ZANCO Journal of Pure and Applied Sciences
The official scientific journal of Salahaddin University-Erbil
https://zancojournals.su.edu.krd/index.php/JPAS

ISSN (print): 2218-0230**, ISSN (online):** 2412-3986**, DOI: http://dx.doi.org/10.21271/zjpas**

RESEARCH PAPER

Smart COVID-19 Prediction System Using Neural Network

Ala Saeed Khudhur[1], Abdulqadir Ismail Abdullah[2]

[1]Computer Science and IT Department, College of Science, Salahaddin University-Erbil, Kurdistan Region, Iraq
[2]Computer Science Department, College of Science, International University of Erbil-Erbil, Kurdistan Region, Iraq

A B S T R A C T:

The pandemic of coronavirus COVID-19 has created a great danger and concern for humanity. Many researchers have done different types of work in this area to provide medical services. In this paper, we proposed a smart Covid-19 diagnosis system by using a Feed Forward Backpropagation Neural Network (FFBNN) and Probabilistic Neural Network (PNN). Based on personal information from patients such as (age, gender, contact with sick person) and five symptoms (headache, fever, cough, sore throat, and shortness of breath) for this purpose we used 510 samples that are collected from different sources, and then compared to previous studies. Results of this work showed that using FFBNN has achieved highest accuracy (98.0%), sensitivity (100%), specificity (94.4%), precision (97.1%), recall (100%) and F1-score (98.52%). But PNN that has accuracy, sensitivity, specificity, precision, recall, F1-score of 90.2%, 92.7%, 87.2%, 89.47%, 92.7% and 91.07% respectively. The most relevant features to positive Covid-19 were fever, shortness of breath, and cough with correlation coefficient of 0.591, 0.495 and 0.488.

KEY WORDS: COVID-19; Artificial Neural Network; FFBNN; PNN.
DOI: http://dx.doi.org/10.21271/ZJPAS.34.5.5
ZJPAS (2022) , 34(5);42-49 .

1. INTRODUCTION:

The novel coronavirus disease (COVID-19) was first reported on December 31st, 2019 in Wuhan, Hubei Province, China. The outbreak of COVID-19 coronavirus, namely SARS-CoV-2 has created a calamitous situation throughout the world (Wu et al., 2020; Lu et al., 2020). The cumulative incidence of COVID-19 is rapidly increasing day by day. The World Health Organization (WHO) has declared the coronavirus outbreak as pandemic, while the virus is continuing to spread (Tuli et al., 2020). In December 6, 2021, a total of 266,215,281 confirmed positive cases have reported in 222 countries and 5,273,301 deaths have recovered [www.worldometers.info/coronavirus].

* **Corresponding Author:**
Ala Saeed Khudhur
E-mail: ala.khudhur@su.edu.krd
Article History:
Received: 26/03/2022
Accepted: 27/06/2022
Published: 20/10 /2022

According to Dashboard Coronavirus (COVID-19) in the Kurdistan Region [www.gov.krd/coronavirus-en/what-you-should-know/] in 14/1/2022), four governorates have confirmed 386,863 positive confirmed cases, 373,660 recovered cases and 7,158 deaths. Recently the outbreak of coronavirus has opened up new challenges for the research community worldwide. Covid-19 is considered a virus disease that patients observe many symptoms such as headache, fever, cough, shortness of breath, sore throat and may be other symptoms in some patients that cause death and 78.8% of patients need for respiratory support (Sudre et al., 2021). But likely according to study (Khozeimeh et al., 2021) number of recoveries are more than the number of deaths and early detections may increase the chances of recovery and getting special treatments.

