lives-not-in-affected-areas, respectively.

Figure (4) shows that 38.33% of lives-in affected-areas has been recognized as infected people with COVID-19 (≈ 282 people out of 595, Figure 5), while only 69.24% of lives-not-in affected-areas has been recognized as affected people with COVID-19 (≈623 people out of 900 people, Figure 5).



	Lives_in_affected_area	COVID output
1	1	0.692437
0	0	0.383333

Fig. 4: COVID-19 infection rate based on LIVES-IN-AFFECTED-AREA

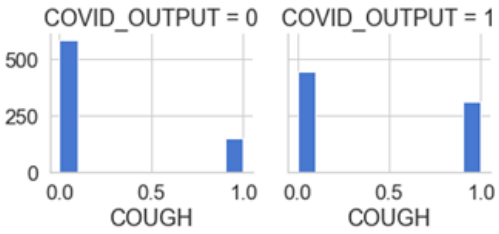


Number of persons diagnosed with LIVES_IN_AFFECTED_AREA: 595

Number of persons not diagnosed with LIVES_IN_AFFECTED_AREA: 900

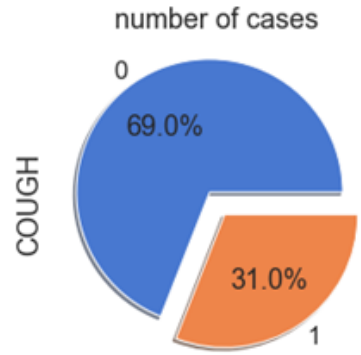
Fig. 5: The percentage of patients (living in the affected area, or not) from the dataset

In the same line, the cough, the third significant factor, can be used to predict of the infection cases. Figure (6) displays the prediction ratio of COVID-19 infection based on cough symptoms and Figure (7) demonstrates the percentages and number of the predicted infected people based on the cough characteristic. Figure (6) exhibits that 66.95% of people with cough symptom has been categorized as infected people with COVID-19 (≈ 691 people out of 1032, Figure 6), while only 43.31% of people with cough symptom has been categorized as affected people with COVID-19 (≈ 200 people out of 463people, Figure 7).



	COUGH	COVID_OUTPUT
1	1	0.669546
0	0	0.433140

Fig. 6: COVID-19 infection rate based on cough characteristic



Number of persons diagnosed with COUGH: 463

Number of persons not diagnosed with COUGH: 1032

Fig. 7: the percentage of patients (with cough, or without)

On the other hand, gender, age, headache, vomiting, and sore throat are the other characteristics that have shown no significant effect on the prediction of the COVID-19 infection. For example, the gender feature has not been seriously considered in our research due to the low rate of prediction of potential infection of the virus as shown in Figure (8). The infected males and females are shown according to the data set of our research. It has been noted that the number of infected and uninfected females is approximately equal, while for males, the number of infected is slightly greater than the uninfected males, as shown in Figure (8).

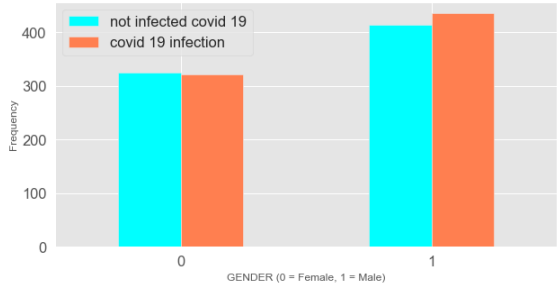


Fig. 8: Percentage of gender/COVID-19-output.

Efficiency/Accuracy of Algorithms

The efficiency and accuracy of the three algorithms applied to the collected dataset are determined. It is found that the efficiency/accuracy of the decision tree reached 91.17%, as shown in Table (1), and the Random Forest algorithm is 91% accuracy as well, as in Table (2). However, using the ANN algorithm shows an accuracy of 85% as shown in Table (3).

