

Use of Artificial Intelligence for Predicting COVID-19 Outcomes: A Scoping Review

Jinyan LYU^{a,1}, Wanting CUI^a and Joseph FINKELSTEIN^a
^a*Icahn School of Medicine at Mount Sinai, New York NY, USA*

Abstract. During the COVID-19 pandemic, artificial intelligence has played an essential role in healthcare analytics. Scoping reviews have been shown to be instrumental for analyzing recent trends in specific research areas. This paper aimed at applying the scoping review methodology to analyze the papers that used artificial intelligence (AI) models to forecast COVID-19 outcomes. From the initial 1,057 articles on COVID-19, 19 articles satisfied inclusion/exclusion criteria. We found that the tree-based models were the most frequently used for extracting information from COVID-19 datasets. 25% of the papers used time series to transform and analyze their data. The largest number of articles were from the United States and China. The reviewed artificial intelligence methods were able to predict cases, death, mortality, and severity. AI tools can serve as powerful means for building predictive analytics during pandemics.

Keywords. Scoping Review, Artificial Intelligence, COVID-19

1. Introduction

In December 2019, the first COVID-19 case was reported in Wuhan, China. Millions of people have been infected with the virus, and billions of people are affected by the pandemic at work and in life [9]. The impact of this pandemic has lasted longer than many people expected. There are broad areas of research dedicated to combating the COVID-19 pandemic. Artificial intelligence (AI) and traditional mathematical prediction models are valuable tools to forecast the spread of disease and mortality rates.

In this paper, we collected all papers with artificial intelligence models that forecast outcomes. The goal of this paper was to perform a scoping review to explore papers that used AI to predict outcomes of the COVID-19 pandemic.

2. Method

We conducted the scoping review using the Arksey and O'Malley framework to identify papers that used artificial intelligence methods to forecast patients diagnosed with COVID-19. We performed a five-step scoping review: identify the research question, identify relevant studies, study selection, chart the data, and collect, summarize, and report the results [2,11]. We searched PubMed, the MEDLINE database, for published and preprint literature with full text between March 12, 2020, and May 24, 2021. There

¹ Corresponding Author, Jinyan Lyu, Icahn School of Medicine at Mount Sinai, 1770 Madison Ave, New York, NY, USA, 10035; E-mail: jinyan.lyu@mssm.edu.

were 1,057 papers in total. Moreover, we researched selected papers' reference lists. Five different search terms were used: forecasting, predicting, artificial intelligence, machine learning, and COVID-19.

Four inclusion criteria were proposed. The paper should use at least one artificial intelligence method to predict. The predictors are COVID-19 case count, mortality, or severity. The prediction must include at least one predictor, and it can be more than one predictor. The data should be identified data, de-identified data, or synonym data. Furthermore, three exclusion criteria were applied. If the data only comes from social media; if the data is not patient-related data; if a paper only proposed new approach with no modeling application.

Article selection made in two steps, first screening title and abstract, second, screening full text. However, when we searched related papers in MEDLINE, A relatively large result was received. And as the search engine would post all the results, including reference or paper suggestions, on that web page, all irrelevant results were deleted. Then, the two authors, Cui and Lyu, screened all abstracts and titles independently, which eliminated papers that did not use forecasting or prediction models. Secondly, they independently reviewed the remaining papers' full texts. Cohen Kappa tests had been conducted to examine the similarity between two reviewers' screening results, the abstract result was 0.86, and the full-text result was 1.

The data extraction form included an article's published date, dataset name and type, country, the start date of the dataset, the end date of the dataset, medical condition of patients, predictor, AI method, and other methods that were used. We focused on AI methods for prediction. AI is defined as a computational method that learned from current data and then given appropriate outputs.

3. Results

There were 1,057 research articles were found from MEDLINE. After removing duplicate results, irrelevant papers, pre-prints and reviews, there were 197 papers remained. By screening the title and abstract, 58 papers were selected, and then full-texts were reviewed. Based on our selection criteria (**Figure 1**), 17 papers were qualified for further analysis. The first AI forecasting paper was published in May 2020 [5], and most articles were published between Aug 2020 and Apr 2021. One paper could use more than one dataset [14,16,17], and also those datasets might have more than one country's data. Thus, there are 31 countries mention in the dataset. Also, datasets from China and the United States were the most popular. Datasets from Asia were the most analyzed among papers.

