

Approach for Training Quantum Neural Network to Predict Severity of COVID-19 in Patients

Engy El-shafeiy¹, Aboul Ella Hassanien², Karam M. Sallam^{3,*} and A. A. Abohany⁴

¹Department of Computer Science, Faculty of Computers and Artificial Intelligence, University of Sadat City, Sadat City, 32897, Egypt

²Faculty of Computers and Artificial Intelligence, Cairo University, 12613, Egypt

³Faculty of Computers and Information, Zagazig University, 44519, Egypt

⁴Faculty of Computers and Information, Kafrelsheikh University, 33516, Egypt

*Corresponding Author: Karam M. Sallam. Email: karam_sallam@zu.edu.eg

Received: 24 July 2020; Accepted: 11 September 2020

Abstract: Currently, COVID-19 is spreading all over the world and profoundly impacting people's lives and economic activities. In this paper, a novel approach called the COVID-19 Quantum Neural Network (CQNN) for predicting the severity of COVID-19 in patients is proposed. It consists of two phases: In the first, the most distinct subset of features in a dataset is identified using a Quick Reduct Feature Selection (QRFS) method to improve its classification performance; and, in the second, machine learning is used to train the quantum neural network to classify the risk. It is found that patients' serial blood counts (their numbers of lymphocytes from days 1 to 15 after admission to hospital) are associated with relapse rates and evaluations of COVID-19 infections. Accordingly, the severity of COVID-19 is classified in two categories, serious and non-serious. The experimental results indicate that the proposed CQNN's prediction approach outperforms those of other classification algorithms and its high accuracy confirms its effectiveness.

Keywords: Predict COVID-19; lymphocytic count; quantum neural network; dynamic change

1 Introduction

Towards the end of 2019, the latest coronavirus disease (COVID-19) emerged in China and spread quickly across the globe due to advanced means of transport. To date, it has claimed hundreds of thousands of lives in China and globally, and is the most powerful virus humanity has had to face since the so-called Spanish flu in 1919. It is expected that advanced technologies will help to overcome it by detecting it in real time and expediting the discovery of a possible treatment for it using supercomputers and advanced machine-learning algorithms [1].

Genomic sequencing is the process by which a lab technician analyzes a person's blood sample and prepares it to sequence a human cell which contains 23 pairs of chromosomes. This structure contains the person's DNA which is coiled in a form called a double helix that can be unwound into a ladder shape



This work is licensed under a Creative Commons Attribution 4.0 International License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

made of 6 billion paired chemical elements called bases. To read these bases, a blood sample is inserted into a sequencing instrument in which a high-frequency sound wave breaks down its DNA, each fragment of which is sequentially copied hundreds of thousands of times, with clusters of identical ones created. Powerful computers combine the individual fragments to reveal the sequence of this DNA and then a medical team can use software to analyze and compare different sequences [2]. Using such technology, the DNA of the current coronavirus was discovered very quickly and provided information that could be used to conduct PCR tests.

Taiwan suffered an outbreak of severe acute respiratory syndrome (SARS) in 2003, recording 346 cases and 37 deaths. Informed by lessons learned, Taiwan has used all its knowledge, specifically in technology, to stop the spread of COVID-19. It has integrated people's recent histories of travel to China from customer and immigration databases to complement cloud-based health records so that health professionals can be aware of patients' recent travels, determine whether they entered infected areas and display a warning if they had visited China, in particular Wuhan, in the previous 3 months [3]. Huang et al. [4] identified a recent pneumonia cluster in Wuhan, China, caused by COVID-19. They documented the radiological, laboratory, clinical and epidemiological treatments implemented and the characteristics and clinical outcomes of the patients. Lu et al. [5] considered that the genetic architecture of this human coronavirus could cause severe pneumonia and also shed light on its origin and receptor-binding properties. The outbreaks of diseases linked to COVID-19 highlight the secret reservoirs of viruses in wild animals and their capabilities to periodically spread to human populations.

Zhu [6] reported an unknown cause of pneumonia in a group of patients associated with a wholesale seafood market in Wuhan, China, with a newly discovered betacoronavirus identified by objective sequencing in samples from these patients with pneumonia. Human airway epithelial cells that have been used to extract COVID-19 generate a clade within the sub-genus sarbecovirus which is a member of the orthocoronavirinae sub-family. Although different from both SARS-CoV and MERS-CoV, COVID-19 is the seventh member of the family of coronaviruses to infect humans. Chan et al. [7] identified a family group with pneumonia associated with COVID-19 that indicated person-to-person transmission. They documented their microbiological, radiological, laboratory, clinical and epidemiological findings for 5 patients in this group who had undiagnosed pneumonia after returning to Shenzhen, Guangdong province, China, after visiting Wuhan, and an additional family member who hadn't traveled to Wuhan.

