Journal of Communicable Diseases
Volume 55, Issue 4 - 2023, Pg. No. 71-79
Peer Reviewed & Open Access Journal

Research Article

A Machine Learning Model to Predict COVID-19 Infection Risk in the United Arab Emirates

Sara AlShaya[1], Maryam Sayed Jaffar[1], Sheik Abdullah Jamal Mohideen[1], Aji Gopakumar[1], Vibhor Mathur[1], Maimoona Saeed[1]

[1]Emirates Health Services, Dubai, United Arab Emirates.
DOI: https://doi.org/10.24321/0019.5138.202358

I N F O

Corresponding Author:
Aji Gopakumar, Emirates Health Services, Dubai, United Arab Emirates.
E-mail Id:
aji.gopakumar@ehs.gov.ae
Orcid Id:
https://orcid.org/0000-0002-7485-8662
How to cite this article:
AlShaya S, Jaffar MS, Mohideen SAJ, Gopakumar A, Mathur V, Saeed M. A Machine Learning Model to Predict COVID-19 Infection Risk in the United Arab Emirates. J Commun Dis. 2023;55(4):71-79.

Date of Submission: 2023-09-25
Date of Acceptance: 2023-12-26

A B S T R A C T

Background: One of the biggest challenges that public health officials face is to identify the infection risk among the population and the same scenario was repeated during the COVID-19 period. It was crucial to identify the patients at high risk who needed immediate care, and accordingly, an optimal machine learning model has been developed for the prediction of infection risk which can be modified and reused.

Objective: The main objective of this article is to predict the COVID-19 infection risk of a patient. The predictive model would help in developing comprehensive plans to respond to both clinical treatments and control of the spread.

Method: A retrospective study was carried out on 23,996 encounters in a nationwide cohort in the United Arab Emirates. Patient information was collected retrospectively from respective physicians and was uploaded in a patient under investigation (PUI) form filled during the COVID-19 screening. The variables were age, gender, body temperature, comorbidities and patient symptoms. This data (age, gender, body temperature, comorbidities and patient symptoms) was fed into AI-based machine learning algorithms to come up with a COVID-19 infection prediction model.

Results: The study found that 3818 (15.9%) cases were COVID-19 positive. Based on model performance, parsimony, and explainability, the Gradient Boosting Model was finally selected. The area under the curve (AUC) for this model was 0.746 on the training dataset and 0.736 on the validation dataset.

Conclusion: A highly interpretable machine learning model comprising multiple patient characteristics and symptoms was developed to predict COVID-19 infection in patients.

Keywords: COVID-19, Prediction Model, Infection Risk, United Arab Emirates

Introduction

The COVID-19 pandemic was disrupting the lives of people around the world and governments and businesses were struggling to deal with this important disruption. With 94 million cases worldwide, one of the biggest challenges that public health officials were facing was to identify the infection risk among the population and predict it among suspected or healthy population groups during the COVID-19 period. Demographics, clinical evaluations, and basic clinical test results can be used to categorise patients who are at a higher risk of contracting COVID-19 and should, therefore, be prioritised for testing and isolation. As the COVID-19 pandemic started gripping the globe in early 2020, like most of the other nations, the UAE healthcare system had to be ready overnight with what is now known as the 'response phase'. The pandemic response required an end-to-end strategy that would not only include clinical support but encapsulate an inter-system mechanism to effectively manage the existing cases whilst taking all required measures to limit the spread of the disease.

The Emirates Health Services (EHS) Establishment of UAE, with a strong vision of providing comprehensive and responsive healthcare to sustain the health of its population, immediately devised a plan aligning with its mission and values. EHS, as one of the largest healthcare providers in UAE (total facilities under EHS–136), consisting of 17 hospitals and several COVID-19 field hospitals with more than 100 primary healthcare centres and public health centres, carried the responsibility of reducing the spread of COVID-19 infection in UAE. UAE's COVID-19 strategies and response have resulted in the nation being ranked among the top countries in the world in controlling the pandemic effectively. The EHS was at the forefront of the national response, including early steps to contain the spread of infection, conduct testing for disease identification, and implement standard operating procedures to contain the outbreak. Utilising its rich data insights, EHS has relied on a scientific, data-backed approach to curate the strategy and communications. Its well-connected system not only enables early case identification, tracking test results, and evaluation of resource consumption and constraints but also has a mechanism to identify at-risk groups and direct cases for treatment.