The purpose of using the artificial intelligence method was to predict cases, death, mortality, and severity. Since one paper could predict more than one variable, among the 17 papers, seven papers forecasted mortality [1,7,11,13-15,19], five papers forecasted daily report COVID-19 test-positive cases and death cases [4-6,12,16], four papers forecasted severity [3,9,17,18], and one paper predicted all of them [20]. Tree-based model was the most common choice, which was used seven times [9,11,12,13,15,16,19]. The neural network method was a new trend, and five papers applied this method [3,4,5,7,20]. Three papers used the regression method [1,6,17]. In addition, two papers used the deep learning method [14,18]. Furthermore, due to the time sensitivity of COVID-19, four papers used time series analysis first, and then applied the artificial

intelligence model [7,12,14,16]. ARIMA model was the most common choice for time series analysis [7,12].

To predict new confirmed or death cases, authors intended to use the r squared and RMSE to check the model fitness. Meanwhile, forecasting mortality and severity, authors would use the AUC-ROC curve and accuracy to test model performance. In the model fitness part, three articles used the r squared as a standard. The statistical neural network model and the long short-term memory networks model could predict COVID-19 confirmed cases and death cases, with the maximum r square score of 0.99. Moreover, the r squared of using artificial neural network model predict death cases was 0.92. To test the model performance, seven papers used the AUC score, and three papers used accuracy. The best AUC score of predicted mortality was 0.963 using Lasso model. The highest AUC score of severity prediction using random forest was 0.93. The highest accuracy rate was 98% on the three days mortality prediction using the time-aware long short-term memory neural network.

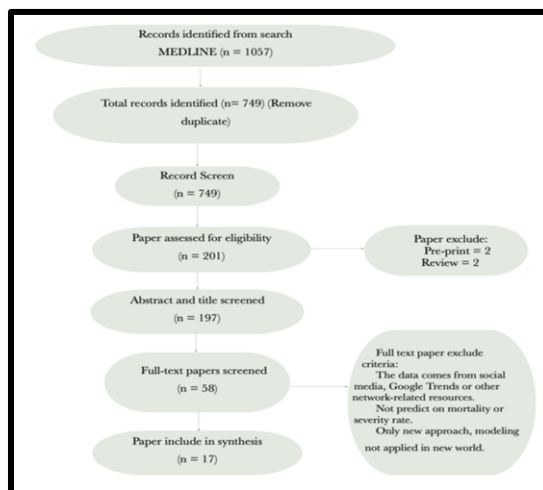


Figure 1. Paper selection process

4. Discussion

The scoping review is a good method to address undefined and broad questions. The majority of papers predicted mortality. In the study of mortality, neural network method was the most popular mortality prediction method. And the Lasso model generated the highest AUC score. Four papers predicted severity, three of which used random forest model, and this model had the best model performance. Five papers predicted daily confirmed cases and death cases. Neural network and random forest were the equally most common used models. Statistical neural network under this case had better fitness. Overall, the random forest model was the most popular model, and the Lasso model always tended to have a better performance. Due to the differences in datasets, it was difficult to compare models across datasets. Besides, we only chosen papers included AI models in forecasting cases, mortality, and severity. This topic was relatively narrow. In the future study, we will design a method to measure model performance and also extend our topics to other predictors that could be predicted by AI.

5. Conclusion

We conducted a scoping review of AI methods that were used to predict COVID-19 cases. We found that the tree-based model was the universal method. And most papers used official resources or EHR datasets. Asia was the most analyzed continent. Thus, scoping review is a valid method to explore the current trends of the use of AI in predicting COVID-19. Future study is warranted.