The proven artificial intelligence (AI) methods can be useful for predicting the risks, parameters and effects of such an epidemic. These predictions may help controlling and preventing the spread of some other diseases. The use of AI and machine learning technologies can minimise time, cost, human expertise and incorrect diagnosis. Artificial Neural Network (ANN) is a type of machine learning that works in a similar way to human brain. To solve medical problems, ANNs have extensively been applied and significantly used in the field of medicine. ANN has also used for the diagnosis and classification of various diseases (Ansari et al., 2011). A sophisticated system is needed to assist doctors in diagnosing diseases accurately and efficiently. In the last few years, ANN methods have successfully applied to predict various diseases, such as heart disease (Krishnn et al., 2021; Rufai and Umar, 2018), breast cancer (Saritas and Yasar, 2019) prediction, as well as to detect brain tumour abnormalities or normal brain (George et al., 2015). Prediction techniques have also proved to be useful in many healthcare applications. By calculating odd ratio and statistical analyses for all risk factors and reported symptoms, Zens et al. (2020) identified loss of smell, chills, fever, nausea and vomiting and shortness of breath as the top five strongest predictors of a COVID-19 infection. Prakash et al. (2020) determined that the age of 20-50 are most likely to become infected with COVID-19. The purposes of this paper are: 1) to use our dataset for the prediction of COVID-19 based on five symptoms and three person’s information such as age, gender, contact to sick person using two types of neural networks: The Feedforward Backpropagation Neural Network (FFNN), and the Probabilistic Neural Network (PNN); 2) perform a comparison between the networks we used, and 3) comparing the results with the previous studies that used the same data for the diagnosis and classification of COVID-19. Rest of the paper is organised into five sections: Section one presents the method and the material used for the prediction of the COVID-19 virus. Section two discusses the architecture of ANN used in the research. Section three provides the discussion on experimental results. Section four provides a comparison of the results with previous studies and Section five finally concludes the paper.

2. MATERIALS AND METHODS

2.1. Use of Artificial Neural Network in different fields

ANN is used to classify or predict different types of problems. In the medical field, Neural Network (NN) is used to predict different types of diseases. Vijayarani et al. (2015) used Support Vector Machine (SVM) and ANN for predicting four types of kidney diseases and they showed better ANN accuracy (87.70%) than SVM accuracy (76.32%). However, Rufai and Umar (2018) used ANN to predict coronary heart disease with accuracy of 92.2%. El_Jerjawi and Abu-Naser (2018) showed the accuracy of 87.3% when they used ANN to predict diabetes. Moreover, Saritas and Yasar (2019) showed that ANN is more powerful than the Nave Bayes for detection of breast cancer. While Krishnan et al. (2021) proposed a hybrid deep learning model for heart disease prediction using recurrent neural networks (RNN) in combination with multiple gated recurrent units (GRU) and long short-term memory (LSTM). Their results revealed that this model has an accuracy of 98.69%.

2.2. Previous studies on COVID-19

Tostmann et al. (2020) showed that anosmia and muscle ache are the strongest predictors by applying Lasso regression. Wu et al. (2020) used random forest (RF) method for identifying the patients who need quarantine and the accuracy of their method was 96.97% for the test set. Zhou et al. (2020) used SVM to predict the progression of illness severity based on measures during the first 12 days, they also indicated that male patients were more likely to be infected with SARSCoV-2 than female patients. However, Menni et al. (2020) performed a stepwise logistic regression, combining forward and backward algorithms and they showed that the strongest predictor was loss of smell and taste, which is specific to COVID-19. However, Zoabi et al. (2021) used gradient-boosting machine model built with a decision-tree base-learner to predicted positive COVID-19 infection in a Real-Time Polymerase Chain Reaction (RT-PCR) test.

In this study we used ANN to predict Covid-19 persons by knowing eight features such as gender, age, contact with ill person, cough, fever, sore throat, shortness of breath, as well as headache, by using a dataset of 510 samples collected from different sources in Kurdistan region. For this purpose, we designed two different neural networks including feed forward backpropagation neural network and probabilistic neural network to determine which neural network is more powerful for determining correct Covid-19 case by evaluating some of metrics like accuracy, sensitivity, specificity, recall, precision and F1-scoure.

