Identifying Features for COVID-19 Hospitalization Risk Prediction

José J. Casado Garrido, Gabriel A. Cruz, Zhiyi Dong, Blake Anderson, May D. Wang

Abstract— **The recent development of monoclonal antibody therapies (mAb) promises to reduce the viral load in individuals that contract COVID-19. Predicting if a patient may need to be hospitalized is of importance in providing these mAbs. In this study, we developed an Artificial Neural Network (ANN) to assess a patient's risk to require medical assistance using Electronic Health Records obtained prior to COVID-19 diagnosis. We used Maximum Relevance Minimum Redundancy (MRMR) and Principal Component Analysis (PCA) to identify predictive features, while reducing the dimensionality of the feature space. We are able to achieve 82.59% accuracy, 74.74% sensitivity and 0.43 F1-Score to predict the patient risk suffering a catastrophic event such as ICU admission, ventilation support, and death. Our model's performance is comparable to that of recently published COVID-19 related neural network models.**

I. Introduction

The prevention of COVID-19 infection has improved with the advent of vaccines, but early treatment is still necessary for precision staging and hospitalization risk. Monoclonal antibody therapies are being rapidly developed and authorized for early treatment [1]. Tools for risk prediction, like artificial neural network (ANN) based models, have demonstrated high accuracy and sensitivity in mortality risk prediction, including systems for COVID-19 prognosis [2-3]. In this study, we analyzed a set of clinical biomarkers through machine learning techniques and built an artificial neural network to assess a patient's risk to require medical assistance and facilitate clinical decisions at the time of diagnosis. The dataset consists of a series of electronic health records (EHR) acquired prior to COVID-19 diagnosis, provided by Dr. Blake Anderson from Emory University.

II. Methodology and System Design

The proposed workflow of our system implementation is shown in Fig. 1. The complete dataset contained 17,806 subjects from Emory University in Atlanta, Georgia USA. 71 variables were collected, and variables with more than 50% missing values were dropped. From the set of remaining features (n=63), 13 were continuous variables (imputed with k-nearest neighbor, k = 5) and 50 were categorical variables (one-hot encoded). The neural network consisted in four main layers, including an input layer, two hidden layers (fully connected layers) and an output layer (sigmoid activation).

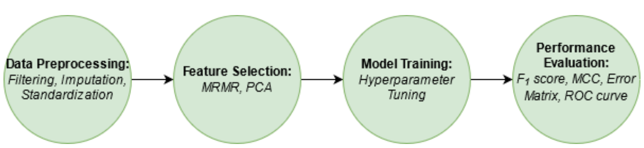

Figure 1. Proposed System Workflow with four main steps: 1) Data Preprocessing, 2) Feature Selection, 3) Model Training, and 4) Performance Evaluation

Hyperparameter tuning was conducted through grid search and 5-fold cross-validation.

III. Results

The most clinically intriguing predictive features from our MRMR analysis were chronic heart failure history and use of smoking cessation agents. Our best performing model obtained a prediction accuracy of 82.59%, specificity of 83.35%, sensitivity of 74.74% and F1-Score of 0.43, which are all comparable to recently published models whose accuracies range 86.25%-93.75% and specificities 85.94%-90.65% [2-3].

IV. Conclusion

We have developed a neural network capable of predicting potential patient hospital admission. This prediction is based on pre-diagnosis EHR which can assist and facilitate clinical decisions at the time of diagnosis for the usage of mAbs. Our final model has a prediction accuracy comparable to that of recently published models and we believe it is capable of use as a tool to predict patient hospitalization risk.

Acknowledgment

We would like to thank Mr. Felipe Giuste and Ms. Wenqi Shi for their help during the initiation and progress of this project.

References

[1] U.S. Food and Drug Administration, Center for Drug Evaluation and Research. "Bamlanivimab and Etesevimab approval letter". FDA. https://www.fda.gov/media/145801/download (accessed March 3, 2021)

[2] A. Abdulaal, A. Patel, E. Charani, S. Denny, N. Mughal, and L. Moore, "Prognostic Modeling of COVID-19 Using Artificial Intelligence in the United Kingdom: Model Development and Validation," Journal of Medical Internet Research, vol. 22, no. 8, 2020.

[3] M. Pourhomayoun and M. Shakibi, "Predicting Mortality Risk in Patients with COVID-19 Using Artificial Intelligence to Help Medical Decision-Making," MedRxiv, 2020.

Target Outcome	**Total Positive Subjects**	
ICU Adm.	9.03 %	
Ventilation	5.51 %	
Death	3.09 %	
Feature	**Total Positive Subjects**	**MRMR Score**
CHF	11.29 %	0.03391
Male	43.21 %	0.03163
Smoking Cessation Agents	0.90 %	0.00189

Table 1 Summary of Target and MRMR top features' distribution in dataset

J. J. Casado is with the Department of Biomedical Engineering, Georgia Institute of Technology, Atlanta, GA 30332 (e-mail: jgarrido7@gatech.edu).

G. A. Cruz is with the Department of Bioinformatics, Georgia Institute of Technology, Atlanta, GA 30332(e-mail: gcruz8@gatech.edu).

Z. Dong is with the Department of Biomedical Engineering, Georgia Institute of Technology,Atlanta,GA 30332(e-mail: zhiyidong3@gatech.edu).

Blake Anderson is with the Division of General Medicine, The Emory Clinic, Atlanta,GA 30322 (e-mail: blake.john.anderson@emory.edu)

May D. Wang, awarded with The Wallace H. Coulter Distinguished Faculty Fellowship, is with the Department of Biomedical Engineering, Georgia Institute of Technology, Atlanta, GA 30332 (e-mail: maywang@gatech.edu).