


Diagnosis of coronavirus disease 2019 and the potential role of deep learning: insights from the experience of Cairo University Hospitals

Marwa M. Ahmed¹, Amal M. Sayed²,
Dina El Abd², Inas T. El Sayed¹,
Yasmine S. Elkholy³, Ahmed H. Fares⁴ and
Samar Fares¹ 

Abstract

Objectives: Early detection of coronavirus disease 2019 (COVID-19) is crucial for patients and public health to ensure pandemic control. We aimed to correlate clinical and laboratory data of patients with COVID-19 and their polymerase chain reaction (PCR) results and to assess the accuracy of a deep learning model in diagnosing COVID-19.

Methods: This was a retrospective study using an anonymized dataset of patients with suspected COVID-19. Only patients with a complete dataset were included (n = 440). A deep analytics framework and dual-modal approach for PCR-based classification was used, integrating symptoms and laboratory-based modalities.

Results: Participants with loss of smell or taste were two times more likely to have positive PCR results (odds ratio [OR] 1.86). Participants with neutropenia, high serum ferritin, or monocytosis were three, four, and five times more likely to have positive PCR results (OR 2.69, 4.18, 5.42, respectively). The rate of accuracy achieved using the deep learning framework was 78%, with sensitivity of 83.9% and specificity of 71.4%.

¹Family Medicine Department, Kasralainy Faculty of Medicine, Cairo University, Cairo, Egypt

²Department of Clinical & Chemical Pathology, Kasralainy, Faculty of Medicine, Cairo University, Cairo, Egypt

³Department of Medical Microbiology & Immunology, Kasralainy Faculty of Medicine, Cairo University, Cairo, Egypt

⁴Computer Science and Engineering Department, Faculty of Engineering, Benha University, Cairo, Egypt

Corresponding author:

Samar Fares, Department of Family Medicine, Cairo University, 12 Shiem El Shafy, Kasr Al Ainy St., Cairo 11562, Egypt.

Email: Samar.fares@kasralainy.edu.eg



Conclusion: Loss of smell or taste, neutropenia, monocytosis, and high serum ferritin should be routinely assessed with suspected COVID-19 infection. The use of deep learning for diagnosis is a promising tool that can be implemented in the primary care setting.

Keywords

Primary care, deep learning, neural network, early detection, coronavirus disease 2019, severe acute respiratory syndrome coronavirus 2

Date received: 4 February 2022; accepted: 7 June 2022

Introduction

Severe acute respiratory syndrome (SARS) and Middle East respiratory syndrome (MERS) emerged in China and the Arabian Peninsula, respectively, with both viruses exhibiting community and hospital-acquired transmission.¹ Mortality rates reached approximately 10% for SARS and 35.6% for MERS.^{1–3} Similar dynamics have been demonstrated with the novel severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), which causes coronavirus disease 2019 (COVID-19) and has led to the ongoing global pandemic. Mortality rates of COVID-19 infection have reached 2%,^{4,5} lower than those of SARS and MERS. Yet, the number of visits to health care facilities, including primary, secondary, and tertiary health care services is higher for infection with SARS-CoV-2, which negatively affects the health systems and economies of countries worldwide.⁶

Initial symptoms of COVID-19 are non-specific, and disease progression and outcomes vary from patient to patient; therefore, triage services are key in providing cost-effective health services.⁷ Reverse transcriptase polymerase chain reaction (RT-PCR) is the cornerstone of COVID-19 diagnosis. Unfortunately, several limitations have been described for this

technique, including cost, turnaround time of the test, and false negative results. PCR is also unavailable in primary care, which necessitates patient referral to a secondary or tertiary care facility.⁸ Early detection of patients with COVID-19 is critical because an initial false negative PCR result could lead to a delay in treatment and an increased risk of COVID-19 transmission. The cost of PCR places a large burden on health care facilities, especially in developing countries with limited resources.⁹ Reliance on chest computed tomography (CT) alone may have limited negative predictive value, with some patients having normal radiological findings in early stages of the disease.¹⁰

Because the use of PCR and CT for COVID-19 diagnosis currently requires a long time and sufficient available resources, a fast and efficient diagnostic tool is urgently needed to improve the early detection of SARS-CoV-2 infection and decrease unnecessary patient referrals. Artificial intelligence could be beneficial in early detection of COVID-19 infection, especially in primary care settings.¹¹ Advanced machine learning methods can be quickly generalized to identify patients with COVID-19 who have minor symptoms and signs. However, such strategies must be implemented promptly for valid and efficient outcomes. In addition to cost-effectiveness, this modeling

approach could help in the detection and management of COVID-19 in communities under quarantine, which would decrease the burden on local health care services in countries worldwide.¹²

In this study, we aimed to correlate clinical and laboratory data of patients with suspected and confirmed COVID-19 to their PCR results during the first epidemic wave and to assess the accuracy of PCR-correlated symptoms and a laboratory-based deep learning model in diagnosing COVID-19. We hypothesized that developing a prediction model that depends on patients' clinical symptoms and laboratory findings can facilitate the diagnosis of COVID-19 in primary care settings, especially in resource-limited countries. Our proposed method offers an inexpensive and rapid tool that can be applied to the entire population with no physical contact, thereby limiting the spread of infection.

Methods

In this retrospective study, we used an anonymized dataset from patients with

suspected COVID-19 infection during the first wave of the pandemic from 1 June through 20 July 2020. The reporting of this study conforms to the TRIPOD (Transparent Reporting of a multivariable prediction model for Individual Prognosis Or Diagnosis) guidelines.¹³ The participants in our study were employees of Cairo University and Cairo University Hospitals and their first-degree relatives.