Like an artificial neural network (ANN), a new, applicable and useful concept introduced recently is the Quantum Neural Network (QNN) [8]. It integrates the quantum computational paradigm with the basics of ANN and is superior to the conventional ANN. It is used to handle big data, function approximation, computer games, etc. and its algorithms are applied in approach-automated control systems, associative memory devices, social networks, etc., Beer et al. [9] suggested a complete quantum analog of classical neurons which constructed quantum feedforward NNs capable of universal quantum computing. They defined the effective training of these networks using fidelity as a cost function which provided both classical and efficient quantum implementations. Their method involved rapid optimization with decreased storage requirements as the number of qubits needed scales with only the width, thereby enabling deep-network optimization. They benchmarked their proposal for the quantum task of learning and found remarkable generalization behavior and striking robustness to noisy training data. Gao et al. [10] presented a novel method for deep learning to determine a person's state of health. Firstly, one deep-learning approach, called stacked denoising auto-encoders, found features in the raw data for retaining the original information and then inserted them into a QNN to classify the dataset, with the loss function of the QNN enhancing its classification performance. Experiments conducted on benchmark datasets showed that their proposed method was more robust and effective than traditional ones. Also, they built an integrated modular avionics degradation approach for changing the probability of occurrence of soft faults during the whole life serves.

This paper presents an approach based on a QNN for predicting the severity of COVID-19 in patients. It uses the serial blood counts performed during their hospitalizations to record their lymphocytic counts from days 1 to 15 which relate to the relapse rate.

This paper is organized as follows: in Section 2, the clinical characteristics of COVID-19 patients are discussed; in Section 3, training QNNs is described; in Section 4, Quick Reduct Feature Selection (QRFS) is explained; in Section 5, the proposed approach for predicting the severity of COVID-19 based on the QNN is described; in Section 6, the simulation results are presented; and, in Section 7, a conclusion is provided.

2 Clinical Characteristics of Datasets of COVID-19 Patients Structure

Zhang et al. [11] obtained the datasets used in this study from the National Health Commission of the People's Republic of China. They reported that 13 patients with non-serious COVID-19 previously diagnosed on admission to the First Affiliated Hospital of Xi'an Jiaotong University between January 22 and February 2, 2020 [11] were included. Tab. 1 describes the main characteristics of these datasets attributes.

Table 1: Characteristics of datasets

ID	Attribute Name	Type	Minimum Value	Maximum Value
1	Time of onset of illness	Date/Time	16/01/2020	31/01/2020
2	Time of admission	Date/Time	22/01/2020	02/02/2020
3	Gender	Text		Male or Female
4	Age	Numeric	22	70
5	Epidemiological history	Text		Wuhan, China
6	Comorbidities	Text	Hypertension, diabetes, IgA nephropathy, hypothyroidism and bronchial asthma	
7	Febrile days	Numeric	1	10
8	Temperature before admission	Numeric	37.3°C	40°C
9	Concomitant symptoms	Text	Pharyngalgia, dry cough, feebleness and shortness of breath	
10	Time body temperature returns to normal	Date/Time	27/01/2020	16/02/2020
11	Oxygen uptake	Boolean		Yes or No
12	Lopinavir/ritonavir	Boolean		Yes or No
13	Arbidol hydrochloride	Boolean		Yes or No
14	Interferon atomization	Boolean		Yes or No
15	WBC	Numeric	2.58	11.75
16	NEUT	Numeric	1.3	9.38
17	LYMPH	Numeric	0.56	1.4
18	NEUT%	Numeric	42.50%	79.80%
19	LYMPH%	Numeric	9.60%	45.70%
20	PLT	Numeric	82	255
21	HGB	Numeric	120	169
22	RBC	Numeric	4.16	5.66
23	ALT	Numeric	6	85
24	AST	Numeric	19	54
25	ALB	Numeric	35.4	52.6
26	TBIL	Numeric	2.3	26.6

(Continued)

Table 1 (continued).

ID	Attribute Name	Type	Minimum Value	Maximum Value
27	DBIL	Numeric	1.2	9.6
28	CRP	Numeric	10	38.9
29	PCT	Numeric	0.022	0.108
30	IL-6	Numeric	1.5	40.9
31	LYMPH0	Numeric	0.56	1.4
32	LYMPH3	Numeric	0.4	1.45
33	LYMPH5	Numeric	0.37	1.65
34	LYMPH7	Numeric	0.37	1.91
35	LYMPH9	Numeric	0.51	2.08
36	LYMPH11	Numeric	0.32	3.06
37	LYMPH13	Numeric	0.55	1.98
38	LYMPH15	Numeric	0.45	2.43

3 Training Quantum Neural Networks

A QNN [8,9] is a computational model based on a set of artificial neuronal units, the behaviors of which are roughly comparable to those observed in the axons of neurons in biological brains. Each neuron is connected to many others, which can increase the activation function of its adjacent neurons, and it operates individually using additional functions. There can be a threshold for each connection to an actual neuron; for example, a signal must exceed a limit before it can propagate to another neuron. These systems learn and train themselves rather than being explicitly programmed, and excel in areas where detecting solutions or features are difficult using conventional programming.

