JMAI JOURNAL OF MEDICAL ARTIFICIAL INTELLIGENCE
AN OPEN ACCESS JOURNAL FOCUS ON MEDICAL ARTIFICIAL INTELLIGENCE

Peer Review File

Article Information: Available at http://dx.doi.org/10.21037/jmai-20-47

Reviewer A

The article "Building a COVID-19 Vulnerability Index" is a noble contribution to identify those individuals who are at the greatest risk for severe complications due to COVID-19 pandemics. It also provides some further analysis of risk prediction and attempts to extract the most important results.
My overall impression of the document is that it is thorough, fairly complete, no major flaws, reflects a regulatory-based perspective and important. It follows scientific method, clinically relevant, analysis is technically logical and most importantly discussion and conclusions flow logically from data.
However, I am concerned about the Clarity of Presentation, appropriate formation for journal and readability for readers who are not expert in medical filed.
Here some general remark to concert:
1. The author should present the information containing in different Figure and Table clearly (Specially for Figure 1 and Table 5) as well as for the section '3. Results and Model Interpretation' and '3.1. Validation' to increase the readability and make the article more understandable for readers who is not expert in this fields.

Reply 1: We have updated the table to make the data presentation clearer and more understandable for readers with less expertise.
Changes in the text: We have updated the table headings for table 5 and the captions for figures and tables throughout.

2. According to the JMAI Instruction for Author, three to five key words should be supplied below the Abstract recommended by the US National Library of Medicine's Medical Subject Headings (MeSH).

Reply 2: We have added keywords
Changes in the text: We have added the following 3 key words: ***COVID-19, artificial intelligence, vulnerable populations***

3. According to the JMAI Instruction for Author, references should be cited using Arabic numerals in round brackets in which they appear consecutively. Use [e.g., "previously identified in humans (1)"; "cardiac arrest, and death (17,18)"] instead of [e.g., "previously identified in humans [1]"; "cardiac arrest, and death [17] [18]"].

Reply 3: Changes have been made to comply with citation instructions
Changes in the text: All references have been updated to reflect round brackets

4. If I was not wrong, I think author missed the percentage (%) sign in the end of section '2.2.3 Combined Population'.

Reply 4: Typo fixed
Changes in the text: Percentage sign has been added to the last word in the section.

Reviewer B

The paper proposes a set of AI models to identify individual susceptible to COVID-19 by using proxy indicators in the absence of primary data sources. In particular, the authors presented results for 3 models predicting the complications, with a tradeoff between increasing predictive effectiveness and decreasing ease of implementation. The authors provided a good narrative of the present COVID-19 situation and explained the necessity for identifying COVID-19 risks.

In general, the solution approach appears appropriate for mitigating the current COVID-19 situation and detecting probable high-risk individuals. However, the merits of the paper are overshadowed by few weaknesses and unappealing facets of its presentation.

- I am not sure why the authors intended to use older data from 2015 and 2016 in the CMS dataset (LDS) to detect COVID-19, which was just recently discovered. It appears that the training data did not consist of true COVID-19 cases.

Reply 1: The 2015 & 2016 data sets were all we had access to at the time the work was completed. Although this was data from before COVID-19 was discovered, our models were built using a proxy endpoint (respiratory infection) that has not changed significantly from 2015.

- It is understandable that the availability of COVID-19 data may have been limited at the time of writing. Perhaps the authors can update the paper with newly available datasets from recent times and conduct new experiments.

Reply 2: We do not have access to a COVID-19 data set of sufficient size and quality to replicate the models at this time. Although such a data set could be assembled in theory given the volume of COVID-19 cases, we do not have access to one and amassing such a data set is a significant effort and beyond the scope of this research.

- Limiting the test and training dataset to Medicare population reduces the papers credibility since they only represent less than a fifth of the US population (https://www.statista.com/statistics/200962/percentage-of-americans-covered-by-medicare/). The authors should not only focus on Medicare members but the whole population in general.

Reply 3: We used a data set from Healthfirst which contained younger individuals and combined that with the Medicare data set to get a representative sampling of the general population. This sample is not perfect and we discuss some of the issues with this approach in a new section of the text.
Changes in text: We have added a paragraph to 2.2.3 explaining how the population we

used differs from the overall US population.

-The "Survey Model" is a rather simple model based on the basic logistic regression. I do not quite see the importance of including this particular model in this paper. There are several reliable solutions available today that takes a survey from users and conclude on the risk factors for COVID-19. Two examples are IBM Return to Workplace Advisor (https://www.itpro.com/marketing-comms/communications/356132/watson-works-helps-employers-make-informed-decisions-about) and Buoy Health (https://www.buoyhealth.com/symptom-checker/?configuration=ma_covid&concern=coronavirus).

