

# Point-by-Point Response to Reviewer's Comments

Debang Ou

**Reviewer: Andrew Resnick, University of Connecticut**

Date: November 14, 2024

## Overview of Changes

I sincerely appreciate the constructive feedback provided by the reviewer and I have made the following major changes in response to the comments:

I added a detailed discussion on fairness in Section 4.2, explicitly addressing False Positive Rate (FPR) and False Negative Rate (FNR) as measures of fairness, and how I failed to further explore fairness in this study.

A new subsection titled "Assumptions of the Models" was added in Section 3.2 to outline the assumptions for each model.

A new subsection titled "Research Hypotheses" was introduced in Section 1, articulating explicit hypotheses for the models.

Below, we provide point-by-point responses to each comment, quoting the changes where applicable.

## Comment 1: Addressing Fairness in Models

**Reviewer Comment:** The reviewer noted that one of the initial research questions involved assessing fairness among the models used, but the actual evaluation of fairness was not comprehensively discussed in the paper.

**Response:** Thank you for pointing out this oversight. I added some explanation of fairness, and I also acknowledged the limitations of using only FPR and FNR across models and suggested future directions to incorporate broader fairness measures, such as demographic parity and calibration. The revised text in Section 4.2 now reads:

*"Limited discussion regarding fairness: We evaluated fairness by comparing FPR and FNR across different models instead of comparing FPR and FNR across subgroups. This comparison provides a basic understanding of fairness by revealing differences in misclassification rates, but a thorough fairness evaluation requires analyzing these metrics across subgroups and including other fairness measures like demographic parity or calibration, which are infeasible in this study."*

## Comment 2: Discussion of Model Assumptions

**Reviewer Comment:** The reviewer suggested adding a discussion about the assumptions of each model.

**Response:** I apologize for not making the assumptions of the models explicit. I have added a new subsection titled *Assumptions of the Models* in Section 3.2, which outlines the assumptions for Logistic Regression, LASSO, and Random Forest. The added text is as follows:

*"Each predictive model has underlying assumptions that impact its suitability and performance: • Logistic Regression: Assumes a linear relationship between predictors and the logodds of the outcome. It also assumes no multicollinearity among predictors and requires a sufficiently large sample size for stable estimates. • LASSO: Similar to logistic regression, LASSO assumes linearity between predictors and the outcome. • Random Forest: Unlike the other two models, Random Forest, as an advanced machine learning algorithm, does not require linear relationships. However, it assumes sufficient variability in predictor features to form meaningful splits, and is generally robust to overfitting given enough trees are used."*

## Comment 3: Inclusion of Hypotheses

**Reviewer Comment:** The reviewer pointed out that the paper did not explicitly outline hypotheses.

**Response:** I appreciate this suggestion. I have added Research Hypotheses in Section 1. This subsection outlines our expectations for model performance, including hypotheses about accuracy, bias, and feature selection. The added text reads:

*"• Hypothesis 1: The Random Forest model will achieve higher predictive accuracy compared to Logistic Regression and LASSO due to its capability to model complex non-linear relationships. • Hypothesis 2: LASSO will effectively reduce overfitting by feature selection, providing a balance between model simplicity and predictive power, but may not achieve the same level of accuracy as Random Forest."*

## Summary

I sincerely thank the reviewer for their valuable comments, which have helped me improve the overall quality and clarity of the manuscript. I have made changes to address each of the concerns, hoping that it made the work clearer and more accessible for everyone.