**Dear Editor-in-chief,**

We appreciate the respectful reviewers for insightful comments against the submitted manuscript entitled “Unlocking the Potential of microRNAs: Machine Learning Identifies Key Biomarkers for Myocardial Infarction Diagnosis” submitted to Cardiovascular Diabetology. The manuscript was carefully revised according to the reviewers ‘comments and changes were highlighted inside the manuscript in yellow. We hope that these changes make our manuscript eligible for publication in Cardiovascular Diabetology.

Sincerely yours

Dr. F. Bani

*Head of Department of Medical Nanotechnology, Faculty of Advanced Medical Sciences, Tabriz University of Medical Sciences, Tabriz, Iran.*

**Reviewer 1**

1. In section of “Introduction”, I think that the review of research on the identification of miRNA biomarkers for MI is not enough.

Thank you for the positive feedback. We completely believe that the suggested items can improve the significance rigor in the introduction section. In this regard, we added two paragraphs summarizing some recent studies related to the application of miRNAs as valid biomarkers for MI diagnosis and prognosis under several cardiovascular conditions.

2. Three datasets used in this article are mRNA expression data. How can miRNA expression profile be obtained? If the miRNA data were obtained through re-annotation, the process of re-annotation should be detailed described. Additionally, I wonder why not directly select miRNA expression profiles for analysis.

Nice comment indeed. The datasets used in this study are associated with the mRNA expression data. However, the GPL6244 platform contains probes for 189 human miRNAs in addition to mRNA transcripts, according to the original annotation data from GEO. Therefore, both mRNA and miRNA expression data were extracted from these datasets after preprocessing. To identify differentially expressed miRNAs, differential expression analysis was conducted on the full dataset, including both mRNAs and miRNAs, between the sample classes. This approach yielded a list of overall differentially expressed genes (DEGs). The obtained list was then filtered to only retain the DEGs corresponding to validated human miRNAs. This subset of DEGs represented the differentially expressed miRNAs. To better explain the process of identifying differentially expressed miRNAs, a new subsection was added under the Differential Expression Analysis section in both Methods and Results, titled "Differentially expressed miRNAs." In the machine learning section, the availability of miRNA expression data was leveraged within the mRNA datasets to investigate miRNAs as potential biomarkers. To be honest, it was suggested that analyzing datasets with direct miRNA profiling could be an alternative approach. However, we prioritized having matched samples across early-stage MI, stable CAD, and healthy groups which were not available in miRNA-specific datasets to our knowledge at the time. The mRNA datasets allowed us to meet our study goals while utilizing miRNA data available within the platform used. As we are focused on miRNAs as valuable markers and also informative biomolecules for MI diagnosis, and to overcome the lack of a proper miRNA dataset for addressing this issue, we have an ongoing project for producing and publishing such a dataset in our research group.

3. Although three datasets used in this article belong to the same platform, they come from different populations and laboratories. In my opinion, after removing the batch effect, there is still batch effect, and the results obtained from these data require further validation.

Thanks again for the precise comment. You raise a valid concern about potential remaining batch effects from integrating datasets across different populations and labs. To the best of our knowledge, fRMA is the only batch effect removal algorithm (for datasets produced with the same platform) that can be applied to datasets with different sample group ratios, since it uses a universal baseline for correction (for example Combat needs nearly same sample group size ratios in datasets). We investigated and tested several batch removal algorithms, but found fRMA performed best for our data. The RLE plot indicates the means are rearranged around 0 after fRMA, suggesting technical biases between datasets are mitigated. However, biological differences between the sample classes remain, as evidenced by interquartile distances above 0.1. Moreover, the answer to the following question can validate the effectiveness of fRMA in eliminating batch effects.

4. The number of samples used in the article includes 51 health samples, 46 CAD samples and 111 MI patients, and the datasets were divided into training and test sets at a ratio of 7:3. I think 5-fold cross validation is more suitable. Additionally, independent validation sets are needed.

According to the reviewer’s suggestion, it is noteworthy to mention that a 10-fold CV was used at the first step but it provides less biased, more reliable performance estimates compared to lower fold numbers like 5-fold. Using 10-fold CV, the training set and resampling subsets sizes are closer compared to 5-fold CV. The smaller difference in set sizes produces a less biased estimate of the model's true performance. Additionally, more folds result in more test set repetitions, allowing the model to be evaluated on more distinct data subsets. This provides more reliable performance estimation across the range of data. The greater number of test sets also enables a more granular assessment of how the model generalizes. We can obtain a more accurate picture of performance across different data samples with a 10-fold CV. Therefore, while a 5-fold CV is also suitable, we chose a 10-fold CV for its advantages in minimizing bias, improving reliability, and allowing more generalized evaluation as discussed.