Table 1: Decision tree algorithm classification report

	Precision	Recall	F1-score	Support
0	0.77	0.79	0.78	201
1	0.75	0.73	0.74	173
Accuracy			0.76	374
Macro avg	0.76	0.76	0.76	374
Weighted avg	0.76	0.76	0.76	374
Accuracy decision tree: 91.17 Accuracy test decision tree: 76.2				

Table 2: Random Forest algorithm classification report

	precision	recall	f1-score	support
0	0.77	0.79	0.78	201
1	0.75	0.73	0.74	173
Accuracy			0.76	374
Macro avg	0.76	0.76	0.76	374
Weighted avg	0.76	0.76	0.76	374
Accuracy Random Forest: 91.17 Accuracy test Random Forest: 76.47				

Table 3: ANN algorithm classification report

	precision	recall	f1-score	support
0	0.80	0.82	0.81	201
1	0.79	0.77	0.78	173
accuracy			0.80	374
macro avg	0.80	0.79	0.80	374
weighted avg	0.80	0.80	0.80	374
Accuracy ANN: 85.28 Accuracy Test ANN: 79.14				

The results obtained from Tables 1-3 are combined and shown in Figure (9) to clarify the value of the accuracy of the Score Training and Score Testing using the three algorithms: ANN, Random Forest and Decision Tree algorithms.

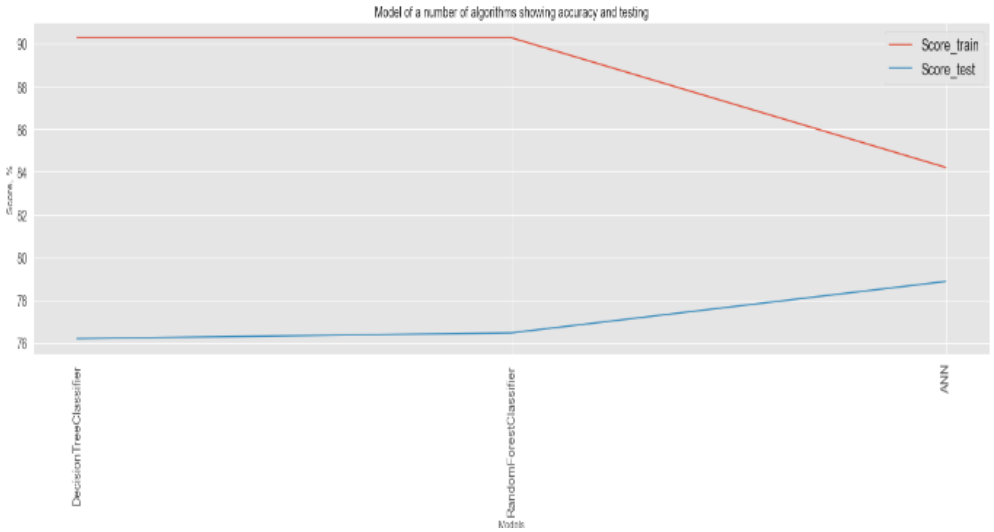


Fig. 9: Accuracy values of (Score train and Score test) using ANN, Random Forest and Decision Tree algorithms

The values of the Score Training & Score Testing of the three algorithms for each mode are shown in Table 4. Table 4 shows the values of the Score Training using Random Forest and Decision Tree are equally represented by 91.17. On the other hand, these values of the Score Testing are found to be 73.8 and 70.32 for the Random Forest and Decision Tree, respectively. In addition, the values of the Score Training and the Score Testing are found to be equal to 84.83 and 78.07, respectively.

Table 4: Values of the Score Training & Score Testing using the three algorithms

	Model	Score Training	Score Testing
1	Random Forest Classifier	91.17	73.80
2	Decision Tree Classifier	91.17	70.32
3	ANN	84.83	78.07

Table 5 shows the difference between the values of the Score Training and the Score Testing, which are 6.76, 17.37 and 20.85 for the ANN, RF, and DT, respectively. It can be seen that the small difference obtained using ANN provides a good indication with close values which means there is no over-fitting model[42].