2.3. Data collection

Samples from 559 persons were collected manually from different sources including online survey and two hospitals in Erbil City-Kurdistan region, including Lalav hospital and Zanco Health Centre. Data and information were taken from questionnaires completed by clinicians during case admission on Covid-19 wards from patients. In the online survey, we asked for eight quotations (gender, age, contact with a sick person, and five symptoms such as: fever, shortness of breath, headache, sore throat and cough) as well as we asked for the results of the COVID-19 test either positive or negative. The positive COVID-19 cases were confirmed by Chest CT and RT-PCR tests in the hospital-approved laboratories.

2.4. Data wrangling

The 559 collected cases were cleaned and the duplicated rows were removed, thus only 510 rows of data were remained without missing values (Table 1). Correlation was used to give information about the relationship between dataset features and offers an important information about the features and their influence on the target value. A value of near to 1 means positive correlation, but a value of near to -1 means negative correlation. The number of correlations in each feature is shown in Figure 1. The dataset consists of nine columns with the data type being "Yes" or "NO", and numeric types. We also have categorical variables such as gender. Since the neural network model requires all the data that passes as an input to be in the numeric form, we performed label-encoding of categorical variables and for other Yes or No values as shown in Table 2.

Table 1: Description of our data and the percentages.

Features		Number of cases	Percentage %
Gender	Male	258	50.6
	Female	252	49.4
Age		(4-90)	
Contact with sick person	Yes	254	49.8
	No	256	50.2
Cough	Yes	259	50.8
	No	251	49.2
Fever	Yes	268	52.55
	No	242	47.45
Sore throat	Yes	238	46.7
	No	272	53.3
Shortness of breath	Yes	193	37.8
	No	317	62.2
Headache	Yes	256	50.2
	No	254	49.8
Covid-19 Test	Positive	286	56.1
	Negative	224	43.9

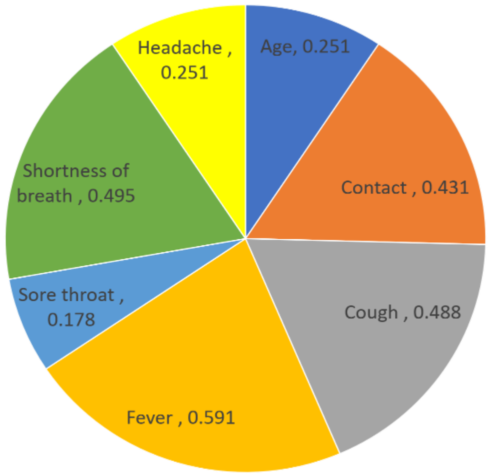

Figure 1: Number of correlation of features.

Table 2: Labelling of data.

Column	Values (for categorical variables)
Gender	Male (1), Female (0)
Age	Number
Contact with sick person	Yes (1), No (0)
Cough	Yes (1), No (0)
Fever	Yes (1), No (0)
Sore Throat	Yes (1), No (0)
Headache	Yes (1), No (0)
Shortness of breath	Yes (1), No (0)
COVID-19 Test	Positive (1), Negative (0)

2.5. Artificial Neural Network

Artificial neural networks (ANN) is a popular machine learning technique which is inspired by the biological neural network in human brain (Bhardwaj and Tiwari, 2015). The common ANN type is Feed forward neural network which sends the weighted artificial neuron values as output to the next layer after processing with inputs from neurons in the previous layer (Bebis and

Georgiopoulos, 1994). Multilayer Perceptron (MLP) is an important class of feed forward neural network and the most widely used MLP training technique is the back-propagation algorithm which changes the weights between neurons to minimize the error. This quite good model in learning patterns and can easily adapt to new values in the data, but the system can show a slow convergence and has risk of a local optimum (Saritas and Yasar, 2019).

2.5.1. Proposed Feed Forward Neural Network

The one type of artificial neural networks is Feedforward neural network, since FFNN with a hidden layer, appropriate transfer function in the hidden layer and the sufficient neurons are able to estimate any function with an arbitrary accuracy. For this aim, we present a structure of FFNN modelling to predict the COVID-19 problem as seen in (Figure 2).

