8th International Scientific Conference
Technics and Informatics in Education
Faculty of Technical Sciences, Čačak, Serbia, 18-20th September 2020

Session 2: *IT Education and Practice*

Professional paper
UDC: 004:616.98

Data Analysis and Artificial Neural Network Modelling of COVID-19 Patients

Dijana Stojić [1], Dejan Vujičić [1*], Đorđe Damnjanović [1], Željko Jovanović [1]
[1] University of Kragujevac, Faculty of Technical Sciences Čačak, Serbia
[*] dejan.vujicic@ftn.kg.ac.rs

Abstract: *The pandemic of novel coronavirus (SARS-CoV-2) causing the COVID-19 infectious disease has changed our life significantly. From the moment of its detection, numerous teams of scientists and medical staff have worked tirelessly to find the cure and vaccine, and to take care of the patients. In this paper, a small contribution in this combat has been made by creating an artificial neural network model for predicting the patients' outcome of the disease. The mean accuracy of the realized model was 87.50%, which can be further improved with a larger dataset being provided.*

Keywords: *artificial neural network; COVID-19*

1. INTRODUCTION

COVID-19 is an infectious disease caused by SARS-CoV-2 (severe acute respiratory syndrome coronavirus 2) virus, a member of the coronavirus family. It was first reported in Wuhan, China, in December 2019, and the World Health Organization (WHO) declared a pandemic of COVID-19 disease on March 11, 2020 [1] – [3].

As of August 3, 2020, there are 17,918,582 total confirmed cases of COVID-19, with 686,703 deaths worldwide [4]. The number of COVID-19 cases by countries is graphically shown in Fig. 1 [4]. The distribution of COVID-19 total cases by WHO regions is given in Table 1 [5]. As for Serbia, the total number of registered COVID-19 patients by August 3, 2020, is 26,451, and the total number of registered deaths caused by COVID-19 disease is 598 [6].

The main symptoms of SARS-CoV-2 infection are fever, dry cough, and tiredness [7]. Other symptoms include loss of taste or smell, aches, nasal congestion, headache, sore throat, diarrhea, skin rash, etc. Around 80% of the patients develop only mild symptoms of the disease, while those with chronic diseases are of great risk of complications [7] – [9].

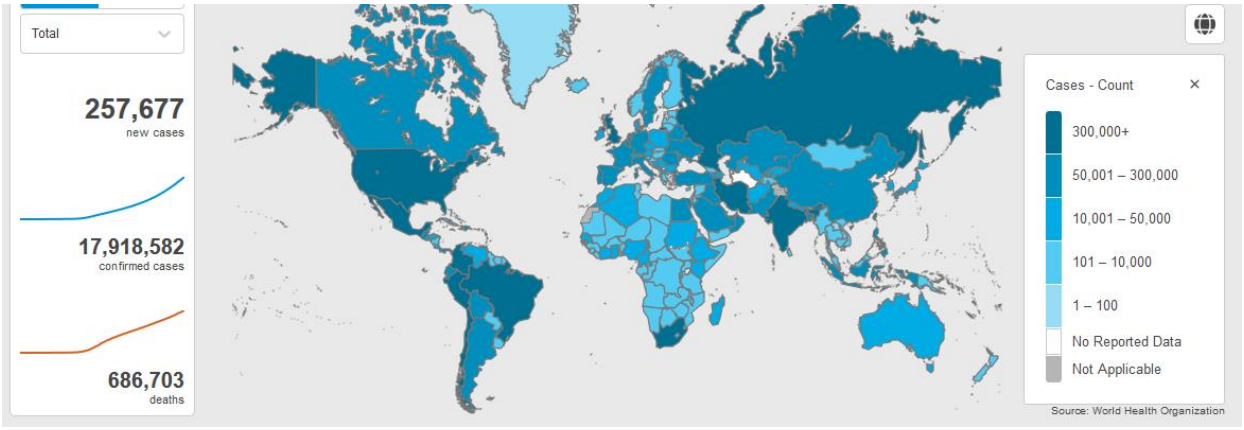

Figure 1. *The number of COVID-19 cases by countries [4]*

The main purpose of this research is to make contribution in the combat with COVID-19 by inspecting possible correlations between patient's age, symptomatology, and medical condition. This is done by creating the artificial neural network that takes these parameters as inputs and tries to determine the patient's outcome (dead or healed).

In [10], the authors created a model for predicting the status of COVID-19 patients in South Korea. The total number of patients records used was 1308. The overall accuracy of recovered cases was 95.7% in training, and 93.8% in testing subset. The overall accuracy of dead cases was 99.6% in training, and 99.5% in testing subset.

