A machine learning approach for identification of gastrointestinal predictors for the risk of COVID-19 related hospitalization

General comments

In this revision, the authors tried to incorporate some of the comments and provided rebuttal for questions that required explanations. This revised version is better than the original article but there are still some problems. Such as, (a) language, abbreviation usage, and punctuations still require editing, (b) the graphs occupied unnecessarily large spaces and on two pages. These figures can be minimized to occupy a single page,, (c) the Figure captions are not explanatory enough as explained in the detailed comments below, and (d) the comments mentioned in the detailed comments below need attention.

Detailed comments:

Lines 176-177: "...O2 saturation, CRP, GMT, AST, Bilirubin…). The abbreviations are mentioned for the first time. These have to be interpreted.

Line 180: ATB is mentioned for the first time without interpretation.

Line 200: " 30 patients were …" Starting sentences with numbers are still present in the manuscript even though the authors corrected some of the paragraphs in the introduction.

Line 227: "... wasalso …" is written as one word. Consider correcting it.

Lines 227-231: The figures for ALS and AST are separate and it is difficult to compare. Perhaps combining the two graphs may provide clear information. Plus there is no figure for Bilirubin but in the statement, " There is no statistically significant difference in the levels of ALT(Figure 2) and Bilirubin."

Lines 247-252: Figure 5 has two dimensions dim1 (~14%) and dim2(~10%). These are to indicate the first pca and second pca respectively. Neither the explanation nor the Figure captions has sufficient explanation about the PCA values.

Lines 309-311: “Further, we focused on identifying the possible predicting factors for hospitalization in COVID-19 patients using advanced artificial intelligence techniques such as the Random Forest (RF) machine learning algorithm.” This sentence implies that the authors have used more than one artificial intelligence technique, while the whole manuscript is written with the application of one.

Lines 342-348: The variable importance curves have both positive and negative values. The author didn’t mention the implications of these values either in the manuscript in those lines or in the Figure captions.