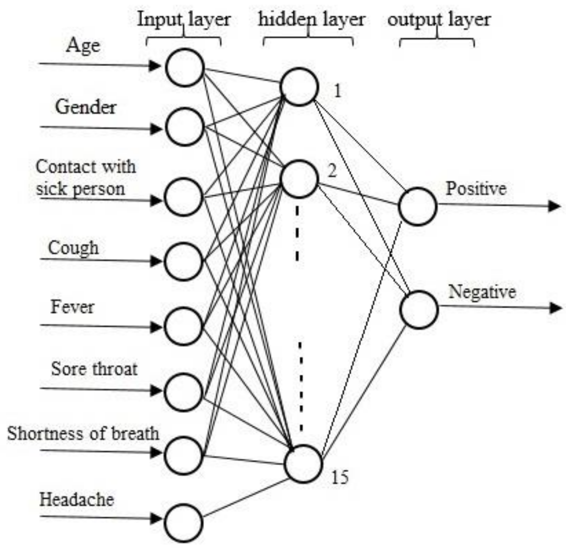

Figure 2: Architecture of FFNN.

In general, we have three kinds of layers, first layer is called input layer that get the raw data and fed to the network. Second layer is hidden layer(s): This layer's (or layers') function is dictated by inputs, weight, their connection, and the hidden layer (s). When a hidden unit needs to be activated the weights between input and hidden units was determined. The last layer is the output layer; the functions of the output unit depend on the activity and weight of the hidden unit and the connection between hidden units and output. The purpose of this experiment was to identify whether the person has Covid-19 or not. For this purpose, we used (nprtool) toolbox called Neural Network Pattern Recognition was performed to determine Covid-19 in Matlab 2019a software. Eight inputs in the input layer with one hidden layer, in the hidden layer used sigmoid transfer function and a softmax activation function in the output layer. To determine the number of neurons in the hidden layer several experiments were tried to choose 15 neurons in the hidden layer and we have two output neurons in the output layer, positive class and negative class. In the next step, the dataset is divided randomly into three sets as follows: 80% is used for training, 10% for validation that the network is generalizing and to stop training before overfitting. Furthermore, 10% is used as a completely independent network generalization test. The network trained with scaled conjugate gradient (SCG) Backpropagation algorithm. A cross-entropy was used to evaluate ANN performance. The minimum cross-entropy occurred at epoch 7 and was equal to 0.21115 the performance graph of ANN model (Figure 3).

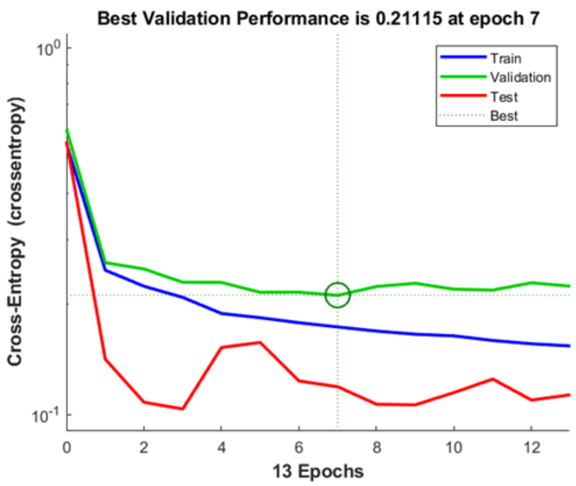

Figure 3: The performance graph of optimized FFNN model.

