

COVID-19 Prediction with Improved Self Organizing Map

Abeer A. Al-Mohdar¹, Saleh O. Albatati²

¹Associate Professor, Department of Information Technology, College of Computers and Information Technology

²Assistant Professor of Cell Biology, Histology, and Genetics, College of Medicine

^{1,2}Hadhramout University, Mukalla, Yemen

Abstract - Coronavirus disease 2019 (COVID-19) is a highly contagious disease that is caused by Severe Acute Respiratory Syndrome Coronavirus 2 (SARS-CoV-2). It is difficult to know exactly how lethal the disease is. However, more than 5 million death were reported. The most important weapon against this virus is stopping its transmission, where early prediction and detection using machine learning techniques can play a crucial role. In this paper we propose a model based on the improved self-organizing map (IMP_SOM) for the prediction of COVID-19 from the symptoms data. With the short training time and capability of processing large amount of data, IMP-SOM showed high accuracy in predicting the disease. These results suggesting the IMP-SOM as an excellent tool to be used in this pandemic.

Key Words: COVID-19, IMP-SOM, Prediction Model, Machine Learning, Coronavirus

1. INTRODUCTION

Coronavirus disease 2019 (COVID-19) was characterized by the World Health Organization (WHO) as a pandemic in March 2020[1]. It is caused by a highly transmissible virus, Severe Acute Respiratory Syndrome Coronavirus 2 (SARS-CoV-2)[2] which emerged in Wuhan, China, at the end of 2019[3]. Since that date, more than 270 million cases and 5 million covid-19 related deaths were registered worldwide[4].

The first member of coronaviruses was discovered in 1931, which was the avian infectious bronchitis virus (IBV) followed by the discovery of other viruses including human severe acute respiratory syndrome coronavirus (SARS-CoV), Middle East respiratory syndrome coronavirus (MERS-CoV) and finally SARS-CoV-2[5]. Furthermore, the evolution of this virus generated different variants of concern in the past two years, including delta and omicron variants, that have higher transmission rate and vaccine evasion capability[6]–[8].

During this pandemic, a massive response in the scientific research community has been tried to minimize its impact in different ways. This includes early classification of the virus and the related disease[2], studying and modeling its epidemiology and diagnosis[9], [10], designing and testing drugs[11]–[13], studying short and long-term complications[14]–[18], and the most important is the very long journey to vaccination[19]. All that effort was supported by computation methods[20]–[23] and artificial intelligence tools[24]–[28].

Self-Organizing Map (SOM) is a unique Machine Learning (ML) technique that directly involved in industry[29], engineering[30], economy[31], and science[32], [33]. In medicine, SOM applications have a growing interest. It is used in medical classification of diseases[34], community health assessment[35], monitoring specific cases in the community[36] or even visualizing the publications in medical sciences[37].

In this work we used the improved self-organizing map to build a model for COVID-19 prediction by analyzing symptoms data. This model is designed to help in early prediction and diagnosis of the disease to control the spreading of the virus as early as possible.

2. METHODOLOGY

Kohonen network is a type of SOM which maps the input vectors close to each other in a discrete map. SOM consists of many of input and output units. The input pattern with their matched weights is fed and connected to the output unit by small random numbers. After many iterations, the winning output node is obtained which is a weight vector that has the smallest distance as Euclidean distance to the input pattern. In addition, the weights of every node in the neighborhood of the winning nodes are updated, and this will move each node in the neighborhood closer to the input pattern (**Figure 1**). At the same time, the learning rate will decrease as time increases. In the last step if the parameters are well defined, SOM will be able to capture the clusters of its owing input data.

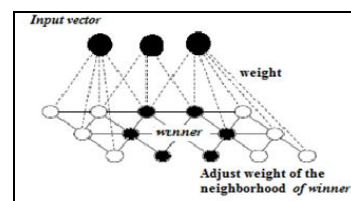


Figure 1: SOM Network

A dataset titled “COVID-19 Symptoms and Presence” were taken from Kaggle. Before applying SOM on these data, it needed to be preprocessed in order to fit the nature of SOM system.

Improved Self Organizing Map (IMP-SOM) Algorithm

IMP-SOM is started by the initialization of random weights to each node. Covid-19 data set is used in the training and

testing steps, it will process the whole data set to the map iteratively in the four steps below (Figure 2):

1) IMP-SOM Initialization: initial weight vectors as random values for the w_j and assign a small positive value to the learning rate parameter α .