During the COVID-19 period, EHS built vast capabilities in a short time to collect and store data that could be analysed in myriad ways to help mitigate the impact of this pandemic. Amongst the several mechanisms to manage this health crisis, a customised surveillance workflow was implemented by the ministries through the Wareed, system of Electronic Medical Records (EMR). Wareed is the programme name for the Cerner millennium EMR used in EHS which centrally connects EHS's medical facilities in order to cater to user needs and obtain up-to-date medical records and clinical information. This workflow included registration, documentation and patient registry that could electronically capture vital data elements for every individual screened in any way at any mass surveillance site. The screening included capturing demographic information of the individuals, outline of their past medical history, travel information, their overall health status and associated high-risk factors. These screening responses were captured at all mass surveillance sites like airports, high-density population areas, primary health centres, and tertiary hospitals. Having such data elements available for a large proportion of the population that was being screened for COVID-19 supported the idea of having a tool for COVID-19 risk identification in the population utilising data insights.

A research report on clinical features of 85 fatal cases from Wuhan revealed that the mean age of patients who developed COVID-19 infection was 65.8 years, and out of the total patients, 72.9% were male.[1] Common symptoms were fever (91.8%), shortness of breath (58.8%), fatigue (58.8%) and dyspnoea (70.6%). Hypertension, diabetes, and coronary heart disease were the most common comorbidities. Another interesting work was about the risk factors of critical and mortal COVID-19 cases.[2] Thirteen studies were included in a meta-analysis involving a total number of 3027 patients with SARS-CoV-2 infection. The analysis revealed that being male, older than 65 years of age, and smoking were the risk factors associated with disease progression in patients with COVID-19. It also confirmed that the proportion of underlying comorbidities such as hypertension, diabetes, cardiovascular disease, and respiratory disease was significantly higher in critical/ mortal patients compared to non-critical patients. Based on a similar meta-analysis that considered 207 studies in research, it was found that 49 factors related to patients' demographics, history and other physical and health examination factors could provide valuable prognostic information on mortality and/ or poor disease prognosis in patients with the COVID-19 infectious disease.[3] It was further concluded that the identified prognostic factors can help clinicians and policymakers in tailoring management strategies for patients with COVID-19 infectious disease while researchers can utilise the findings to develop multivariable prognostic models that could eventually facilitate decision-making and improve patient-related outcomes. In another study, Argenziano et al. found that out of the 1000 patients covered in the analysis, 150 presented to the emergency department, 614 were admitted to hospital and 236 were admitted or transferred to intensive care units.[4] The most common presenting symptoms were cough (73.2%), fever (72.8%), and dyspnoea (63.1%). Patients in hospitals, particularly those treated in intensive care units, often had baseline comorbidities,

ISSN: 0019-5138
DOI: https://doi.org/10.24321/0019.5138.202358

including hypertension, diabetes, and obesity. Patients admitted to the intensive care units were older (comprising 95% of people aged more than 35 years as compared to 92% for the sample), predominantly male (66.9%), and had longer lengths of stay (staying for 23 days in the hospital as compared to 6 days for the sample). Around 78% of these people developed acute kidney injury and 35.2% needed dialysis. Only 4.4% (6/136) of patients who required mechanical ventilation were first intubated more than 14 days after symptom onset. Another empirical study suggested that elderly patients with COVID-19 are more likely to progress to severe disease.[5] There is research mentioning that fatigue and expectoration are signs of severe COVID-19 infection.[6] A multicenter study conducted in Wuhan tried to identify patient symptoms and characteristics to construct an effective model for early identification of cases at high risk of progression to severe COVID-19.[7] Of the 372 patients in the cohort, around 19.4% developed severe COVID-19. Older age; higher serum lactate dehydrogenase, C-reactive protein, coefficient of variation of red blood cell distribution width, blood urea nitrogen, and direct bilirubin; and lower albumin were associated with severe COVID-19. A nomogram was generated which indicated high clinical net benefit. A study conducted in Yicheng City built an effective prediction model for COVID-19 severity by combining radiological outcomes with clinical biochemical indices and forecasted the severity based on the results of patients' laboratory tests where the model yielded 81% accuracy.[8] A similar type of study used an AI framework with predictive analytics (PA) and found that alanine transaminase (ALT), presence of body pain, and elevated haemoglobin (RBC) are the clinical features that are the most accurate indicators of severity with achievement of 70% to 80% accuracy in predicting severe cases.[9] Overall, the diverse nature of previous studies emphasises that there is no golden standard of a predictive model and the most appropriate method remains to be identified.

Objectives

As the pandemic marked its entry into the region and there was a surge of cases, it was imperative to develop the infection risk identification mechanism for different groups of individuals within the UAE population. With laboratories already strained, a scientific approach was required for the clinical evaluation at the time of screening itself to predict the possibility of infection in an individual which would facilitate the future direction of moving the person to a quarantine facility or only a home quarantine until either the COVID-19 confirmatory results are available, or the individual is out of the risk period.