2.5.2. Probabilistic Neural Network

A probabilistic neural network was used for the prediction of COVID-19, which was performed by Matlab 2019a software. The PNN structure consists of a single hidden layer (radial basis layer) of locally tuned units which are fully interconnected to an output layer (Figure 4). In this system, the real-valued input vector is the feature vector, and the two outputs are the indices of two classes, positive and negative. All hidden units simultaneously receive the 8-dimensional real-value input vector. The input vector of the network is passed to the hidden layer nodes via unit connection weights. The hidden layer consists of a set of radial basis functions (Figure 3). In this method, the data was randomly divided for training and testing into two separate sets, with 80% for training and 20% for testing. A width or spread parameter is used as a global parameter that determines the width of the kernel. This is the only parameter used to optimize the performance of PNN (Ahmadlou and Adeli, 2010). In this work, the value of the spread is 1 get a good result of accuracy. After that, we test the network on the designed input vectors.

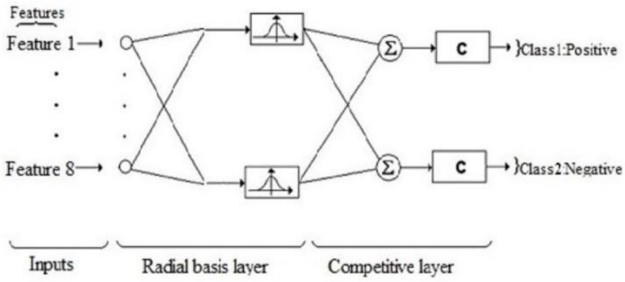

Figure 4: Implementation of PNN for COVID-19 prediction.

2.5.3. Evaluation models

For evaluation in this work, we used these metrics: accuracy, sensitivity, specificity, recall, precision and F1 score by using the equations (1-6). It is worth mentioning that the number of false positives indicates FP and the number of true negatives indicates TN, while the number of true positives indicates TP and the number of false positives indicates FP (Patterson *et al*., 2021).

2.5.3.1. Accuracy

It is the metric used to determine how well a classifier work.

Accuracy = (TP + TN) / (TP+TN+FP+FN) × 100 ……. (1).

2.5.3.2. Sensitivity

It calculates the proportion of positives that are truly identified.

Sensitivity = TP / (TP + FN) × 100 ………… (2).

2.5.3.3. Specificity

It calculates the proportion of negatives that are truly identified.

Specificity = TN / (TN + FP) × 100………………. (3).

2.5.3.4. Precision

The proportion of positive identifications is calculated by precision that are actually correct.

Precision = TP / (TP + FP) × 100 ………………. (4).

2.5.3.5. Recall

It calculates the proportion of actual positives that were correctly identified.

Recall = TP / (TP + FN) × 100 ………………… (5).

2.5.3.6. F1-Score

It is the harmonic mean of precision and recall that is the way of combining precision and recall.

F1-Score = 2 × (precision × recall) / (precision + recall) …………… (6)

3. RESULTS

Based on the eight inputs: age, gender, contact with is ill persons, cough, fever, headache, shortness of breath, sore throat for classifying Covid-19 positive or Negative cases using two neural networks; the results showed that the classification obtained by FFBNN is more accurate than the PNN. Thus, FFBNN is more suitable for classification of Covid-19 data. As shown in Table 3, it is obvious that the obtained results during the present study by using FFBNN was better than the results obtained by using the other methods. So the accuracy was 98%, sensitivity was 100% and specificity was 94.4% by FFBNN model, whereas by the PNN model the accuracy was 90.2%, sensitivity was 92.7% and specificity was 87.2%.

Furthermore, Figure 5 explains the results of our two models that used some metrics including accuracy, sensitivity, specificity, precision, recall and F1-Score. It is obvious that the achievement of FFBNN showed the highest accuracy, sensitivity, specificity, precision, recall and F1-score with values of 98.0%, 100%, 94.4%, 97.1%, 100% and 98.52% respectively when compared with PNN that has an accuracy, sensitivity, specificity, precision, recall and F1-score of 90.2%, 92.7%, 87.2%, 89.47%, 92.7% and 91.07% respectively. Moreover, the most relevant features to positive Covid-19 were fever, shortness of breath, and cough with correlation coefficient of 0.591, 0.495 and 0.488.

