https://doi.org/10.26682/csjuod.2023.26.2.54

Journal of University of Duhok., Vol. 26, No.2 (Pure and Engineering Sciences), Pp 605 - 612, 2023
4th International Conference on Recent Innovations in Engineering (ICRIE 2023) (Special issue)

PREDICTING LONG-TERM COVID-19 SYMPTOMS USING MACHINE LEARNING: A CASE STUDY IN KURDISTAN REGION OF IRAQ

AVEEN **K**AKAMEN **M**USTAFA[*] and **I**BRAHIM **I**SMAEL **H**AMARASH[**]
[*]Computer Science, University of Kurdistan Hewler. Kurdistan Region - Iraq
[**]Computer Engineering, Salahaddin University-Erbil, University of Kurdistan Hewler, Kurdistan Region - Iraq

(Accepted for Publication: November 27, 2023)

ABSTRACT

The COVID-19 pandemic has introduced substantial challenges to individuals, communities, and healthcare systems worldwide. While initial responses primarily addressed the acute impact of the virus, emerging evidence highlights a noteworthy portion of individuals grappling with persistent symptoms even after recuperating from the acute phase. This research delves into the domain of algorithms and their application to the context of COVID-19. Specifically, we employ Machine Learning (ML) techniques to formulate a robust model for assessing the likelihood of enduring long-term COVID-19 symptoms among individuals in the recovery phase. Our investigation revolves around a comprehensive dataset drawn from 3,500 patients residing in the Kurdistan Region of Iraq, all of whom had previously contracted COVID-19. Employing a combination of hospital records and direct/mobile interviews, we systematically capture information pertaining to six prevalent long-term symptoms. Rigorous preprocessing techniques are then applied to the collected data, ensuring standardization and mitigating any inherent inconsistencies or biases. To achieve our objective, we harness the capabilities of the TensorFlow and Keras libraries, leveraging a deep learning algorithm. This algorithm plays a pivotal role in predicting the probability of sustained COVID-19 symptoms among recovered patients. This endeavor demonstrates the potential of deep learning, especially when harnessed within a well-structured dataset and coupled with adept preprocessing methodologies. Consequently, our findings underscore the viability of utilizing deep learning algorithms as potent tools for forecasting the propensity of long-term symptom manifestation in individuals previously diagnosed with COVID-19.

KEYWORD: Long-Term, COVID 19, Pandemic, Machine Learning, Deep Learning, Healthcare Systems, Tenser Flow ,keras

1. INTRODUCTION

The emergence of the COVID-19 pandemic in December 2019, originating in Wuhan City, Hubei Province, China, marked a critical global event. The causal agent, severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), prompted the classification of the outbreak as a Public Health Emergency of International Concern on January 30, 2020. During this period, 49,053 laboratory-confirmed cases and 1,381 fatalities were reported [1]. The repercussions of the pandemic extended beyond the immediate realm of critical care, leading to an increased demand for noncritical care disciplines. COVID-19's clinical manifestations exhibit remarkable heterogeneity, with respiratory complications emerging as a recurrent theme. Distinctively, SARS-CoV-2 has manifested in a subset of individuals as a prolonged viral challenge. This prolonged phase of COVID-19 is associated with an array of consequences, including but not limited to lung fibrosis, venous thromboembolism, arterial thromboses, cardiac thrombosis, stroke, cognitive impairments often referred to as "brain fog," dermatological complexities, and mood dysfunctions. A meticulous review of the respiratory dynamics pertaining to SARS-CoV-2 revealed that severe COVID-19 in adults is characterized by enduring and significant shedding of the virus within the lower respiratory tract (LRT). Intriguingly, an emerging body of research delves into the intricate details of COVID-19 transmission dynamics and post-recovery consequences. Notably, investigations into severe cases have unveiled a curious trend: an uptick in upper respiratory tract (URT) shedding following symptom onset. Curiously, this heightened shedding coexists with comparable rates of viral

