ABSTRACT TITLE:

An Artificial Intelligence (AI) Model for prediction of COVID 19 infection risk in the UAE population

Sara Alshaya, Sheik Abdullah Jamal Mohideen, Aji Gopakumar, Vibhor Mathur

RELEVANCE:

One of the biggest challenges Public health officials are facing is predicting the infection risk for COVID-19 among a specific population. It was imperative to develop the infection risk identification mechanism for different groups as laboratories were 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 homecare or individual is out of the risk period. To support this high priority requirement and improved health care in UAE, a prediction tool was needed to alert physicians about the patient at risk. The diverse nature of the previous studies emphasizes that there is no golden standard of a predictive model and the most appropriate method remains to be identified.

CONTEXT AND AIMS:

COVID-19 hit the world heavily against the existing healthcare information and its vast capabilities in a short period. 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. To support this high priority requirement, a prediction tool was needed to alert clinicians about the risk of a patient testing positive to 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. In this context, Machine learning techniques were used to predict the COVID-19 infection risk for a patient in the UAE population. The main aim of the study was to identify the patients at high infection risk of COVID-19 and determine the incidence rate among different population groups. Accordingly, an infection risk score model was developed and Data scientists/clinicians worked together on the development of this machine model. The study found it has great potential to supplement COVID-19 care.

METHODS AND FINDINGS:

A retrospective cohort study was carried out on 23996 encounters in a nationwide cohort of UAE. Ethical Approval for the research was obtained from the Ministry of Health and Prevention (MoHAP) and Administrative Approval was issued by EHS. The participants were the patients who visited various EHS facilities in UAE and included entire 4-months data as a population-based study. Patient details were retrospectively collected from the Wareed platform and there was no direct interaction between the investigators and the patients. The research developed an effective AI model for predicting the potential patient characteristics that are associated with covid-19 infection. The model comprised of multiple variables such as age, sex, body temperature, comorbidities, patient symptoms and their clinical test results. The domain 'disease symptoms' included admit mode of the patient, encounter type, abdominal/stomach pain, abdominal bleeding, body ache, diarrhea, difficulty in breathing, fatigue, headache, muscle pain, health history, infected by infectious disease, history of dry cough, history of hospitalization, nausea, recent exposure to Covid-19 positive patient, shortness of breath, running nose, sore throat, vomiting, weakness, travel history, diabetes, lung disease, lever disease, neuro disease etc. Variables were checked for data quality and were discussed with healthcare Subject Matter Experts (SME)and Physicians to shortlist the candidate variables to be included in machine learning algorithms. A high-level infection risk screening, which captured the data elements in PUI form that were used in the machine learning model. This data was finally fed into AI based machine learning algorithms to come up with a COVID-19 infection prediction model. Complete case analysis was performed, and input variables were explored for data quality and exploratory data analysis. The incidence rate covid-19 infection was calculated by identifying the total number of new infected cases and dividing that by the population at risk. Artificial intelligence-based multiple machine learning algorithms such as decision tree-based model, Random Forest, and neural network were used. 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 of the model. All these algorithms were used to develop predictive models, which were then compared based on their discriminative power. Gradient Boosting Model was finally selected based on performance parsimony and explainability. The data was split into training and test dataset. A model was developed on training dataset and its performance was evaluated in test dataset. The discriminatory power of the model was calculated using the receiver operating characteristic curve (AUC- 0.75 on the training dataset and 0.74 on the validation dataset). 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 more prone to covid-19 infection. Patients with fever were 2.8 times more likely to be infected than patients without fever. In the gender analysis the male patients are 2 times more likely to be infected with COVID- 19 when compared to female patients. Similarly, patients in the age group 46-60 were 1.5 times more likely be infected compared to patients in other age groups.

CONCLUSION:

A highly interpretable machine learning model was developed comprising of symptoms and multiple patient characteristics to predict the COVID-19 infection risk in the population. It was concluded that identified prognostic factors can help clinicians and policy makers in tailoring management strategies for patients with COVID-19 infectious disease while researchers can utilize the findings to develop multivariable prognostic models that could eventually improve patient important outcomes. EHS gets more than 2.5 million patient visits every year which generates a huge amount of valuable data and therefore AI technology-driven solution assisted in deriving accurate research results and bringing timely / efficient delivery of health care services. The trend, patterns and current infection rates with visualized facts were made accessible through MANARA platform which gives dashboard facilities to end-users. The study results lead to the revision of the covid-19 instructional manual and precautionary/preventive measures in EHS facilities. As an extension of this research, the model was revised based on a pilot study including patients' cultural and ethnic differences since certain nationalities were found impacted most. Another machine learning-based severity prediction model is under development that will help to predict the level of severity once infected.

REFERENCES:

1. 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 Sep 20;16(9):e0257234. doi: 10.1371/journal.pone.0257234. PMID: 34543294; PMCID: PMC8452016.
2. Annwesha Banerjee Majumder, Somsubhra Gupta2 ,Dharmpal Singh3 and Sourav Majumde. An Intelligent System for Prediction of COVID-19 Case using Machine Learning Framework-Logistic Regression. J. Phys.: Conf. Ser. 2021, 1797 012011.
3. Soudeh Ghafouri-Fard, Hossein Mohammad-Rahimi, Parisa Motie, Mohammad A.S. Minabi, Mohammad Taheri, Saeedeh Nateghinia. Application of machine learning in the prediction of COVID-19 daily new cases: A scoping review. Heliyon 2021; 7(10): e08143, ISSN 2405-8440, https://www.sciencedirect.com/science/article/pii/S2405844021022465
4. Jayanthi Devaraj, Rajvikram Madurai Elavarasan, Rishi Pugazhendhi, G.M. Shafiullah, Sumathi Ganesan, Ajay Kaarthic Jeysree, Irfan Ahmad Khan, Eklas Hossain. Forecasting of COVID-19 cases using deep learning models: Is it reliable and practically significant? Results in Physics 2021, 21; 103817, ISSN 2211-3797. https://doi.org/10.1016/j.rinp.2021.103817.
5. Zoabi, Y., Deri-Rozov, S. & Shomron, N. Machine learning-based prediction of COVID-19 diagnosis based on symptoms. npj Digit. Med. 4, 3 (2021). https://doi.org/10.1038/s41746-020-00372-6.