As shown in Fig. 1, σ is the non-linear function, h_i the value of neuron i in each hidden layer, $R(h_i)$ the hidden layer and y the final prediction value generated from the hybrid network.

4 Reducing Set of Attributes

A simple approach can be used to identify some attribute values that are not necessary for the dataset. Then, reducing the number of them in a set based on the reduced information policy is a separate process known as shorthand, whereby no more attributes can be deleted without losing information in the dataset. As discounts are minimal sub-groups that do not have any irreplaceable features, a reduction should be capable of classifying objects without changing the form of knowledge representation [12,13]. Subset A of group D is called an abbreviation if, and only if, it includes the following characteristics:

$$\frac{U}{A} = \frac{U}{D} \quad (1)$$

$$\frac{U}{(A - \{a\})} \neq \frac{U}{D}, \text{ for all } a \in A \quad (2)$$

$$[x]_{A-\{a\}} \neq [x]_D, \text{ for all } a \in A \quad (3)$$

$$\text{DF of subset A} = \text{DF of set D} \quad (4)$$

The substance is necessary to represent the knowledge or rules and is the main part section of all reductions that is:

$$\text{Core (dataset)} = \cap \text{reduct (attribute in dataset)} \quad (5)$$

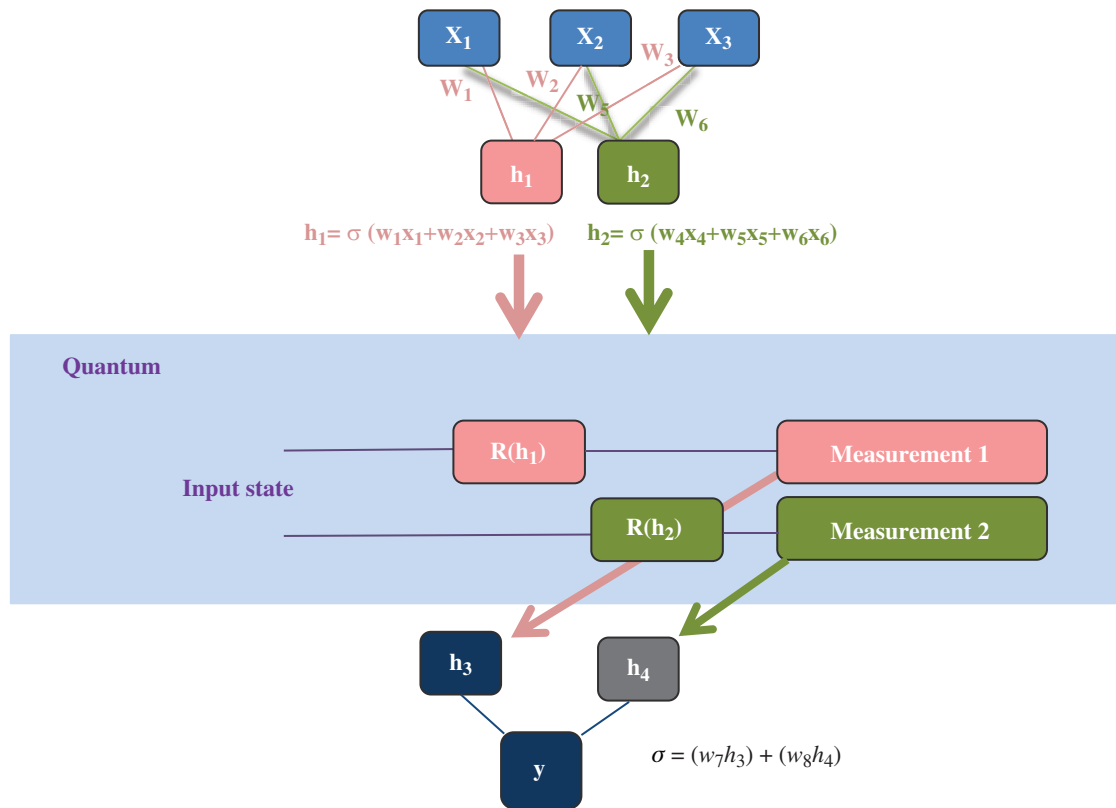


Figure 1: Quantum neural network

5 Proposed CQNN Approach for Predicting Severity of COVID-19 in Patients

Since there are no reliable risk-splitting methods for determining patients with serious COVID-19 infections, we aimed to build an effective means of identifying early cases at risk of becoming mild to severe.

