



Cardiovascular Disease Risk Improves COVID-19 Patient Outcome Prediction

Diego Machado Reyes¹ , Hanqing Chao¹ , Fatemeh Homayounieh² ,
Juergen Hahn¹ , Mannudeep K. Kalra² , and Pingkun Yan¹  

¹ Department of Biomedical Engineering, Rensselaer Polytechnic Institute Troy, New York, USA

{machad, chaoh, hahnj, yanp2}@rpi.edu

² Department of Radiology, Massachusetts General Hospital, Boston, MA, USA
mkalra@mgh.harvard.edu

Abstract. The pandemic of coronavirus disease 2019 (COVID-19) has severely impacted the world. Several studies suggest an increased risk for COVID-19 patients with underlying cardiovascular diseases (CVD). However, it is challenging to quantify such risk factors and integrate them into patient condition evaluation. This paper presents machine learning methods to assess CVD risk scores from chest computed tomography together with laboratory data, demographics, and deep learning extracted lung imaging features to increase the outcome prediction accuracy for COVID-19 patients. The experimental results demonstrate an overall increase in prediction performance when the CVD severity score was added to the feature set. The machine learning methods obtained their best performance when all categories of the features were used for the patient outcome prediction. With the best attained area under the curve of 0.888, the presented research may assist physicians in clinical decision-making process on managing COVID-19 patients.

Keywords: COVID-19 · Machine learning · Cardiovascular disease · Chest CT · Severity score

1 Introduction

Coronavirus disease 2019 (COVID-19), caused by severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), has had a global impact like no other pandemic or disease in modern times. COVID-19 patients present a clinical picture similar to the other two coronaviruses that in previous years have caused pandemic diseases (SARS-CoV, MERS-CoV) [10]. Patients initially manifest a respiratory infection that can lead to viral pneumonia, which can culminate into acute respiratory distress syndrome (ARDS) [15, 20]. Early prediction of disease

This work was partially supported by National Institute of Biomedical Imaging and Bioengineering (NIBIB) under award R21EB028001 and National Heart, Lung, and Blood Institute (NHLBI) under award R56HL145172.

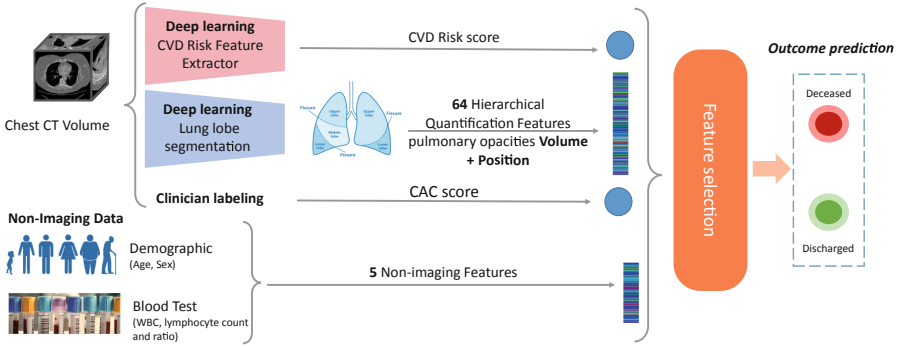


Fig. 1. The proposed framework utilizing both imaging and non-imaging features including lung severity and CVD risk for COVID-19 outcome prediction.

progression and severity assessment in at-risk patients can help determine and plan healthcare needs related to hospital admission, mechanical ventilation, and intensive care monitoring.

Several approaches have been taken to implement machine learning methods for COVID-19 patient outcome prediction [14]. For example, Chao et al. [2] developed a framework integrating various features extracted from chest CT scans with patient demographic information, vital signs, and laboratory blood exams to assess disease severity and predicting intensive care unit (ICU) admission. In Tang et al. [18], a random forest model was trained using quantitative features from the lungs to assess the severity of the disease. Said features were obtained from a deep learning-based chest CT analysis tool. Other approaches have focused on developing similar deep learning-based tools to extract relevant information from chest CT-scans and then use the obtained features for disease severity assessment. Fang et al. [4] employed a deep learning approach to segment the lung lobes and pulmonary opacities and computed severity scores from those. Afterward, random forest, SVM, and logistic regression models were trained with the severity scores to predict the patient outcome. However, the existing works mainly focus on the imaging features from the lungs. COVID-19 severity prognosis has been associated with multiple comorbidities through systematic reviews and meta-analyses and retrospective cohorts [6–8, 12, 17, 19]. Among the comorbidities, cardiovascular disease (CVD) was found to have a risk ratio ranging from 2.25 to 3.15 [6, 12, 17] and a prevalence ranging from 5.8% to 25% [7, 8, 19]. It is thus important to take CVD risk into account when assessing patient prognosis outcomes.