Table 5: Differences between the Score Training & Score Testing using the three algorithms

	Model	Score Training	Score Testing	Score different
3	ANN	84.83	78.07	6.76
1	RF	91.17	73.8	17.37
2	DT	91.17	70.32	20.85

Figure (10) indicates the accuracy of the ANN algorithm in making a predictive estimation of the data, while the loss indicates the amount of error the algorithm made in the predictive guess. Figure (11) shows the confusion matrix of ANN. Applying the algorithm, show how the values are constantly changing throughout the training period, in which the algorithm is learning. The accuracy refers to how accurately an algorithm guesses previously seen data. The loss indicates the amount of error the algorithm made in the predictive guess. As a result, the higher the precision value and the lower the loss value, the better the performance of the algorithm [43].

Fig. 10: Train accurate and Train loss

Therefore, the Training accuracy and Testing accuracy are 85.11 and 79.67, respectively, and the Training loss and Testing loss are 36.19 and 51.60, respectively, which are greatly good.

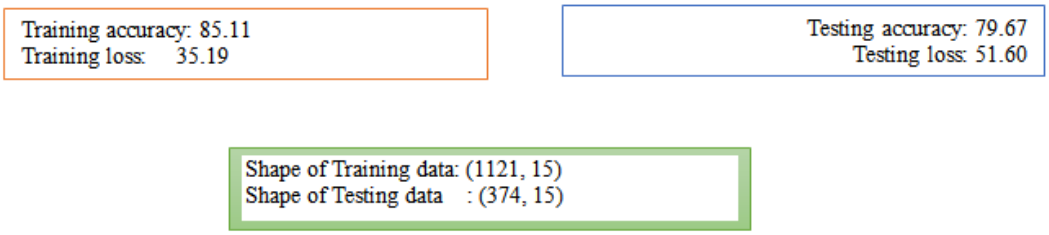


Fig. 11: The predicted values as Positive and Negative and actual values as True and False

Figure (12) presents the confusion matrix of ANN which indicates that for the True Positive: 459: the prediction is positive and it's true, the for a True Negative = 492, the prediction is negative and it's true; the False Positive = 97, the prediction is positive and it's false; and for the False Negative = 73, the prediction is negative and it's false.

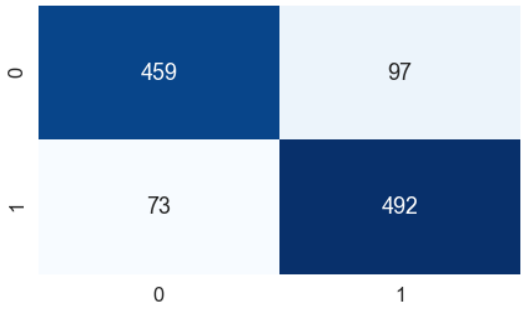


Fig. 12: Confusion matrix of ANN

It is known that the accuracy and efficiency of the algorithm can be changed by increasing the data set or changing its structure. However, the 85% accuracy rate of the ANN applied in the estimated search is considered highly good. In addition, the accuracy value of both the Random forest and Decision Tree algorithms, which are considered the best among the machine learning algorithms, are very high estimated by 91% for each.

Conclusion

These algorithms, ANN, Random Forest, and Decision Tree, are being applied to the medical dataset of COVID-19 patients. Based on the Python platform using libraries (pandas, NumPy and matplotlib) to process and analyze the dataset collected from Kaggle. One hidden layer containing 20 neurons was used, batch size = 150, epochs = 1500, and the data is divided into two sets of training data (75%) and test data (25%). Disease-related symptoms were used to predict infection. This algorithm is characterized by its ability to accommodate different data sets with their branches and large sizes. The results show high accuracy and low loss function, saving time, effective diagnosis and reducing cost. This research can be considered as a tool to facilitate the final medical decision.

References

[1] M. A. Shereen, S. Khan, A. Kazmi, N. Bashir, and R. Siddique, "COVID-19 infection: emergence, transmission, and characteristics of human coronaviruses," Journal of Advanced Research, vol. 24, pp. 91–98, 2020.View at: Publisher Site | Google Scholar.