The proposed CQNN approach consists of two stages; in the first, the most distinctive subset of features is selected using the QRFS method to improve its classification performance and, in the second, the QNN predicts the levels of severity of COVID-19 in patients. It does this by learning the classifications of patients with COVID-19, with the training sub-step contributing to the forecasts according to its best feature selection using a test sample. For QRFS, 24 laboratory parameters, that is, WBC, NEUT, LYMPH, NEUT%, LYMPH%, PLT, HGB, RBC, ALT, AST, ALB, TBIL, DBIL, CRP, PCT, IL-6, LYMPH0, LYMPH3, LYMPH5, LYMPH7, LYMPH9, LYMPH11, LYMPH13 and LYMPH15, and 6 clinical ones, that is, oxygen uptake, age, febrile days, temperature before admission, concomitant symptoms and epidemiological history, are recommended.

One of the criteria for diagnosing a COVID-19 infection [14,15] is a normal laboratory parameter such as ‘normal/decreased number of leukocytes’ or ‘decreased number of lymphocytes’, as shown in Tab. 2.

Table 2: Normal values for leukocyte count and total WBC count

Differential WBC counts of adult	
Leukocyte count, total (WBC count)	3600–11,200/mm ³
WBC count, total	3600–11,200/mm ³
Neutrophils	45–79%
Lymphocytes	16–47%

The analysis conducted of this dataset suggests that a lower LYMPH number is a potentially more reliable laboratory predictor of a SARS-CoV-2 infection than the recommended ‘lymphocytic counts’ and ‘lymphopenia’. As long as the two clinical manifestations of ‘fever and/or respiratory symptoms’ and ‘normal or reduced number of white blood cells or reduced number of lymphocytes at the onset of symptoms’ are observed, individuals may be considered suspect cases according to the diagnosis and treatment of pneumonia caused by COVID-19 [16].

The training rule, which is the major aspect of the CQNN, modifies the weights to eliminate the mean square error (MSE). The general CQNN mathematical equations are as follows based on Tab. 3 parameters.

$$\delta_k = (t_k - O_k)O_k(1 - O_k) \{k = 1, 2, 3, \dots, K\} \quad (6)$$

$$\delta_j = O_j(1 - O_j) \sum_k^K W_{kj}^{New} \delta_k \{j = 1, 2, 3, \dots, J\} \{k = 1, 2, 3, \dots, K\} \quad (7)$$

$$W_{kj}^{New} = W_{kj}^{old} + \eta \delta_k O_j \{j = 1, 2, 3, \dots, J\} \{k = 1, 2, 3, \dots, K\} \quad (8)$$

$$W_{ji}^{New} = W_{ji}^{old} + \eta \delta_j O_i \{i = 1, 2, 3, \dots, I\} \{j = 1, 2, 3, \dots, J\} \quad (9)$$

$$SGM(x) = \left(\frac{1}{(1 + \exp(-x))} \right) \quad (10)$$

$$SGM(x) = \sum_{r=1}^{n_x} \left(\frac{1}{(1 + \exp(-x + \theta_r))} \right) \quad (11)$$

$$\theta = \theta_r + \Delta \quad (12)$$

Table 3: Parameters used in CQNN approach

Parameters used			
θ_r	Quantum interval at r	O_k	Output layer
n_s	Number of levels	W_{ij}	Weights between input and hidden layers
δ_k	Function error (output layer)	W_{kj}	Weights between hidden and output layers
δ_j	Function error (hidden layer)	$y(t)$	Target value
n_i	Features for input layer		

The hidden layer in the ANN CQNN uses non-linear activation functions ($\mathbf{y(t)}$). The complete approach presented in the previous section is tested in terms of accuracy based on the dataset given in [11] to classify COVID-19 patients. In the experimental environment, the data are separated into 23% for testing and 77% for training using our NN structures.

Fig. 2 shows the structure of the QNN. It uses 30 neurons as an input sample layer, one for each input feature in the dataset’s vector, 10 in the hidden layer and 2 in the output layer, one for each class of COVID-19 (serious and non-serious). After partitioning the data into two groups (training and testing), as shown in Fig. 2, the approach is built based on the training dataset which contains 13 cases, with 0 and 1 representing non-serious and serious ones, respectively. Fig. 3 shows the Flowchart of the proposed approach.