aveen.kakamen@ukh.edu.krd
Ibrahim.hamad@ukh.edu.krd

clearance, adding an extra layer of complexity to the virus's behavior [2]. Expanding the scope, a comprehensive study involving a cohort of 76 patients alongside 40 healthy individuals embarked on a multifaceted exploration. Through the ingenious application of log-mel spectrograms and the innovative realm of transfer learning, this study unveiled a striking revelation: the post-COVID-19 syndrome casts its shadow over a substantial 10-20% of patients who have seemingly recovered. Remarkably, this enigmatic constellation of persistent symptoms was found to exhibit a predilection for more pronounced manifestation in severe cases and among the female demographic [3]. However, despite remarkable strides within the scientific community towards unraveling the enigma of long-term COVID-19, the challenge at hand persists as an unresolved puzzle. Enter the realm of artificial intelligence (AI) and its potent subfield, machine learning (ML). These cutting-edge methodologies have now taken center stage as potent tools that can potentially unlock the insights concealed within vast troves of historical medical data. In doing so, they offer a tantalizing promise: the ability to foresee the contours of future disease patterns with unprecedented accuracy and foresight [4]. Within this intricate tapestry of scientific exploration, the present study emerges as a significant contributor. By harnessing the formidable might of machine learning (ML) and diving into the depths of deep learning (DL), the focus turns towards the prediction of long-term COVID-19 outcomes. Yet, the narrative acquires a distinct geographical hue as we venture into the Kurdistan region of Iraq—a landscape rich in culture and history, now intersecting with the cutting-edge realm of medical research. It is here that the study's gaze is cast, seeking to unravel the trajectory of COVID-19 persistence among patients who once faced the diagnosis. With AI and DL as guiding beacons, the study aspires to illuminate uncharted avenues of comprehension while also offering pathways for mitigating the lingering specter of COVID-19-related symptoms. As the world watches with bated breath, the fusion of technology and medical inquiry stands poised to reveal insights that could reshape our understanding of post-recovery outcomes, not just in Kurdistan but potentially on a global scale.

2. RELATED WORKS

Long Covid describes the condition of not recovering for many weeks or months following acute SARS-CoV2 infection. It was first described and named as an umbrella term through a social media movement in Spring 2020[5]. Evidence describing the condition is scarce but is starting to emerge on the long-term health impairment and organ damage following COVID-19. A review has concluded there is insufficient evidence to provide a precise definition of Long Covid symptoms and prevalence. Many of those infected in spring 2020 did not have access to testing and therefore have struggled to receive recognition, diagnosis, and support [6],[7]. People who have been exposed to the virus COVID-19 may develop a variety of new, recurring, or ongoing health issues. These disorders are known as post-COVID conditions. The development of post-COVID problems might be observed at least four weeks after infection because most COVID-19 patients recover within a few days to a few weeks after infection [8]. post COVID symptoms can affect anyone who contracted the infection. However, some people who later developed post-COVID illnesses did not know when they became infected. Many people with post-COVID disorders showed symptoms days after discovering they had COVID-19. There is no test that can tell you whether COVID-19 is the cause of your symptoms or disease [9].

Post-COVID-19 symptoms and diseases appeared on many survivors from COVID-19 which are like that of the post-severe acute respiratory syndrome (SARS) fatigue.[10] A study aims to investigate and characterize the manifestations which appear after eradication of the coronavirus infection and its relation to disease severity About 287 survivors from COVID-19 were included in the study, each received a questionnaire divided into three main parts starting from subjects' demographic data, data about the COVID-19 status and other comorbidities of the subject, and finally data about post-COVID-19 manifestations [11]. Response surface plots were produced to visualize the link between several factors. Only 10.8% of all subjects have no manifestation after recovery from the disease while a large percentage of subjects suffered from several symptoms and diseases. The most common symptom reported was fatigue (72.8%), more

critical manifestations like stroke, renal failure, myocarditis, and pulmonary fibrosis were reported by a few percent of the subjects. There was a relationship between the presence of other comorbidities and severity of the disease. Also, the severity of COVID-19 was related to the severity of postCOVID-19 manifestations [12]

Long-term COVID recognized as a public health concern by many research studies, however, datasets are still limited to support a clear map of the problem, over the last two years researches have begun using AI, ML and DL to analyze the available datasets to predict the long COVID suffering in individuals diagnosed with COVID-19. most of the datasets are ordinary information collected from the hospitals and health care systems worldwide.[13],[14],[15],[16]. The utilization of ML in the field of COVID-19 holds promising potential in understanding the factors contributing to the development of long COVID. to the knowledge dataset to the long covid in the literature.