[2] A. E. Gorbalenya, S. C. Baker, R. S. Baric et al., "The species severe acute respiratory syndrome-related coronavirus: classifying 2019-nCoV and naming it SARS-COVID-19," Nature Microbiology, vol. 5, no. 5, pp. 536–544, 2020.View at: Publisher Site.

[3] World Health Organization, Director-General's Opening Remarks at the Media Briefing on COVID, WHO, Geneva, Switzerland, 2020, https://www.who.int/director-general/speeches/detail/who-director-general-s-opening-remarks-at-the-media-briefing-on-COVID-19---11-march-2020.

[4] World Health Organization, COVID-19 Disease 2019 (COVID-19): Weekly Epidemiological and Operational Updates, WHO, Geneva, Switzerland, 2020, https://www.who.int/emergencies/diseases/novel-coronavirus-2019/situation-reports.

[5] RANI, K. Usha. Analysis of heart diseases dataset using neural network approach. *arXiv preprint arXiv:1110.2626*, 2011.

[6] Zhang S, Diao M, Yu W, Pei L, Lin Z, Chen D. Estimation of the reproductive number of novel coronavirus (COVID-19) and the probable outbreak size on the diamond princess cruise ship: a data-driven analysis. Int J Infect Dis. 2020;93:201–4. https://doi.org/10.1016/j.ijid.2020.02.033. Article PubMed PubMed Central CAS

[7] Fanelli D, Piazza F. Analysis and forecast of COVID-19 spreading in China, Italy and France. Chaos Solitons Fractals. 2020;134:109761. https://doi.org/10.1016/j.chaos.2020.109761.

[8] Yang J, Zheng Y, Gou X, Pu K, Chen Z, Guo Q, et al. Prevalence of comorbidities in the novel Wuhan coronavirus (COVID-19) infection: a systematic review and meta-analysis. Int J Infect Dis. 2020. https://doi.org/10.1016/j.ijid.2020.03.017.

[9] Nishiura H, Kobayashi T, Yang Y, Hayashi K, Miyama T, Kinoshita R, et al. The rate of under ascertainment of novel coronavirus (2019-nCoV) infection: estimation using Japanese passengers data on evacuation flights; 2020. https://doi.org/10.3390/jcm9020419.

[10] Nouvellet P, Cori A, Garske T, Blake IM, Dorigatti I, Hinsley W, et al. A simple approach to measure transmissibility and forecast incidence. Epidemics. 2018;22:29–35.https://doi.org/10.1016/j.epidem.2017.02.012.

[11] Data classification. In Wikipedia. (n.d.). http://en.wikipedia.org/wiki/Data classification

[12] Vinod, D. N., & Prabaharan, S. R. S. (2020). Data science and the role of Artificial Intelligence in achieving the fast diagnosis of Covid-19. *Chaos, Solitons & Fractals*, *140*, 110182.

[13] NASSER, Ibrahim M.; ABU-NASER, Samy S. Lung cancer detection using artificial neural network. *International Journal of Engineering and Information Systems (IJEAIS)*, 2019, 3.3: 17-23.

[14] Mesgarpour, M., Abad, J. M. N., Alizadeh, R., Wongwises, S., Doranehgard, M. H., Ghaderi, S., & Karimi, N. (2021). Prediction of the spread of Corona-virus carrying droplets in a bus-A computational based artificial intelligence approach. *Journal of Hazardous Materials*, *413*, 125358.

[15] Wang, S., Kang, B., Ma, J., Zeng, X., Xiao, M., Guo, J., ... & Xu, B. (2021). A deep learning algorithm using CT images to screen for Corona Virus Disease (COVID-19). *European radiology*, 1-9.

[16] R. Das, I. Turkoglu and A. Sengur, "Diagnosis of Valvu-lar Heart Disease through Neural Networks Ensembles," Computer Methods & Programs in Biomedicine, Vol. 93, No. 2, 2009, pp. 185-191.

[17] S. Mehrabi, M. Maghsoudloo, H. Arabalibeik, R. Noor-mand and Y. Nozari, "Application of Multilayer Percep-tron and Radial Basis Function Neural Networks in Differentiating between Chronic Obstructive Pulmonary and Congestive Heart Failure Diseases," Expert Systems with Applications, Vol. 36, No. 3, 2009, pp. 6956-6959.

