184626_1686640017

by Kevin Barnes

General metrics

35,646	5,448	311	21 min 47 sec	41 min 54 sec
characters	words	sentences	reading time	speaking time

Score

This text scores better than 76% of all texts checked by Grammarly

Writing Issues

345	59	286
Issues left	Critical	Advanced

Plagiarism

This text hasn't been checked for plagiarism

Writing Issues

138	**Correctness**	
14	Misspelled words	
4	Faulty tense sequence	
52	Punctuation in compound/complex sentences	
1	Citation style options	
14	Determiner use (a/an/the/this, etc.)	
2	Improper formatting	
4	Wrong or missing prepositions	
4	Comma misuse within clauses	
7	Misplaced words or phrases	
4	Faulty subject-verb agreement	
1	Unknown words	
3	Confused words	
7	Incorrect noun number	
14	Misuse of semicolons, quotation marks, etc.	
2	Conjunction use	
1	Misuse of modifiers	
3	Incorrect verb forms	
1	Modal verbs	
39	**Engagement**	
39	Word choice	
164	**Clarity**	
26	Wordy sentences	
103	Passive voice misuse	

34	Unclear sentences	
1	Intricate text	
4	**Delivery**	
3	Tone suggestions	
1	Incomplete sentences	

Unique Words

Measures vocabulary diversity by calculating the percentage of words used only once in your document

18%

unique words

Rare Words

Measures depth of vocabulary by identifying words that are not among the 5,000 most common English words.

44%

rare words

Word Length

Measures average word length

5.1

characters per word

Sentence Length

Measures average sentence length

17.5

words per sentence

184626_1686640017

Machine Learning Deriven[1] Set of microRNAs as a Novel Biomarker Set for Myocardial Infarction Diagnosis

2023-06-13

MicroRNAs (miRNAs) play a crucial role in regulating adaptive and maladaptive responses in cardiovascular diseases, making them attractive targets for potential biomarkers. However, their potential as novel biomarkers for diagnosing cardiovascular diseases requires systematic evaluation. In this study, we aimed to identify a key set of miRNA biomarkers using integrated bioinformatics and machine learning analysis. We combined and analyzed three gene expression datasets from the Gene Expression Omnibus (GEO) database, which contains peripheral blood mononuclear cells (PBMCs) samples from individuals with myocardial infarction (MI), stable coronary artery disease (CAD), and healthy individuals. Additionally, we selected a set of miRNAs based on their area under the receiver operating characteristic curve (AUC-ROC) for separating the CAD and MI samples. We designed a two-layer architecture for sample classification, in which the first layer is isolating[2] healthy samples[3] from not-healthy ones, and the second layer is classifying[4,5] stable CAD and MI samples. We trained different machine learning models using both biomarker sets and evaluated their performance on a test set. We identified miR-21, miR-186, and miR-32 as the only miRNAs among the differentially expressed genes, and a set including miR-186, miR-21, miR-197, miR-29A, and miR-296 as the optimum set of miRNAs selected by their AUC-ROC. Both biomarker sets could distinguish healthy from not-healthy samples with complete accuracy. The best performance for the classification of CAD and

MI was achieved with an SVM model trained using the biomarker set selected by AUC-ROC, with an AUC-ROC of 0.96, and an accuracy of 0.94 on the test data. Our study demonstrated that miRNA signatures derived from PBMCs could serve as valuable novel biomarkers for cardiovascular diseases.

1 Introduction

At present, cardiovascular diseases (CVDs) are the leading cause of human mortality with 32% of all global deaths. It is estimated that about 85% of CVDs mortality were diagnosed with myocardial infarction (MI) ("Cardiovascular Diseases (CVDs)" n.d.). MI is an acute coronary syndrome with sudden blockage and stenosis of the coronary artery, and subsequent myocardial ischemia, leading to extensive cardiomyocyte damage and necrosis (Yap et al. 2023). Over the last 50 years, numerous attempts have been collected to use biomarkers to facilitate diagnosis, assess risk, follow-up therapy, and determine therapeutic efficacy in CVDs candidates. Based on the released guidelines, cardiac troponins (cTns) are used as a highly-sensitive and accurate approach for the detection of MI. Despite the inherent advantages, the high-rate sensitivity of cTn-based assays has also led to more false positive results (Thygesen et al. 2018), which do necessitate the advent and development of new modalities with pathological values. To improve diagnostic value upon existing MI biomarkers, the combination of complementary biological markers, such as microRNAs (miRNAs) and other genetic factors, is proposed. Previous researches support the notion that miRNAs exhibit the great potential to be used as alternative biomarkers in CVDs detection and follow-up (Schulte et al. 2020). It is suggested that miRNAs possess 18-22 nucleotides and can play a crucial role in the regulation of gene expression. Evidence point to the fact that miRNAs are involved in the pathogenesis of cardiac tissue injury (Schulte, Karakas, and Zeller 2017). Several biological activities such as angiogenesis,

cardiomyocyte growth and contractility, lipid metabolism, plaque formation, and cardiac rhythm are regulated by miRNAs (Kalayinia et al. 2021).These elements can easily circulate in biofluids and could be considered as theranostics targets in terms of CVDs (Schulte, Karakas, and Zeller 2017). It is postulated that the function and diagnostic properties of miRNAs are beyond the myocardium in CVD patients. To be specific, the expression of miRNAs can vary in different biofluids and cell components such as serum and peripheral blood mononuclear cells (PBMCs) (Soler-Botija, Gálvez-Montón, and Bayés-Genís 2019).

PBMCs are a fraction of white blood cells, including monocytes, lymphocytes, macrophages, and other cells belonging to the immune system (Gao et al. 2020). Emerging data have indicated that PBMCs can be used as a valid source of biomarkers for monitoring various pathological conditions. Of note, the alteration of mRNAs and miRNAs under pathological conditions gives us valuable information about different kinds of disorders. PBMCs could recapitulate the conditions of the target tissues, thus, providing a highly sensitive and specific source of biomarkers (Mosallaei et al. 2022).

Commensurate with these conditions, these cells are repositories of dysregulated genes and miRNAs expression profiles in CVDs (Gao et al. 2020; Mosallaei et al. 2022).

In recent years, the advent and applicarion of machine learning (ML) is an exciting prospect for advancing scientific discoveries. Although the concept of ML and its initial algorithms were conceived many years ago, recent improvements in computing power and access to vast amounts of data have shown that ML techniques outperform classical statistical methods in various fields. Furthermore, the progress made in omics technologies has enabled the analysis of massive and intricate biological data sets, consisting of hundreds to

thousands of samples, which makes it possible for ML to extract valuable biological insight and information from such data (Torun et al. 2023). As a result, machine learning (ML) provides innovative methods for merging and interpreting diverse types of omics data, leading to the identification of new biomarkers. These biomarkers can aid in precise disease prediction, patient stratification, and the development of novel therapeutic approaches (Reel et al. 2021).

In this study, we aimed to identify potential miRNA biomarkers for MI patients by combining and analyzing three different microarray datasets from PBMCs. It is suggested that the integration of omics data with bioinformatics and ML techniques could be a promising tool in the discovery of new and more accurate biomarkers for monitoring MI. Besides, this approach can deepen our vision into the underlying mechanisms of MI and aid in the development of valid diagnostic biomarkers, and patient stratification.