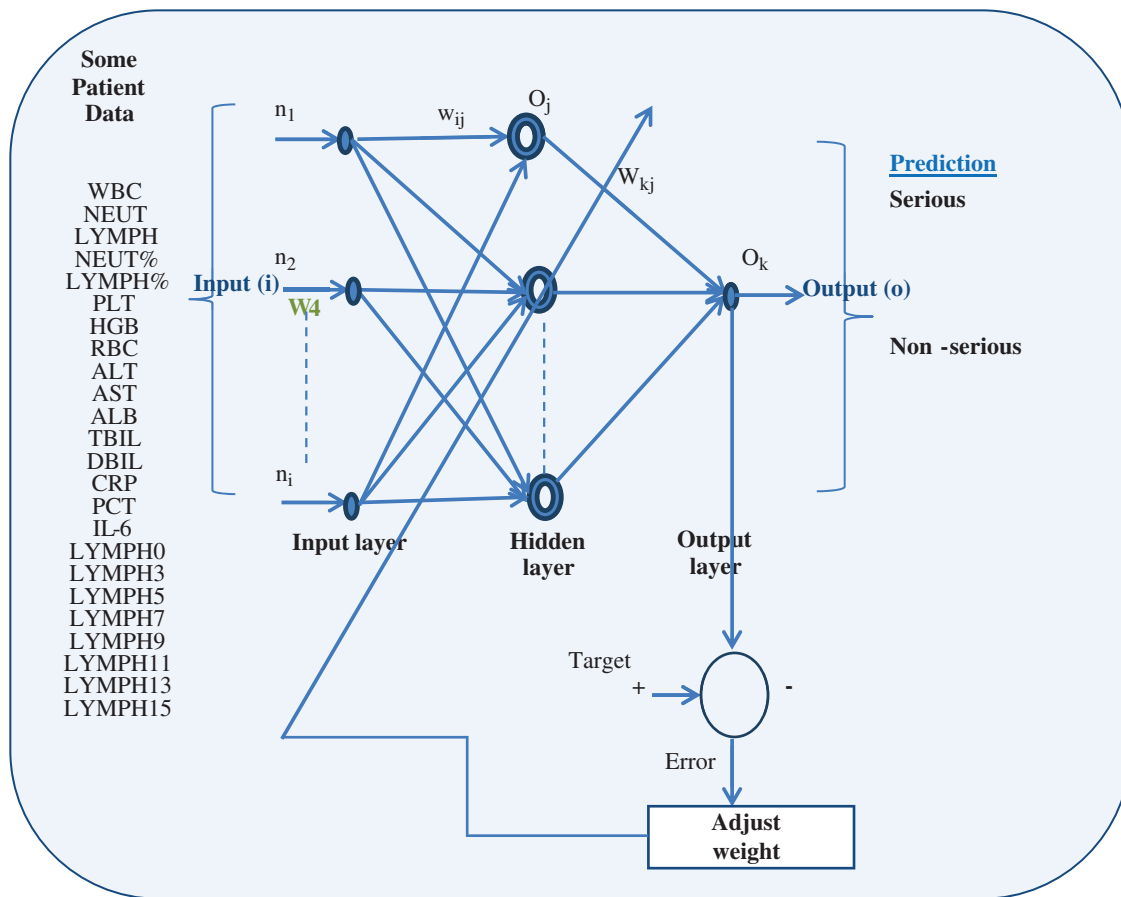


Figure 2: Details of quantum neural network approach

6 Results

In the Covid-19 dataset, of 12 acute non-acute patients, 91.7% showed abnormal or low white blood cell (WBC) counts, 6 had lymphocytes, 2 a decrease in the number of platelets, 4 an increase in CRP and 5 an increase in IL-6. 13 patients had normal levels of PCT and 3 an increase in ALT and AST upon admission. The data used the effect of several parameters, such as blood culture, blood counts (HGB, WBC and PLT), NEUT% and LYMPH%, with the statistical correlations among them shown in Fig. 4. During the first 3 days, the number of COVID-19 lymphocytes did not change significantly but vibrated beneath normal before the disease became severe.

All patients' data were recorded within 2 days of the onset of the disease, with their numbers of lymphocytes appearing not to increase 2 weeks later which suggested that they were very likely to be incubating severe COVID-19 because a dynamic change in the number of lymphocytes early can be an important indicator of the development of serious diseases. This easily commensurable parameter can assist clinicians to identify patients at severe risk of COVID-19 very quickly. Lymphocytosis is common in seriously ill patients (with MERS-CoV or SARS-CoV) because the infestation of viral particles destroys lymphocytes. We assume that SARS-CoV affects mainly lymphocytes which can suppress the cellular immune function in the body and cause a cytokine storm which worsens the disease in some patients. Several papers have stated that lymphocytosis may be a determining factor in the severity of COVID-19. We also performed sub-group analyses of the patients with non-serious COVID-19.