[18] Schmidhuber J 2015. Deep learning in neural networks: An overview, Neural Networks, vol 61, pp 85–117.

[19] Hamamoto I, Okada S, Hashimoto T, Wakabayashi H, Maeba T, Maeta H: Prediction of the early prognosis of the hepatectomized patient with hepatocellular carcinoma with a neural network. Comput Biol Med, 1995; 25(1): 49-59.

[20] Ravdin PM, Clark GM, Hilsenbeck SG et al.: A demonstration that breast cancer recurrence can be predicted by neural network analysis. Breast Cancer Res Treat, 1992; 21(1): 47-53.

[21] Al-Najjar H, Al-Rousan N. A classifier prediction model to predict the status of coronavirus COVID-19 patients in South Korea. Eur Rev Med Pharmacol Sci. 2020;24(6):3400–3.

[22] TAMANG, S. K.; SINGH, P. D.; DATTA, B. Forecasting of Covid-19 cases based on prediction using artificial neural network curve fitting technique. Global Journal of Environmental Science and Management, 2020, 6.Special Issue (Covid-19): 53-64.

[23] SABA, Amal I.; ELSHEIKH, Ammar H. Forecasting the prevalence of COVID-19 outbreak in Egypt using nonlinear autoregressive artificial neural networks. Process safety and environmental protection, 2020, 141: 1-8.

[24] ILIYAS, Mrs Mehdi Khundmir; SHAIKH, Mr Imran Sadekh; STUDENT, M. C. A. Prediction of Heart Disease Using Decision Tree. Allana Inst of Management Sciences, Pune, 2019, 9: 1-5.

[25] MAJI, Srabanti; ARORA, Srishti. Decision tree algorithms for prediction of heart disease. In: Information and communication technology for competitive strategies. Springer, Singapore, 2019. p. 447-454.
[26] PRAKASH, Kolla Bhanu, et al. Analysis, prediction and evaluation of covid-19 datasets using machine learning algorithms. International Journal, 2020, 8.5.
[27] Charu C. Aggarwal, data mining the textbook p 326, IBM T.J. Watson Research Center Yorktown Heights New York the USA.
[28] Gurney, K. (1997). An Introduction to Neural Networks (1st ed.). CRC Press. https://doi.org/10.1201/9781315273570.
[29] N. Ganesan, K. Venkatesh, M. A. Rama Application of Neural Networks in Diagnosing Cancer Disease Using Demographic Data. 2010 International Journal of Computer Applications (0975 - 8887) Volume 1 – No.26.
[30] Carlos Hernandez Espinosa Mercedes Fernandez Redondo 2001 Universidad Jaume I Campus de Riu Sec Spain IEEE 2001.
[31] KARSOLIYA, Saurabh. Approximating number of hidden layer neurons in multiple hidden layer BPNN architecture. *International Journal of Engineering Trends and* Technology, 2012, 3.6: 714-717.
[32] Farhad Malik. (2019, May 21). What Are Hidden Layers. Medium. https://medium.com/fintechexplained/what-are-hidden-layers-4f54f7328263.
[33] ASTHANA, Stuti; HANEEF, Farha; BHUJADE, Rakesh K. Handwritten multiscript numeral recognition using artificial neural networks. *International Journal of Soft Computing and Engineering*, 2011, 1.1: 1-5.
[34] Algorithmia. (2018, April 30). Introduction to loss functions. https://algorithmia.com/blog/introduction-to-loss-functions#whats-a-loss-function .
[35] Iwendi, C., Bashir, A. K., Peshkar, A., Sujatha, R., Chatterjee, J. M., Pasupuleti, S., ... & Jo, O. (2020). COVID-19 patient health prediction using boosted random forest algorithm. Frontiers in public health, 8, 357.
[36] N, Ganesan & K, Venkatesh & M A, Rama & Palani, Malathi. (2010). Application of Neural Networks in Diagnosing Cancer Disease using Demographic Data. International Journal of Computer Applications. 1. 10.5120/476-783.
[37] AGGARWAL, Charu C. Data mining: the textbook. Springer, 2015.
[38] Chen, X., Yang, B. & Lin, Z. A random forest learning assisted "divide and conquer" approach for peptide conformation search. Sci Rep 8, 8796 (2018). https://doi.org/10.1038/s41598-018-27167-w.
[39] Lavanya, D., & Rani, K. U. (2011). Performance evaluation of decision tree classifiers on medical datasets. International Journal of Computer Applications, 26(4), 1-4.
[40] Filippo Amato, Alberto López, Eladia María Peña-Méndez, Petr Vaňhara, Aleš Hampl, Josef Havel, Artificial neural networks in medical diagnosis, Journal of Applied Biomedicine, Volume 11, Issue 2, 2013, Pages 47-58, ISSN 1214-021X.
[41] Alex Strebeck. (2018, June 5). Training and Testing Machine Learning Models. Medium. https://medium.com/@alexstrebeck/training-and-testing-machine-learning-models-e1f27dc9b3cb.
[42] Sumit Banik. (2020, June 16) COVID19-Identification-using-ML. GitHub. https://github.com/iSumitBanik/COVID19-Identification-using-ML.
[43] Aditya Mishra. (2018, Feb 24). Metrics to Evaluate your Machine Learning Algorithm. Towards Data Science. https://towardsdatascience.com/metrics-to-evaluate-your-machine-learning-algorithm-f10ba6e38234.