The identification of a prediction tool is needed to assist in early risk assessment and support the healthcare system's response to COVID-19. It may help to recognise the evolving nature of the virus and the required public health measures. In other words, this prediction tool serves as a valuable addition to the public health toolkit, as it can provide early insights to support decision-making in allocating resources and quarantine measures effectively. By identifying individuals at higher risk, the tool can contribute to early isolation measures and contact tracing, to limit the spread of the virus effectively.

To support this high-priority requirement, a prediction tool was needed to alert clinicians about the risk of a patient testing positive for the virus and make this part of an end-to-end process that goes beyond any single facility to on-the-ground mass screening efforts.

Materials and Methods

The study adopted an observational cohort study design, conducted over a period of six months in the year 2023. Ethical approval was obtained from the Research Ethics Committee (REC) of the Ministry of Health and Prevention (MoHAP) in June 2022, with an approval number of MOHAP/ DXB-REC/ MJJ/No. 58/2022. Data were retrospectively collected from January to May 2021. Since this was a retrospective study with no direct interaction with participants, an informed consent procedure was not involved. Waiver of consent was submitted to REC and it was approved on the basis of the justification provided.

Machine learning techniques are used to aggregate the collective experiences of thousands of patients to generate an infection risk score. As a result, these machine learning methods have great potential to supplement COVID-19 care. The infection risk score model evaluated different variables and grouped them into exposure risk factors, demographic variables, clinical findings, and clinical test results. Statistical analysis was performed using the SAS software, and multiple machine learning algorithms were applied. Ultimately, the Gradient Boosting Model was selected for analysis.

In this article, we describe how an association of data scientists, healthcare Subject Matter Experts (SMEs) and clinical doctors worked on the development of a machine learning model to predict the risk infection score for a person in the UAE population. This collaboration focused on improving the identification of individuals at risk of infection within the UAE population. While this is a valuable contribution to public health, it is crucial to recognise that the model may have limitations, including the possibility of missed-out cases and the inability to identify asymptomatic individuals. It is thus important that the model should be used as a complementary tool in a broader public health strategy that includes regular testing, contact tracing, and other preventive measures.

ISSN: 0019-5138
DOI: https://doi.org/10.24321/0019.5138.202358

Study Setting

This model was created utilising identified data entered into the EMR (Wareed) of EHS facilities.

Inclusion Criteria

We included adult patients who had a documented encounter with COVID-19 PCR test recorded in Wareed within the past 5 months.

Exclusion Criteria

Participants with incomplete data were excluded from the study.

Sampling Technique

A random sample of 23,996 encounters accounting for 22,865 people who got tested for COVID-19 was analysed to build the model.

Input Variables

A number of variables were analysed such as patient factors, type of visit, and disease symptoms, including admit mode of the patient, encounter type, abdominal/ stomach pain, abdominal bleeding, body ache, diarrhoea, difficulty in breathing, fatigue, headache, muscle pain, health history including any infectious disease, history of dry cough, history of hospitalisation, nausea, recent exposure to a COVID-19 positive patient, shortness of breath, running nose, sore throat, vomiting, weakness, travel history, diabetes, lung disease, liver disease, neurological disease, and any COVID-19 symptom.

Figure 1 shows the process of screening for risk of infection and Figure 2 shows a data flow diagram of the complete process. The average age of persons under consideration for the analysis was 34 years and they were largely male (67%).

The risk infection event is defined as positive based on the COVID-19 PCR test result. Among the 23,996 individuals, 3,818 were found to be infected with COVID-19. Though most of the people (~80%) did not show any symptoms of COVID-19 infection, a minor population showing severe and critical symptoms (~0.27%) had a very high COVID-19 infection rate. In the cohort population, males were more prone to COVID-19 infection having a much higher probability as compared to females. Patients who were asked to go for home quarantine, or home care, or who went for isolation experienced a much higher COVID-19 infection rate as compared to others. Patients suffering from fever or other infections were found to be more susceptible to COVID-19 infection. In line with other studies carried out previously, people in the age group of 46–60 years and those experiencing worsening cough, body ache or headache were more susceptible to COVID-19 infection.

Modelling Techniques

Some machine learning algorithms like Logistic Regression, Random Forest, Gradient Boosting and Neural Network were used to predict the COVID-19 infection risk in our sample. The importance of variables as per Logistic Regression was as follows:

- Encounter type of the patient
- Fever or other infectious disease
- Gender
- COVID-19 symptom severity
- Age group
- Body ache
- Symptoms of worsening cough
- Headache

A complete case analysis was performed, and input variables were explored for data quality and exploratory