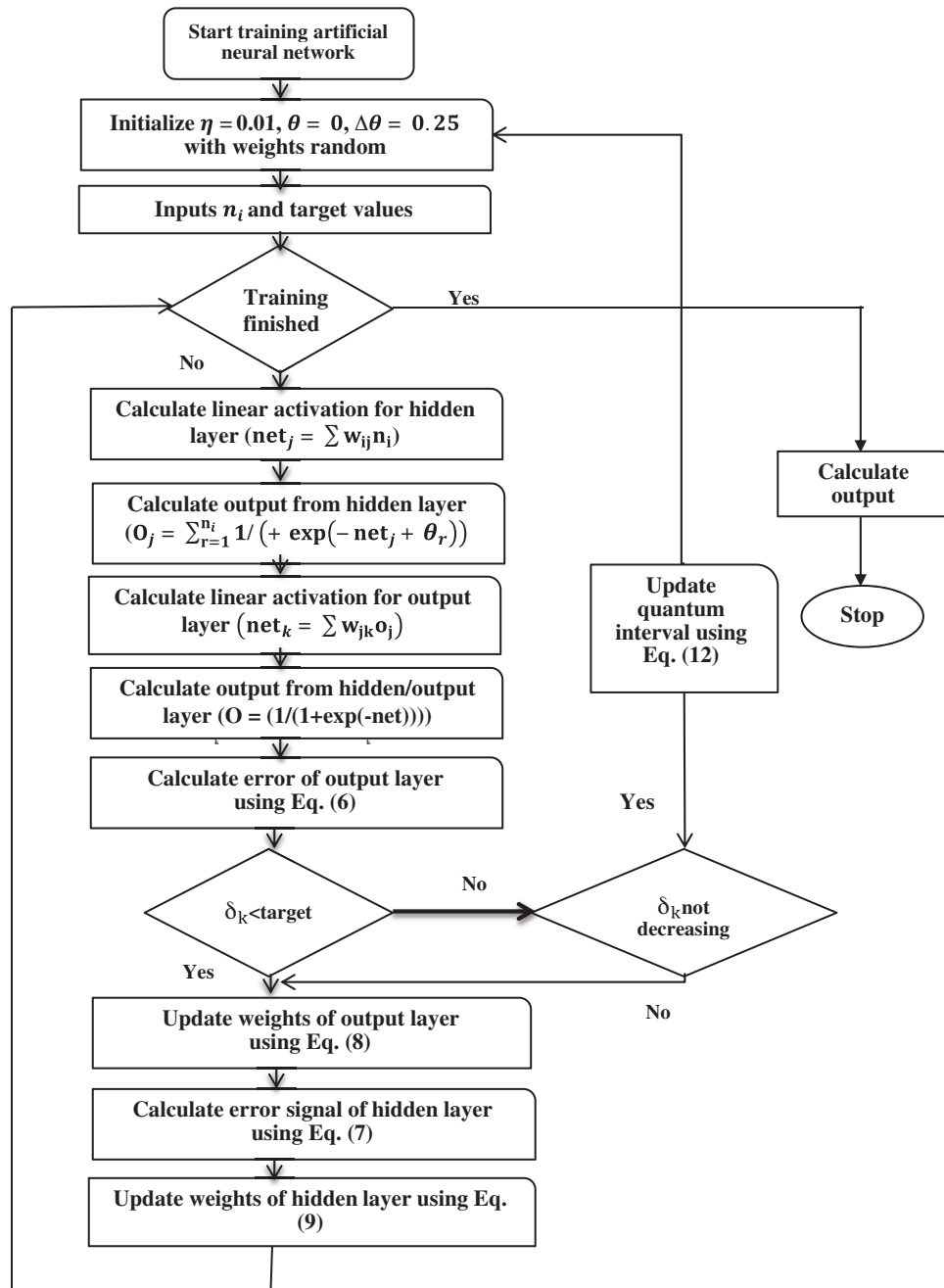


Figure 3: Flowchart of QNN approach

In the proposed approach, the 2 patients in group one took longer than those in group 1 (means of 15 and 5 days, respectively, $P = 0.05$) to normalize their body temperatures, as shown in Fig. 5. These results indicated that patients with fewer than the normal number of lymphocytes in the early stages of the disease should have more severe lung lesions and, therefore, be likely to recover more slowly than those with normal numbers of lymphocytes.

The results demonstrated that a low lymphocytic level was a strong predictor of the detection of COVID-19.

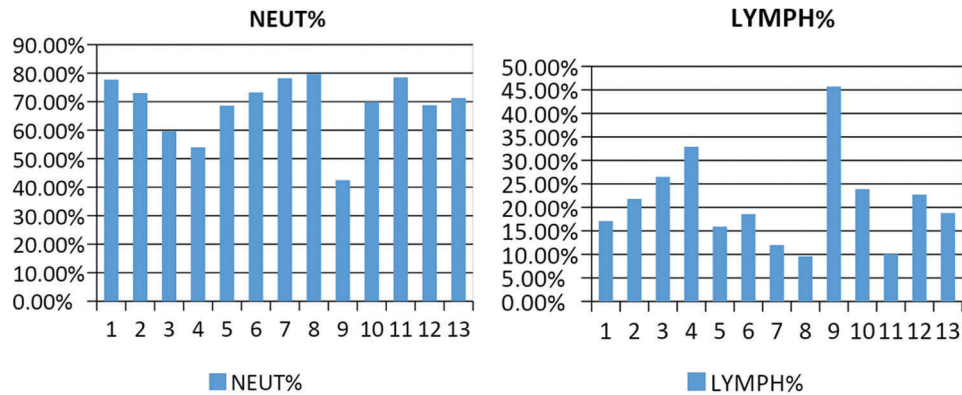


Figure 4: Statistical correlations of NEUT and LYMPH values for patients in dataset