During the first epidemic wave at Cairo University Hospitals, the diagnosis of COVID-19 comprised several phases, depicted in Figure 1. The first phase included detection of patients with suspected infection using a phone triaging system, with referral of the patient for confirmatory testing when necessary. Skilled family physicians conducted this phase. The second phase included confirmation using RT-PCR assay of a nasal or pharyngeal swab sample and laboratory testing. Confirmed cases were defined as patients with a positive result of RT-PCR. The third phase included assessing disease severity in each patient and providing appropriate management. This included home

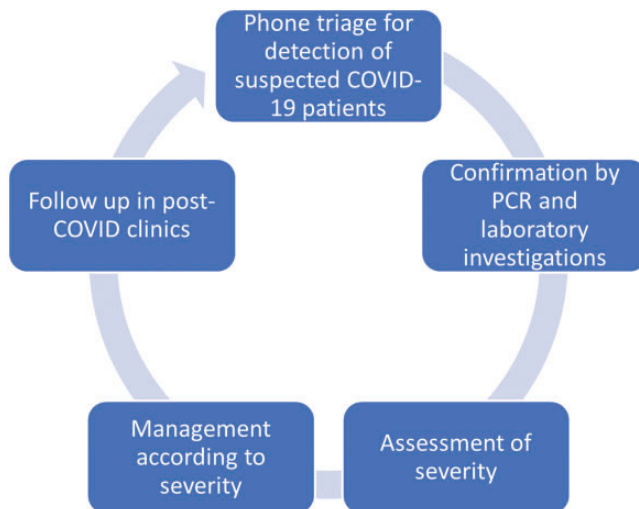


Figure 1. Cycle of diagnosis and management of COVID-19-positive patients at Cairo University Hospitals. COVID-19, coronavirus disease 2019; PCR, polymerase chain reaction.

isolation and regular follow-up for mild cases via telephone and hospital admission for patients with moderate-to-severe illness. The fourth phase involved patient assessment during the post-COVID-19 phase and ordering a follow-up PCR test in the post-COVID-19 clinic. Data from the first and second phases were included in the present study.

Data extraction

We retrieved datasets from the departments of family medicine and clinical pathology as well as the infection control unit at Cairo University Hospitals. A team of experienced family physicians reviewed and cross-checked the data. Only patients with a complete dataset, including medical history and laboratory and PCR results, were included in the study. The medical history entailed exposure history, COVID-19 symptoms, and risk factors. Laboratory investigations comprised complete blood picture, liver and kidney function, C-reactive protein (CRP), lactate dehydrogenase (LDH), total creatine kinase (CK), ferritin, and international normalized ratio. The researchers compared symptoms and laboratory findings between patients with suspected and confirmed COVID-19 infection.

With the extensive advancement and application of deep learning, increasingly more models that use deep learning are proposed to model the prognosis of patients with COVID-19 infection using clinical¹⁴ and laboratory data,¹⁵ imaging data,^{16,17} or a combination of both. We applied a deep analytics framework and dual-modal approach for PCR-based classification by integrating symptoms with laboratory-based modalities. More precisely, our proposed deep learning framework comprised three phases: a dual-modal phase, feature extraction phase, and classification phase. Consequently, the dual-modal phase is aimed at integrating both symptoms and

laboratory-based data into a common feature space. The feature extraction phase involves encoding the preprocessed dual-modal common feature space into PCR-dependent features. The classification phase is aimed at classifying the encoded PCR-dependent features.

A structural illustration of the proposed deep analytics framework and dual-modal approach, including the parameters settings, is shown in Figure 2. The dual-modal phase takes the raw data of symptoms and laboratory-based data as inputs and encodes these using dual-modal layers. Each consists of five operations: Dence, Batch Normalization, LeakyReLU, Dropout, and Flatten operations. These layers extract a common feature space and feed this to the feature extraction phase, which uses a sequence of two layers, each consisting of four operations: Dence, Batch Normalization, LeakyReLU, and Dropout. This sequence of layers is used to extract class-dependent features; these are fed to the classification stages, which use the SoftMax activation function. The SoftMax function converts the extracted feature vector to a vector with two categorical probabilities, namely, positive PCR and negative PCR.

Ethical approval

This study was approved by the ethical committee of the Faculty of Medicine, Cairo University (ID: N-17-2021). Informed patient consent was waived because of the retrospective nature of the study.

Statistical analysis

Summary statistics are presented as number and percent or median and interquartile range for categorical and numeric variables, respectively. Associations of PCR results with categorical variables were assessed using the chi-square test and using the

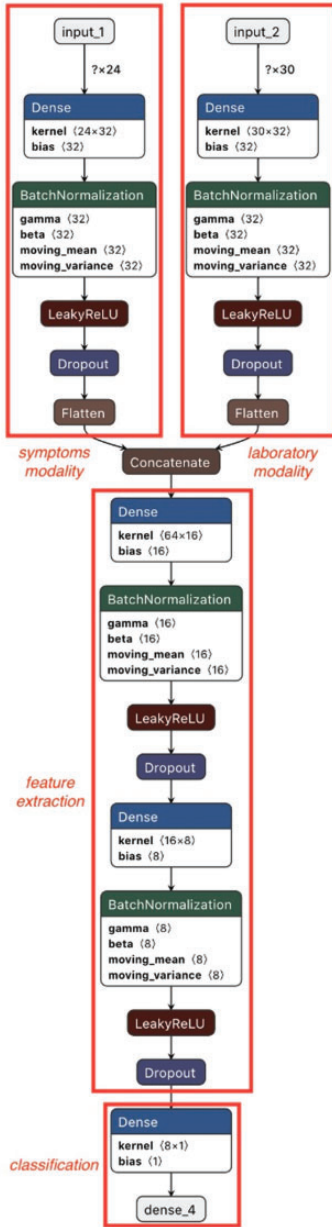


Figure 2. Dual-modal and analytics learning framework.

Mann–Whitney (Wilcoxon rank-sum) test in the case of continuous variables. Univariate logistic regression was performed for all significant variables.