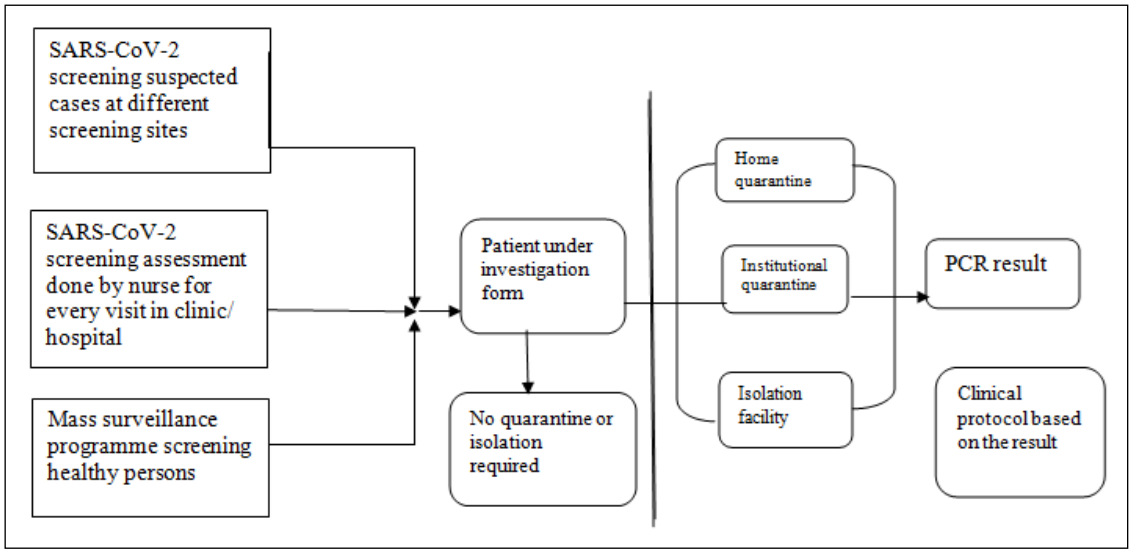

Figure 1.Screening for Risk of Infection at a High Level which Captured the Data Elements in PUI Form that were used in the Machine Learning Model

ISSN: 0019-5138
DOI: https://doi.org/10.24321/0019.5138.202358

data analysis. Multiple machine learning models such as decision tree-based models and neural network models were considered. Classification-based algorithms are intuitive and easy to interpret, and problems that can be described in a linear manner would be best solved by these algorithms. Decision tree-based algorithms consist of multiple decision-based algorithms comprising multiple true or false conditions for input variables. A neural network comprises layers of interconnected artificial neurons that are designed based on a biological neuron. These artificial neurons receive multiple inputs that are multiplied by weights. Neural network models are difficult to interpret but can give better predictive power to the model. All these algorithms were used to develop predictive models, which were then compared based on their discriminative power. Variables were checked for data quality and were discussed with healthcare SMEs and physicians to shortlist the candidate variables to be included in machine learning algorithms. The data were split into training and test datasets. A model was developed on the training dataset and its performance was evaluated in the test dataset. The discriminatory power of the model was calculated using the receiver operating characteristic (ROC) curve in the derivation and validation groups.

Results

A random sample of 23,996 encounters accounting for 22,865 people across the UAE was considered for the development of the machine learning models. 3,818 individuals came out to COVID-19 positive from the given sample. Data were split into training (70%) and validation (30%) datasets. The bivariate distribution of some of the variables is shown in Table 1.

Statistical analysis was performed to analyse the predictor variables. The significant findings from the analysis showed that males, people in the age group of 46–60 years, those who experienced worsening cough, body ache/ headache, and people at home quarantine/ homecare were found to be more prone to high risk of infection. Patients with COVID-19 who were hospitalised and had comorbidities suffered from significant or severe morbidity