The authors of [11] created a support vector regression model for predicting COVID-19 cases in India. Their model had above 97% accuracy in predicting number of dead, recovered and total number of cases, and 87% accuracy in predicting daily new COVID-19 cases in India.

Similar research was conducted in [12], where artificial neural networks were used in forecasting COVID-19 cases in several countries.

Table 1. *Total number of COVID-19 cases by WHO regions [5]*

Region	Total number of cases
Africa	815,996
Americas	9,630,598
Eastern Mediterranean	1,564,836
Europe	3,391,779
South-East Asia	2,187,015
Western Pacific	327,617
Worldwide	17,918,582

2. DATASET DESCRIPTION AND ANALYSIS

The dataset used is consisted of 26,529 patients and it is available at [13]. Of these patients, 54.93% are male, and female the rest. The distribution of patients' countries is given in Fig. 3.

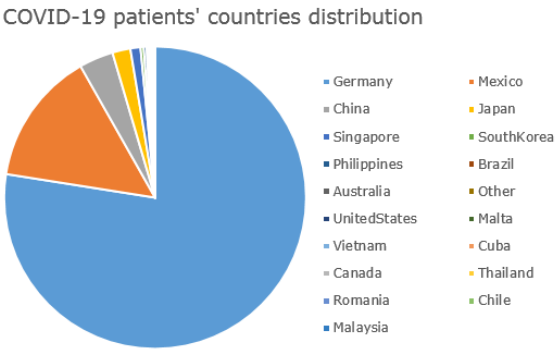

Figure 3. *The distribution of patients' countries*

The average age of patients was 44.22 years. The average age of male patients was 44.72 years, and of female patients was 43.63 years.

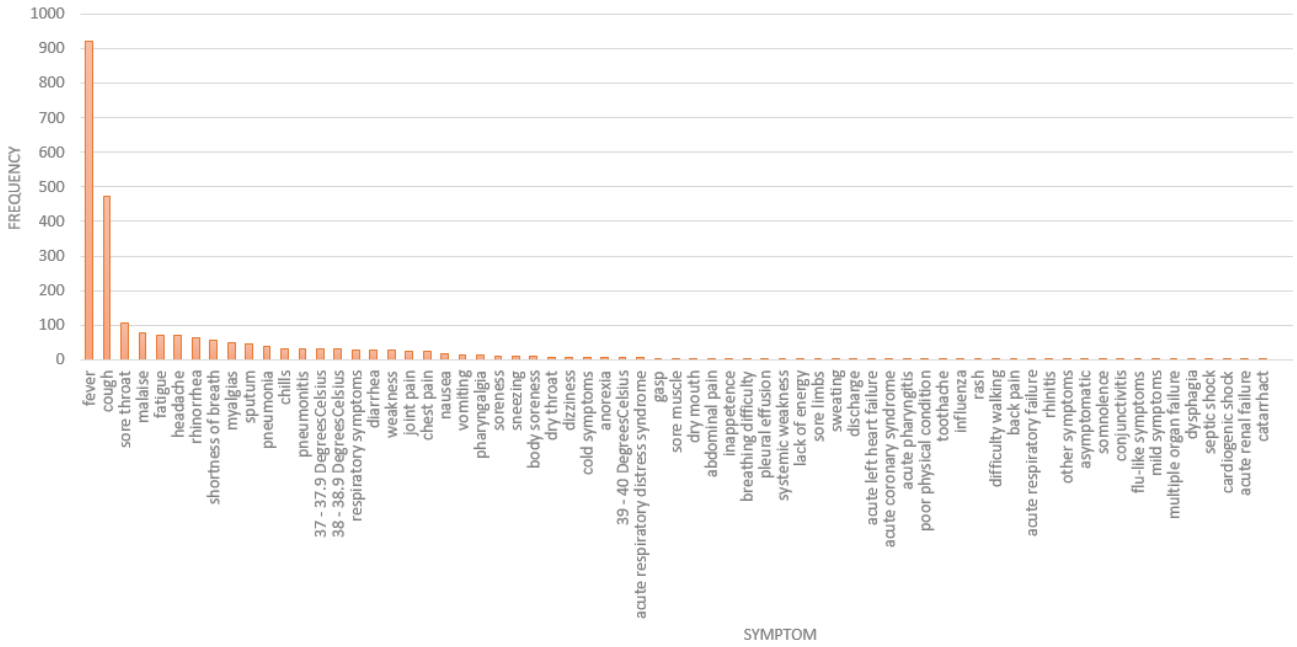

Figure 2. *The frequency of patients' symptoms in the dataset*

The number of recorded patients with symptoms was 1220. These symptoms, along with their frequency is shown in Fig. 2. From Fig. 2, it can be seen that fever and cough are the two most frequent symptoms among COVID-19 patients. Far from those, there are sore throat, malaise, fatigue, headache, rhinorrhea, shortness of breath, etc.