2) Compute Euclidean distance: input vector X and output class d_j to activate the IMP-SOM in supervised model, and find Best Matching Unit (BMU) neuron X_i at iteration p , using the normalization of smallest, "Euclidean Distance", usual measure as in "Equ (1)",

$$E = \min_j \|X - W_j(p)\| = \sqrt{\sum_{i=1}^n [X_i - W_{ij}(p)]^2} \quad (1)$$

$j = 1, 2, \dots, m$

Where n is the number of neurons in the input layer, and m is the number of neurons in the IMP-SOM layer.

3) Updating IMP-SOM: update weight by "Equ (2)",

$$\Delta w_{ij}(p+1) = \begin{cases} \alpha [x_i - w_{ij}(p)], & j \in \Lambda_j(p) \\ 0, & j \notin \Lambda_j(p) \end{cases} \quad (2)$$

Where Θ is restraint due to distance from BMU usually called the neighborhood function, $\alpha(t)$ is the learning rate, and $W_{ij}(P)$ is the weight repairing in p^{th} iteration.

4) Repeating: return to step 2 until changing of feature map stops or no changes occur in the map. After processing all of the input, the result should be a spatial organization of the input data organized into similar regions.

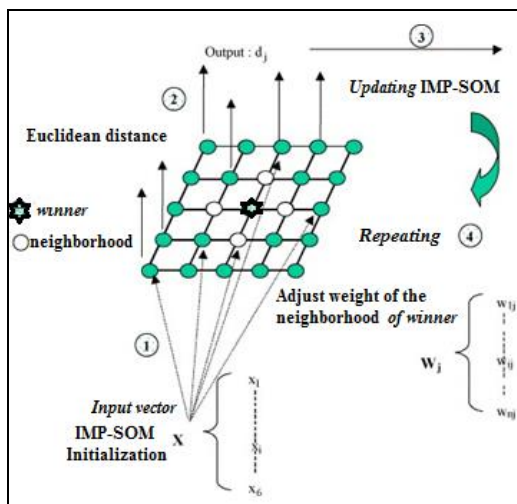


Figure 2: Proposed Model (IMP-SOM)

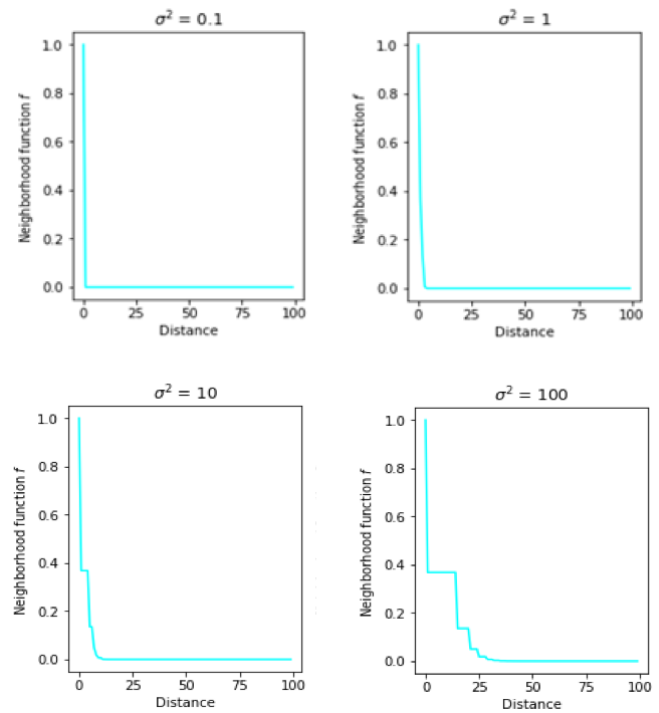


Figure 3: Representation of Training Learning Rate $\alpha(t)$

3. RESULTS AND DISCUSSION

Self-Organizing Map is a powerful tool in data classification with the fast training of the network and the high accuracy reached. These features are given the priority when selecting the ML technique for the processing of data related to human diseases to ensure a fast response with the very minimum rate of mistakes. But in COVID-19, we continuously receive an enormous amount of data, as a result of the recurrent waves of the disease. Consequently, SOM gives an additional favor when compared to many other ML techniques as it can overcome the limitation of data amount to be used.[38]

In our model (Figure 3, Table 1), IMP-SOM reached high accuracy in the detection of COVID-19 cases from symptoms data set (98.91%). This value is much higher than what achieved by two other ML classifiers used by Painuli *et al.*, random forest and extra tree classifier, with the latter got the best accuracy of 93.62%[39]. Villavicencio *et al.*[40] applied different ML techniques on COVID-19 prediction including J48 Decision Tree, Random Forest, Support Vector Machine, K-Nearest Neighbors and Naïve Bayes algorithms and reached accuracy of 98.81%. Another study by Alnajjar *et al.*[41] also applied different models and all reached an accuracy less than 90%.