Also, we originally intended to use dataset GSE62646 for validation, as it contained MI and CAD samples profiled on the same platform. However, after batch removal with fRMA, we noticed nearly all the CAD and MI samples overlapped with our training data from GSE59867. Upon contacting the authors, they confirmed GSE59867, unfortunately, reused the same samples in GSE62646, making it unsuitable for validation. Regrettably, we could not obtain another matched validation set with samples profiled and processed in the same way. While not ideal, this limitation highlights the difficulty of finding independent validation data meeting all criteria. This limitation is another reason that led us to produce our miRNA dataset. Thank you again for the thoughtful feedback on this critical need for validation. On a positive note, identifying the sample duplication reinforced the ability of fRMA batch removal to detect true biological signals. The distinct clustering of the "repeated" samples post-fRMA indicates effective batch effect removal.



67 stands for GSE59867 and 46 stands for GSE62646 datasets.

5. For identifying two different types of biomarkers for MI, the authors chose different strategies, please provide the reason.

The number of differentially expressed miRNAs between MI and CAD was very limited, with only one miRNA meeting the statistical significance threshold. However, other miRNAs may still hold diagnostic value for discriminating between MI and CAD, even if they do not meet the criteria for differential expression due to the inherent biological similarity of the conditions. Therefore, to capture miRNAs with the best separability between MI and CAD beyond those deemed statistically differentially expressed, we supplemented the DE analysis with a targeted selection based on AUC. By picking miRNAs with AUC > 0.8 for classifying MI vs CAD, we identified additional miRNAs with strong diagnostic discrimination ability, unconstrained by differential expression thresholds. To better explain the reason, we have added an explanation in the Methods section. Please let me know if you would like me to modify or expand this explanation further.

6. Page 5, “To handle the severe imbalance……the sample weights for the healthy and the not-healthy samples were set to 1 and 0.5, respectively”, please provide the reason.

Thank you for catching this. We should have justified setting the sample weights to handle class imbalance in the first layer models. The reason we set the sample weights to 1 for healthy and 0.5 for non-healthy is that there was a large imbalance between the number of healthy samples (51) compared to the combined CAD and MI samples (157 total). Given that there were fewer healthy samples, we assigned them a higher weight (1) compared to the not-healthy samples (0.5). This strategy is a common practice in machine learning to offset the impact of class imbalance on model learning. The specific weights of 1 and 0.5 were chosen empirically through experimentation to produce a good model performance on both classes. We will make sure to explain this more explicitly in our revised manuscript.

7. Are miRNAs studied in the article precursor or mature miRNAs? The writing of “miR-29A” is incorrect. Hsa-mir-29a has two mature miRNAs of “hsa-miR-29a-3p” and “hsa-miR-29a-5p” in the miRbase database.

Thank you for catching the incoherent miRNA naming. To address this, we have double-checked that all the miRNAs mentioned in the manuscript text and figures specifically refer to the mature miRNA sequences, not the precursors. We have also verified that the probe sequences used on the microarray platform align with the intended mature miRNA targets. For clarity, we will update the manuscript to use the complete mature miRNA names (e.g. "hsa-miR-29a-5p") rather than abbreviated versions. Replacing the shorthand names with the full nomenclature will unambiguously indicate the specific mature miRNAs studied.

**Reviewer 2**

1. Abstract: The authors should clarify that miR-21, miR-186, and miR-32 are differentially expressed miRNAs, not genes.

Thanks for your comment. The abstract section was revised accordingly.

2. Data Collection: The authors should provide a clear explanation of why control samples from different GSE IDs were used for analysis. This information will help in understanding the rationale behind the selection of control samples and any potential implications for the study.

To clarify the rationale for using healthy control samples from different datasets, we have added a statement in the Methods section explaining that combining samples from GSE56609 and GSE54475 enabled a sufficiently large healthy control group for robust comparison against the sizable MI and CAD sample sets available in GSE59867. By integrating these three datasets, we aimed to ensure all sample classes had adequate statistical power for the comparative analyses.