The total number of patients who died was 45 with an average age of 70.44 years. Out of all patients who died, 14 were women, and 31 were men. Of these 45 deceased patients, 30 had chronic diseases, with 17 of them having two or more chronic diseases. These chronic diseases with the number of occurrences among the deceased patients are given in Table 2.

Table 2. *Chronic diseases with a number of occurrences among deceased patients*

Chronic disease	Number of occurrences
Asthma	1
Cerebral infarction	1
Chronic bronchitis	2
Chronic kidney disease	2
Chronic obstructive pulmonary disease	1
Chronic renal insufficiency	1
Colon cancer	1

Coronary heart disease	4
Coronary stent	1
Diabetes	15
Encephalomalacia	1
Frequent ventricular premature beat	1
Hemorrhage of the digestive tract	1
Hip replacement	1
Hypertension	20
Parkinson's disease	2
Stenocardia	1
Tuberculosis	1
Valvular heart disease	1

As can be seen from Table 2, the two most frequent chronic diseases are hypertension and diabetes, with coronary heart disease far in third place.

The symptoms that these patients had before admitting to the hospital are given in Table 3.

Table 3. *Onset symptoms among deceased patients*

Symptom	Number of occurrences
Acute renal failure	1
Acute respiratory distress syndrome	5
Acute respiratory failure	2
Cardiogenic shock	1
Chest pain	4
Chills	1
Cough	18
Diarrhea	1
Dizziness	1
Fatigue	6
Fever	27
Gasp	4
Headache	2
Multiple organ failure	1
Myalgias	2
Pneumonia	6
Septic shock	1
Shortness of breath	4
Somnolence	1
Sputum	2
Vomiting	1
Weakness	1

As can be seen from Table 3, the top two symptoms among deceased patients are fever and cough.

The number of patients who were healed was 194. Their average age was 45.68 years, and 116 of them were men, while 78 were women. The number of discharged patients with chronic diseases was 21, with hypertension and diabetes being the most frequent. Also, the two topmost symptoms among them were fever and cough.

3. ARTIFICIAL NEURAL NETWORK MODEL AND ANALYSIS

The main goal of the research was to determine if the artificial neural network (ANN) could be used for the prediction of patients' outcome (dead or healed). Firstly, the original dataset was modified in a way that only patients that were categorized as dead or healed were included. From the original 26,529 patients, the comprised dataset consisted of 241 patients, of which 45 were deceased, and the rest were discharged from the hospital as healthy.

The ANN model was created in Python programming language, using Keras and Sci-Kit Learn libraries. The ANN consisted of three layers, one input layer of 82 neurons, one hidden layer of 41 neurons, and one output layer of one neuron. The input variables were age, sex, symptomatology with individual symptoms, presence of chronic diseases, and the number of chronic diseases. The output variable was the determination if the patient had died.

The Adam optimizer was used with a uniform kernel initializer. Also, the activation function used was the Rectified Linear Unit in the hidden layer and Sigmoid in the output layer. For the loss function, the Binary cross-entropy was chosen. For ANN training, 80% of the dataset was used, and the rest for ANN testing [14].

The initial run of the ANN model gave an accuracy of 91.15% on the training data subset. The accuracy of the testing subset was obtained with the confusion matrix, Fig. 4.

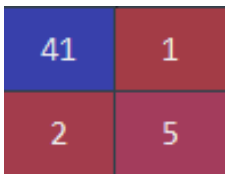

Figure 4. *The confusion matrix*

The confusion matrix is composed of True Positive (TP) cases (top left), False Negative (FN) cases (top right), True Negative (TN) cases (bottom right), and False Positive (FP) cases (bottom left). The accuracy from confusion matrix is calculated as [15]:

$$accuracy = \frac{TP+TN}{TP+TN+FP+FN} \quad (1)$$

As can be seen from Fig. 4 and Eq. (1), the accuracy of the ANN model on the testing subset was 93.88%. To establish model validity, 10-fold cross-validation was performed. Based on cross-validation results, the mean value of accuracies was 87.50% with an 8.5% variance, which means that the ANN model has significant variations in the accuracy. This can be explained by the lack of data since it was trained on only 241 patients' data.