Many similar papers have reviewed different approaches for applying machine learning models for predicting the long-term effects of COVID-19. However, these researches have all shown a few flaws. A number of these studies had insufficient feature sets in their datasets, which would have resulted in a depiction of the long-term COVID-19 results that was incomplete and convoluted. Furthermore, the volume of patient data has presented difficulties for a number of studies, it could compromise their models' generalizability and dependability. Another major issue is how to handle missing variables. Some studies use too simplistic approaches like mean imputation or even accept null values, which could introduce bias or mistakes into their predictions. These drawbacks highlight the necessity for a thorough and robust strategy that takes into account a variety of variables, a sizeable amount of data, and an efficient management of missing data in order to produce more precise and trustworthy forecasts for the long-term effects of COVID-19.

In this study a standard dataset including 40 features and nine target labels for 3000 patients who have been diagnosed with COVID-19 is built. in Kurdistan region of Iraq. deep learning technique is used to develop a prediction model for long COVID based on the collected data. The importance of the study rests in its ability to pinpoint early warning signs and COVID risk factors, which will help medical practitioners provide individualized patient care and prompt intervention. Our work intends to improve the comprehension and management of long-term COVID prediction by adding to the expanding body of research in this area.

3. METHODOLOGY

In heath related fields Data collection is challenging due to privacy concerns and potential patient deaths. Our new dataset was created by direct interaction of patients with (Dr. Rebaz H. Salh), a COVID expert physician at Par International Hospital. 37 attributes and nine target labels represent the dataset, which is divided into six categories: demographic, prior medical history, diagnosis, symptoms during COVID, current disease six months after COVID, and clinical parameter. The research seeks to increase the data's precision and dependability for machine learning applications. Data is gathered from a variety of sources, including hospital records and resident doctors. By contacting each person individually, more information and symptoms are personally gathered. To add Long COVID characteristics to the dataset, the raw data, which includes 3645 observations, is also used. The necessary data is retrieved by the system. Medical tests could be required if the patient's health is poor.

figure 1 shows the categorization of the features of the dataset and a screenshot of the first and last five observations of the dataset is shown in figure 2.

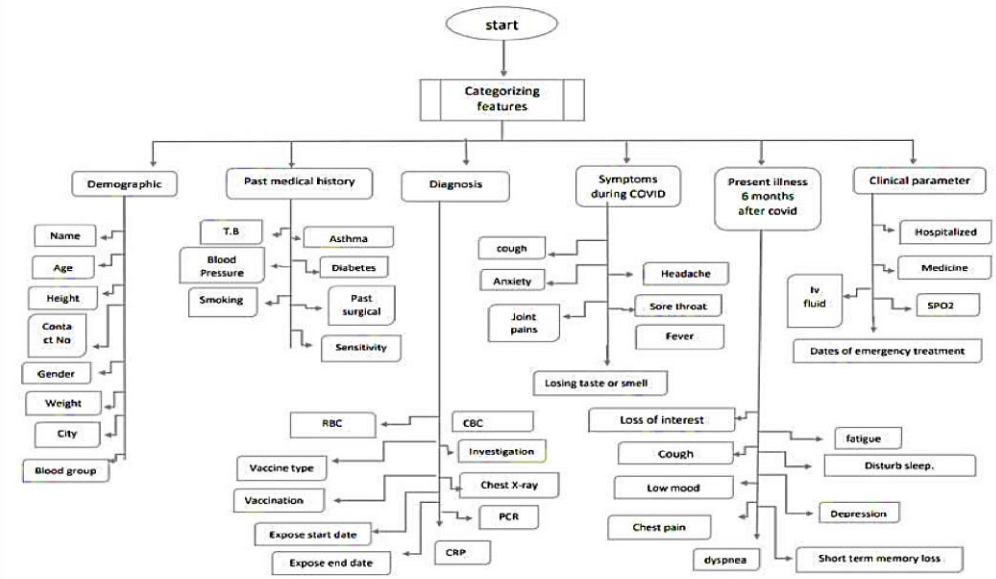

Fig.(1):- Schematic of a data gathering system

.