2 Materials and Methods

2.1 Microarray data collection

Microarray datasets were obtained from the Gene Expression Omnibus (GEO) database (https://www.ncbi.nlm.nih.gov/geo/). To obtain sufficient classification power between MI, healthy and CAD samples, a relatively large sample size was required. Therefore, GSE59867 for MI and CAD samples, and GSE56609 and GSE54475 for healthy samples were selected. All samples were produced using Affymetrix Human Gene 1.0 ST Array (GPL6244) platform. Only healthy, CAD and early-stage MI samples were selected from these datasets for further analyses. The basic information for the three datasets evaluated in the current study is provided in Table 2.1. The bioinformatics sections were fully conducted on R, ver. 4.2.0, (R Core Team 2022), using RStudio (RStudio Team

2020), and all plots and graphics of these sections were created using ggplot2[70] R package (Wickham 2016).[14]

Table 2.1: Basic information of[71] the GEO microarray datasets.

Dataset	Platform	Healthy	CAD	MI	Refrence[72]
GSE59867	GPL6244	-	46	111	(Maciejak et al. 2015)
GSE56609	GPL6244	46	-	-	(Matone et al. 2015)
GSE54475	GPL6244	5	-	-	(Canali et al. 2014)[14]

2.2 Pre-processing

The raw data in the form of CEL files from all datasets were obtained from the GEO. To prepare the data for analysis, we utilized the fRMA package (M. N. McCall, Bolstad, and Irizarry 2010), which facilitated the pre-processing of individual microarray samples and their consistent combination. For each dataset, background correction was applied using the RMA algorithm, followed by quantile normalization based on the reference distribution. To account for probe-specific effects , batch effects were eliminated during summarization, and gene expression variances were estimated accordingly. In cases where multiple probe sets matched the same gene, the mean log fold change was retained. Consequently, fRMA acn serve as a technique for removing batch effects across diverse datasets generated by identical microarray platforms (Lazar et al. 2013). To ensure the effectiveness of batch effect removal, we employed principal component analysis (PCA) and relative log expression (RLE) plots to visualize the data before and after applying fRMA.

2.3 Differential expression analysis

The barcode algorithm, as introduced by McCall et al. (Matthew N. McCall et al. 2011), aimed to convert actual expression values into binary barcode values. Extensive sample collections were gathered, and normalization was performed using fRMA across multiple platforms, including the Affymetrix Human Gene 1.0 ST Array (GPL6244) platform. By utilizing these normalized datasets, the distribution of observed intensities, both for expressed and unexpressed genes, was estimated. The determination of whether a gene was expressed or unexpressed was based on the following equation, where a value of 1 indicated expression and a value of 0 indicated non-expression:

xij=1if xij>=μne+C×σne0otherwise

In the barcode algorithm, the normalized intensity of gene i in sample j is denoted by xij. A user-defined parameter, C, is introduced, along with the standard deviation (σne) and mean (μne) of the non-expressed distribution. Based on these values, the barcode representation of a sample is generated as a vector consisting of ones and zeros, representing the estimated expression (ones) and non-expression (zeros) of each gene. The barcode function within the R fRMA package was employed for implementing the barcode algorithm, utilizing the default value of C.

To assess the differences in expressed ratios between the MI group and the healthy control group, Fisher's exact test was performed on the barcode values of individual genes. Genes that exhibited a false discovery rate (FDR) below 0.05, calculated using the Benjamini-Hochberg procedure to account for multiple testing issues, were identified as differentially expressed genes (DEGs). The same procedures were applied to the CAD versus healthy control comparison, as well as the MI versus CAD group, in order to identify DEGs specific to each comparison.

2.4 Functional and pathway enrichment analyses

The R clusterProfiler package (Yu et al. 2012) was utilized to perform Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway enrichment analysis and Gene Ontology (GO) functional annotation on the set of DEGs. The GO analysis encompassed three categories: biological process (BP), cellular component (CC), and molecular function (MF). For statistical significance, an adjusted p-value threshold of less than 0.05 was employed. The enrichment analyses were conducted separately for the DEGs specific to the MI-healthy and CAD-healthy comparisons. In these analyses, all default parameters provided by the package were used.

2.5 ML procedure

The ML analysis was performed using Python software, ver. 3.9, Numpy (Harris et al. 2020), pandas (McKinney 2010), and Scikit-Learn packages (Pedregosa et al. 2011). Whenever hyper-tuning was needed, the scikit-opt package (Head et al. 2021) was used. In all ML analyses, the datasets were divided into train and test sets by a 0.7:0.3 ratio and all reported results are the average of 10-fold cross-validation.

Two different approaches were used for selecting miRNAs for model training. The first approach was using the miRNAs that are differentially expressed. In the second approach, miRNAs with individual AUC-ROC over 0.8 for separating MI from CAD were selected. Having the result of these two different approaches can provide an informative comparison between the predictive capabilities of sets of miRNAs selected with different logics.

2.5.1 miRNAs in DEGs

In this approach, a two-layer architecture was deployed to the data to maximize the prediction values. The first layer predicted whether a sample is healthy or not, and the second layer separated MI from CAD in the samples which were predicted as not healthy in the first layer. To this end, a distinct ML model was trained for each layer. Since there is a limited number of miRNAs in DEGs, both layers were trained with all of them. For further comparison with the models' performance, the ROC curve of each miRNA for classifying healthy and not-healthy, as well as CAD and MI, were generated using a Logistic Regression model.

2.5.1.1 First layer for isolation of healthy and not-healthy samples:

A support vector machine (SVM) model using RBF kernels was trained and hyper-tuned using all miRNAs in DEGs. To handle the severe imbalance in the number of samples (51 for the healthy and 157 for the not-healthy groups), the sample weight for the healthy and the not-healthy samples were set to 1 and

0.5, respectively. The ROC curve and confusion matrix for the model were reported.

2.5.1.2 Second layer for seperating MI and CAD samples:

For the sake of reaching the highest classification performance, different models were investigated. To do so, SVM (with linear, polynomial, and RBF kernels), Logistic Regression (LR), Random Forests (RF), k-Nearest Neighbor (kNN), Gradient Boosting (GB), XGBoost (XGB) and Decision Tree (DT) models were trained. All models were trained with their pre-set parameters with 10-fold cross-validation. The criteria for choosing the best model were the highest accuracy and AUC-ROC on the test set. The best model was hyper-tuned with the scikit-opt package (Head et al. 2021) to get the best classification performance. The ROC curve and confusion matrix for the best model were reported.

2.5.2 miRNAs with the highest AUC-ROC

Like the previous approach, a two-layers strategy was conducted. The first layer classified samples into healthy and not-healthy, and the separated MI and CAD samples. However, to keep the number of miRNAs as low as possible miRNAs were selected from the second layer, and then their performance wase evaluated in the first layer. AUC-ROC of all miRNAs for classifying MI and CAD samples were calculated and the miRNAs with the AUC-ROC over 0.8 were selected. The ROC curves for each selected miRNA for separating healthy samples from not-healthy and MI from CAD samples were also plotted for further comparison.