In this work, we combine features of CVD analysis, chest CT image, and non-imaging data to predict the outcome of COVID-19 patients through machine learning, as shown in Fig. 1. By leveraging the correlation between CVD and COVID-19 severity, our model can more accurately separate patients with high mortality risks. The rest of this paper is organized as follows. First, the datasets

Table 1. Demographic statistics of the dataset.

Features	Deceased	Discharged
Gender (M:F)	55% : 45%	55% : 45%
Age (years)	72.9 \pm 12.5	64.5 \pm 18
WBC	2902 \pm 3940	4148 \pm 2799
Lym	341 \pm 450.6	849.4 \pm 676.8
Lym ratio (%)	14.6 \pm 12.9	19.2 \pm 9.9
CAC score (No:Mild:Moderate:Severe)	23% 38% 5% 34%	43% 28% 9% 20%
CVD score	0.511 \pm 0.13	0.424 \pm 0.15

employed are presented, then an exploratory analysis of the variables is introduced. Next, the feature selection and machine learning methods used for outcome prediction are discussed. Then the experimental results are presented, which show the performance improvement with including the CVD severity score. Finally, the findings and conclusions of this paper are discussed.

2 Materials and Methods

2.1 Datasets

The data employed in this work was obtained from two hospitals, Firoozgar Hospital (Tehran, Iran), and Massachusetts General Hospital (MGH) (Boston, MA, USA). These datasets were comprised of patients' demographics (age, sex), laboratory blood tests (white blood cell count, lymphocyte count, lymphocyte ratio), coronary artery calcification (CAC) score, and the outcome of the patients (discharged, deceased). Moreover, non-contrast chest CT scans without intravenous contrast injection were provided. The Firoozgar Hospital dataset included data from 113 patients, while the MGH dataset was comprised of 125 patients. Both datasets were combined into one larger dataset comprised of 238 patients. CT scans were manually inspected for lines, tubes, and imaging artifacts, the ones including such artifacts were removed from the dataset. A resulting total of 208 patients were available for outcome prediction. From the combined dataset, 108 patients were discharged and the remaining 100 were deceased. Other relevant demographic statistics of the patient population are presented in Table 1.

2.2 Outcome Prediction Methods

In order to perform the outcome prediction, both the imaging and metadata features were divided into three categories. 1) First, all the non-imaging features from demographic, vital signs, and blood examination (DVB) were placed together. These features being age, sex, white blood cell count, lymphocyte count, and lymphocyte ratio. 2) The second category consisted of 64 hierarchical

lobe-wise quantification features (HLQ) calculated with the approach introduced in [2]. The five lung lobes and pulmonary opacities are first segmented by a deep neural network [5]. Then, in 8 regions of interest (ROIs) (whole lung, left & right lung, and 5 lung lobes), the volumes of pulmonary opacities grouped under 4 Hounsfield unit (HU) ranges and their ratios to the corresponding ROIs are calculated ($8 \times 4 \times 2 = 64$). 3) Furthermore, a third category was the collection of CVD severity scores that were calculated from the chest CT-scans.

In this paper, we studied two kinds of CVD scores. One is the deep learning-based CVD risk score automatically calculated based on the model introduced in [3]. The other one is the CAC score manually assigned by radiologists. A radiologist from MGH categorized all 208 CT images into 4 CAC levels, i.e., normal: no calcification, minimal: calcification less than 1/3 of the length of coronary arterial length, moderate: calcification over 1/3 to 2/3 of the coronary arteries, and heavy: calcification over 2/3 of the arterial length.

As a first step, an exploratory analysis was performed to examine the relevance of the CVD risk and CAC scores towards the differentiability of the patients. Next, the sets of non-imaging (DVB) and imaging (HLQ) features were used as inputs for machine learning models to predict the patient outcome. In further detail, experiments were run with the non-imaging features (DVB) and lung imaging features (HLQ) individually and together to establish a baseline. Afterward, the CVD or the CAC scores were added in the following experiments to check for performance improvement. Three well-known machine learning algorithms – random forest classifier, logistic regression, and support vector machine (SVM) – were used to perform the outcome prediction. Several configurations of their parameters were tested and the one with the best performance was chosen.