We controlled for potential confounders (age, sex, presence of comorbidities, LDH, and all variables in univariate regression) using multivariate logistic regression. Post-estimation tests were conducted to examine the accuracy of our regression model, with area under the receiver operating characteristic curve (AUC) equal to 0.792. In addition, to assess whether the model fit our data, we used the Hosmer–Lemeshow goodness-of-fit test with a p value equal to 0.389. We tested for multicollinearity using the variance inflation factor (VIF) with a mean VIF of 1.68. Accuracy testing for the deep learning framework was performed and we determined the sensitivity, specificity, AUC, and positive and negative predictive values.

Results

As presented in Table 1, we analyzed the data of 440 patients, with most comprising female patients (56.36%) and individuals younger than 45 years old (63.08%). The most prevalent factor was contact with a patient positive for COVID-19 infection who had respiratory symptoms (79.73%). The most common symptoms among study participants were fever, cough, sore throat, and body aches. Loss of smell or taste sensations were found in 30.91% of participants. The presence of comorbidities such as diabetes, hypertension, cardiovascular, pulmonary conditions, or chronic kidney diseases was identified in 34.77% of patients.

Associations of COVID-19 PCR results with patients’ demographic characteristics, symptoms, and risk factors using the chi-square test are presented in Table 1. Contact with a case that had respiratory symptoms, spending time in a location where diagnosed COVID-19 cases were present, presence of sore throat, loss of smell or taste, and smoking were factors showing a significant association with PCR results

Table 1. Associations of demographic data, symptoms, and risk factors with COVID-19 PCR results using Pearson chi-square test (n = 440).

	Total n (%)	PCR		χ^2	p value*
		Negative n (%)	Positive n (%)		
• Demographic factors					
Age group	409 (100)				
16–45 years	258 (63.08)	134 (51.94)	124 (48.06)	0.327	0.567
46–85 years	151 (36.92)	74 (49.01)	77 (50.99)		
Sex	440 (100)				
Female	248 (56.36)	131 (52.82)	117 (47.18)	0.345	0.557
Male	192 (43.64)	96 (50.00)	97 (50.00)		
• Factors and symptoms related to COVID-19					
Contacted a case with respiratory symptoms	439 (100)				
No	89 (20.27)	36 (40.45)	53 (59.55)	5.438	0.020
Yes	350 (79.73)	190 (54.29)	160 (45.71)		
Spending time in a location with COVID-19 cases present	435 (100)				
No	213 (48.97)	98 (46.01)	115 (53.99)	5.459	0.019
Yes	222 (51.03)	127 (57.21)	95 (42.79)		
Working in health care or isolation facility	439 (100)				
No	214 (48.75)	104 (48.60)	110 (51.40)	1.618	0.203
Yes	225 (51.25)	123 (54.67)	102 (45.33)		
Fever >38°C	438 (100)				
No	105 (23.97)	57 (54.29)	48 (45.71)	0.470	0.493
Yes	333 (76.03)	168 (50.45)	165 (49.55)		
Severe or persistent cough	438 (100)				
No	151 (34.47)	77 (50.99)	74 (49.01)	0.034	0.854
Yes	287 (65.53)	149 (51.92)	138 (48.08)		
Sore throat	438 (100)				
No	166 (37.90)	73 (43.98)	93 (56.02)	6.218	0.013
Yes	272 (62.10)	153 (56.25)	119 (43.75)		
Vomiting or diarrhea	437 (100)				
No	249 (56.98)	123 (49.40)	126 (50.60)	1.505	0.220
Yes	188 (43.02)	104 (55.32)	84 (44.68)		
Myalgia or body aches	438 (100)				
No	70 (15.98)	41 (58.57)	29 (41.43)	1.518	0.218
Yes	368 (84.02)	186 (50.54)	182 (49.46)		
Loss of smell or taste	440 (100)				
No	304 (69.09)	168 (55.26)	136 (44.74)	5.311	0.021
Yes	136 (30.91)	59 (43.38)	77 (56.62)		
• Risk factors and comorbidities					
Smoking	418 (100)				
No	366 (87.56)	179 (48.91)	187 (51.09)	6.170	0.013
Yes	52 (12.44)	35 (67.31)	17 (32.69)		

(continued)

Table 1. Continued.

	Total n (%)	PCR		χ^2	p value*
		Negative n (%)	Positive n (%)		
Pregnancy	243 (100)				
No	238 (97.94)	124 (52.10)	114 (47.90)	0.122	0.726
Yes	5 (2.06)	3 (60.00)	2 (40.00)		
Comorbidities	440 (100)				
No	287 (65.23)	154 (53.66)	133 (46.34)	1.413	0.235
Yes	153 (34.77)	73 (47.71)	80 (52.29)		
Immunodeficient conditions or drugs	434 (100)				
No	418 (96.31)	217 (51.91)	201 (48.09)	0.411	0.521
Yes	16 (3.69)	7 (43.75)	9 (56.25)		

*Significant at $p < 0.05$.

PCR, polymerase chain reaction; COVID-19, coronavirus disease 2019.

($p = 0.020$, 0.019 , 0.013 , 0.021 , and 0.013 , respectively).

Among all participant results for complete blood count (CBC), several factors showed significant median differences between COVID-19 PCR-positive and PCR-negative groups, as shown in Table 2. Factors with a significant difference between groups were mean corpuscular volume, platelet count, white blood cell (WBC) count, neutrophil count, lymphocyte count, monocyte count, eosinophil count, and basophil count ($p = 0.020$, <0.001 , <0.001 , $=0.002$, <0.001 , $=0.005$, <0.001 , and <0.001 , respectively). Median levels of these factors were significantly lower in the PCR-positive group.

Table 3 shows the median levels for participants' blood chemistry results and differences between the two groups. Median levels of LDH, CK, and serum ferritin were significantly different between PCR-positive and PCR-negative groups ($p < 0.001$, $=0.033$, and $=0.004$, respectively). LDH and serum ferritin levels were significantly higher in the PCR-positive group.