2.5.2.1 First layer for isolation of healthy and not-healthy samples:

An SVM model with an RBF kernel was trained using the selected set of miRNAs. Additionally, the model was hyper-tuned to find the hyper-parameters for the highest AUC-ROC and accuracy. The same sample weights as previous

approach (1 for healthy and 0.5 for not-healthy samples) were used. The ROC curve and confusion matrix for the model were reported.

2.5.2.2 Second layer for separating MI and CAD samples:

The selected miRNAs set was used to train different algorithms to find the best model. Similar to the previous approach, SVM (with linear, polynomial, and RBF kernels), LR, RF, kNN, GB, XGB, and DT model were trained. All models were trained with their pre-set parameters using 10-fold cross-validation. The models with the highest AUC-ROC and accuracy on the test set were selected and hyper-tuned using the scikit-opt package (Head et al. 2021). The ROC curve and confusion matrix for the best model were reported.

3 Results

3.1 Pre-processing

The PCA plots of the samples are shown in Figure 3.1A and B. As shown, healthy samples were separated from CAD or MI samples in primary data and also after conducting fRMA. In the RLE plot, there was a distinct difference between dataset means for all samples before conducting fRMA (Figure 3.1C). All datasets were rearranged around 0 in the RLE plot after conducting fRMA (Figure 3.1D). Moreover, there was a clear change in inter-quantile distances, but the values still were over 0.1.

Figure 3.1: principal component analysis plots for (A) primary data and (B) the data after fRMA; and the relative log expression plots for (C) primary data and (D) the data after fRMA.

3.2 Differential expression analysis

Table 3.1: Total, up-, and down-regulated DEGs and differentially expressed miRNAs.

According to the cutoff criterion of FDR<0.05, there were 860 DEGs between the MI and the healthy samples. Among them, 323 were up-regulated, and 537 were down-regulated in MI compared to the healthy controls. In CAD and healthy groups comparison, we found 670 DEGs, of which 262 and 408 DEGs were up- and down-regulated, respectively in CAD samples. In the MI and CAD groups, the number of DEGs was 260, and the number of up- and down-regulated genes in MI samples were 144 and 116 in comparison with CAD samples, respectively. These data are summarized in Table 3.1.

Figure 3.2: Venn diagram for DEGs in CAD/Healthy, MI/Healthy, and MI/CAD comparison.

The venn diagram in Figure 3.2 shows that CAD and MI samples shared a majority of their DEGs. From 860 DEGs of MI/healthy and 670 DEGs of CAD/healthy, 531 genes were common which is 62% of MI/healthy DEGs and 79% of CAD/healthy DEGs.

3.3 GO and KEGG enrichment analyses of the DEGs

To explore the biological classification of the DEGs, we performed GO and KEGG pathway enrichment analyses on MI/healthy and CAD/healthy DEGs. For MI/healthy, GO enrichment analysis in the BP category, suggested that the DEGs were enriched in "immune response-regulating signaling pathway", "lymphocyte differentiation", "immune response-regulating cell surface receptor signaling pathway", and "leukocyte activation involved in immune response" (Figure 3.3A). In the CC category, the DEGs were enriched in "secretory granule membrane", "azurophil granule", "ficolin-1-rich granule", "tertiary granule", and "ficolin-1-rich granule membrane" (Figure 3.3B). In the MF category, the DEGs were involved in "cadherin binding" and "MHC class I protein binding" (Figure 3.3C). KEGG pathway analysis indicated that the DEGs

were related to the following pathways: "Chemokine signaling pathway",[196] "Lipid and atherosclerosis",[197] and "Hematopoietic cell lineage" (Figure 3.3D).

Figure 3.3: Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathways enriched with the MI and healthy DEGs. (A) Biological process terms. (B) Cellular component terms. (C) Molecular function terms. (D) KEGG analysis.

The enrichment results for CAD/healthy DEGs were as follows. In the BP category, GO enrichment suggested that the DEGs were enriched in "positive regulation of defense response",[198] "positive regulation of innate immune response",[199] "mononuclear cell differentiation",[200] and "positive regulation of response to external stimulus" (Figure 3.4A). In the CC category, the DEGs were enriched[201] in "azurophil granule",[202] "ficolin-1-rich granule", and "ficolin-1-rich granule membrane" (Figure 3.4B). In the MF category, the DEGs were involved in "lipoprotein particle receptor binding" and "NF-κB binding" (Figure 3.4C). KEGG pathway analysis showed that the DEGs were related to the following pathways: "Chemokine signaling pathway",[203] "Lipid and atherosclerosis",[204] and "Hematopoietic cell lineage" (Figure 3.4D).

Figure 3.4: Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathways enriched with the CAD and healthy DEGs. (A) Biological process terms. (B) Cellular component terms. (C) Molecular function terms. (D) KEGG analysis.

3.4 Machine Learning

3.4.1 miRNAs in DEGs

Among all DEGs,[205] just miR-186, miR-32, and miR-21 were detected[206] as differentially expressed miRNAs. The expression profile of these three miRNAs

is presented in Figure 3.5. Additionally, The ROC curves of each miRNA for each layer are presented in Figure 3.6. Using the logistic regression model the AUC-ROC of miR-21, miR-32, and miR-186 for separating healthy and not-healthy samples was 0.98, 0.99, and 0.90 respectively (Figure 3.6A). Besides, the accuracy of each miRNA for classifying the samples into healthy and not-healthy groups on the test set was 0.92, 0.98, and 0.89 for miR-21, miR-32, and miR-186, respectively. Moreover, the ROC curve of each miRNA for classifying MI and CAD samples was presented in Figure 3.6B. The AUC-ROC and accuracy for miR-21, miR-32, and miR-186 on the test set were 0.85; 0.70; and 0.86, and 0.78; 0.67; and 0.74, respectively.

Table 3.2: Investigated miRNAs log fold-change and adjusted p-values for CAD samples relative to healthy, MI samples relative to healthy, and MI samples relative to CAD.

Figure 3.5: Expression profile of all miRNAs in two approach in different sample classes.

Figure 3.6: ROC curve for single miRNAs on test set classification for (A) healthy and not-healthy samples and (B) CAD and MI samples.

3.4.1.1 First layer for healthy not-healthy isolation:

Although single miRNAs had acceptable performance for this layer, their predictive value could be improved even further by using them as a set. The ROC curve for the SVM model with an RBF kernel trained with all three miRNAs is presented in Figure 3.7A. The model had a better performance in classification than single miRNAs.The AUC-ROC for the model is 1, and its accuracy on the test set was also 1. In Figure 3.8A, the confusion matrix for the model is presented.

Figure 3.7: ROC curve for the model trained with miRNAs in DEGs on test set classification; (A) An SVM model with RBF kernel for healthy and not-healthy and (B) An SVM model with linear kernel for CAD and MI samples classification.

Figure 3.8: Confusion matrix for the model trained with miRNAs in DEGs on test set classification; (A) An SVM model with RBF kernel for healthy and not-healthy and (B) An SVM model with linear kernel for CAD and MI samples classification.