Table 1: Results of The Proposed Model (IMP-SOM)

	Precision (%)	Recall (%)	F1-Score (%)	Support
Class 0	98.5	98.5	97.5	1051
Class 1	99	99	99	4382

Micro avg	98	98	98	5433
Macro avg	97	97	97	5433
Weighted avg	98	98	98	5433
Accuracy (%)	98.91			

4. CONCLUSIONS

The high transmission rate of COVID-19 causing virus (SARS-Cov-2) requires the health system response to be quick enough to avoid the spread of the disease. For that, machine learning predictive models are recommended programs. In this study, IMP-SOM based model showed very high accuracy in predicting COVID-19 in patients from symptoms data, suggesting this machine learning technique to be used. In countries like Yemen, where PCR tests are used at a very low level, we recommend health officials to increase the awareness about COVID-19 and encourage people to contact health authorities with any of COVID-19 related symptoms. The self-reported symptoms with the application of ML models will fill the gap and help in reduce the dependence on the expensive testing tools.

REFERENCES

[1] "WHO Director-General's opening remarks at the media briefing on COVID-19 - 11 March 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> (accessed Dec. 12, 2021).

[2] Coronaviridae Study Group of the International Committee on Taxonomy of Viruses, "The species Severe acute respiratory syndrome-related coronavirus: classifying 2019-nCoV and naming it SARS-CoV-2," *Nat. Microbiol.*, vol. 5, no. 4, pp. 536–544, Apr. 2020, doi: 10.1038/s41564-020-0695-z.

[3] B. Hu, H. Guo, P. Zhou, and Z.-L. Shi, "Characteristics of SARS-CoV-2 and COVID-19," *Nat. Rev. Microbiol.*, vol. 19, no. 3, pp. 141–154, Mar. 2021, doi: 10.1038/s41579-020-00459-7.

[4] "COVID-19 Map," Johns Hopkins Coronavirus Resource Center. <https://coronavirus.jhu.edu/map.html> (accessed Dec. 12, 2021).

[5] P. V'kovski, A. Kratzel, S. Steiner, H. Stalder, and V. Thiel, "Coronavirus biology and replication: implications for SARS-CoV-2," *Nat. Rev. Microbiol.*, vol. 19, no. 3, pp. 155–170, Mar. 2021, doi: 10.1038/s41579-020-00468-6.

[6] E. J. Rubin, L. R. Baden, S. S. Abdool Karim, and S. Morrissey, "Audio Interview: The Omicron Variant of SARS-CoV-2," *N. Engl. J. Med.*, vol. 385, no. 23, p. e96, Dec. 2021, doi: 10.1056/NEJMe2118839.

[7] J. Lopez Bernal et al., "Effectiveness of Covid-19 Vaccines against the B.1.617.2 (Delta) Variant," *N. Engl. J. Med.*, vol. 385, no. 7, pp. 585–594, Aug. 2021, doi: 10.1056/NEJMoa2108891.

[8] D. Singh and S. V. Yi, "On the origin and evolution of SARS-CoV-2," *Exp. Mol. Med.*, vol. 53, no. 4, pp. 537–547, Apr. 2021, doi: 10.1038/s12276-021-00604-z.

[9] T. Šušteršič et al., "Epidemiological Predictive Modeling of COVID-19 Infection: Development, Testing, and Implementation on the Population of the Benelux Union," *Front. Public Health*, vol. 9, p. 727274, Oct. 2021, doi: 10.3389/fpubh.2021.727274.

[10] M. Locquet et al., "A systematic review of prediction models to diagnose COVID-19 in adults admitted to healthcare centers," *Arch. Public Health*, vol. 79, no. 1, p. 105, Dec. 2021, doi: 10.1186/s13690-021-00630-3.

[11] G. Culletta, M. R. Gulotta, U. Perricone, M. Zappalà, A. M. Almerico, and M. Tutone, "Exploring the SARS-CoV-2 Proteome in the Search of Potential Inhibitors via Structure-Based Pharmacophore Modeling/Docking Approach," *Computation*, vol. 8, no. 3, p. 77, Sep. 2020, doi: 10.3390/computation8030077.