Table 4 shows bivariate associations of COVID-19 PCR results with the results of CBC and blood chemistry after categorization according to low, normal, or high levels. The categorization was according to reference values of Kasralainy Faculty of Medicine's central lab, where the samples were analyzed. WBCs, neutrophils, lymphocytes, alanine aminotransferase (ALT), and serum ferritin categories showed a significant association with final PCR results ($p < 0.001$, <0.001 , <0.001 , $=0.039$, and <0.001 , respectively).

In Table 5, we present findings of the regression model for predictors of COVID-19 PCR results. We controlled for potential confounders (age, sex, presence of comorbidities, LDH, and all variables in univariate analysis) in the multivariate logistic regression. Participants with a loss of smell or taste were two times more likely to have a positive PCR result than participants who did not lose their sense of taste or smell (odds ratio [OR] 1.86; 95% confidence interval [CI] 1.04–3.35, $p = 0.038$). Participants with neutropenia or monocytosis were three and five times more likely,

Table 2. Associations of complete blood picture results with COVID-19 PCR results using Mann–Whitney test (n = 440).

	n	Median	IQR	z	p value*
Red blood cell count ($10^6/\mu\text{L}$)					
Negative PCR	223	5.1	4.9–5.5	−0.360	0.719
Positive PCR	203	5.2	4.7–5.6		
Hemoglobin (g/dL)					
Negative PCR	223	13.9	12.7–15.0	0.418	0.676
Positive PCR	203	13.7	12.4–14.9		
Hematocrit (%)					
Negative PCR	223	41.7	38.4–44.8	1.276	0.202
Positive PCR	203	40.7	37.6–44.2		
MCV (fL)					
Negative PCR	223	81.1	76.4–84.9	2.330	0.020
Positive PCR	203	79.8	75.2–83.1		
MCH (pg)					
Negative PCR	223	27.2	24.8–28.9	1.103	0.270
Positive PCR	203	27.0	24.8–28.4		
MCHC (g/dL)					
Negative PCR	223	33.3	32.3–34.4	−1.770	0.077
Positive PCR	203	33.6	32.4–34.4		
RDW (%)					
Negative PCR	223	13.9	13.0–15.4	1.330	0.184
Positive PCR	203	13.7	12.8–15.1		
Platelet count ($10^3/\mu\text{L}$)					
Negative PCR	223	240	198–283	3.671	<0.001
Positive PCR	203	219	172–251		
MPV (fL)					
Negative PCR	202	9.5	8.3–10.6	−0.079	0.938
Positive PCR	187	9.7	8.3–10.6		
White blood cell count ($10^3/\mu\text{L}$)					
Negative PCR	223	7.4	6.0–9.2	4.252	<0.001
Positive PCR	203	6.3	4.9–8.21		
Neutrophil count ($10^3/\mu\text{L}$)					
Negative PCR	217	3.79	2.62–5.21	3.041	0.002
Positive PCR	199	3.2	1.96–4.63		
Lymphocyte count ($10^3/\mu\text{L}$)					
Negative PCR	222	2.87	2.21–3.45	4.823	<0.001
Positive PCR	202	2.37	1.89–2.97		
Monocyte count ($10^3/\mu\text{L}$)					
Negative PCR	222	0.57	0.42–0.67	2.785	0.005
Positive PCR	202	0.48	0.37–0.65		
Eosinophil count ($10^3/\mu\text{L}$)					
Negative PCR	222	0.13	0.07–0.22	4.761	<0.001
Positive PCR	202	0.08	0.04–0.17		
Basophil count ($10^3/\mu\text{L}$)					
Negative PCR	217	0.09	0.05–0.12	4.753	<0.001
Positive PCR	199	0.06	0.04–0.09		

(continued)

Table 2. Continued.

	n	Median	IQR	z	p value*
Neutrophil to lymphocyte ratio					
Negative PCR	217	1.4	1.0–1.8	0.454	0.650
Positive PCR	199	1.3	0.8–2.1		

*Significant with $p < 0.05$.

COVID-19, coronavirus disease 2019; PCR, polymerase chain reaction; IQR, interquartile range; MCV, mean corpuscular volume; MCH, mean corpuscular hemoglobin; MCHC, mean corpuscular hemoglobin concentration; RDW, red blood cell distribution width; MPV, mean platelet volume.

Table 3. Associations of blood chemistry with COVID-19 PCR results using Mann–Whitney test ($n = 440$).

	n	Median	IQR	z	p value
AST (U/L)					
Negative PCR	226	25	20–32	–1.311	0.190
Positive PCR	213	26	21–35		
ALT (U/L)					
Negative PCR	226	27	17–42	–1.859	0.063
Positive PCR	213	33	19–45		
Blood urea (mg/dL)					
Negative PCR	226	25.5	20–31	0.085	0.932
Positive PCR	213	25	20–31		
Serum creatine (mg/dL)					
Negative PCR	226	0.87	0.76–1.01	–0.839	0.402
Positive PCR	213	0.87	0.76–1.01		
LDH (U/L)					
Negative PCR	226	280	238–333	–3.484	<0.001
Positive PCR	213	313	259–372		
CK total (U/L)					
Negative PCR	226	90	65–124	2.132	0.033
Positive PCR	213	80	55–117		
CRP (mg/L)					
Negative PCR	227	4–55	2.2–10.3	–0.060	0.952
Positive PCR	212	4–11	1.42–13–3		
Ferritin (ng/L)					
Negative PCR	181	75.3	31.3–140.2	–2.883	0.004
Positive PCR	192	103.4	34.9–224.3		
INR					
Negative PCR	220	1.01	0.96–1.04	0.556	0.579
Positive PCR	206	1.0	0.97–1.06		

*Significant with $p < 0.05$.

COVID-19, coronavirus disease 2019; PCR, polymerase chain reaction; IQR, interquartile range; AST, aspartate aminotransferase; ALT, alanine aminotransferase; LDH, lactate dehydrogenase; CK, creatine kinase; CRP, C-reactive protein; INR, international normalized ratio.

Table 4. Associations of complete blood picture and blood chemistry with COVID-19 PCR results using Pearson chi-square test (n = 440).