3.4.1.2 Second layer for separating MI samples from CAD:

Different models were trained[223] using expression values for three differentially[224] miRNAs. The models' AUC-ROC and accuracy on the test set are reported[225] in Figure 3.9. The best model from both AUC-ROC and accuracy point-of-view was the SVM model with linear[226] kernel[227]. The AUC-ROC and accuracy for this model with its pre-set values were 0.93 and 0.82 respectively[228]. The model was hyper-tuned for C and gamma hyper-parameters, and therefore[229] the model showed better performance. The ROC curve of the hyper-tuned model is presented[230] in Figure 3.7B. For this model[231] the AUC-ROC reached 0.95 and[232] the accuracy improved to 0.85 (Table 3.3). Moreover, the sensitivity and specificity for the model on the test set were 0.91 and 0.71 respectively[233]. The confusion matrix for the hyper-tuned model is illustrated[234] in Figure 3.8B.

Figure 3.9: Area under the receiver operating characteristic curve and accuracy of different models trained with three miRNAs in DEGs.

Table 3.3: AUC-ROC and accuracy for SVM with linear[235] kernel as the best model trained with differentially expressed miRNAs on the train and test set before and after hyper-tuning

3.4.2 AUC-ROC approach

After calculating the AUC-ROC for each miRNA for the classification of MI and CAD samples, the miRNAs with AUC-ROC over 0.8 were selected. The selected miRNAs were miR-29a, miR-197, miR-186, miR-21, and miR-296. The expression values of these miRNAs in healthy, CAD, and MI samples are presented in Figure 3.5. The ROC curve of the selected miRNAs for both layers are illustrated in Figure 3.6.

3.4.2.1 First layer for healthy not-healthy isolation:

Using the selected set , an SVM model with an RBF kernel was trained to separate healthy from not-healthy samples. The ROC curve for the model is presented in Figure 3.10A and the confusion matrix is illustrated in Figure 3.11A. Both AUC-ROC and accuracy for the model on the test set were equal to 1.

Figure 3.10: ROC curve for model trained with the set of miRNAs selected by AUC-ROC on test set classification; (A) SVM with RBF kernel for healthy and not-healthy samples classification. (B) Logistic regression model for CAD and MI samples classification. (C) SVM with polynomial kernel for CAD and MI samples classification.

Figure 3.11: Confusion matrix for model trained with the set of miRNAs selected by AUC-ROC on test set classification; (A) SVM with RBF kernel for healthy and not-healthy samples classification. (B) Logistic regression model for CAD and MI samples classification. (C) SVM with polynomial kernel for CAD and MI samples classification.

3.4.2.2 Second layer for separating MI samples from CAD:

To find the best model for training the set, different models were trained using their pre-set values. Their AUC-ROC and accuracy results on the test set are presented in Figure 3.12. The best model from the AUC-ROC point-of-view was the SVM with linear kernel and from the accuracy point-of-view, it was the SVM model with an RBF kernel. For the SVM-linear model the AUC-ROC and accuracy were 0.93 and 0.82, respectively; and for the SVM-RBF, the values were 0.92 and 0.84, respectively. Both models were hyper-tuned and the ROC curve for their best performance is presented in Figure 3.10B and C. The AUC-ROC and accuracy for the SVM-linear model modified to 0.92 and 0.88, respectively. For the SVM-RBF, these values increased to 0.96 and 0.94, respectively (Table 3.4). The sensitivity for the SVM-linear and SVM-RBF models were 0.91 and 0.97, respectively; and the specificity for them was 0.79 and 0.86, respectively. The confusion matrix for both models is illustrated in Figure 3.11B and C.

Figure 3.12: Area under the receiver operating characteristic curve and accuracy of different models trained with AUC-selected miRNAs.

Table 3.4: AUC-ROC and accuracy for SVM with the linear kernel as the best model trained with miRNAs selected based on their individual AUC-ROC on the train and test set before and after hyper-tuning

4 Discussion

The prevalence of MI can lead to high-rate mortality in the clinical setting. However, early diagnosis and application of suitable treatment protocols can reduce mortality and improve MI prognosis ("Cardiovascular Diseases (CVDs)" n.d.; Thygesen et al. 2018; Tsao et al. 2022). Studies have suggested that changes in miRNA expression may play a significant role in the progression of MI and the subsequent remodeling (Laggerbauer and Engelhardt 2022). It is

believed that the expression of miRNAs is altered during the various biological processes correlated with MI within the myocardium or other related tissues (Khan, Gupta, and Mahapatra 2022). Although several researches have been concentrated on examining free circulating miRNAs in the serum samples for the detection of cardiac tissue injuries (Kaur et al. 2020), more information is needed to fully comprehend the miRNAs found in different blood sub-components like plasma, platelets, and PBMCs. Based on previous findings, PBMCs play a crucial role in the destabilization and rupture of plaques, and also the initial inflammatory reactions in individuals experiencing a myocardial infarction (MI). (Mosallaei et al. 2022; Hapke et al. 2022). Moreover, PBMCs have specific miRNA profile that is altered under certain pathological conditions which are great candidates as disease biomarkers (Mosallaei et al. 2022).

PBMCs can respond to several insulting conditions such as MI in the least possible time with prominent changes in their miRNA profile (Mosallaei et al. 2022). Considering the regulatory roles, subtle changes in the transcription of miRNAs can be monitored even before alteration in the levels of mRNAs and proteins (Schulte et al. 2020). These features make the miRNAs an early-stage valid diagnostic tool for the detection of minor and major cell injuries. To date, few studies have been performed to compare the miRNA profiles in PBMCs belonging to MI patients and other CADs and healthy samples to find a robust set of identical miRNAs to differentiate these pathological conditions.

In this study, we combined three GEO datasets of healthy, CAD, and MI samples. Having these samples set alongside bioinformatics analysis and ML means, enabled us to identify potential biomarker sets and also effective therapeutic targets. The results of the DEG analysis (Table 3.1 and Figure 3.2) are proof of the close relationship between the MI and CAD samples.

Interestingly, functional enrichment analysis demonstrated that DEGs in both CAD/healthy and MI/healthy were strongly correlated to immune cell response which is a major cellular part of PBMCs. Here, two different sets of miRNAs were selected as biomarker sets for sample classification. miR-21; miR-32; and miR-186 were selected as differentially expressed miRNAs, and miR-186, miR-21; miR-29a; miR-197; and miR-296 were selected according to their AUC-ROC values. As shown in Figure 3.6, all miRNAs selected with both approaches had AUC-ROC over 0.9 for the isolation of healthy and not-healthy samples except for miR-296 and miR-29a. Data confirmed that the real challenge is to classify CAD and MI samples because of close overlap. Of 6 miRNAs under investigation in both approaches except for miR-32, all miRNAs had an AUC-ROC over 0.8 for the discrimination of CAD and MI samples. As expected, the high AUC-ROC values of miRNAs confirms their high potential as biomarkers. ML models when trained with miRNA sets selected by both DEG and AUC-ROC approaches showed better performance in the classification than each miRNA. To avoid unwanted complexity and poor predictive values, a two-layer architecture was also designed. The first layer was for the discrimination of healthy from not-healthy samples, and the second layer for separating the CAD from MI candidates. As expected, in both approaches a hyper-tuned SVM model could flawlessly separates healthy from not-healthy samples using distinct miRNAs sets. The ML models were also capable of effectively separating CAD from MI patients. Although both miRNA sets had nearly the same AUC-ROC with their best model, the accuracy, sensitivity, and specificity were different. The model trained with AUC-selected miRNAs had better performance in all predictive values, which is logical because of more miRNAs in the set. Numerous studies have reported different biological processes can affect the expression of miRNAs in PBMCs. However, there are still controversies

regarding the exact role of miRNAs in the function of immune cells and the correlation of specific pathological conditions with miRNA profiles. Several studies have proved the activation of specific miRNA types in PBMCs under cardiovascular events H. Li et al. (2018). For instance, there is evidence that the elevation of miR-186 suppresses the expression of Cystathionine-γ-lyase, leading to the subsequent secretion of pro-inflammatory cytokines and cellular lipid accumulation. Besides, macrophage-derived miR-186 may promote atherosclerotic plaques (Yao et al. 2016). In line with this claim, we found that miR-186 is up-regulated in both CAD and MI candidates related to control counterparts. Surprisingly, the obtained data indicated that the expression of miR-186 is higher in CAD patients in comparison to MI (Figure ??). To be specific , miR-186 is the only differentially expressed miRNA between CAD and MI, with a clear up-regulation in CAD, indicating its main role in the promotion of atherosclerosis.

