Detailed response to **#REVIEWER 1:**

1. The main cohort used in this study was constructed by AGP. It involves hundreds of host parameters, such as diet and antibiotics, but no related work addressed these confounders. Many key factors have not been examined in this study as far as I can see. In particular, I saw no assessment of other factors known to alter the microbiome. Furthermore, authors are unaware of the age and demographic. If these subjects of specific dataset also significantly differ in age from the test subject's microbiome composition will be significantly different regardless of subtypes.

We appreciate the reviewer´s comments. To address this question, about the possible effects of different confounders, we have conducted an exploratory analysis. This analysis is presented as supplementary material in the new version of the manuscript (Figures S1 and S2).

Both train and test sets of the AGP were analysed. External validation data could not be included in the analysis due to missing clinical variables. A correlation analysis was performed between the model predictions and the most disease-related confounders. Specifically, the variables age, BMI, alcohol consumption, antibiotic and probiotics intake, appendix removal and sex were included in this analysis. In the train set, the variables that were found to be significant were: alcohol consumption, antibiotic and probiotic intake and appendix removal. In the test, only alcohol consumption was significant.

This analysis also aims to answer the age-related question formulated by the reviewer. No stratification by age was performed in this study, with the intention that the model generated could extract information without taking into account the age of the patients. As can be seen in the figures, our model is independent of patient age.

On the other hand, the reason for not including confounders in the model was the unavailability of such confounders in the external validation data. If cofounders had been included, the prediction of the model could not have been externally validated. Although we agree with the reviewer's assessment, we find it impossible to include these variables, as our priority is to validate the model externally. Fortunately, as seen in the present analysis, metagenomic variables are associated with certain clinical factors, so indirectly, such information is correlated with the predictions made by our model.

2. There have been significant developments in our understanding of determining different microbiome patterns of IBD in the last few years. The authors have ignored some important recent literature.

For example, published work have reported the comprehensive analysis of molecular profiles of host and microbial activity in IBD

Lloyd-Price, J., Arze, C., Ananthakrishnan, A. N., Schirmer, M., Avila-Pacheco, J., Poon, T. W.,& Huttenhower, C. (2019). Multi-omics of the gut microbial ecosystem in inflammatory bowel diseases. Nature, 569(7758), 655-662.

Franzosa, E. A., Sirota-Madi, A., Avila-Pacheco, J., Fornelos, N., Haiser, H. J., Reinker, S.,& Xavier, R. J. (2019). Gut microbiome structure and metabolic activity in inflammatory bowel disease. Nature microbiology, 4(2), 293-305.

Thank you for your suggestions. We have read both articles and indeed consider it necessary to include them as references in this new version of the manuscript.

3. There was no description of control(non-IBD), which criteria are considered?

Control patients were randomly selected from among those without the disease. Furthermore, we consider that this issue needs to be addressed in more depth. As noted in the response of reviewer 2, the AGP cohort is a very heterogeneous cohort, in which the selection of suitable controls is complicated. If only controls without any comorbidity were selected, it would be impossible to obtain a balanced subgroup of class-positive patients.

In order to address this aspect, one option is to train the model with the external validation data and validate the prediction in the AGP cohort. In this way, the selection of controls is much more appropriate and focused on the problem. This experiment was done but the results were not satisfactory. As expected, the performance on the train and test data are similar to those obtained in the original version, but failed to be validated in the AGP cohort, as this cohort is much noisier than the others. This is why we considered it appropriate to keep the same structure to train the model. Furthermore, we consider that one strength of this work is the model's ability to identify patients with IBD, among many others who may have co-morbidities.

Based on the reviewer's question, we thought it appropriate to add a new table to the manuscript where we describe the patients according to age, sex and presence of disease (see Table 1). In addition, in this new version, we have published all the code to ensure the reproducibility of the work, and you can see the step-by-step preprocessing of the patient selection process.

3. The results section should include the direct findings, information that does not present the direct findings or outcome of the study should be left out of this section. Results normally refer to direct answers to your research questions E.g. Line 218 “Figure 1D shows the performances achieved in both the 219 train and test sets. “ Line 235 “Figure 1C shows the variable importance of best model in the training set.” Readers would like to to understand the performance and variable results directly.

Thank you for this appreciation. We agree with what you have said, and have rewritten some paragraphs of the results.

1 Line 223 “Features identified by differential abundance and LDM do not show satisfactory performance during training, as it is shown in Figure ??D.” Correct Figure numbers

Thank you, we have fixed this typo in the new version.

2 Verb Tense was not consistent throughout the results section. The results section usually requires the past tense to detail the results obtained

Also in relation to question 3, we have carefully reviewed and corrected this section.

2. Statistical significance of the comparison should be included and annotated in figures (e.g. Fig2B)

Significance has been added in figure 2B.