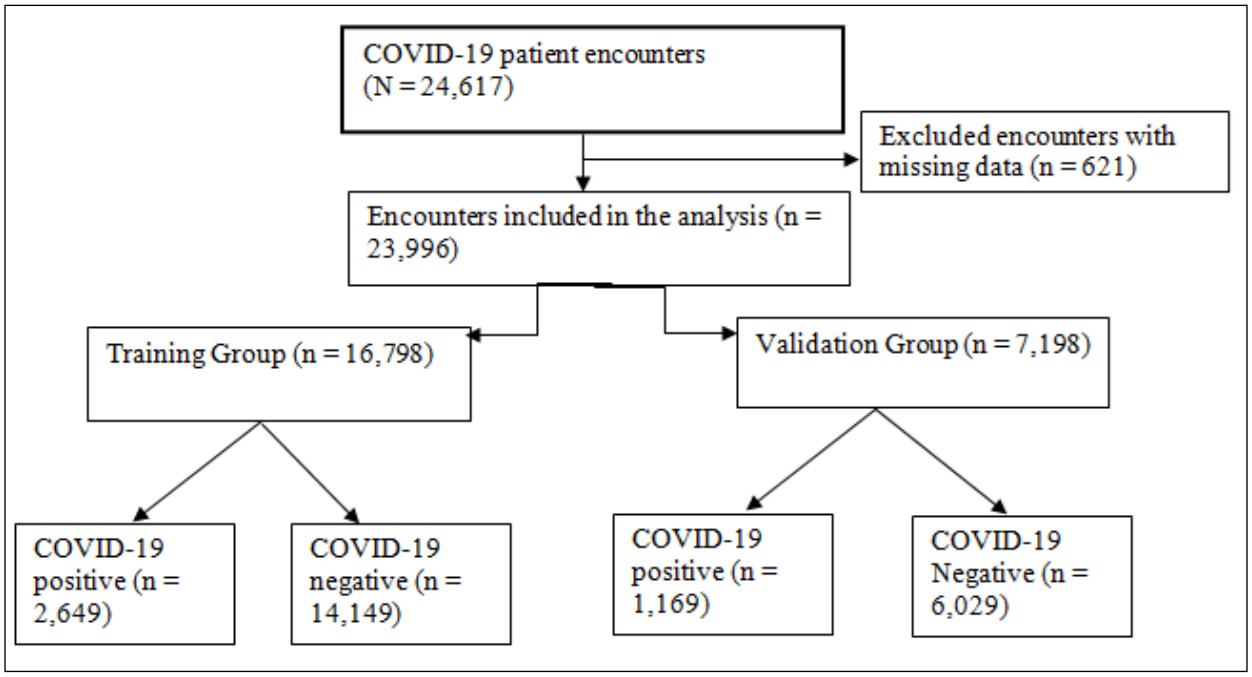

Figure 2.Data Flow Diagram of the Process

Table 1.Bivariate Distribution of Variables

Variables	Total Encounters	Training	Validation
COVID-19 infections			
COVID-19 negative	20178 (84.1)	14149 (84.2)	6029 (83.8)
COVID-19 positive	3818 (15.9)	2649 (15.8)	1169 (16.2)
Fever or infectious disease			
Unknown	546 (2.3)	377 (2.2)	169 (2.3)

ISSN: 0019-5138
DOI: https://doi.org/10.24321/0019.5138.202358

No	18671 (77.8)	13108 (78.0)	5563 (77.3)
Yes	4779 (19.9)	3313 (19.7)	1466 (20.4)
Gender			
Female	7840 (32.7)	5555 (33.1)	2285 (31.7)
Male	16122 (67.2)	11223 (66.8)	4899 (68.1)
Unknown	34 (0.1)	20 (0.1)	14 (0.2)
Body ache			
Unknown	1291 (5.4)	898 (5.3)	393 (5.5)
No	20070 (83.6)	14048 (83.6)	6022 (83.7)
Yes	2635 (11.0)	1852 (11.0)	783 (10.9)
New or worsening cough			
Unknown	803 (3.3)	557 (3.3)	246 (3.4)
No	19947 (83.1)	14021 (83.5)	5926 (82.3)
Yes	3246 (13.5)	2220 (13.2)	1026 (14.3)
Headache			
Unknown	715 (3.0)	477 (2.8)	238 (3.3)
No	20138 (83.9)	14078 (83.8)	6060 (84.2)
Yes	3143 (13.1)	2243 (13.4)	900 (12.5)
COVID-19 symptoms			
Unknown	18775 (78.2)	13177 (78.4)	5598 (77.8)
Critical/ severe	14 (0.1)	10 (0.1)	4 (0.1)
Mild	2003 (8.3)	1378 (8.2)	625 (8.7)
Asymptomatic	3204 (13.4)	2233 (13.3)	971 (13.5)
Encounter type			
Home care, home quarantine	1627 (6.8)	1136 (6.8)	491 (6.8)
Outpatient, e-clinic	15673 (65.3)	10982 (65.4)	4691 (65.2)
Emergency, inpatient	4092 (17.1)	2858 (17.0)	1234 (17.1)
Dialysis, surveillance, daycare, newborn, recurring outpatient	2144 (8.9)	1497 (8.9)	647 (9.0)
Isolation	460 (1.9)	325 (1.9)	135 (1.9)
Age (years)			
< 18	2390 (10.0)	1659 (9.9)	731 (10.2)
18–30	6786 (28.3)	4760 (28.3)	2026 (28.1)
31–45	10343 (43.1)	7286 (43.4)	3057 (42.5)
46–60	3440 (14.3)	2378 (14.2)	1062 (14.8)
61–80	945 (3.9)	651 (3.9)	294 (4.1)
> 80	92 (0.4)	64 (0.4)	28 (0.4)

ISSN: 0019-5138
DOI: https://doi.org/10.24321/0019.5138.202358

Table 2.AUC on the Test and Validation Groups

Model	AUC: Training	AUC: Validation
Gradient boosting	0.745861	0.736096
Random forest	0.745441	0.735430
Logistic regression	0.722580	0.723881
Neural network	0.722888	0.721473
Decision tree	0.643122	0.657723

Patients with fever were 2.8 times more likely to be infected than patients without fever. In the gender analysis, male patients were found to be 2 times more likely to be infected with COVID-19 when compared to female patients. Similarly, patients in the age group of 46–60 years were 1.5 times more likely to be infected as compared to patients in other age groups.