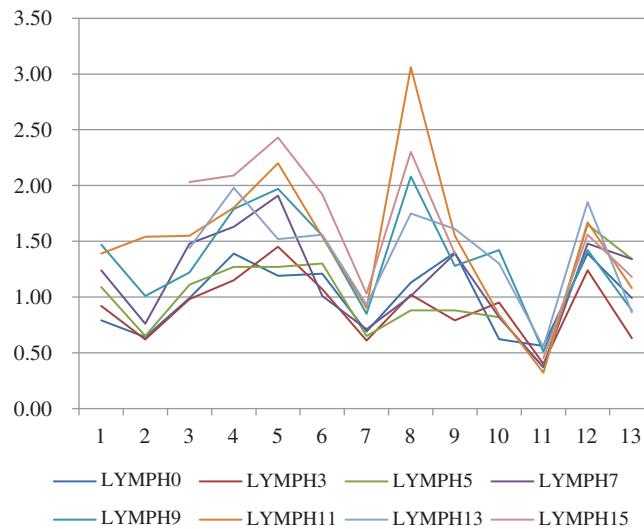


Figure 5: LYMPH% values over 15 days for 13 patients

A decrease in the lymph index for Covid-19 was observed mainly at $n = 15$. Only a few cases overlapped between the two groups of patients (non-serious and serious), as shown in [Figs. 6](#) and [7](#), respectively.

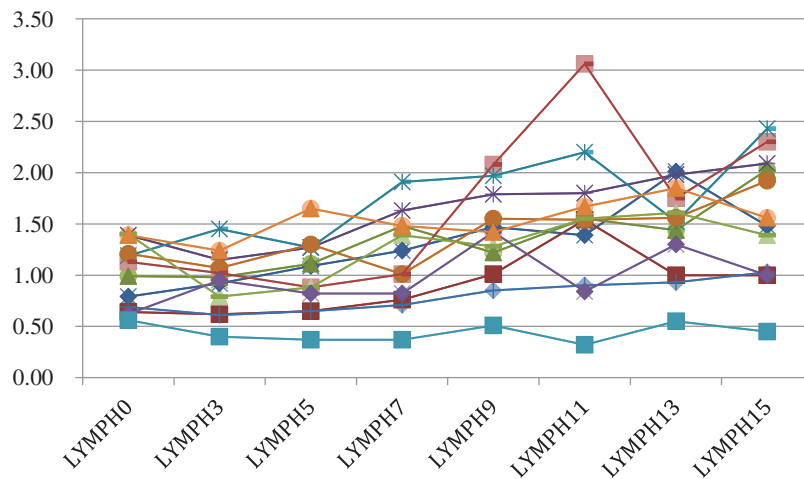


Figure 6: LYMPH values over 15 days for 12 patients (non-serious)

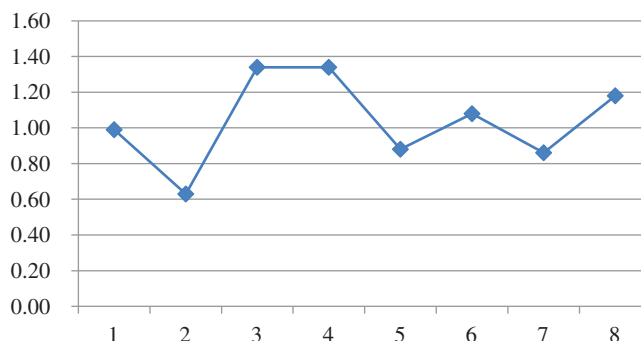


Figure 7: LYMPH values over 15 days for 1 patient (serious)

The number of lymphocytes in the first diagnosis is very significant, with a gradual decrease in its percentage indicating the severity of the pathological condition.

Based on [Tab. 4](#), the CQNN approach obtained better results for the training dataset than the logistic regression and NN approaches, and ID3. It also achieved an accuracy of 92.33% for the testing dataset, better than those of the other methods.

Table 4: Comparison of training and testing results obtained from CQNN and other approaches

Methods	Training Set		Testing Set	
	Accuracy	MSE	Accuracy	MSE
CQNN	0.97143	0.1174	0.92334	0.1374
Logistic regression	0.69228	0.4778	0.59228	0.4789
Neural network	0.76945	0.4279	0.66945	0.3688
ID3	0.84657	0.6983	0.64657	0.5880

As shown in [Tab. 5](#), the CQNN approach had (x) input nodes and its input samples could be defined as a vector.