2.3 Exploratory Statistical Analysis

While CVDs have been widely associated with increased risk ratios and prevalence among the patients with fatal outcomes [6–8, 12, 17, 19], a key initial step was to test if the deep learning-based CVD risk scores and the CAC scores showed significant differences between the patients with fatal outcome and the discharged patients. First, the normality of the distribution of the CVD risk scores and CAC scores of each group were tested through a Shapiro-Wilk test ($\alpha = 0.05$). Due to the non-normal distribution of the CVD scores, Mann-Whitney U tests were employed to determine significance between the CVD risk scores and CAC scores of two groups ($\alpha = 0.05$). Second, to enhance the explainability of the developed model, an exploratory factor analysis [16] was performed to examine the relevance of the features and further granularities in the data. First, the loadings for each factor were calculated to determine relations between the features present in the datasets. Second, groups of features were obtained for each factor; thus providing an improved understanding of which features share the most similar information between each other, and especially to the features of interest – the CVD risk score and the CAC scores. The factor analysis was implemented through the FactorAnalyzer python library, using a varimax rotation matrix. Bartlett’s sphericity test [1] and Kaiser-Meyer-Olkin test [9] were used to test for adequacy of applying factor analysis for the datasets.

2.4 Feature Selection

As the ratio of available features to patients was not favorable, two feature reduction strategies were implemented depending on the machine learning method. For the SVM and logistic regression, a leave-one-out strategy was implemented to select the best set of features. The second feature reduction approach ranks features based on their Gini importance [11]. Because of the randomness of Random Forest, the Gini importance of each feature might vary in different runs. To alleviate such randomness, we calculate the features' Gini importance 100 times with different random seeds. Each time the ranks of each feature were recorded. The final rank was determined by sorting the total summed rank of each feature. Based on the final ranking, the model was trained and tested multiple times on the top $K \in [1,100]$ features and its performances were recorded. The combination of features that rendered the highest AUC was found. Furthermore, due to the small sample size a traditional train-test split was not ideal for independent assessment; therefore, a 5-fold cross-validation was implemented for all experiments to evaluate the methods performances differences. Finally, significance in performance increase was tested using Mann–Whitney U tests.

3 Results

3.1 Exploratory Analysis

The Mann–Whitney U tests revealed a significant difference between the CVD risk scores from the outcome groups in the Iran, MGH and combined datasets; while it revealed no significant difference between the CAC scores in any of the datasets. Therefore, for the individual datasets and the combined version of these, it can be understood that the CVD risk score is a valuable feature to differentiate patients' possible outcomes.

The factor analysis revealed granularities in the dataset with especial interest in the CVD risk and CAC scores. For the Iran data set, the factor analysis indicated that the CVD risk score – labeled as softmax_pred on the right end of the first subplot of Fig. 2 – has a strong correlation with age, and a mild correlation with features such as white blood cell count and the infection volume and ratios in lobes 2 and 3. These results are expected, as patients of older age have a higher prevalence of cardiovascular disorders [13]. The relation between the CVD risk score and the lung features could be explained in part due to the closeness of the heart and lobe 2 (middle lobe of the right lung). In the low dose CT-scans employed for the HLQ and CVD risk score features calculation parts of the lung and heart, in their respective cases, could appear in the field of view of the neural networks producing the respective quantitative measurements. On the other hand, for the MGH data set relationships were found between the CVD risk score and the infection volumes and ratios of multiple sections from the lobes 3 and 4. This can be seen on the right plot of Fig. 2. Similar to the Iran dataset factor analysis results, lobe 4 (superior lobe of the left lung) shares boundaries with the left atrium and aorta, which are major areas of interest

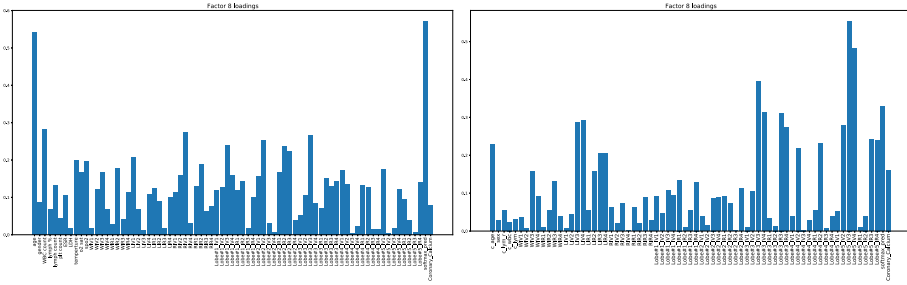


Fig. 2. Factor loadings plots for Iran (left) and MGH (right) datasets

for the CVD risk estimator network as calcification in the aorta is a key driver of this score. Furthermore, the factor loadings also showed a mild correlation between the CVD risk score and the age following a similar pattern as the Iran dataset results.