	Total n (%)	PCR		χ^2	p value*
		Negative n (%)	Positive n (%)		
• Complete blood picture					
Red blood cell count	426 (100)				
Erythrocytopenia	7 (1.64)	2 (28.57)	5 (71.43)	2.058	0.357
Normal	79 (18.54)	39 (49.37)	40 (50.63)		
Erythrocytosis	340 (79.81)	182 (53.53)	158 (46.47)		
Platelet count	426 (100)				
Thrombocytopenia	63 (14.79)	30 (47.62)	33 (52.38)	0.726	0.695
Normal	347 (81.46)	184 (53.03)	163 (46.97)		
Thrombocytosis	16 (3.76)	9 (56.25)	7 (43.75)		
White blood cell count	426 (100)				
Leucopenia	27 (6.34)	6 (22.22)	21 (77.78)	15.737	<0.001
Normal	337 (79.11)	175 (51.93)	162 (48.07)		
Leucocytosis	62 (14.55)	42 (67.74)	20 (32.26)		
Neutrophil count	426 (100)				
Neutropenia	58 (13.94)	17 (29.31)	41 (70.69)	15.282	<0.001
Normal	324 (77.88)	178 (54.94)	146 (45.06)		
Neutrophilia	34 (8.17)	22 (64.71)	12 (35.29)		
Lymphocyte count	426 (100)				
Lymphopenia	13 (3.07)	3 (23.08)	10 (76.92)	16.925	<0.001
Normal	270 (63.68)	127 (47.04)	143 (52.96)		
Lymphocytosis	141 (33.25)	92 (65.25)	49 (34.75)		
Monocyte count	424 (100)				
Monocytopenia	18 (4.25)	3 (16.67)	15 (83.33)	11.875	0.003
Normal	340 (80.19)	189 (55.59)	151 (44.41)		
Monocytosis	66 (15.57)	30 (45.45)	36 (54.55)		
Eosinophil count	424 (100)				
Eosinopenia	54 (12.74)	16 (29.63)	38 (70.37)	15.253	<0.001
Normal	360 (84.91)	198 (55.00)	162 (45.00)		
Eosinophilia	10 (2.36)	8 (80.00)	2 (20.00)		
Neutrophil to lymphocyte ratio	426 (100)				
<3.1	392 (94.23)	209 (53.32)	183 (46.68)	3.619	0.057
≥3.1	24 (5.77)	8 (33.33)	16 (66.67)		
• Blood chemistry					
AST	439 (100)				
Normal	359 (81.78)	191 (53.20)	168 (46.80)	2.341	0.126
High	80 (18.22)	35 (43.75)	45 (56.25)		
ALT	439 (100)				
Normal	271 (61.73)	150 (55.35)	121 (44.65)	4.246	0.039
High	168 (38.27)	76 (45.24)	92 (54.76)		
Blood urea	439 (100)				
Low	38 (8.66)	22 (57.89)	16 (42.11)	0.843	0.656
Normal	167 (38.04)	83 (49.70)	84 (50.30)		
High	234 (53.30)	121 (51.71)	113 (48.29)		

(continued)

Table 4. Continued.

	Total n (%)	PCR		χ^2	p value*
		Negative n (%)	Positive n (%)		
Serum creatine	439 (100)				
Normal	423 (96.36)	221 (97.79)	202 (94.84)	2.721	0.099
High	16 (3.64)	5 (31.25)	11 (68.75)		
LDH	439 (100)				
Normal	47 (10.71)	28 (59.57)	19 (40.43)	2.832	0.243
High	392 (89.29)	198 (50.51)	194 (49.49)		
CK total	439 (100)				
Normal	383 (87.24)	196 (51.17)	187 (48.83)	0.112	0.738
High	56 (12.76)	30 (53.57)	26 (46.43)		
CRP	439 (100)				
Normal	264 (60.14)	135 (51.14)	129 (48.86)	0.087	0.768
High	175 (39.86)	92 (52.57)	83 (47.43)		
Ferritin	439 (100)				
Low	27 (7.24)	15 (55.56)	12 (44.44)	16.973	<0.001
Normal	232 (62.20)	129 (55.60)	103 (44.40)		
High	114 (30.56)	37 (32.46)	77 (67.54)		

*Significant with $p < 0.05$.

COVID-19, coronavirus disease 2019; PCR, polymerase chain reaction; AST, aspartate aminotransferase; ALT, alanine aminotransferase; LDH, lactate dehydrogenase; CK, creatine kinase; CRP, C-reactive protein.

respectively, to have positive PCR results compared with patients who had normal leucocyte and monocyte counts (OR 2.69, 5.4; 95% CI 1.06–6.83, 1.97–14.89; $p = 0.037$, 0.001, respectively). Participants with high serum ferritin levels were four times more likely to have positive PCR results (OR 4.18; 95% CI 1.27–13.78; $p = 0.019$).

As shown in Table 5, participants who contacted cases with respiratory symptoms were 66% less likely to have positive PCR results compared with those who did not contact such cases (OR 0.34; 95% CI 0.16–0.75; $p = 0.007$). Compared with participants without sore throat, those with sore throat were 47% less likely to have positive PCR results (OR 0.53; 95% CI 0.30–0.96; $p = 0.035$). Participants with leucocytosis, lymphocytosis, or eosinophilia were 77%, 53%, and 92% less likely to

have positive PCR results compared with participants who had normal levels (OR 0.23, 0.47, 0.08; 95% CI 0.05–0.97, 0.25–0.89, 0.01–0.92; $p = 0.046$, 0.021, 0.043, respectively).

Regarding the deep learning framework, along with the classification precision rate, we determined the chance-level and PCR precision rates. The chance-level precision rate for two-class classification was 50%, and the precision rate achieved using the proposed framework was 78%. The proposed framework achieved a sensitivity of 83.9%, specificity of 71.4%, AUC of 0.78, positive predictive value of 76.5%, and a negative predictive value of 80.0%, as illustrated in Figure 3.