	age	gender	weight(cm)	height(meters)	address	bloodGroup	T.B	asthma	smoking	blood pressure	...	CBC	PCR	chest-x-ray	investigation	vaccinat
0	84	0	50	155	0	6	0	0	0	1	...	0	1	0	1	1
1	72	0	61	159	0	6	0	0	0	0	...	0	0	0	1	1
2	55	1	69	168	0	6	1	1	0	1	...	0	0	0	1	1
3	39	0	64	178	1	6	0	0	0	0	...	0	1	0	1	1
4	34	1	70	155	1	6	0	1	0	1	...	0	0	0	1	1
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
3058	48	0	83	168	0	6	0	0	0	0	...	0	0	0	1	1
3059	51	1	85	178	0	0	0	0	0	1	...	0	0	0	1	0
3060	45	0	47	167	0	6	0	0	0	0	...	0	0	0	1	1
3061	63	0	40	155	0	6	0	0	0	1	...	0	0	0	1	1
3062	21	1	54	168	0	6	0	0	0	0	...	0	0	0	1	0

Fig.(2):- First/last five **rows** of prepared KLTCD

Data preprocessing is a crucial aspect of machine learning, as it significantly impacts a project's success. MS Excel as better data "beats fancier algorithms." Thorough data cleaning can yield decent results, especially when dealing with large datasets. Outliers, which are extreme figures close to the data range or deviating from the trend, can indicate issues with data input and can cause unstable findings in statistical techniques. Eliminating outliers is essential for improving machine learning modeling and model quality. Eliminating outliers should only be done if there is a compelling reason, and valid justification, such as suspicious measurements. The data preprocessing at an observation level resulted in a relation of observation for 3645 to 3063 records.

4. DATA ANALYSIS AND FEATURE SELECTION

Machine learning refers to a system's ability to learn from problem-specific training data to automate the process of constructing analytical models and solving associated tasks. Deep learning is an artificial neural network-based machine learning concept. Deep learning models outperform shallow machine learning models and traditional data analysis methodologies in many situations. Keras is TensorFlow's official high-level API for constructing and training deep learning models. It has an easy-to-use interface that abstracts away the complexities of TensorFlow's lower-level operations. Keras in TensorFlow is meant to be modular, adaptable, and simple to use, making it appropriate for both novice and professional deep learning practitioners.

Real-world data is frequently skewed, contradictory, and inaccurate. Additionally, it is usually difficult to find precise attribute values or

aveen.kakamen@ukh.edu.krd
Ibrahim.hamad@ukh.edu.krd

trends. It is critical that we preprocess the data before introducing it to our model since both the quality of the data and the information that can be obtained from it have a direct impact on how effectively our model can learn [17].

```
Index(['age', 'gender', 'weight(cm)', 'height(meters)', 'address',
       'bloodGroup', 'T.B', 'asthma', 'smoking ', 'DIABETES',
       'past-surgirical', 'sensitivity ', 'fever ', 'cough ', 'Anxiety',
       'Headache', 'sore throat', 'joint pain ', 'PC-loss -of-intresert 1',
       'P-C, Disturb sleep ', 'P-C,Dyspnea 0', 'P-C,Cough 3', 'P-C,Chest pain',
       'P-C,Fatigue 6', 'target', 'IV.fluid', 'medicine ', 'SPO2',
       'Hospitalized', 'No-of-Dates-of-Treatment', 'chest pain', 'CBC', 'PCR',
       'investigation ', 'NO-OF-TESTS', 'CRB'],
      dtype='object')
```

Fig.(3):- dataset feature names, a screenshot from Jupyter notebook

A total of 36 characteristics are included across 3105 patient records in the collected data set. After data cleaning, there were 3063 records left, which is a respectable number considering how carefully the data was put into the dataset and the fact that the researcher didn't include a patient who had recovered if there were more than three Null results. Due to ethical issues the name and phone number features were removed from the dataset. Figure 3 shows a screenshot of Jupyter notebook which illustrates the names of all the attributes. The collected data consists of both numerical and category data types. Age, height, weight, dates of treatment, SPO2, expose start date, and expose end date are the seven attributes out of 46 (including target labels) that provide numerical features with continuous values. Additionally, we just have one text string attribute. The remaining are categorial data and yes/no (1/2) choices for features like smoking, fever, cough, diabetes, and all other attributes.