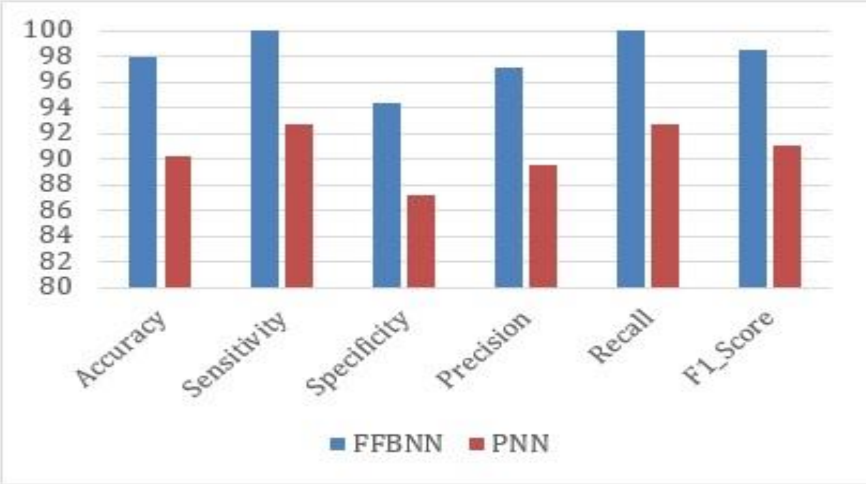

Figure 5: Comparison results of the proposed models.

Table 3: Result of our models compared to previous studies.

Author	Model	Accuracy %	Sensitivity%	Specificity %	Dataset
Tostmann *et al*., 2020	Lasso regression	-	91.20	55.6	Symptoms questionnaire
Zoabi *et al*., 2021	Gradient boosting With decision tree	-	87.3	71.98	WHO
Wu *et al.,* 2020	random forest (RF)	95.95	95.12	96.97	multiple sources
Zhou *et al.,* 2020	SVM Threshold 0.45	94.0	70.0	99.0	Clinical data
Menni *et al.,* 2020	stepwise logistic regression	-	65.0	78.0	UK smartphone-app reported data
Our model	FFBNN PNN	98.0 90.2	100 92.7	94.4 87.2	Our dataset

4. DISCUSSION

In this research paper we used 8 input features that are described in Table 1 for the determination of positive or negative cases of COVID-19, the first model we used for pattern recognition, the feedforward backpropagation algorithm, could achieve better results. The accuracy of 98% in the test set means that the model was able to accurately identify most cases, and sensitivity of 100% means that the model could identify all cases that have positive COVID-19 test. Moreover, the specificity of 94.4% identifies the true negative, and the F1 score is a harmonic mean between precision and recall. This score takes both false positives and false negatives into account with a value of 98.52%. The second proposed model, PNN, could achieve an accuracy 90.2% less than the first model, and the sensitivity of 92.7%, but the specificity was 87.2%. The first model is more powerful than PNN in all metrics that evaluated by our model. By using lasso regression with seven symptoms by the authors Tostmann *et al*. (2020), they got a sensitivity of 91.2%, which was less than both of our models (Table 3), while the value of specificity was 55.6%, which is very low in comparing with our two models. Another difference is fever, in which the study showed that fever is not the strongest predictor of positive COVID-19 test, while in our study the most related symptom to indicate a positive COVID-19 test was fever with a value of correlation of 0.591. Also, the FFBNN model in terms of accuracy and sensitivity is better than the Random Forest that was used by (Wu *et al*., 2020) to know which patient need quarantine. The RF was able to achieve a good result, with 95.95% accuracy and 95.12% sensitivity, but in terms of specificity, the RF could achieve a higher score than the first model, with 96.97%, while the result of our PNN is less than the RF in all three metrics (Zhou *et al*., 2020). SVM could achieve an accuracy of 86% in the test set, which was less than the present two models, but sensitivity was 70% and specificity was 99%. There was a great difference between the values of specificity and sensitivity, which means that 30% is wrong in identifying true positives, and this is a high percentage. But in terms of specificity, it was higher than in both models, and they used 0.45 as a threshold value, whereas in our study the threshold value was 0.5. Moreover, Menni *et al*., (2020), determined that loss of smell and taste is more associative with COVID-19 by using logistic regression, but in our study, loss of smell and taste doesn't have this feature, but fever has a greater value of correlation, 0.591, indicating that there is more relation to COVID-19 in our study than in logistic regression. The sensitivity and specificity of our two models were greater than the logistic regression. The difference in Zoabi *et al.*'s (2021) work was age features that are used as a binary data > 60 indicates yes, otherwise no, while in our data age is a number and the performance in their model was less than our two models in terms of sensitivity and specificity (87.3% and 71.98%).