Discussion

Given the complex interplay of factors that determine the outcome of COVID-19 and

Table 5. Results of univariate and multivariate logistic regression for prediction of COVID-19 PCR results.

	Univariate regression		Multivariate regression	
	OR (95% CI)	p value*	OR (95% CI)	p value*
Contacted a case with respiratory symptoms				
Yes	0.57 (0.36–0.91)	0.021	0.34 (0.16–0.75)	0.007
Spending time in a location with COVID-19 cases present				
Yes	0.64 (0.44–0.93)	0.020	0.99 (0.57–1.74)	0.269
Sore throat				
Yes	0.61 (0.41–0.90)	0.013	0.53 (0.30–0.96)	0.035
Loss of smell or taste				
Yes	1.61 (1.07–2.42)	0.022	1.86 (1.04–3.35)	0.038
Smoking				
Yes	0.46 (0.25–0.86)	0.015	0.48 (0.19–1.28)	0.145
White blood cells				
Leucopenia	3.78 (1.49–9.60)	0.005	1.01 (0.27–3.78)	0.988
Leucocytosis	0.51 (0.29–0.910)	0.023	0.23 (0.05–0.97)	0.046
Neutrophils				
Neutropenia	2.94 (1.60–5.39)	<0.001	2.69 (1.06–6.83)	0.037
Neutrophilia	0.66 (0.32–1.39)	0.278	1.33 (0.21–8.17)	0.760
Lymphocytes				
Lymphopenia	2.96 (0.80–10.99)	0.105	1.06 (0.21–5.28)	0.936
Lymphocytosis	0.47 (0.31–0.72)	<0.001	0.47 (0.25–0.89)	0.021
Monocytes				
Monocytopenia	6.26 (1.78–22.02)	0.004	4.63 (0.87–24.77)	0.073
Monocytosis	1.50 (0.88–2.55)	0.132	5.42 (1.97–14.89)	0.001
Eosinophils				
Eosinopenia	2.90 (1.56–5.39)	0.001	1.51 (0.58–3.91)	0.399
Eosinophilia	0.30 (0.06–1.46)	0.137	0.08 (0.01–0.92)	0.043
ALT				
High	1.50 (1.02–2.21)	0.040	0.98 (0.55–1.76)	0.962
Ferritin				
Normal	0.99 (0.45–2.22)	0.996	1.19 (0.43–3.32)	0.739
High	2.60 (1.11–6.11)	0.028	4.18 (1.27–13.78)	0.019

*Significant with $p < 0.05$.

COVID-19, coronavirus disease 2019; PCR, polymerase chain reaction; ALT, alanine aminotransferase; OR, odds ratio; CI, confidence interval.

the non-specificity of the initial symptoms, which are mainly fever, dry cough, and fatigue, a cost-effective and reliable diagnostic tool is crucial for early detection. Accordingly, in this study, we aimed to correlate the clinical and laboratory data of suspected and confirmed COVID-19 cases with PCR results, using available data and artificial intelligence, in the early detection of COVID-19. We aimed to find an easy,

rapid, and effective method for definitive diagnosis of COVID-19 or at minimum, rapid identification of patients highly suspicious for COVID-19 infection, to enhance outcomes and prevent infections at our hospital.

In our study, contact with cases that had respiratory symptoms and spending time in a location with confirmed COVID-19 cases present showed a significant association

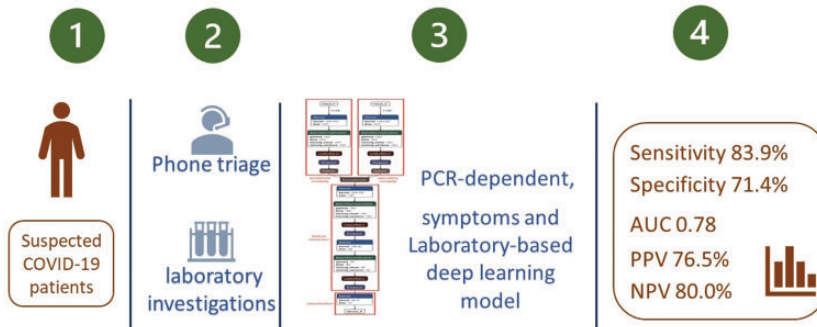


Figure 3. Summary of model performance.

COVID-19, coronavirus disease 2019; PCR, polymerase chain reaction; AUC, area under the receiver operating characteristic curve; PPV, positive predictive value; NPV, negative predictive value.

with the results of COVID-19 RT-PCR. Surprisingly, however, contact with cases that had respiratory symptoms and spending time in a location where COVID-19 cases were present had an inverse impact on rates of infection. Our study participants were 60% less likely to have positive PCR results than those without these risk factors (OR 0.40, 95% CI 0.19–0.85, 0.018). In Qatar, a survey among 393 health care workers found that 5% of the study population reported acquiring infection at a COVID-19-designated facility and 95% at a non-COVID-19 facility, where incidental exposure to an asymptomatic colleague or patient was the main source of infection. In the same study, adherence to the use of full personal protective equipment was 82% at COVID-19-designated facilities and 68% at non-COVID-19 facilities.¹⁸ Cairo University Hospitals comprises one of the largest hospital facilities in Egypt, and participants in our study were mainly health care workers at this institution and their first-degree relatives. The situation described in this study and the study in Qatar highlight the importance of adherence to infection control measures under all conditions throughout the COVID-19 pandemic.