Correlation matrix Is calculated to examine the relationships between the features in the dataset and their observations with target labels i.e., considering features- features and features-target correlations. Each variable has a perfect correlation with itself, hence the diagonal of the heatmap will consist entirely of 1. The Long-Term COVID syndromes had been eliminated from the dataset after adding the target column, counting, and identifying the LONG term COVID syndromes. Others had at least one of the Long-Term COVID symptoms, and we got 559 negative results. The classification in this instance is multiclass. a few characteristics are correlated with one another, which is a good thing for feature selection because it eliminates the need for us to make a decision.

We used a code that selects highly linked features using a correlation matrix, generating a highly_correlated_mask based on absolute correlation values (in our case greater than or equal to 0.5). Evaluating highly connected features using a threshold can help identify and delete correlated features. This can enhance machine learning model performance and interpretability. The outcome of this process is a new version of the dataset with 31 attributes as shown in figure 4.

```
Index(['age', 'gender', 'weight(cm)', 'height(meters)', 'address',
       'bloodGroup', 'T.B', 'asthma', 'smoking ', 'blood pressure', 'DIABETES'
       'past-surgirical', 'sensitivity ', 'fever ', 'cough ', 'Anxiety',
       'Headache', 'sore throat', 'joint pain ', 'target ', 'IV.fluid',
       'medicine ', 'SPO2', 'Hospitalized', 'No-of-Dates-of-Treatment',
       'chest pain', 'CBC', 'PCR', 'investigation ', 'NO-OF-TESTS', 'CRB'],
```

Fig.(4):- selected features of KLTCD

Figure 3 presents an age distribution histogram, a crucial visualization of age groups within a dataset. This process transforms continuous age values into discrete intervals or groups, providing a structured representation of age-related patterns. By observing the distribution of individuals across these discretized age groups, valuable insights can be gained about the prevalence of long-term COVID-19 effects within specific age cohorts. The distinct bars in the histogram assess the population's age composition and the potential impact of different age ranges on prolonged COVID-19 symptoms. This helps in understanding the dataset's age distribution and elucidating potential age-related trends, which can inform predictive models' outcomes and recommendations.

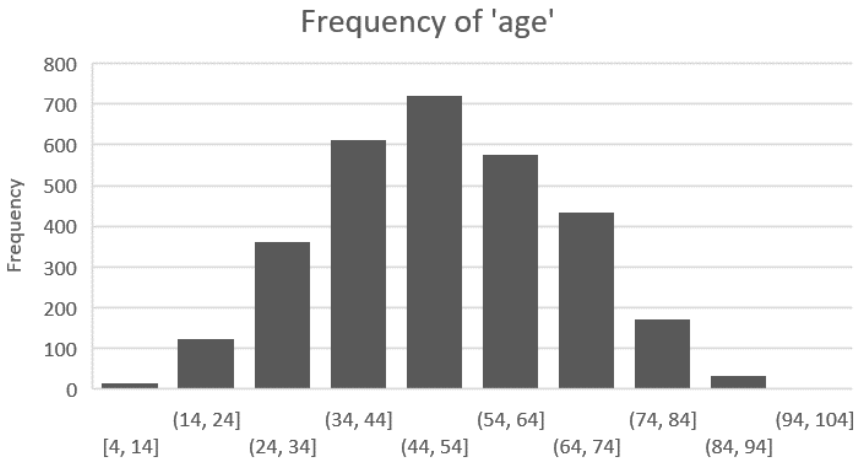

Fig.(5):- Discretization of Age Attribute with histogram

5. EXPERIMENTAL SETTING AND PERFORMANCE ANALYSIS

Before starting the experimentation, we applied some more preprocessing to the dataset to be prepared for the implementation phase categorial data represent in the original dataset were converted into numerical attributes manually. The resulted dataset severed as foundation for our classification model. the TensorFlow keras library within Jupyter notebook environment in the implementation phase. during the experimentation we explained two scenarios.