Multiple machine learning algorithms like Logistic Regression, Gradient Boosting, Random Forest, Neural Network and decision tree were used to predict COVID-19 infection. Based on model performance, parsimony, and explainability, the Gradient Boosting Model was finally selected. AUC values for different models have been shown in Table 2.

The model was also able to validate the statistical findings and variable importance complimented the statistical analysis.

The model concluded that the prediction scoring had sufficient predictive value to identify and categorise the individuals who were at a higher probability of having a COVID-19 infection and should be considered to stratify at-risk populations for laboratory testing, isolation and contact tracing measures. To use the model, a confusion matrix was created based on the judgement of data scientists, health SMEs and clinic doctors. Model drivers are similar to the earlier published works such as those of Argenziano et al., Gong et al., and others.[4,7,10] As an extension of this work, we are working to include the nationality of the patient and test if there are certain nationalities that are impacted most. Also, a machine learning-based severity prediction model is a work in progress that will help predict the severity based on current and historical medical records.

Discussion and Future Work

Literature shows that it is of utmost importance to focus efforts on the development of technological tools that allow optimal use of healthcare resources and researchers have carried out few studies in the related areas. One of the recent studies presented an effective machine learning algorithm for the identification of high-risk patients, who present with COVID-19 symptoms.[11] This technology enables rapid identification of high-risk patients at four different clinical stages, ranging from the onset of COVID-19 to ICU admission. An AI-driven research proposed a machine learning method to identify whether a patient has the risk of COVID-19 using the Logistic Regression model, considering multiple symptoms.[12] The model includes independent variables such as the symptoms of pneumonia, diabetes, chronic obstructive pulmonary disease, asthma, hypertension, cardiovascular disease, renal disease, obesity, tobacco-consuming habits, and contact with other COVID-19-positive cases and predicts with 92% model accuracy. A scoping review conducted in the year 2021 compared the performance of several machine learning methods in the prediction of COVID-19 spread and found that these models are capable of identification of learning parameters that affect dissimilarities in COVID-19 spread across various regions or populations.[13] There are a number of studies that have introduced various AI models to forecast COVID-19 cases using a deep learning approach including features like temperature, population, geographical area, population density etc. to find the best suitable model.[14] Another study also explored a research question on the prediction of COVID-19 diagnosis based on various symptoms and an AI model was developed that detects COVID-19 cases by simple information accessed by asking basic questions.[15] The framework helped the health providers and epidemiologists to prioritise testing for COVID-19 when testing resources are limited. The literature shows no studies that addressed the research question of predicting infection risk in the general population. Therefore, the current study focused on determining the incidence rate of COVID-19 infection in the UAE population and testing the significance of observed infection risk. Accordingly, the researchers have hypothesised that there are many patient-related factors associated with COVID-19 infection risk. The developed AI model is used as a decision-support tool in collaboration with healthcare professionals. The final diagnosis and clinical decisions involve the expertise of medical practitioners who can interpret the AI model's output in the context of the patient's overall health and medical history. We acknowledge that the accuracy of machine learning models, including the Logistic Regression model used in our study, can significantly improve with more extensive and diverse datasets. We encourage further research collaborations to pool international data on COVID-19 patients from different populations. The inclusion of data from various regions and populations can lead to a more robust and generalizable model.

The significant contribution of this study can be described from two aspects: machine learning and data perspective in association with clinical relevance in the healthcare domain. This research has not only dealt with the clinical biomarkers but has also focused on identifying an

ISSN: 0019-5138
DOI: https://doi.org/10.24321/0019.5138.202358

appropriate approach for building a machine learning model to predict the patients at risk of COVID-19 infection. The final model proposed by this study has good accuracy and other evaluation metrics and this novel method is also very simplistic. This addresses one of the challenges that is in healthcare AI implementation, the interpretation and explainability of the machine language models to the clinical community in coherence with medical relevance. The predictors identified could be considered as the candidate predictors for new models. As the strategy to manage and control the COVID-19 pandemic moved from the response phase to the recovery phase and confirmatory laboratory testing became a priority, the prediction of infection risk led to the next set of challenges to predict the severity of risk as well as hospitalisation risk. The results from the model can be further used to build a model for predicting the infection severity and triaging population groups who may need acute care during recovery. The reviewed literature and the discussed problem statement imply the significance of the identified key features, which had a very high degree of statistical relevance and clinical significance. The outcomes of this study have practical and clinical findings that may lead to improvements in clinical guidelines and targeted screening approaches. The model could be used by the EMR/ EHR (Electronic Health Record) system to build alerts and registries for patients at a higher risk of any infection and could be used as a first-line screening tool. The model being discussed uses demographic and clinical data as a foundational reference point. For a model that includes information that varies over time and observations collected at different time points, it would give deep insights as it analyses trends, patterns, and changes over time. The data such as laboratory test results taken over time, or symptoms observed during the course of treatment etc. allows for the generation of profound insights when developing future models. The outcome needs to be evaluated and tested on various geographical cohorts. The models developed in this study and the approach could be easily replicated and customised for the same. The features in the data are limited to a few clinical markers and demographics; additional clinical markers from imaging and diagnostics can be included in future models. The medication and pharmacy data with all the time series clinical data needs to be studied and applied to a broader population.