3.2 Inclusion of CVD Risk Score in Patient Classification

As mentioned before, multiple parameter configurations were tested for the implemented machine learning methods. All three machine learning methods used Scikit-Learn package implementation. For the parameters used, in the case of the SVM, the best results were accomplished when using a RBF kernel. As for the logistic regression, the best performance was achieved by using a stochastic average gradient descent solver. Furthermore, the random forest classifier obtained its best results when setting the number of estimators to 300.

Table 2 reports the results from the 5-fold cross-validation using each best-performing method and features used as input for prediction. Most of the experiments showed an increased performance, reflected on the AUC after the CVD severity score was included on the feature set. Nevertheless, after testing for significance with a Mann–Whitney U test, 5 were found to be significant, mainly improving when all features were introduced. Moreover, while the inclusion of the CAC scores resulted in increased performance in some of the experiments, mainly in the HLQ+CAC category, none of the improvements were found to be significant by the Mann–Whitney U test. The lack of significant improvement when including the CAC score as a feature concurs with the results from the exploratory analysis in which no significant differences were found in the distributions of the CAC scores between the deceased and discharged patients.

Table 2 and Fig. 3 show consistently increased performance for all the three models when the CVD severity score was added to the DVb+HLQ configuration, with the increase on the random forest and the logistic regression being validated as significant by the Mann–Whitney U test. Moreover, this feature configuration also achieves the best performances on each method, with the highest performing experiment being accomplished by the logistic regression using all 3 sets of features. Furthermore, similar increases in performance can be seen for the

4 Discussion and Conclusion

Although the majority of patients with COVID-19 infection are either asymptomatic or present minor symptoms, most healthcare systems from the most to the least developed countries in the world were overwhelmed, because the pandemic nature of COVID-19 pneumonia results in the substantial volume of patients with COVID-19 infection. Patients present with rapid deterioration requiring hospital admission and mechanical ventilation. Therefore, understanding and predicting patients at risk of severe and potentially life-threatening infection is key to resource planning and prognosis prediction. One commonly recognized comorbidity associated with adverse COVID-19 outcomes is the presence of cardiovascular disease. In this work, we explored the inclusion of a deep learning-based CVD severity score to improve the prediction accuracy of patient outcome.

Some key results from the experiments showed a significant increase in performance, which was seen in the random forest classifier after including the CVD severity score as a feature for prediction. Concurrently, a significant performance improvement was observed in the logistic regression when using all features and the CVD severity score with the latter model being the best performing of all configurations. The results confirmed that the inclusion of the CVD severity score benefits the models by increasing their ability to predict the patients' outcome. Additionally, while the sample size used was relatively small for machine learning models, the presented results were strengthened as the experiments were validated on a combination of the datasets. Thus, showing the relevance of including the CVD severity score for COVID-19 patient outcome prediction regardless of the patient population. This concurs with the clinical studies, presented in the introduction, that show cardiovascular diseases as relevant comorbidities for COVID-19 patients and suggest to consider cardiovascular diseases when assigning treatment to a COVID-19 patient.

Moreover, while several of the tested features and model configurations resulted in significant performance increases, these increases are not overwhelmingly high. As discussed in the exploratory analysis results, this could be due to shared information across the multiple variables employed for the outcome prediction as shown by the factor loadings. Moreover, while the manual CAC score was also tested, the CVD severity score outperformed the CAC score results across all experiments as expected from the non-significant differences in the distributions of the CAC scores. This could be due to mainly two reasons, one being, the CVD severity score might include additional relevant information of the presence of the comorbidity extracted from the CT-scans, rather than just the calcification of the coronary artery. Furthermore, as CAC is a categorical variable, contrary to the continuous values of the CVD severity score, the machine learning methods could be hindered in their prediction performance.

We hope that this work can shed more light on assigning a better prognosis for COVID-19 patients, as well as for improved clinical resource management. The main takeaway is to add evidence into a series of clinical and computational investigations, as discussed in the introduction, that indicate a significance on the impact of cardiovascular diseases as a risk factor for COVID-19 patients.