In the first scenario with the nine target labels, we opted for a binary classification approach, we combined all nine target labels into single binary labels of yes and no, classification weather the patient is affected by long term COVID or not. we used deep learning algorithm to build the model deep leering algorithm implemented with the TensorFlow keras library. The result was not satisfactory with the accuracy of 79.11% as shown in table 1. This leads to selected alternative approaches to improve the performance of the model.

Table (1):- confusion matrix of all features

	TN	FP
FN	9	105
TP	1	498
ACCURACY	79.11%	

In the second scenario we applied the algorithm on the selected features to increase the accuracy. Using confusion matrix and its related accuracy index, the accuracy achieved was 99.83%. the confusion matrix is shown in table 2. note that TN (True Negative): The count is 113. These are instances where the model correctly predicted the negative class, and the actual class was also negative. FP (False Positive): The count is 1. This represents cases where the model

predicted the positive class, but the actual class was negative. FN (False Negative): The count is 0. This indicates that there were no instances where the model predicted the negative class, but the actual class was positive. TP (True Positive): The count is 499. These are instances where the model correctly predicted the positive class, and the actual class was indeed positive. Accuracy: The accuracy is calculated as (TP + TN) / (TP + TN + FP + FN), which in this case is (499 + 113) / (499 + 113 + 1 + 0) = 99.83%

Table (2):- confusion matrix of selected features

	TN	FP
FN	113	1
TP	0	499
ACCURACY	99.83%	

The confusion matrices reveals that all cases in the training set are properly identified, resulting in zeros in both the false positive and false negative counts. This is an ideal categorization. The erroneous negative count implies that just one positive case was misclassified as negative, while the rest were correctly classified. The training was carried out over 100 epochs. The loss value is the loss function value at the conclusion of each epoch. The model attained an accuracy of roughly 99.83% on the testing. That is, the model correctly identified nearly all the cases in the training set.

6. CONCLUSIONS

This study presented the development of the Kurdistan Long-Term COVID Dataset (KLTCD), which comprises comprehensive data from 3,500 COVID recovery patients in Erbil, Kurdistan, Iraq. The dataset includes 37 features, such as demographic information, clinical parameters, medical history, COVID diagnosis, symptoms during COVID, and nine long-term symptoms observed six months after recovering from COVID-19.Through a preprocessing and feature selection phases, the dataset was prepared for analysis. Various visualization techniques were employed to gain valuable insights. Notably, the highly_correlated_mask approach was used to identify the most influential features and remove highly correlated ones.

Deep learning was then applied to predict long-term COVID outcomes, considering two scenarios. The first scenario involved using all target labels during analysis, while the second scenario involved combining all target labels into a single binary label. The results demonstrated the effectiveness of deep learning in predicting both normal and abnormal instances in such datasets.

Long-term COVID research are crucial for healthcare practitioners, policymakers, and researchers to establish effective methods for monitoring and treating COVID-19 patients' long-term health requirements. This study can help identify individuals more likely to suffer long-term consequences and provide tailored treatments and assistance.

Aknowledgment

Special thanks to Dr. Rebaz hamza Salih for his expert guidance and unwavering encouragement throughout this research journey. His invaluable insights and constructive feedback significantly enhanced the quality of this thesis on COVID-19. Additionally, I extend my heartfelt appreciation to Par Hospital for their generous cooperation and provision of essential data, which proved instrumental in conducting this study.

REFERENCE

H. Zhang *et al.*, "Data-driven identification of post-acute SARS-CoV-2 infection subphenotypes," *Nat Med*, vol. 29, no. 1, pp. 226–235, Jan. 2023, doi: 10.1038/s41591-022-02116-3.

P. Z. Chen, N. Bobrovitz, Z. Premji, M. Koopmans, D. N. Fisman, and F. X. Gu, "SARS-COV-2 shedding dynamics across the respiratory tract, sex, and disease severity for adult and pediatric COVID-19," *Elife*, vol. 10, Aug. 2021, doi: 10.7554/eLife.70458.