We believe that this model is not just limited to one case study, rather this approach could be evaluated and modified to be applied for other pandemics where initial response times are very critical and mass screening methods are yet to be developed, thus helping the healthcare community improve clinical care and patient outcomes. The prediction model for COVID-19 can augment medical decision-making at a time when it is urgently required.

Limitations

While this study has made significant strides in exploring the potential of AI in addressing larger pandemics, it is important to acknowledge certain limitations that may impact the broader applicability and effectiveness of the proposed AI-based strategies. These limitations highlight the need for further improvement and research in this domain. One of the primary limitations of this study is the relatively small dataset used for training and testing the AI models. The accuracy achieved in this study, while promising, may not necessarily generalize to larger populations or more diverse datasets. The UAE population pyramid is very heterogeneous with about 195 nationalities hence the models developed aimed at a generalised study rather than focusing on accuracy which may lead to overfitting of the models.

The diseases targeted in this study are associated with high mortality rates, making it imperative to achieve extremely high accuracy in AI-based strategies. However, even a small margin of error can have significant consequences in the context of diseases with high mortality rates. While the AI models may show promise, they need to be further refined to minimise false negatives and false positives. The study also acknowledges that the use of AI for pre-screening may miss some cases of epidemiological importance in the transmission dynamics of the diseases. While the study provides a valuable foundation, it is essential to recognise that AI-based pre-screening should not replace comprehensive diagnostic methods entirely and the limitations of using AI as the sole method of disease detection should be considered. In addition, the performance of AI models may differ when applied to other diseases with varying characteristics. Therefore there is a need for caution when extending the findings of this study to diseases with different epidemiological profiles.

Conclusion

The impact of COVID-19 was profound, heavily straining our healthcare information systems. However, it also exposed our extensive capabilities using AI during emergencies. The pandemic response required a strategy that would not only include clinical support but encapsulate an inter-system mechanism to effectively manage the existing cases. It is crucial for any public health system to identify which patients will need immediate care. In this context, machine learning techniques were used to predict the COVID-19 infection risk for a patient in the UAE. An infection risk score model was developed and data scientists/ clinical doctors worked together on the development of this machine model and found that it has great potential to supplement COVID-19 care. This highly interpretable machine learning model helped to predict the COVID-19 infection risk in various categories of the UAE population. The model gave

ISSN: 0019-5138
DOI: https://doi.org/10.24321/0019.5138.202358

insights into the patients who were in need of immediate care and at high risk. Gradient Boosting model was built with a high discriminatory power that facilitated the future direction of moving the person to hospital admission, or quarantine. Males, patients in the age group of 46–60 years and with specific comorbidities and worsening symptoms were found more prone to infection. This led to the revision of the COVID-19 instruction manual and precautionary/ preventive measures in the EHS facilities.

Acknowledgements

We acknowledge all the healthcare workers involved in the diagnosis and treatment of patients with COVID-19 in the UAE. We thank the Emirates Health Services, the Ministry of Health and Prevention, UAE, and the medical representatives of hospitals across the UAE for their efforts in collecting the medical records.