It is now well accepted that different strains of SARS-CoV-2 circulate in different countries, and this has had a large impact on prevalent symptoms, outcomes, and infectivity among patients. The clinical presentation of COVID-19 ranges from an asymptomatic state to respiratory failure and multi-organ dysfunction. Common clinical features include fever, cough, sore throat, headache, fatigue, myalgia, dyspnea, and conjunctivitis.¹⁹ Published viral sequencing results have demonstrated that SARS-CoV-2 strains share a similar gene sequence, with a few changes that seem to be directed toward the evolution of mutants that cause milder symptoms.²⁰ The most prevalent symptoms among our study population were sore throat and loss of smell or taste. Other studies have reported different symptoms among patients. A recent systematic review of 54 studies that included hundreds of patients reported that the most common symptoms found were fever in 81.2% of cases, cough in 58.5%, and fatigue in 38.5%.²¹

To our knowledge, there is an extreme paucity of existing data regarding the prevalent strains of SARS-CoV-2 in Egypt. However, the relatively low number of cases and deaths in the country suggest

the circulation of less virulent strains.²² This assumption is supported by the findings of our study, where the symptoms prevalent among our patients differed from those reported internationally.

The gold standard for the diagnosis of SARS-CoV-2 infection is RT-PCR. However, many factors can interfere with the sensitivity of this test. Thus, a negative test result in a patient with a typical clinical picture should not be the basis for excluding the possibility of COVID-19 infection. In such cases, routine laboratory testing can help to guide clinicians and infection prevention specialists in making sound decisions regarding patient management. In our population, symptomatic patients presenting with leukopenia, neutropenia, and lymphocytopenia were more likely to have a positive PCR result for SARS-CoV-2. A recently published non-systematic review reported that lymphopenia and eosinopenia have been linked to disease severity and worse prognosis.²³ Similarly, other studies have reported leukopenia and lymphopenia among the notable laboratory findings seen in patients with COVID-19.^{5,24}

As for blood chemistry, our results showed that elevated LDH and serum ferritin levels were statistically significant in PCR-positive patients (p values <0.001 and 0.004 , respectively). A recently published systematic review and meta-analysis that included a total of 4663 patients reported that increased LDH was found among 46.2% of patients.²⁵ According to the Centers for Disease Control and Prevention, the most prominent laboratory findings among patients with COVID-19 are elevated CRP, ALT, aspartate aminotransferase, LDH, and ferritin levels.²⁶ In our study, although not all of the aforementioned laboratory findings were significantly different between patients with and without COVID-19, except for CRP, all were elevated among patients with COVID-19 infection.

The proposed model provides a potentially useful COVID-19 diagnostic tool that is based on symptoms and laboratory findings. The precision rate achieved by the proposed framework was 78%, which considerably outperforms the chance-level precision rate and is comparable to the PCR precision rate, which is approximately 73.3%.²⁷

The proposed framework achieved a sensitivity of 83.9%, specificity of 71.4%, AUC of 0.78, positive predictive value of 76.5%, and negative predictive value of 80.0%. A systemic review in 2020 assessing the role of deep learning in detecting and diagnosing COVID-19 using CT and chest X-ray concluded that the specificity for COVID-19 diagnosis using CT-based deep learning was more than 92%. In many studies, the reported sensitivity was also higher than or equal to that of traditional diagnostic modalities. Whereas the sensitivity and specificity of CT-based deep learning are higher than those of our proposed model, the use of CT is not feasible in the primary care setting.²⁸

Another study conducted in 2020 developed a machine learning model to detect COVID-19 cases among hospitalized patients according to symptoms and laboratory findings.¹⁵ The model included age and sex; the presence of any COVID-19 symptoms including dyspnea, sore throat, and cough; and laboratory findings. The accuracy of that model ranged from 83% to 88%, and the sensitivity and specificity ranged from 76% to 89% and 84% to 91%, respectively. The AUC ranged from 0.83 to 0.90. In another study conducted in 2020, two machine learning models based on laboratory findings were developed; the sensitivity of the two models ranged from 92% to 95%, and the accuracy ranged from 82% and 86%.²⁹ Our proposed model differed from these models in that it integrates clinical data, risk factors, and routine laboratory test results.

RT-PCR is considered the gold standard for the diagnosis of COVID-19. Although chest CT has high sensitivity, it is not feasible to be used as a screening tool. The hematochemical values of routine laboratory tests represent a cheaper and faster alternative. Several machine learning models have been proposed for COVID-19 diagnosis using radiological findings, but few are based on symptoms and hematochemical values.

SARS-CoV-2 has posed an unprecedented threat to health care and infection control systems all over the world. Widespread public panic has resulted in massive pressure on the health care sector, as has the sharp decline in the global strategic stocks of personal protective equipment and the spread of infection among health care workers. Together, these factors have made it crucial to find rapid, effective alternatives to deal with this public health crisis. Deep learning models used in disease prediction are among the novel tools that can meet this need. Our proposed model is based on patients' clinical data, risk factors, and the results of routine laboratory tests. The use of PCR-correlated symptoms and a deep learning model based on laboratory findings could provide an excellent solution for the detection of patients with COVID-19 in primary care settings, especially in resource-limited countries, offering an inexpensive and fast tool that can be applied in the whole population.

Limitations of this study

In this study, we only used the data of patients with a complete dataset, including medical history and laboratory and PCR results. We excluded more than 500 patients with incomplete data. Use of a large dataset can help to increase accuracy of the model by increasing its machine-learning capability.

Conclusion

Using the proposed deep learning approach, we found that participants with a loss of smell or taste sensations, neutropenia, monocytosis, or high serum ferritin were more likely to have positive COVID-19 PCR results. Routine assessment of these four criteria should be done for each patient with suspected COVID-19. Because RT-PCR and chest CT currently require a long time and the availability of sufficient resources, use of a deep learning model is a promising tool with an encouraging precision rate for the diagnosis of COVID-19. Our proposed framework can be implemented to prioritize the referral of patients with suspected COVID-19 for PCR testing, especially in developing countries with limited resources.

Declaration of conflicting interest

The authors declare that there is no conflict of interest.

Data availability

For the convenience of research dissemination, we have made the source codes publicly available for downloading at <https://github.com/ahfares/dual-modal-and-analytics-learning-framework>.

Funding

This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.