References

1. Bartlett, M.S.: Properties of sufficiency and statistical tests. *Proc. Roy. Soc. London Ser. A Math. Phys. Sci.* **160**(901), 268–282 (1937)
2. Chao, H., et al.: Integrative analysis for COVID-19 patient outcome prediction. *Med. Image Anal.* **67**, 101844 (2020)
3. Chao, H., et al.: Deep learning predicts cardiovascular disease risks from lung cancer screening low dose computed tomography. [arXiv:2008.06997](https://arxiv.org/abs/2008.06997) [cs, eess] (2020)
4. Fang, X., et al.: Association of AI quantified COVID-19 chest CT and patient outcome. *Int. J. Comput. Assist. Radiol. Surg.* **16**(3), 435–445 (2021). <https://doi.org/10.1007/s11548-020-02299-5>
5. Fang, X., Yan, P.: Multi-organ segmentation over partially labeled datasets with multi-scale feature abstraction. *IEEE Trans. Med. Imaging* **39**(11), 3619–3629 (2020). <https://doi.org/10.1109/TMI.2020.3001036>
6. Figliozzi, S., et al.: Predictors of adverse prognosis in COVID-19: a systematic review and meta-analysis. *Eur. J. Clin. Invest.* **50**(10) (2020)
7. Guan, et al.: Clinical characteristics of coronavirus disease 2019 in China. *N. Engl. J. Med.* **382**(18), 1708–1720 (2020). <https://doi.org/10.1056/NEJMoa2002032>
8. Jain, V., Yuan, J.-M.: Predictive symptoms and comorbidities for severe COVID-19 and intensive care unit admission: a systematic review and meta-analysis. *Int. J. Public Health* **65**(5), 533–546 (2020). <https://doi.org/10.1007/s00038-020-01390-7>
9. Kaiser, H.F., Rice, J.: Little jiffy, mark iv. *Educ. Psychol. Measur.* **34**(1), 111–117 (1974). <https://doi.org/10.1177/001316447403400115>
10. Liu, J., et al.: A comparative overview of COVID-19, MERS and SARS: review article. *Int. J. Surg.* **81**, 1–8 (2020). <https://doi.org/10.1016/j.ijso.2020.07.032>
11. Moore, D.H.: Classification and regression trees, by Leo Breiman, Jerome H. Friedman, Richard A. Olshen, and Charles J. Stone. Brooks/Cole Publishing, Monterey, 1984, 358 Pages, \$27.95. *Cytometry* **8**(5), 534–535 (1987)
12. Nishiga, M., Wang, D.W., Han, Y., Lewis, D.B., Wu, J.C.: COVID-19 and cardiovascular disease: from basic mechanisms to clinical perspectives. *Nat. Rev. Cardiol.* **17**(9), 543–558 (2020). <https://doi.org/10.1038/s41569-020-0413-9>
13. North, B.J., Sinclair, D.A.: The intersection between aging and cardiovascular disease. *Circ. Res.* **110**(8), 1097–1108 (2012)
14. Shi, F., et al.: Review of artificial intelligence techniques in imaging data acquisition, segmentation, and diagnosis for COVID-19. *IEEE Rev. Biomed. Eng.* **14**, 4–15 (2021). <https://doi.org/10.1109/RBME.2020.2987975>
15. Siddiqi, H.K., Mehra, M.R.: COVID-19 illness in native and immunosuppressed states: a clinical-therapeutic staging proposal. *J. Heart Lung Transplant.* **39**(5), 405–407 (2020). <https://doi.org/10.1016/j.healun.2020.03.012>
16. Spearman, C.: “general intelligence,” objectively determined and measured. *Am. J. Psychol.* **15**(2), 201 (1904)
17. Ssentongo, P., Ssentongo, A.E., Heilbrunn, E.S., Ba, D.M., Chinchilli, V.M.: Association of cardiovascular disease and 10 other pre-existing comorbidities with COVID-19 mortality: a systematic review and meta-analysis. *PLOS ONE* **15**(8), e0238215 (2020). <https://doi.org/10.1371/journal.pone.0238215>
18. Tang, Z., et al.: Severity assessment of coronavirus disease 2019 (COVID-19) using quantitative features from chest CT images. [arXiv:2003.11988](https://arxiv.org/abs/2003.11988) [cs, eess] (2020)

19. Yang, J., et al.: Prevalence of comorbidities and its effects in patients infected with SARS-CoV-2: a systematic review and meta-analysis. *Int. J. Infect. Dis.* **94**, 91–95 (2020). <https://doi.org/10.1016/j.ijid.2020.03.017>
20. Zhou, F., et al.: Clinical course and risk factors for mortality of adult inpatients with COVID-19 in Wuhan, China: a retrospective cohort study. *Lancet* **395**(10229), 1054–1062 (2020). [https://doi.org/10.1016/S0140-6736\(20\)30566-3](https://doi.org/10.1016/S0140-6736(20)30566-3)