التنبؤ بعدوى فيروس كورونا باستخدام التعلم العميق والتعلم الآلي

ليث إبراهيم صالح

قسم هندسة الحاسوب ، كلية الهندسة ، جامعة اوكان / تركيا

الملخص

مع الانتشار السريع لفيروس كورونا حول العالم، أصبح من الضروري إيجاد حلول لعلاج هذا الوباء. نقدم في هذا البحث حلاً واعداً للتنبؤ بالحالات المصابة بناءً على الأعراض أو السمات الشديدة التي تم تسجيلها باستخدام طرق غير إكلينيكية مثل الذكاء الاصطناعي. يوفر الذكاء الاصطناعي العديد من الأدوات لتحليل البيانات والتحليل الإحصائي والبحث الذكي. قدم العالم الرقمي مؤخرًا طرقًا واقعية نسبيًا لاستخدامها في مختلف المجالات العلمية. في هذه الورقة، نركز على التنبؤ بعدوى COVID-19 التي انتشرت إلى جميع البلدان في جميع أنحاء العالم، باستخدام برامج / برامج التعلم العميق التي تمثلها الشبكات العصبية الاصطناعية (ANN) والتعلم الآلي. يتم تطبيق خوارزميات عالية الدقة وهي Randomforest : و Decision Tree، لتحليل مجموعات البيانات الطبية بشكل فعال، بناءً على الأعراض الأكثر شيوعًا والحادة ، مثل السعال والحمى والصداع والإسهال والعيش في المناطق المصابة والآلام وضيق التنفس، قشعريرة ، احتقان بالأنف ، وبعض أعراض المرض الأخرى. يتم استخدام مجموعة بيانات تقابل 1496 مريضًا لتحديد ما إذا كان المريض مصابًا بالفيروس أم لا. تنقسم مجموعة البيانات إلى 75٪ من بيانات التدريب و25٪ من بيانات الاختبار بعد تطبيق خوارزميات التعلم الآلي والتعلم العميق. يتم أيضًا استخدام مكتبات Python مثل pandas و NumPyو matplotlib بالإضافة إلى استخدام sklearn و .Keras تظهر نتائج البحث دقة عالية جدًا باستخدام التعلم الآلي المشار إليه بنسبة 91٪ من RandomForestمع المقدرات = 200 و91٪ من شجرة القرار. في التعلم العميق، تبلغ دقة الشبكة العصبية الاصطناعية 85٪. وبالتالي، يوفر هذا البحث الجديد مؤشرًا مهمًا للتنبؤ المحتمل بإصابة COVID-19 باستخدام برامج التعلم العميق.