3. Early-stage MI Samples: The authors should provide a justification for collecting only early-stage MI samples. Clarifying the reason behind this specific sample selection will add value to the study.

Thank you for your insightful comment. We would like to clarify that the rationale for focusing on early-stage MI samples extends beyond the reasons previously mentioned.

Primarily, the early stage of myocardial infarction presents a unique window of opportunity for intervention. Identifying reliable biomarkers at this stage can significantly improve the timely diagnosis and prognosis of patients. Moreover, it allows us to capture the earliest molecular changes that may be crucial in understanding the onset and progression of the disease.

In addition to these reasons, early-stage MI samples can offer more relevant insights into the initial response mechanisms that the body triggers in response to the acute event. It enables us to better understand the pathophysiology and signaling pathways that get activated immediately after myocardial injury, leading to possible new therapeutic targets.

Finally, the heterogeneity of late-stage MI samples, due to varying treatment regimens and diverse patient responses, can confound our data analysis. In contrast, early-stage samples are more likely to provide a clear and consistent picture of the molecular changes that can serve as potential biomarkers.

We hope this clarifies why we specifically selected early-stage MI samples for our study. We appreciate your interest and willingness to delve deeper into our research methods.

4. Clarification of Bioinformatics Analysis: The sentence "Bioinformatics analyses were conducted using R, ver. 4.2.0 [12], RStudio" should be rewritten to provide more specific details on the types of bioinformatics analyses performed using R and RStudio. This will help readers understand the specific methodologies employed.

The items were revised according to the reviewer’s comments.

5. Differential Expression Analysis: The authors should provide a clear explanation of what they mean by "expression (ones) and non-expression (zeros)" in the context of differential expression analysis. Clarifying whether it refers to overexpression or underexpression will enhance understanding.

To clarify, the 1s and 0s generated by the barcode algorithm refer to binary calls of whether or not a gene is estimated to be expressed (1) or not expressed (0) in each sample. The 1s and 0s do not directly indicate whether a gene is over- or under-expressed between the sample groups. They are simply a discrete representation of the estimated detection of expression above background noise (1) or lack of detectable expression (0) within each sample. We then compared the ratios of 1s vs 0s for each gene between sample groups using Fisher’s exact test to determine genes that are differentially expressed. But the 1/0 calls themselves do not specify the direction of differential expression.

6. Filtering of miRNAs: The authors should provide a clear methodology for filtering miRNAs from the list of differentially expressed genes (DEGs). It is important to explain the specific criteria and approach used to identify the differentially expressed miRNAs separately from the DEGs. This clarification will address the concern raised about mixing the terms DEGs and differentially expressed miRNAs.

You raise a good point. The list of DEGs indeed includes both differentially expressed mRNAs and miRNAs. To separate out the differentially expressed miRNAs from this list, we used the annotation data provided with the microarray platform used in the original studies. This platform included probes for 189 human miRNAs. Therefore, we isolated those DEGs that corresponded to these miRNA probes, yielding a list of diferentially expressed miRNAs. To better explain the process of identifying differentially expressed miRNAs, we have added a new subsection under Differential Expression Analysis section in the both Methods and Results, titled "Differentially expressed miRNAs."

7. Manuscript Flow: The authors should restructure the manuscript to follow a logical flow, including the steps of data collection◊ preprocessing◊ identification of differentially expressed genes (DEGs) and differentially expressed miRNAs (DEMs)◊ machine learning analysis◊  functional enrichment. Aligning the methods and results sections with this flow will enhance the clarity and organization of the manuscript.

Thank you for the suggestion to restructure the manuscript to follow a more logical flow. To better delineate the steps, we have added a subsection under Differential Expression Analysis specifically describing the identification of the differentially expressed miRNAs. However, we believe keeping the machine learning analysis before functional enrichment maintains a suitable flow. The functional enrichment serves to validate that the DEGs and differentially expressed miRNAs relate to PBMC biology after they have been analyzed. Therefore, it seems appropriate to conduct ML using the differentially expressed miRNAs first, and then do an enrichment analysis to confirm their relevance. Please advise if you think reordering these sections by putting enrichment first would significantly enhance the flow. We are happy to rearrange sections as you recommend. We aim to structure the manuscript in the clearest way possible for readers. Please let us know your thoughts on if the current order needs to be modified.