As mentioned before, miR-21 was also up-regulated in both MI and CAD patients in comparison to healthy controls. Moreover, the expression value of miR-21 was significantly higher in MI than that of the CAD group (Table 3.2). It is thought that the up-regulation of miRNA-21 in PBMCs is a compensatory reaction to reduce Treg lymphocyte number in response to the reduction of TGFβ1 secretion into the plasma through a TGFβ1/smad-independent pathway. In line with previous and present data, miR-21 can modulate the activity of PBMCs following the occurrence of cardiovascular diseases (S. Li et al. 2015). Recent data have supported the elevation of miR-32 in CAD samples with the calcification of coronary artery. It is worth noting that miR-32 promotes vascular smooth muscle calcification in mice by controlling the activity of several proteins, including bone morphogenetic protein-1, runt-related transcription factor-2 (RUNX2), osteopontin, and bone-specific phosphoprotein

matrix GLA protein. (Liu et al. 2017). Likewise, there are some reports associated with the activity of miR-32 in PBMCs under several pathologies (Zeng et al. 2021; Wang et al. 2020). The exact role of miR-32 in PBMCs after cardiovascular events remained to be elucidated.

Molecular analyses have indicated the regulatory role of miRNAs selected by the AUC-ROC approach in PBMCs after a cardiovascular event. Two common miRNAs in DEGs and AUC-ROC approachs, miR-21 and miR-186, were covered already. Based on numerous reports mir-29a can be activated in different diseases (Horita, Farquharson, and Stephen 2021).Data analysis indicated that miR-29a is significantly up-regulated in CAD patients in comparison to healthy and MI groups (Table 3.2). Increased miR-29a is associated with the progression of atherosclerosis, and the combination of miR-29a and ox-LDL was offered as a valid biomarker set for paraclinical classification (Huang et al. 2016). However, the role of miR-29a in the function of PBMCs in CAD patients has not been completely examined.

Data indicated that miR-197 is also significantly up-regulated in both CAD/healthy and MI/healthy groups. Previous research has demonstrated that miR-197 may play a crucial role in controlling the anti-inflammatory response of IL-35 by influencing the secretion of cytokines that can either promote or suppress inflammation, the ratio of M1/M2 macrophages, and the proliferation of Treg lymphocytes, which are responsible for suppressing immune responses (Bhansali et al. 2022). Alongside with our findings, it can be concluded that miR-197 could be a useful diagnostic tool for predicting adverse cardiovascular events.

The findings of this study demonstrate the potential of miR-296 as a biomarker with high discriminatory power for distinguishing between samples from individuals with MI and CAD. MiR-296 has been identified as a key regulator in

the development and advancement of atherosclerosis by controlling the expression of target genes associated with various biological processes including angiogenesis, cholesterol metabolism, inflammation, cellular proliferation, hypertension, and apoptosis (H. Li et al. 2018). In a study it has been shown that miR-296 expression levels are significantly increased in the PBMCs of CAD patients compared to healthy controls, suggesting its involvement in regulating pro-inflammatory cytokines like IL-6 and TNF-a (Fard et al. 2020). These findings suggest that miR-296 may have a significant impact on the pathogenesis of atherosclerosis and could potentially serve as a diagnostic biomarker for CAD or MI.

5 Conclusion

In summary, we derived a set of miRNA biomarkers by comparing MI samples to both healthy and CAD samples. We found that the SVM model performed best in both the first layer, which separated healthy and not-healthy samples, and the second layer, which classified MI/CAD samples. The set of miRNAs selected based on their AUC-ROC values had better performance in the second layer. Overall, our two-layer structure achieved a accuracy of 0.96. This demonstrates the potential for combining bioinformatics and machine learning techniques to identify novel biomarkers and gain a deeper understanding of myocardial infarction.

1.	~~Deriven~~ → Driven	Misspelled words	Correctness
2.	~~is isolating~~ → isolates	Faulty tense sequence	Correctness
3.	~~samples~~ → pieces	Word choice	Engagement
4.	~~is classifying~~ → classifies	Faulty tense sequence	Correctness
5.	~~is classifying~~ → classifies	Wordy sentences	Clarity
6.	*The best performance for the classification of CAD and MI was achieved*	Passive voice misuse	Clarity
7.	0.96,	Punctuation in compound/complex sentences	Correctness
8.	*At present, cardiovascular diseases (CVDs) are the leading cause of human mortality with 32% of all global deaths.*	Unclear sentences	Clarity
9.	, with	Punctuation in compound/complex sentences	Correctness
10.	*is estimated*	Passive voice misuse	Clarity
11.	~~CVDs~~ → CVD	Misspelled words	Correctness
12.	artery,	Punctuation in compound/complex sentences	Correctness
13.	, 2023	Punctuation in compound/complex sentences	Correctness
14.	*(Yap et al. 2023); (Thygesen et al. 2018); (Schulte et al. 2020); (Kalayinia et al. 2021); (Gao et al. 2020); (Mosallaei et al. 2022); (Gao et al. 2020; Mosallaei et al. 2022); (Torun et al. 2023); (Reel et al.*	Citation style options	Correctness

	2021); (Wickham 2016); (Canali et al. 2014); (Lazar et al. 2013); (Yu et al. 2012); (Har...		
15.	*MI is an acute coronary syndrome with sudden blockage and stenosis of the coronary artery, and subsequent myocardial ischemia, leading to extensive cardiomyocyte damage and necrosis (Yap et al. 2023).*	Unclear sentences	Clarity
16.	*been collected*	Passive voice misuse	Clarity
17.	~~the detection of~~ → detecting	Wordy sentences	Clarity
18.	necessitating	Wordy sentences	Clarity
19.	*To improve diagnostic value upon existing MI biomarkers, the combination of complementary biological markers, such as microRNAs (miRNAs) and other genetic factors, is proposed.*	Unclear sentences	Clarity
20.	~~the~~ great	Determiner use (a/an/the/this, etc.)	Correctness
21.	~~CVDs~~ → CVD	Misspelled words	Correctness
22.	, 2020	Punctuation in compound/complex sentences	Correctness
23.	*is suggested*	Passive voice misuse	Clarity
24.	, such	Punctuation in compound/complex sentences	Correctness
25.	rhythm,	Punctuation in compound/complex sentences	Correctness
26.	*are regulated*	Passive voice misuse	Clarity