[12] D. M. Weinreich et al., "REGEN-COV Antibody Combination and Outcomes in Outpatients with Covid-19," *N. Engl. J. Med.*, vol. 385, no. 23, p. e81, Dec. 2021, doi: 10.1056/NEJMoa2108163.

[13] M. Wang et al., "Remdesivir and chloroquine effectively inhibit the recently emerged novel coronavirus (2019-nCoV) in vitro," *Cell Res.*, vol. 30, no. 3, pp. 269–271, Mar. 2020, doi: 10.1038/s41422-020-0282-0.

[14] D. A. Berlin, R. M. Gulick, and F. J. Martinez, "Severe Covid-19," *N. Engl. J. Med.*, vol. 383, no. 25, pp. 2451–2460, Dec. 2020, doi: 10.1056/NEJMcp2009575.

[15] S. A. Azer, "COVID-19: pathophysiology, diagnosis, complications and investigational therapeutics," *New Microbes New Infect.*, vol. 37, p. 100738, Sep. 2020, doi: 10.1016/j.nmni.2020.100738.

[16] T. M. Drake et al., "Characterisation of in-hospital complications associated with COVID-19 using the ISARIC WHO Clinical Characterisation Protocol UK: a prospective, multicentre cohort study," *The Lancet*, vol. 398, no. 10296, pp. 223–237, Jul. 2021, doi: 10.1016/S0140-6736(21)00799-6.

[17] K. Yang et al., "Complication and Sequelae of COVID-19: What Should We Pay Attention to in the Post-Epidemic Era," *Front. Immunol.*, vol. 12, p. 711741, Sep. 2021, doi: 10.3389/fimmu.2021.711741.

[18] S. SeyedAlinaghi et al., "Late Complications of COVID-19; a Systematic Review of Current Evidence," *Arch.*

Acad. Emerg. Med., vol. 9, no. 1, p. e14, Jan. 2021, doi: 10.22037/aaem.v9i1.1058.

[19] "COVID-19 vaccine - Wikipedia." https://en.wikipedia.org/wiki/COVID-19_vaccine (accessed Dec. 15, 2021).

[20] Z. Qiao, H. Zhang, H.-F. Ji, and Q. Chen, "Computational View toward the Inhibition of SARS-CoV-2 Spike Glycoprotein and the 3CL Protease," *Computation*, vol. 8, no. 2, p. 53, May 2020, doi: 10.3390/computation8020053.

[21] I. A. Muhammad et al., "A Computational Study to Identify Potential Inhibitors of SARS-CoV-2 Main Protease (Mpro) from Eucalyptus Active Compounds," *Computation*, vol. 8, no. 3, p. 79, Sep. 2020, doi: 10.3390/computation8030079.

[22] S. Sarv Ahrabi, M. Scarpiniti, E. Baccarelli, and A. Momenzadeh, "An Accuracy vs. Complexity Comparison of Deep Learning Architectures for the Detection of COVID-19 Disease," *Computation*, vol. 9, no. 1, p. 3, Jan. 2021, doi: 10.3390/computation9010003.

[23] W. Zeng, A. Gautam, and D. H. Huson, "On the Application of Advanced Machine Learning Methods to Analyze Enhanced, Multimodal Data from Persons Infected with COVID-19," *Computation*, vol. 9, no. 1, p. 4, Jan. 2021, doi: 10.3390/computation9010004.

[24] N. Arora, A. K. Banerjee, and M. L. Narasu, "The role of artificial intelligence in tackling COVID-19," *Future Virol.*, vol. 15, no. 11, pp. 717–724, Nov. 2020, doi: 10.2217/fvl-2020-0130.

[25] M. Abdulkareem and S. E. Petersen, "The Promise of AI in Detection, Diagnosis, and Epidemiology for Combating COVID-19: Beyond the Hype," *Front. Artif. Intell.*, vol. 4, p. 652669, May 2021, doi: 10.3389/frai.2021.652669.

[26] M. Khan et al., "Applications of artificial intelligence in COVID-19 pandemic: A comprehensive review," *Expert Syst. Appl.*, vol. 185, p. 115695, Dec. 2021, doi: 10.1016/j.eswa.2021.115695.

[27] Q.-V. Pham, D. C. Nguyen, T. Huynh-The, W.-J. Hwang, and P. N. Pathirana, "Artificial Intelligence (AI) and Big Data for Coronavirus (COVID-19) Pandemic: A Survey on the State-of-the-Arts," *IEEE Access*, vol. 8, pp. 130820–130839, 2020, doi: 10.1109/ACCESS.2020.3009328.