Reply 4: The survey model is included because it demonstrates the tradeoff between model complexity and accuracy. While other tools to evaluate risk factors exist, the survey fits in the context of this paper to demonstrate how those tools compare to alternate approaches that utilize more data. In specific reference to the models mentioned by the reviewer, the Buoy Health survey noted checks symptoms, but does not assess vulnerability once someone is infected. The IBM tool is not publicly available and was launched in June, well after our CV19 index survey was made available.

- Sections 1.1-1.3 are rather too short. They should be merged under Section 1. Introduction.

Reply 5: The sections have been merged.
Changes in the text: The Introduction section, which is Section 1, now combines the subsections in the initial submission.

-The elaboration of the term "ICD" as International Classification of Diseases in section 2.2.1 should appear earlier (i.e., in section 2.1).

Reply 6: Edit made.
Changes in the text: Expanding ICD-10 now appears in Section 2.1. Section 2.2.1 expands on ICD-9, when it is first used. The original elaboration, which is not where ICD-9 or ICD-10 was first used, now just uses the abbreviation.

- The details of the "Full Model" in section 2.3.3 are not provided. I suggest the authors at least provide a brief description of the mechanism or algorithm for the AI model incorporated in the "Full Model".

Reply 7: While we did note that the Full Model uses the same modeling algorithm as the Open Source Model in Table 3, this was not entirely clear in Section 2.3.3. The details of gradient boosted trees (XGBoost) are described in section 2.3.2 along with additional information on the full model.
Changes in the text: Section 2.3.3 expanded to specify the modeling method used by the Full Model.

-The title of Section 3 could be renamed to a more fitting title such as "Evaluation Criteria" or "Testing Methodology".

Reply 8: In the process of reformatting the paper to comply with journal guidelines this entire section has been renamed to "Results".
Changes in the text: Changed section header from "Results and Model Interpretation" to just "Results"

-The naming scheme for the models (i.e., survey model, open source model, full model) appear generic. They could be improved to better reflect the underlying AI algorithm/technique used.

Reply 9: Thanks for the comment. Since the primary differentiator for the models is the data used, we have renamed the models based on that. The new names for the models are "Survey Risk Factors Model", "Diagnosis History Model", and "Expanded Feature Model"
Changes in the text: Model names have been changed throughout.

-Few typos exist. I suggest the authors carefully revise the paper before submission. Some examples are:
i) Section 2.3.1 - "The purpose of the survey to to ⋯" ->" "The purpose of the survey is to ⋯" "
ii) Section 2.3.4 - "CLosedLoop" ->"ClosedLoop"

Reply 10: Typos fixed.
Changes in the text: The two typos found by the reviewer in addition to other typos after revision have been corrected.

Reviewer C
1- Please explain how the combination of the two mentioned datasets in your study can be a good representative of the US population? Is any statistical analysis done to show their resemblance?

Reply 1:
In section 2.2.3, we describe the factors that were considered when merging the two datasets, namely making sure the age distribution
Changes in the text:
i.) Section 2.2 "a demographic profile" -> "an age profile"
ii.) Section 2.2.3 -> Added paragraph 2, recognizing that other demographic profiles, particularly racial/ethnic makeup of New York is not representative of the whole of the United States.

2- Please explain what type of encoding is done on your categorical features.

Reply 2: We added this in.
Changes in the text: Added section 2.2.4 "Description of data pipeline"

3- Was there any specific reason to use logistic regression and XGBoost? Why not other predictive models?
Reply 3: Discussion of why XGBoost is well suited for this task is included in section 2.3.2. We acknowledge that the comparison of the relative performance of machine learning models is of general interest to the community, but the primary focus of this study was on comparing a range of models with different difficulty of implementation, since this was the practical determining factor for being able to quickly roll these models out in real world use.

Reply 3: We added this in.
Changes in the text: Section 2.3.1 expanded to include motivation for the use of the use of logistic regression

4- Please explain more how your algorithms are trained using the training set. How it is guaranteed that overfitting is prevented while measuring the outcomes on the testset?

Reply 4: We added this in.
Changes in the text: Section 2.3.4 added detailing cross validation

5- The data is not readily available for COVID-19 cases. Do you believe by proper imputation methods (for data shrinkage prevention), your model can be used for real-work COVID-19 cases to identify individuals at higher risks?

Reply 5: Yes, we believe that these models are suitable for risk stratifying health populations. Section 3.1 details how these models worked when considering the health outcomes of real COVID-19 patients.
Changes in the text: Section We have provided additional detail in section 3.1 on how the validation process demonstrates applicability to real world cases.