References

- [1] An C, Lim H, et al. Machine learning prediction for mortality of patients diagnosed with COVID-19: a nationwide Korean cohort study, *Scientific reports* vol (2020), 10,1 18716. 30.
- [2] Arksey H, O'Malley L, Scoping studies: towards a methodological framework, *Int J Soc Res Methodol* (2005), 8(1):19-32.
- [3] Assaf D, Gutman Y, et al. Utilization of machine-learning models to accurately predict the risk for critical COVID-19, *Intern Emerg Med* (2020),15(8):1435-1443.
- [4] Braga MB, Fernandes RDS, et al, Artificial neural networks for short-term forecasting of cases, deaths, and hospital beds occupancy in the COVID-19 pandemic at the Brazilian Amazon, *PLoS One* (2021), 16(3):e0248161.
- [5] Chatterjee A, Gerdes MW, Martinez SG. Statistical Explorations and Univariate Timeseries Analysis on COVID-19 Datasets to Understand the Trend of Disease Spreading and Death, *Sensors (Basel)* (2020), 20(11):3089.
- [6] da Silva CC, de Lima CL, et al. Covid-19 Dynamic Monitoring and Real-Time Spatio-Temporal Forecasting. *Front Public Health* (2021), 9:641253.
- [7] Dhamodharavadhani S, Rathipriya R, Chatterjee JM. COVID-19 Mortality Rate Prediction for India Using Statistical Neural Network Models, *Front Public Health* (2020), 8:441.
- [8] Dong E, Du H, Gardner L. An interactive web-based dashboard to track COVID-19 in real time, *Lancet Infect Dis* (2020), [https://doi.org/10.1016/S1473-3099\(20\)30120-1](https://doi.org/10.1016/S1473-3099(20)30120-1).
- [9] Heldt FS, Vizcaychipi MP, et al. Early risk assessment for COVID-19 patients from emergency department data using machine learning, *Sci Rep* (2021), 11(1):4200.
- [10] Pham MT, Rajić A, et al. A scoping review of scoping reviews: advancing the approach and enhancing the consistency, *Res Synth Methods* (2014), 5(4):371-385.
- [11] Rechtman E, Curtin P, et al. Vital signs assessed in initial clinical encounters predict COVID-19 mortality in an NYC hospital system, *Sci Rep* (2020), 10(1):21545.
- [12] Saba T, Abunadi I, Shahzad MN, Khan AR. Machine learning techniques to detect and forecast the daily total COVID-19 infected and deaths cases under different lockdown types, *Microsc Res Tech* (2021), 84(7):1462-1474.
- [13] Stachel A, Daniel K, et al. Development and validation of a machine learning model to predict mortality risk in patients with COVID-19, *BMJ Health Care Inform* (2021), 28(1):e100235.
- [14] Sun C, Hong S, Song M, Li H, Wang Z. Predicting COVID-19 disease progression and patient outcomes based on temporal deep learning, *BMC Med Inform Decis Mak* (2021), 21(1):45.
- [15] Vaid A, Somani S, et al. Machine Learning to Predict Mortality and Critical Events in a Cohort of Patients With COVID-19 in New York City: Model Development and Validation, *J Med Internet Res* (2020), 22(11):e24018.
- [16] Watson GL, Xiong D, et al. Pandemic velocity: Forecasting COVID-19 in the US with a machine learning & Bayesian time series compartmental model, *PLoS Comput Biol* (2021), 17(3):e1008837.
- [17] Wu G, Yang P, et al. Development of a clinical decision support system for severity risk prediction and triage of COVID-19 patients at hospital admission: an international multicentre study, *Eur Respir J* (2020), 56(2):2001104.
- [18] Yaşar Ş, Çolak C, Yoloğlu S. Artificial Intelligence-Based Prediction of Covid-19 Severity on the Results of Protein Profiling, *Comput Methods Programs Biomed* (2021), 202:105996.
- [19] Yu L, Halalau A, et al. Machine learning methods to predict mechanical ventilation and mortality in patients with COVID-19, *PLoS One* (2021), 16(4):e0249285.
- [20] Zawbaa HM, El-Gendy A, et al. A study of the possible factors affecting COVID-19 spread, severity and mortality and the effect of social distancing on these factors: Machine learning forecasting model, *Int J Clin Pract* (2021), 75(6):e141116.