[28] F. Shi et al., "Review of Artificial Intelligence Techniques in Imaging Data Acquisition, Segmentation, and Diagnosis for COVID-19," *IEEE Rev. Biomed. Eng.*, vol. 14, pp. 4–15, 2021, doi: 10.1109/RBME.2020.2987975.

[29] D. Galvan, L. Effting, L. Torres Neto, and C. A. Conte-Junior, "An overview of research of essential oils by self-organizing maps: A novel approach for meta-analysis

study," *Compr. Rev. Food Sci. Food Saf.*, vol. 20, no. 4, pp. 3136–3163, Jul. 2021, doi: 10.1111/1541-4337.12773.

[30] M. Pacella, A. Grieco, and M. Blaco, "On the Use of Self-Organizing Map for Text Clustering in Engineering Change Process Analysis: A Case Study," *Comput. Intell. Neurosci.*, vol. 2016, pp. 1–11, 2016, doi: 10.1155/2016/5139574.

[31] Y. Cai, X. Wang, and L. Xiong, "Difference Analysis of Regional Economic Development Based on the SOM Neural Network with the Hybrid Genetic Algorithm," *Comput. Intell. Neurosci.*, vol. 2021, pp. 1–9, Sep. 2021, doi: 10.1155/2021/6734345.

[32] R. Rallo et al., "Self-Organizing Map Analysis of Toxicity-Related Cell Signaling Pathways for Metal and Metal Oxide Nanoparticles," *Environ. Sci. Technol.*, vol. 45, no. 4, pp. 1695–1702, Feb. 2011, doi: 10.1021/es103606x.

[33] M. Weber et al., "Practical application of self-organizing maps to interrelate biodiversity and functional data in NGS-based metagenomics," *ISME J.*, vol. 5, no. 5, pp. 918–928, May 2011, doi: 10.1038/ismej.2010.180.

[34] M. Zribi, Y. Boujelbene, I. Abdelkafi, and R. Feki, "The self-organizing maps of Kohonen in the medical classification," in *2012 6th International Conference on Sciences of Electronics, Technologies of Information and Telecommunications (SETIT)*, Sousse, Tunisia, Mar. 2012, pp. 852–856. doi: 10.1109/SETIT.2012.6482027.

[35] H. G. Basara and M. Yuan, "Community health assessment using self-organizing maps and geographic information systems," *Int. J. Health Geogr.*, vol. 7, no. 1, p. 67, 2008, doi: 10.1186/1476-072X-7-67.

[36] J. Tuckova, "The possibility of kohonen self-organizing map applications in medicine," in *2013 IEEE 11th International Workshop of Electronics, Control, Measurement, Signals and their application to Mechatronics, Toulouse Cedex 7, France, Jun. 2013*, pp. 1–6. doi: 10.1109/ECMSM.2013.6648946.

[37] A. Skupin, J. R. Biberstine, and K. Börner, "Visualizing the Topical Structure of the Medical Sciences: A Self-Organizing Map Approach," *PLoS ONE*, vol. 8, no. 3, p. e58779, Mar. 2013, doi: 10.1371/journal.pone.0058779.

[38] T. Kohonen, "The self-organizing map," *Proc. IEEE*, vol. 78, no. 9, pp. 1464–1480, Sep. 1990, doi: 10.1109/5.58325.

[39] D. Painuli, D. Mishra, S. Bhardwaj, and M. Aggarwal, "Forecast and prediction of COVID-19 using machine learning," in *Data Science for COVID-19*, Elsevier, 2021, pp. 381–397. doi: 10.1016/B978-0-12-824536-1.00027-7.

[40] C. N. Villavicencio, J. J. E. Macrohon, X. A. Inbaraj, J.-H. Jeng, and J.-G. Hsieh, "COVID-19 Prediction Applying

Supervised Machine Learning Algorithms with Comparative Analysis Using WEKA,” *Algorithms*, vol. 14, no. 7, p. 201, Jun. 2021, doi: 10.3390/a14070201.

[41] D. Al-Najjar, H. Al-Najjar, and N. Al-Rousan, “Evaluation of the prediction of CoVID-19 recovered and unrecovered cases using symptoms and patient’s meta data based on support vector machine, neural network, CHAID and QUEST Models,” *Eur. Rev. Med. Pharmacol. Sci.*, vol. 25, no. 17, pp. 5556–5560, Sep. 2021, doi: 10.26355/eurrev_202109_26668.