5. CONCLUSIONS

In this study, we concluded that FFBNN, in terms of accuracy, sensitivity, specificity, and F1-score, showed a good result when compared to PNN, which makes this model able to select the Covid-19 cases with 98% accuracy. To make it positive or negative, PNN could test COVID-19 with an accuracy of 90.2%. Furthermore, this study was conducted on 510 data points, so if the study was conducted on grater data set, it may possibly show better results. This model can become an assistant doctor's tool and provide quick help for the doctor to select whether the patient has COVID-19 or no, and this will be a good service for the health and medical field.

References

AHMADLOU, M. and ADELI, H. 2010. Enhanced probabilistic neural network with local decision circles: A robust classifier. *Integrated Computer-Aided Engineering*. **17**(3), pp.197–210.

ANSARI, S., SHAFI, I., ANSARI, A., AHMAD, J. and SHAH, S.I. 2011. Diagnosis of liver disease induced by hepatitis virus using artificial neural networks *In*: *2011 IEEE 14th international multitopic conference*. IEEE, pp.8–12.

BEBIS, G. and GEORGIOPOULOS, M. 1994. Feed-forward neural networks. *IEEE Potentials*. **13**(4), pp.27–31.

BHARDWAJ, A. and TIWARI, A. 2015. Breast cancer diagnosis using genetically optimized neural network model. *Expert Systems with Applications*. **42**(10), pp.4611–4620.

El_JERJAWI, N.S. and ABU-NASER, S.S. 2018. Diabetes prediction using artificial neural network. *International Journal of Advanced Science and Technology*. **121**.

GEORGE, D.N., JEHLOL, H.B. and OLEIWI, A.S.A. 2015. Brain tumor detection using shape features and machine learning algorithms. *International Journal of Advanced Research in Computer Science and Software Engineering*. **5**(10), pp.454–459.

KHOZEIMEH, F., SHARIFRAZI, D., IZADI, N.H., JOLOUDARI, J.H., SHOEIBI, A., ALIZADEHSANI, R., GORRIZ, J.M., HUSSAIN, S., SANI, Z.A. and MOOSAEI, H. 2021. Combining a convKhozeimeh, F., Sharifrazi, D., Izadi, N.H., Joloudari, J.H., Shoeibi, A., Alizadehsani, R., Gorriz, J.M., Hussain, S., Sani, Z.A. and Moosaei, H. 2021. Combining a convolutional neural network with autoencoders to predict the survival chan. *Scientific Reports*. **11**(1), pp.1–18.

KRISHNN, S., MAGALINGAM, P. and IBRAHIM, R. 2021. Hybrid deep learning model using recurrent neural network and gated recurrent unit for heart disease prediction. *International Journal of Electrical & Computer Engineering (2088-8708)*. **11**(6).

LU, H., STRATTON, C.W. and TANG, Y. 2020. Outbreak of pneumonia of unknown etiology in Wuhan, China: The mystery and the miracle. *Journal of medical virology*. **92**(4), p.401.