M. Heightman *et al.*, "Post-COVID-19 assessment in a specialist clinical service: A 12-month, single-centre, prospective study in 1325 individuals," *BMJ Open Respir Res*, vol. 8, no. 1, Nov. 2021, doi: 10.1136/bmjresp-2021-001041.

J. Dong *et al.*, "Machine learning model for early prediction of acute kidney injury (AKI) in pediatric critical care," *Crit Care*, vol. 25, no. 1, Dec. 2021, doi: 10.1186/s13054-021-03724-0.

T. Aishwarya and V. Ravi Kumar, "Machine Learning and Deep Learning Approaches to Analyze and Detect COVID-19: A Review," *SN Computer Science*, vol. 2, no. 3. Springer, May 01, 2021. doi: 10.1007/s42979-021-00605-9.

S. Lopez-Leon *et al.*, "More than 50 long-term effects of COVID-19: a systematic review and meta-analysis," *Sci Rep*, vol. 11, no. 1, Dec. 2021, doi: 10.1038/s41598-021-95565-8.

N. DeLuca *et al.*, "Experiences with COVID-19 case investigation and contact tracing: A qualitative analysis," *SSM - Qualitative Research in Health*, vol. 3, Jun. 2023, doi: 10.1016/j.ssmqr.2023.100244.

Adel Abdel Moneim, Marwa A Radwan, and Ahmed I Yousef, "COVID-19 and cardiovascular disease: manifestations, pathophysiology, vaccination, and long-term implication," Jul. 2022.

D. Assaf *et al.*, "Utilization of machine-learning models to accurately predict the risk for critical COVID-19," *Intern Emerg Med*, vol. 15, no. 8, pp. 1435–1443, Nov. 2020, doi: 10.1007/s11739-020-02475-0.

T. Chakraborty, R. F. Jamal, G. Battineni, K. V. Teja, C. M. Marto, and G. Spagnuolo, "A review of prolonged post-covid-19 symptoms and their implications on dental management," *International Journal of Environmental Research and Public Health*, vol. 18, no. 10. MDPI, May 02, 2021. doi: 10.3390/ijerph18105131.

H. Göker *et al.*, "The effects of blood group types on the risk of COVID-19 infection and its clinical outcome," *Turk J Med Sci*, vol. 50, no. 4, pp. 679–683, 2020, doi: 10.3906/sag-2005-395.

E. A. Troyer, J. N. Kohn, and S. Hong, "Are we facing a crashing wave of neuropsychiatric sequelae of COVID-19? Neuropsychiatric symptoms and potential immunologic mechanisms," *Brain, Behavior, and Immunity*, vol. 87. Academic Press Inc., pp. 34–39, Jul. 01, 2020. doi: 10.1016/j.bbi.2020.04.027.

A. Pavli, M. Theodoridou, and H. C. Maltezou, "Post-COVID Syndrome: Incidence, Clinical Spectrum, and Challenges for Primary Healthcare Professionals," *Archives of Medical Research*, vol. 52, no. 6. Elsevier Inc., pp. 575–581, Aug. 01, 2021. doi: 10.1016/j.arcmed.2021.03.010.

L. E. Boulware *et al.*, "Combating Structural Inequities — Diversity, Equity, and Inclusion in Clinical and Translational Research," *New England Journal of Medicine*, vol. 386, no. 3, pp. 201–203, Jan. 2022, doi: 10.1056/nejmp2112233.

N. Subramanian, O. Elharrouss, S. Al-Maadeed, and M. Chowdhury, "A review of deep learning-based detection methods for COVID-19," *Computers in Biology and Medicine*, vol. 143. Elsevier Ltd, Apr. 01, 2022. doi: 10.1016/j.compbiomed.2022.105233.

R. Karthikeyan, *et al*., "A fractional order model for the novel coronavirus (COVID-19) outbreak", Nonlinear Dynamics, 24 June 2020. doi.org/10.1007/s11071-020-05757-6.

B. A. S. Al-rimy, M. A. Maarof, and S. Z. M. Shaid, "Ransomware threat success factors, taxonomy, and countermeasures: A survey and research directions," *Comput Secur*, vol. 74, pp. 144–166, May 2018, doi: 10.1016/J.COSE.2018.01.001.