Table 5: Input sample nodes and accuracy levels of CQNN

Input Nodes	Accuracy of CQNN	Input Nodes	Accuracy of CQNN
8	0.72445	18	0.88527
10	0.65723	20	0.72877
13	0.76945	24	0.91238
15	0.79836	30	0.92334

7 Conclusion

In this paper, the CQNN method for predicting the severity of COVID-19 is presented. It identifies important features, including the serial blood counts performed during patients' stays in hospital, that is, their numbers of lymphocytes on days 1 to 15 after admission which are correlated with the rate of relapse. This proposed non-linear approach was used to assess the extent of the COVID-19 infection in

two groups of patients (serious and non-serious), with the results obtained showing that it achieved 92.33% accuracy at the prediction stage. In future, we intend to investigate using the proposed method on large datasets to test its effectiveness and then automatically upload the results to a cloud-based epidemiological and early-warning monitoring platform which can process them and upload them to cloud platforms for disease surveillance and medical monitoring in institutions at all levels of government. This will provide early warnings, situation analyses and support for decision-making regarding the state of the pandemic.

Funding Statement: The authors received no specific funding for this study.

Conflicts of Interest: The authors declare that they have no conflicts of interest to report regarding the present study.

References

- [1] World Health Organization, "N. Coronavirus, Situation report, 22," 2019. [online]. Available: <https://www.who.int/docs/default-source/coronaviruse/situation-reports/20200211-sitrep-22-ncov.pdf>.
- [2] L. L. Ren, Y. M. Wang, Z. Q. Wu, Z. C. Xiang, L. Guo *et al.*, "Identification of a novel coronavirus causing severe pneumonia in human: A descriptive study," *Chinese Medical Journal*, vol. 133, no. 9, pp. 1015–1024, 2020.
- [3] C. Lin, W. E. Braund, J. Auerbach, J. H. Chou and J. H. Teng, "Policy decisions and use of information technology to fight coronavirus disease," *Taiwan Emerging Infectious Diseases*, vol. 26, no. 7, pp. 1506, 2020.
- [4] C. Huang, Y. Wang, X. Li, L. Ren, J. Zhao *et al.*, "Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China," *Lancet*, vol. 395, no. 10223, pp. 497–506, 2020.
- [5] R. Lu, X. Zhao, J. Li, P. Niu and B. Yang, "Genomic characterisation and epidemiology of 2019 novel coronavirus: Implications for virus origins and receptor binding," *Lancet*, vol. 395, no. 10224, pp. 565–574, 2020.
- [6] N. Zhu, D. Zhang, W. Wang, X. Li, B. Yang *et al.*, "A novel coronavirus from patients with pneumonia in China," *2019 New England Journal of Medicine*, vol. 382, no. 8, pp. 727–733, 2020.
- [7] J. F. W. Chan, S. Yuan, K. H. Kok, K. K. W. To and H. Chu, "A familial cluster of pneumonia associated with the 2019 novel coronavirus indicating person-to-person transmission: A study of a family cluster," *Lancet*, vol. 395, no. 10223, pp. 514–523, 2020.
- [8] S. Jeswal and S. Chakraverty, "Recent developments and applications in quantum neural network: A review," *Archives of Computational Methods in Engineering*, vol. 26, no. 4, pp. 793–807, 2019.
- [9] K. Beer, D. Bondarenko, T. Farrelly, T. J. Osborne, R. Salzman *et al.*, "Training deep quantum neural networks," *Nature Communications*, vol. 11, no. 1, pp. 1–6, 2020.
- [10] Z. Gao, C. Ma, Y. Luo and Z. Liu, "IMA health state evaluation using deep feature learning with quantum neural network," *Engineering Applications of Artificial Intelligence*, vol. 76, pp. 119–129, 2018.
- [11] X. Zhang, Y. Du, L. Shi, T. Chen, Y. Zhao *et al.*, "Dynamic change in lymphocyte count in the early stage is a potential predictor of the severity of COVID-19," Research Square AJE LLC, 2020.
- [12] O. Emam, M. Farhan and A. Abohany, "Analysis of faults fixing system using rough sets: A case study in large-scale telecommunication company (Telecom Egypt)," in *13th Int. Computer Engineering Conf.*, IEEE, Egypt, pp. 399–404, 2017.
- [13] O. Emam, M. Farhan and A. Abohany, "Faults repairing analysis using rough sets after implementation of labor force redistribution algorithm: A case study in Telecom Egypt," *Information Sciences Letter*, vol. 6, no. 3, pp. 39–48, 2017.
- [14] R. Mardani, A. A. Vasmehjani, F. Zali, A. Gholami and S. D. M. Nasab, "Laboratory parameters in detection of COVID-19 patients with positive RT-PCR; a diagnostic accuracy study," *Archives of Academic Emergency Medicine*, vol. 8, no. 1, pp. e43, 2020.
- [15] L. Tan, Q. Wang, D. Zhang, J. Ding and Q. Huang, "Lymphopenia predicts disease severity of COVID-19: A descriptive and predictive study," *Signal Transduction and Targeted Therapy*, vol. 5, no. 1, pp. 1–3, 2020.
- [16] G. Lippi and M. Plebani, "Laboratory abnormalities in patients with COVID-2019 infection," *Clinical Chemistry and Laboratory Medicine*, vol. 58, no. 7, pp. 1131–1134, 2020.