MENNI, C., VALDES, A.M., FREIDIN, M.B., SUDRE, C.H., NGUYEN, L.H., DREW, D.A., GANESH, S., VARSAVSKY, T., CARDOSO, M.J. and MOUSTAFA, J.S.E.-S. 2020. Real-time tracking of self-reported symptoms to predict potential COVID-19. *Nature medicine*. **26**(7), pp.1037–1040.

PATTERSON, B.K., GUEVARA-COTO, J., YOGENDRA, R., FRANCISCO, E.B., LONG, E., PISE, A., RODRIGUES, H., PARIKH, P., MORA, J. and MORA-RODRÍGUEZ, R.A. 2021. Immune-based prediction of COVID-19 severity and chronicity decoded using machine learning. *Frontiers in immunology*. **12**, p.2520.

PRAKASH, K.B., IMAMBI, S.S., ISMAIL, M., KUMAR, T.P. and PAWAN, Y.N. 2020. Analysis, prediction and evaluation of covid-19 datasets using machine learning algorithms. *International Journal*. **8**(5).

RUFAI, A.U.S. and UMAR, M. 2018. Using artificial neural networks to diagnose heart disease *In*: *IJCAI*., pp.1–6.

SARITAS, M.M. and YASAR, A. 2019. Performance analysis of ANN and Naive Bayes classification algorithm for data classification. *International Journal of Intelligent Systems and Applications in Engineering*. **7**(2), pp.88–91.

SUDRE, C.H., LEE, K.A., LOCHLAINN, M.N., VARSAVSKY, T., MURRAY, B., GRAHAM, M.S., MENNI, C., MODAT, M., BOWYER, R.C.E. and NGUYEN, L.H. 2021. Symptom clusters in COVID-19: A potential clinical prediction tool from the COVID Symptom Study app. *Science Advances*. **7**(12), p.eabd4177.

TOSTMANN, A., BRADLEY, J., BOUSEMA, T., YIEK, W.-K., HOLWERDA, M., BLEEKER-ROVERS, C., TEN OEVER, J., MEIJER, C., RAHAMAT-LANGENDOEN, J. and HOPMAN, J. 2020. Strong associations and moderate predictive value of early symptoms for SARS-CoV-2 test positivity among healthcare workers, the Netherlands, March 2020. *Eurosurveillance*. **25**(16), p.2000508.

TULI, SHRESHTH, TULI, SHIKHAR, TULI, R. and GILL, S.S. 2020. Predicting the growth and trend of COVID-19 pandemic using machine learning and cloud computing. *Internet of Things*. **11**, p.100222.

VIJAYARANI, S., DHAYANAND, S. and PHIL, M. 2015. Kidney disease prediction using SVM and ANN algorithms. *International Journal of Computing and Business Research (IJCBR)*. **6**(2), pp.1–12.

WU, J., ZHANG, P., ZHANG, L., MENG, W., LI, J., TONG, C., LI, Y., CAI, J., YANG, Z. and ZHU, J. 2020. Rapid and accurate identification of COVID-19 infection through machine learning based on clinical available blood test results. *MedRxiv*.

WU, W., WANG, A. and LIU, M. 2020. Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China. *Lancet*. **395**(10223), pp.497–506.

ZENS, M., BRAMMERTZ, A., HERPICH, J., SÜDKAMP, N. and HINTERSEER, M. 2020. App-based tracking of self-reported COVID-19 symptoms: analysis of questionnaire data. *Journal of medical Internet research*. **22**(9), p.e21956.

ZHOU, K., SUN, Y., LI, L., ZANG, Z., WANG, J., LI, J., LIANG, J., ZHANG, F., ZHANG, Q. and GE, W. 2020. Eleven Routine Clinical Features Predict COVID-19 Severity. *medRxiv*.

ZOABI, Y., DERI-ROZOV, S. and SHOMRON, N. 2021. Machine learning-based prediction of COVID-19 diagnosis based on symptoms. *npj digital medicine*. **4**(1), pp.1–5.