ORCID iD

Samar Fares  <https://orcid.org/0000-0002-3438-1329>

References

1. Yin Y and Wunderink RG. MERS, SARS and other coronaviruses as causes of pneumonia. *Respirology* 2018; 23: 130–137.
2. Ksiazek TG, Erdman D, Goldsmith CS, et al. A novel coronavirus associated with

- severe acute respiratory syndrome. *N Engl J Med* 2003; 348: 1953–1966.
3. Luk HK, Li X, Fung J, et al. Molecular epidemiology, evolution and phylogeny of SARS coronavirus. *Infect Genet Evol* 2019; 71: 21–30.
 4. Chen N, Zhou M, Dong X, et al. Epidemiological and clinical characteristics of 99 cases of 2019 novel coronavirus pneumonia in Wuhan, China: a descriptive study. *Lancet* 2020; 395: 507–513.
 5. Huang C, Wang Y, Li X, et al. Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China. *Lancet* 2020; 395: 497–506.
 6. Khalid A, Ali S. COVID-19 and its Challenges for the Healthcare System in Pakistan. *Asian Bioethics Review*. 2020; 12: 551–564.
 7. Ge Y, McKay BK, Sun S, et al. Assessing the impact of a symptom-based mass screening and testing intervention during a novel infectious disease outbreak: the case of COVID-19. *MedRxiv* 2020. Available from: <http://dx.doi.org/10.1101/2020.02.20.20025973>.
 8. Guo L, Ren L, Yang S, et al. Profiling early humoral response to diagnose novel coronavirus disease (COVID-19). *Clin Infect Dis* 2020; 71: 778–785.
 9. Lippi G, Simundic AM and Plebani M. Potential preanalytical and analytical vulnerabilities in the laboratory diagnosis of coronavirus disease 2019 (COVID-19). *Clin Chem Lab Med* 2020; 58: 1070–1076.
 10. Ai T, Yang Z, Hou H, et al. Correlation of chest CT and RT-PCR testing for coronavirus disease 2019 (COVID-19) in China: a report of 1014 cases. *Radiology* 2020; 296: E32–E40.
 11. Mei X, Lee HC, Diao KY, et al. Artificial intelligence-enabled rapid diagnosis of patients with COVID-19. *Nat Med* 2020; 26: 1224–1228.
 12. Tekkeşin A. Artificial intelligence in healthcare: past, present and future. *Anatol J Cardiol* 2019; 22: 8–9.
 13. Collins GS, Reitsma JB, Altman DG, et al. Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD): The TRIPOD statement. *Br J Surg* 2015; 102: 148–158. Available from: <https://www.equator-network.org/reporting-guidelines/tripod-statement/>.
 14. Struyf T, Deeks JJ, Dinnes J, et al. Signs and symptoms to determine if a patient presenting in primary care or hospital outpatient settings has COVID-19. *Cochrane Database Syst Rev* 2020; 7: CD013665.
 15. Cabitza F, Campagner A, Ferrari D, et al. Development, evaluation, and validation of machine learning models for COVID-19 detection based on routine blood tests. *Clin Chem Lab Med* 2021; 59: 421–431.
 16. Kulkarni AR, Athavale AM, Sahni A, et al. Deep learning model to predict the need for mechanical ventilation using chest X-ray images in hospitalized patients with COVID-19. *BMJ Innov* 2021; 7: 261–270.
 17. Feng YZ, Liu S, Cheng ZY, et al. Severity assessment and progression prediction of COVID-19 patients based on the LesionEncoder framework and chest CT. *Information* 2021; 12: 471.
 18. Alajmi J, Jeremijenko AM, Abraham JC, et al. COVID-19 infection among healthcare workers in a national healthcare system: the Qatar experience. *Int J Infect Dis* 2020; 100: 386–389.
 19. Singhal T. A review of coronavirus disease-2019 (COVID-19). *Indian J Pediatr* 2020; 87: 281–286.
 20. Lu R, Zhao X, Li J, et al. Genomic characterization and epidemiology of 2019 novel coronavirus: implications for virus origins and receptor binding. *Lancet* 2020; 395: 565–574.
 21. Alimohamadi Y, Sepandi M, Taghdir M, et al. Determine the most common clinical symptoms in COVID-19 patients: a systematic review and meta-analysis. *J Prev Med Hyg* 2020; 61: E304–E312.
 22. Medhat MA and El Kassas M. COVID-19 in Egypt: uncovered figures or a different situation? *J Glob Health* 2020; 10: 010368.
 23. Goudouris ES. Laboratory diagnosis of COVID-19. *J Pediatr (Rio J)* 2021; 97: 7–12.
 24. Hui DS, Azhar EI, Madani TA, et al. The continuing 2019-nCoV epidemic threat of novel coronaviruses to global health—The latest 2019 novel coronavirus outbreak in

- Wuhan, China. *Int J Infect Dis* 2020; 91: 264–266.
25. Zhang ZL, Hou YL, Li DT, et al. Laboratory findings of COVID-19: a systematic review and meta-analysis. *Scand J Clin Lab Invest* 2020; 80: 441–447.
 26. Centers for Disease Control and Prevention. Coronavirus disease (COVID-19): Interim clinical guidance for management of patients with confirmed coronavirus disease (COVID-19) [Internet]. 2021 [cited 2021 Apr 10]. Available from: <http://cdc.gov/coronavirus/2019-ncov/hcp/clinical-guidance-management-patients.htm>.
 27. Böger B, Fachi MM, Vilhena RO, et al. Systematic review with meta-analysis of the accuracy of diagnostic tests for COVID-19. *Am J Infect Control* 2021; 49: 21–29.
 28. Ghaderzadeh M and Asadi F. Deep learning in the detection and diagnosis of COVID-19 using radiology modalities: a systematic review. *J Healthc Eng* 2021; 2021: 6677314.
 29. Brinati D, Campagner A, Ferrari D, et al. Detection of COVID-19 infection from routine blood exams with machine learning: a feasibility study. *J Med Syst* 2020; 44: 135.