Comparing the obtained results with relevant one from the literature [10], it is clear that the realized model has somewhat lower accuracies, but that can be due to the substantial differences in datasets used in both cases. The dataset used in [10] is over five times larger than the one used in this research.

4. CONCLUSION

COVID-19 is an infectious disease caused by the SARS-CoV-2 virus. It was first detected in China in December 2019, and in a relatively short amount of time, has widespread across the Earth. It caused many deaths, with quarantine being in effect in almost every country that has encountered it. It changed our way of life and habits and has the potential to break even the largest economies.

In this paper, the ANN model was created for predicting COVID-19 patients' outcome. It was shown that even with a relatively small amount of data, the realized model can be used for prediction of whether the patient will die or recover from it. Of course, with the availability of a larger dataset, the model could be further improved and its validity verified.

ACKNOWLEDGEMENTS

The work presented in this paper was supported by the Ministry of Education, Science and Technological Development of the Republic of Serbia, and these results are parts of the Grant No. 451-03-68/2020-14/200132 with University of Kragujevac - Faculty of Technical Sciences Čačak.

REFERENCES

[1] CDC COVID-19 Response Team. (2020). Severe Outcomes Among Patients with Coronavirus Disease 2019 (COVID-19) — United States, February 12–March 16, 2020. *Morbidity and Mortality Weekly Report, 69*, 1-4.

[2] Zheng, Y., Ma, Y., Zhang, J. & Xie, X. (2020). COVID-19 and the cardiovascular system. *Nature Reviews Cardiology, 17*, 259–260. doi: 10.1038/s41569-020-0360-5

[3] Sohrabi, C. et al. (2020). World Health Organization declares global emergency: A review of the 2019 novel coronavirus (COVID-19). *International Journal of Surgery, 76*, 71-76, doi: j.ijsu.2020.02.034

[4] World Health Organization. WHO Coronavirus Disease (COVID-19) Dashboard. Available on: https://covid19.who.int/. Last accessed on: August 3, 2020.

[5] World Health Organization. Coronavirus disease (COVID-19) Situation Report – 196. Available on: https://www.who.int/docs/default-source/coronaviruse/situation-reports/20200803-covid-19-sitrep-196-cleared.pdf?sfvrsn=8a8a3ca4_4. Last accessed on: August 3, 2020.

[6] Statistical data on COVID-19 in the Republic of Serbia. Available on: https://covid19.rs/. Last accessed on: August 3, 2020.

[7] World Health Organization. Q&A on coronaviruses (COVID-19). Available at: https://www.who.int/emergencies/diseases/novel-coronavirus-2019/question-and-answers-hub/q-a-detail/q-a-coronaviruses. Last accessed on: August 3, 2020.

[8] Cascella, M., Rajnik, M., Cuomo, A., Dulebohn, S. & Di Napoli, R. (2020). Features, Evaluation and Treatment Coronavirus (COVID-19), *In: StatPearls [Internet]. Treasure Island (FL): StatPearls Publishing*. Available at: https://www.ncbi.nlm.nih.gov/books/NBK554776/

[9] Cortegiani, A., Ingoglia, G., Ippolito, M., Giarratano, A. & Einav, S. (2020). A systematic review on the efficacy and safety of chloroquine for the treatment of COVID-19. *Journal of Critical Care, 57*, 279-283, doi: 10.1016/j.jcrc.2020.03.005

[10] Al-Najjar, H., & Al-Rousan, N. (2020). A classifier prediction model to predict the status of Coronavirus CoVID-19 patients in South Korea. *European Review for Medical and Pharmacological Sciences, 24*, 3400-3403.

[11] *Parbat, D. & Chakraborty, M. (2020).* A python based support vector regression model for prediction of COVID19 cases in India, *Chaos, Solitons & Fractals, 138*.

[12] 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.

[13] Wolfram Research (2020). Patient Medical Data for Novel Coronavirus COVID-19. *Wolfram Data Repository*. doi: 10.24097/wolfram.11224.data. Available at: https://datarepository.wolframcloud.com/resources/Patient-Medical-Data-for-Novel-Coronavirus-COVID-19

[14] Ratanamahatana, C. A., & Gunopulos, D. (2003). Feature selection for the naive bayesian classifier using decision trees. *Applied artificial intelligence*, *17*(5-6), 475-487.

[15] Visa, S., Ramsay, B., Ralescu, A. L., & Van Der Knaap, E. (2011). Confusion Matrix-based Feature Selection. *MAICS*, *710*, 120-127.