27.	, 2021	Punctuation in compound/complex sentences	Correctness
28.	. These	Improper formatting	Correctness
29.	*be considered*	Passive voice misuse	Clarity
30.	~~as~~	Wrong or missing prepositions	Correctness
31.	*These elements can easily circulate in biofluids and could be considered as theranostics targets in terms of CVDs (Schulte, Karakas, and Zeller 2017).*	Unclear sentences	Clarity
32.	*is postulated*	Passive voice misuse	Clarity
33.	*It is postulated that the function and diagnostic properties of miRNAs are beyond the myocardium in CVD patients.*	Unclear sentences	Clarity
34.	, 2020	Punctuation in compound/complex sentences	Correctness
35.	*be used*	Passive voice misuse	Clarity
36.	~~a valid~~ → a good	Word choice	Engagement
37.	*Of note, the alteration of mRNAs and miRNAs under pathological conditions gives us valuable information about different kinds of disorders.*	Unclear sentences	Clarity
38.	~~conditions~~ → needs, requirements	Word choice	Engagement
39.	a compassionate, a susceptible	Word choice	Engagement
40.	, 2022	Punctuation in compound/complex sentences	Correctness
41.	~~applicarion~~ → application	Misspelled words	Correctness

42.	*were conceived*	Passive voice misuse	Clarity
43.	~~amounts of~~	Wordy sentences	Clarity
44.	sets,	Punctuation in compound/complex sentences	Correctness
45.	~~biological~~ → physical	Word choice	Engagement
46.	, 2023	Punctuation in compound/complex sentences	Correctness
47.	~~the development of~~ → developing	Wordy sentences	Clarity
48.	, 2021	Punctuation in compound/complex sentences	Correctness
49.	*is suggested*	Passive voice misuse	Clarity
50.		Tone suggestions	Delivery
51.	*It is suggested that the integration of omics data with bioinformatics and ML techniques could be a promising tool in the discovery of new and more accurate biomarkers for monitoring MI.*	Unclear sentences	Clarity
52.	~~the development of~~ → developing	Wordy sentences	Clarity
53.	biomarkers,	Punctuation in compound/complex sentences	Correctness
54.	*were obtained*	Passive voice misuse	Clarity
55.	*To obtain sufficient classification power between MI, healthy and CAD samples, a relatively large sample size was required.*	Unclear sentences	Clarity

56.	~~obtain~~ → get	Word choice	Engagement
57.	, and	Comma misuse within clauses	Correctness
58.	samples,	Punctuation in compound/complex sentences	Correctness
59.	~~samples~~ → models	Word choice	Engagement
60.	*were selected*	Passive voice misuse	Clarity
61.	~~samples~~ → models, pieces	Word choice	Engagement
62.	*were produced*	Passive voice misuse	Clarity
63.	healthy,	Punctuation in compound/complex sentences	Correctness
64.	, and	Comma misuse within clauses	Correctness
65.	*Only healthy, CAD and early-stage MI samples were selected*	Passive voice misuse	Clarity
66.	~~selected~~ → chosen	Word choice	Engagement
67.	*is provided*	Passive voice misuse	Clarity
68.	*were fully conducted*	Passive voice misuse	Clarity
69.	4.2.0,	Punctuation in compound/complex sentences	Correctness
70.	the ggplot2	Determiner use (a/an/the/this, etc.)	Correctness
71.	~~of~~ → on	Wrong or missing prepositions	Correctness
72.	~~Refrence~~ → Reference	Misspelled words	Correctness
73.	~~the form of~~	Wordy sentences	Clarity

74.	*were obtained*	Passive voice misuse	Clarity
75.	~~which facilitated~~ → facilitating	Wordy sentences	Clarity
76.	*was applied*	Passive voice misuse	Clarity
77.	*To account for probe-specific effects*	Misplaced words or phrases	Correctness
78.	*To account for probe-specific effects, batch effects were eliminated during summarization, and gene expression variances were estimated accordingly.*	Unclear sentences	Clarity
79.	*batch effects were eliminated*	Passive voice misuse	Clarity
80.	*gene expression variances were estimated*	Passive voice misuse	Clarity
81.	*the mean log fold change was retained*	Passive voice misuse	Clarity
82.	~~fRMA~~ → PhRMA	Misspelled words	Correctness
83.	~~acn~~ → can	Misspelled words	Correctness
84.	~~serve~~ → serves	Faulty subject-verb agreement	Correctness
85.	~~serve as~~ → is	Wordy sentences	Clarity
86.	*Extensive sample collections were gathered*	Passive voice misuse	Clarity
87.	*was performed*	Passive voice misuse	Clarity
88.	*By utilizing these normalized datasets, the distribution of observed intensities, both for expressed and unexpressed genes, was estimated.*	Unclear sentences	Clarity
89.	*was expressed*	Passive voice misuse	Clarity
90.	*was based*	Passive voice misuse	Clarity

91.	~~ɨ~~ → I	Misspelled words	Correctness
92.	*is denoted*	Passive voice misuse	Clarity
93.	~~xij~~ → xi	Misspelled words	Correctness
94.	*A user-defined parameter, C, is introduced, along with the standard deviation (σne) and mean (μne) of the non-expressed distribution.*	Unclear sentences	Clarity
95.	*is generated*	Passive voice misuse	Clarity
96.	*To assess the differences in expressed ratios between the MI group and the healthy control group, Fisher's exact test was performed on the barcode values of individual genes.*	Unclear sentences	Clarity
97.	*To assess the differences in expressed ratios between the MI group and the healthy control group*	Misplaced words or phrases	Correctness
98.	*was performed*	Passive voice misuse	Clarity
99.	~~to account~~ → for accounting	Wrong or missing prepositions	Correctness
100.	~~procedures~~ → methods, techniques	Word choice	Engagement
101.	*The same procedures were applied*	Passive voice misuse	Clarity
102.	~~in order to~~ → to	Wordy sentences	Clarity
103.	*The same procedures were applied to the CAD versus healthy control comparison, as well as the MI versus CAD group, in order to identify DEGs specific to each comparison.*	Unclear sentences	Clarity
104.	*The R clusterProfiler package (Yu et al. 2012) was utilized*	Passive voice misuse	Clarity
105.	the Kyoto	Determiner use	Correctness

		(a/an/the/this, etc.)	
106.	*For statistical significance, an adjusted p-value threshold of less than 0.05 was employed.*	Unclear sentences	Clarity
107.	*was employed*	Passive voice misuse	Clarity
108.	*The enrichment analyses were conducted*	Passive voice misuse	Clarity
109.	*all default parameters provided by the package were used*	Passive voice misuse	Clarity
110.	*was performed*	Passive voice misuse	Clarity
111.	*Whenever hyper-tuning was needed, the scikit-opt package (Head et al. 2021) was used.*	Unclear sentences	Clarity
112.	*scikit-opt*	Unknown words	Correctness
113.	*the scikit-opt package (Head et al. 2021) was used*	Passive voice misuse	Clarity
114.	*were divided*	Passive voice misuse	Clarity
115.	, and	Punctuation in compound/complex sentences	Correctness
116.	*were used*	Passive voice misuse	Clarity
117.	*are differentially expressed*	Passive voice misuse	Clarity
118.	*In the second approach, miRNAs with individual AUC-ROC over 0.8 for separating MI from CAD were selected.*	Unclear sentences	Clarity
119.	*were selected*	Passive voice misuse	Clarity
120.	*a two-layer architecture was deployed*	Passive voice misuse	Clarity
121.	~~is~~ → was	Faulty tense sequence	Correctness