Source of Funding: None

Conflict of Interest: None

References

1. Du Y, Tu L, Zhu P, Mu M, Wang R, Yang P, Wang X, Hu C, Ping R, Hu P, Li T, Cao F, Chang C, Hu Q, Jin Y, Xu G. Clinical features of 85 fatal cases of COVID-19 from Wuhan. A retrospective observational study. Am J Respir Crit Care Med. 2020 Jun 1;201(11):1372-9. [PubMed] [Google Scholar]
2. Zheng Z, Peng F, Xu B, Zhao J, Liu H, Peng J, Li Q, Jiang C, Zhou Y, Liu S, Ye C, Zhang P, Xing Y, Guo H, Tang W. Risk factors of critical & mortal COVID-19 cases: a systematic literature review and meta-analysis. J Infect. 2020 Aug;81(2):e16-25. [PubMed] [Google Scholar]
3. Izcovich A, Ragusa MA, Tortosa F, Marzio MA, Agnoletti C, Bengolea A, Ceirano A, Espinosa F, Saavedra E, Sanguine V, Tassara A, Cid C, Catalano HN, Agarwal A, Foroutan F, Rada G. Prognostic factors for severity and mortality in patients infected with COVID-19: a systematic review. PLoS One. 2020 Nov 17;15(11):e0241955. [PubMed] [Google Scholar]
4. Argenziano MG, Bruce SL, Slater CL, Tiao JR, Baldwin MR, Barr RG, Chang BP, Chau KH, Choi JJ, Gavin N, Goyal P, Mills AM, Patel AA, Romney ML, Safford MM, Schluger NW, Sengupta S, Sobieszczyk ME, Zucker JE, Asadourian PA, Bell FM, Boyd R, Cohen MF, Colquhoun MI, Colville LA, de Jonge JH, Dershowitz LB, Dey SA, Eiseman KA, Girvin ZP, Goni DT, Harb AA, Herzik N, Householder S, Karaaslan LE, Lee H, Lieberman E, Ling A, Lu R, Shou AY, Sisti AC, Snow ZE, Sperring CP, Xiong Y, Zhou HW, Natarajan K, Hripcsak G, Chen R. Characterization and clinical course of 1000 patients with coronavirus disease 2019 in New York: retrospective case series. BMJ. 2020;369:m1996. [PubMed] [Google Scholar]
5. Daoust JF. Elderly people and responses to COVID-19 in 27 countries. PLoS One. 2020;15(7):e0235590. [PubMed] [Google Scholar]
6. Li J, Chen Z, Nie Y, Ma Y, Guo Q, Dai X. Identification of symptoms prognostic of COVID-19 severity: multivariate data analysis of a case series in Henan province. J Med Internet Res. 2020 Jun 30;22(6):e19636. [PubMed] [Google Scholar]
7. Gong J, Ou J, Qiu X, Jie Y, Chen Y, Yuan L, Cao J, Tan M, Xu W, Zheng F, Shi Y, Hu B. A tool for early prediction of severe Coronavirus Disease 2019 (COVID-19): a multicenter study using the Risk Nomogram in Wuhan and Guangdong, China. Clin Infect Dis. 2020;71(15):833-40. [PubMed] [Google Scholar]
8. Li D, Zhang Q, Tan Y, Feng X, Yue Y, Bai Y, Li J, Li J, Xu Y, Chen S, Xiao SY, Sun M, Li X, Zhu F. Prediction of COVID-19 severity using chest computed tomography and laboratory measurements: evaluation using a machine learning approach. JMIR Med Inform. 2020;8(11):e21604. [PubMed] [Google Scholar]
9. Jiang X, Coffee M, Bari A, Wang J, Jiang X, Huang J, Shi J, Dai J, Cai J, Zhang T, Wu Z, He G, Huang Y. Towards an artificial intelligence framework for data-driven prediction of coronavirus clinical severity. Comput Mater Contin. 2020;63(1):537-51. [Google Scholar]
10. Ma X, Li A, Jiao M, Shi Q, An X, Feng Y, Xing L, Liang H, Chen J, Li H, Li J, Ren Z, Sun R, Cui G, Zhou Y, Cheng M, Jiao P, Wang Y, Xing J, Shen S, Zhang Q, Xu A, Yu Z. Characteristic of 523 COVID-19 in Henan Province and a Death Prediction Model. Front Public Health. 2020;8:475. [PubMed] [Google Scholar]
11. Quiroz-Juárez MA, Torres-Gómez A, Hoyo-Ulloa I, León-Montiel RJ, U'Ren AB. Identification of high-risk COVID-19 patients using machine learning. PLoS One. 2021;16(9):e0257234. [PubMed] [Google Scholar]
12. Majumder AB, Gupta S, Singh D, Majumder S. An intelligent system for prediction of COVID-19 case using machine learning framework-logistic regression. J Phys Conf Ser. 2021;1797:012011. [Google Scholar]
13. Ghafouri-Fard S, Mohammad-Rahimi H, Motie P, Minabi MA, Taheri M, Nateghinia S. Application of machine learning in the prediction of COVID-19 daily new cases: a scoping review. Heliyon. 2021;7(10):e08143. [PubMed] [Google Scholar]
14. Devaraj J, Elavarasan RM, Pugazhendhi R, Shafiullah GM, Ganesan S, Jeysree AK, Khan IA, Hossain E. Forecasting of COVID-19 cases using deep learning models: is it reliable and practically significant? Results Phys. 2021;21:103817. [PubMed] [Google Scholar]
15. Zoabi Y, Deri-Rozov S, Shomron N. Machine learning-based prediction of COVID-19 diagnosis based on symptoms. NPJ Digit Med. 2021;4(1):3. [PubMed] [Google Scholar]