122.	~~samples~~ → pieces, examples, instances	Word choice	Engagement
123.	*were predicted*	Passive voice misuse	Clarity
124.	*were predicted*	Passive voice misuse	Clarity
125.	~~predicted~~ → expected, indicated, heralded	Word choice	Engagement
126.	~~healthy~~ → fit, beneficial	Word choice	Engagement
127.	*The first layer predicted whether a sample is healthy or not, and the second layer separated MI from CAD in the samples which were predicted as not healthy in the first layer.*	Unclear sentences	Clarity
128.	*a distinct ML model was trained*	Passive voice misuse	Clarity
129.	*were trained*	Passive voice misuse	Clarity
130.	~~trained~~ → prepared, acquainted	Word choice	Engagement
131.	~~not-healthy~~ → not healthy	Confused words	Correctness
132.	*was trained*	Passive voice misuse	Clarity
133.	*To handle the severe imbalance in the number of samples (51 for the healthy and 157 for the not-healthy groups)*	Misplaced words or phrases	Correctness
134.	~~healthy~~ → fit	Word choice	Engagement
135.	~~samples~~ → pieces	Word choice	Engagement
136.	~~were~~ → was	Faulty subject-verb agreement	Correctness
137.	*were reported*	Passive voice misuse	Clarity
138.	~~seperating~~ → separating	Misspelled words	Correctness

139.	*different models were investigated*	Passive voice misuse	Clarity
140.	*To do so*	Misplaced words or phrases	Correctness
141.	*SVM (with linear, polynomial, and RBF kernels), Logistic Regression (LR), Random Forests (RF), k-Nearest Neighbor (kNN), Gradient Boosting (GB), XGBoost (XGB) and Decision Tree (DT) models were trained*	Passive voice misuse	Clarity
142.	*All models were trained*	Passive voice misuse	Clarity
143.	~~trained~~ → introduced	Word choice	Engagement
144.	~~scikit-opt~~ → sci-kit-opt	Misspelled words	Correctness
145.	~~to get~~ → for	Wordy sentences	Clarity
146.	*were reported*	Passive voice misuse	Clarity
147.	*a two-layers strategy was conducted*	Passive voice misuse	Clarity
148.	possible,	Punctuation in compound/complex sentences	Correctness
149.	*were selected*	Passive voice misuse	Clarity
150.	~~wase~~ → was	Confused words	Correctness
151.	The AUC-ROC	Determiner use (a/an/the/this, etc.)	Correctness
152.	*were calculated*	Passive voice misuse	Clarity
153.	, and	Punctuation in compound/complex sentences	Correctness
154.	~~samples~~ → pieces	Word choice	Engagement
155.	*were also plotted*	Passive voice misuse	Clarity

156.	*was trained*	Passive voice misuse	Clarity
157.	the previous	Determiner use (a/an/the/this, etc.)	Correctness
158.	*The same sample weights as previous approach (1 for healthy and 0.5 for not-healthy samples) were used*	Passive voice misuse	Clarity
159.	*were reported*	Passive voice misuse	Clarity
160.	~~miRNAs~~ → miRNA	Incorrect noun number	Correctness
161.	*was used*	Passive voice misuse	Clarity
162.	~~model~~ → models	Incorrect noun number	Correctness
163.	~~trained~~ → introduced	Word choice	Engagement
164.	~~scikit-opt~~ → sci-kit-opt	Misspelled words	Correctness
165.	*were reported*	Passive voice misuse	Clarity
166.	*are shown*	Passive voice misuse	Clarity
167.	~~Figure~~ → Figures	Incorrect noun number	Correctness
168.	*were separated*	Passive voice misuse	Clarity
169.	~~also~~	Wordy sentences	Clarity
170.	*In the RLE plot, there was a distinct difference between dataset means for all samples before conducting fRMA (Figure 3.1C).*	Unclear sentences	Clarity
171.	~~samples~~ → models, pieces	Word choice	Engagement
172.	~~conducting~~ → running, acting, completing, performing	Word choice	Engagement
173.	~~conducting~~ →	Word choice	Engagement

	running, completing, working		
174.	~~a clear~~ → an apparent, an evident	Word choice	Engagement
175.	~~fRMA;~~ → fRMA,	Punctuation in compound/complex sentences	Correctness
176.	*were up-regulated*	Passive voice misuse	Clarity
177.	~~healthy~~ → nutritional	Word choice	Engagement
178.	~~groups~~ → group	Incorrect noun number	Correctness
179.	, in	Punctuation in compound/complex sentences	Correctness
180.	~~up- and~~ → up-and	Confused words	Correctness
181.	~~in comparison with~~ → compared to	Wordy sentences	Clarity
182.	*are summarized*	Passive voice misuse	Clarity
183.	~~venn~~ → Venn	Misspelled words	Correctness
184.	~~a majority~~ → most	Wordy sentences	Clarity
185.	, which	Punctuation in compound/complex sentences	Correctness
186.	*To explore the biological classification of the DEGs, we performed GO and KEGG pathway enrichment analyses on MI/healthy and CAD/healthy DEGs.*	Unclear sentences	Clarity
187.	category,	Punctuation in compound/complex sentences	Correctness
188.	*were enriched*	Passive voice misuse	Clarity

189.	", → ,"	Misuse of semicolons, quotation marks, etc.	Correctness
190.	", → ,"	Misuse of semicolons, quotation marks, etc.	Correctness
191.	", → ,"	Misuse of semicolons, quotation marks, etc.	Correctness
192.	*were enriched*	Passive voice misuse	Clarity
193.	", → ,"	Misuse of semicolons, quotation marks, etc.	Correctness
194.	", → ,"	Misuse of semicolons, quotation marks, etc.	Correctness
195.	", → ,"	Misuse of semicolons, quotation marks, etc.	Correctness
196.	", → ,"	Misuse of semicolons, quotation marks, etc.	Correctness
197.	", → ,"	Misuse of semicolons, quotation marks, etc.	Correctness
198.	", → ,"	Misuse of semicolons, quotation marks, etc.	Correctness
199.	", → ,"	Misuse of semicolons, quotation marks, etc.	Correctness
200.	", → ,"	Misuse of semicolons, quotation marks, etc.	Correctness
201.	*were enriched*	Passive voice misuse	Clarity
202.	", → ,"	Misuse of semicolons, quotation marks, etc.	Correctness
203.	", → ,"	Misuse of semicolons, quotation marks, etc.	Correctness

204.	~~",~~ → ,"	Misuse of semicolons, quotation marks, etc.	Correctness
205.	*Among all DEGs, just miR-186, miR-32, and miR-21 were detected as differentially expressed miRNAs.*	Unclear sentences	Clarity
206.	*were detected*	Passive voice misuse	Clarity
207.	*is presented*	Passive voice misuse	Clarity
208.	*are presented*	Passive voice misuse	Clarity
209.	~~presented~~ → shown	Word choice	Engagement
210.	, the	Punctuation in compound/complex sentences	Correctness
211.	, respectively	Punctuation in compound/complex sentences	Correctness
212.	*the ROC curve of each miRNA for classifying MI and CAD samples was presented*	Passive voice misuse	Clarity
213.	~~and~~	Conjunction use	Correctness
214.	~~and~~	Conjunction use	Correctness
215.	~~relative~~ → close	Word choice	Engagement
216.	~~healthy~~ → fit	Word choice	Engagement
217.	~~relative~~ → close	Word choice	Engagement
218.	~~approach~~ → approaches	Incorrect noun number	Correctness
219.	an acceptable	Determiner use (a/an/the/this, etc.)	Correctness
220.	*their predictive value could be improved*	Passive voice misuse	Clarity

221.	*is presented*	Passive voice misuse	Clarity
222.	*is presented*	Passive voice misuse	Clarity
223.	*were trained*	Passive voice misuse	Clarity
224.	~~differentially~~ → differential	Misuse of modifiers	Correctness
225.	*are reported*	Passive voice misuse	Clarity
226.	a linear	Determiner use (a/an/the/this, etc.)	Correctness
227.	*The best model from both AUC-ROC and accuracy point-of-view was the SVM model with linear kernel.*	Unclear sentences	Clarity
228.	, respectively	Punctuation in compound/complex sentences	Correctness
229.	*The model was hyper-tuned for C and gamma hyper-parameters, and therefore the model showed better performance.*	Unclear sentences	Clarity
230.	*is presented*	Passive voice misuse	Clarity
231.	model,	Comma misuse within clauses	Correctness
232.	, and	Punctuation in compound/complex sentences	Correctness
233.	, respectively	Punctuation in compound/complex sentences	Correctness
234.	*is illustrated*	Passive voice misuse	Clarity
235.	the linear	Determiner use (a/an/the/this, etc.)	Correctness
236.			

	to classify	Wordy sentences	Clarity
237.	*are presented*	Passive voice misuse	Clarity
238.	~~are~~ → is	Faulty subject-verb agreement	Correctness
239.	*are illustrated*	Passive voice misuse	Clarity
240.	*Using the selected set*	Misplaced words or phrases	Correctness
241.	*Using the selected set, an SVM model with an RBF kernel was trained to separate healthy from not-healthy samples.*	Unclear sentences	Clarity
242.	*an SVM model with an RBF kernel was trained*	Passive voice misuse	Clarity
243.	*is presented*	Passive voice misuse	Clarity
244.	, and	Punctuation in compound/complex sentences	Correctness
245.	*is illustrated*	Passive voice misuse	Clarity
246.	a model	Determiner use (a/an/the/this, etc.)	Correctness
247.	a model	Determiner use (a/an/the/this, etc.)	Correctness
248.	*To find the best model for training the set, different models were trained using their pre-set values.*	Unclear sentences	Clarity
249.	*To find the best model for training the set*	Misplaced words or phrases	Correctness
250.	*were trained*	Passive voice misuse	Clarity
251.	*are presented*	Passive voice misuse	Clarity

252.	a linear	Determiner use (a/an/the/this, etc.)	Correctness
253.	, and	Punctuation in compound/complex sentences	Correctness
254.	*The best model from the AUC-ROC point-of-view was the SVM with linear kernel and from the accuracy point-of-view, it was the SVM model with an RBF kernel.*	Unclear sentences	Clarity
255.	model,	Comma misuse within clauses	Correctness
256.	~~respectively;~~ → respectively,	Punctuation in compound/complex sentences	Correctness
257.	, and	Punctuation in compound/complex sentences	Correctness
258.	*is presented*	Passive voice misuse	Clarity
259.	~~. The~~ → —the	Incomplete sentences	Delivery
260.	were modified	Incorrect verb forms	Correctness
261.	~~respectively;~~ → respectively,	Punctuation in compound/complex sentences	Correctness
262.	*is illustrated*	Passive voice misuse	Clarity
263.	~~individual~~	Wordy sentences	Clarity
264.	Engelhardt,	Punctuation in compound/complex sentences	Correctness
265.	*is believed*	Passive voice misuse	Clarity
266.	~~expression~~ → word	Word choice	Engagement

267.	*is altered*	Passive voice misuse	Clarity
268.	, 2022	Punctuation in compound/complex sentences	Correctness
269.	~~researches~~ → types of research, pieces of research, kinds of research	Incorrect noun number	Correctness
270.	~~been~~	Incorrect verb forms	Correctness
271.	*been concentrated*	Passive voice misuse	Clarity
272.	~~the~~ serum	Determiner use (a/an/the/this, etc.)	Correctness
273.	plaques,	Punctuation in compound/complex sentences	Correctness
274.	*Based on previous findings, PBMCs play a crucial role in the destabilization and rupture of plaques, and also the initial inflammatory reactions in individuals experiencing a myocardial infarction (MI).*	Unclear sentences	Clarity
275.	*is altered*	Passive voice misuse	Clarity
276.	, which	Punctuation in compound/complex sentences	Correctness
277.	, 2022	Punctuation in compound/complex sentences	Correctness
278.	, such	Punctuation in compound/complex sentences	Correctness
279.	, in	Punctuation in compound/complex sentences	Correctness

280.	~~prominent~~ → major, noticeable, notable	Word choice	Engagement
281.	, 2022	Punctuation in compound/complex sentences	Correctness
282.	*subtle changes in the transcription of miRNAs can be monitored*	Passive voice misuse	Clarity
283.	~~valid~~ → useful, good	Word choice	Engagement
284.	~~major~~ → significant	Word choice	Engagement
285.	*been performed*	Passive voice misuse	Clarity
286.	*In this study, we combined three GEO datasets of healthy, CAD, and MI samples.*	Unclear sentences	Clarity
287.	means,	Punctuation in compound/complex sentences	Correctness
288.	~~also~~	Wordy sentences	Clarity
289.	*were strongly correlated*	Passive voice misuse	Clarity
290.	, which	Punctuation in compound/complex sentences	Correctness
291.	*Interestingly, functional enrichment analysis demonstrated that DEGs in both CAD/healthy and MI/healthy were strongly correlated to immune cell response which is a major cellular part of PBMCs.*	Unclear sentences	Clarity
292.	~~different~~	Wordy sentences	Clarity
293.	*two different sets of miRNAs were selected*	Passive voice misuse	Clarity

294.	~~the isolation of~~ → isolating	Wordy sentences	Clarity
295.	, except	Punctuation in compound/complex sentences	Correctness
296.	~~confirms~~ → confirm	Faulty subject-verb agreement	Correctness
297.	models,	Punctuation in compound/complex sentences	Correctness
298.	~~when~~	Wordy sentences	Clarity
299.	approaches,	Punctuation in compound/complex sentences	Correctness
300.	*To avoid unwanted complexity and poor predictive values, a two-layer architecture was also designed.*	Unclear sentences	Clarity
301.	*To avoid unwanted complexity and poor predictive values*	Misplaced words or phrases	Correctness
302.	*was also designed*	Passive voice misuse	Clarity
303.	was for	Incorrect verb forms	Correctness
304.	, a	Punctuation in compound/complex sentences	Correctness
305.	~~separates~~ → separate	Modal verbs	Correctness
306.	~~miRNAs~~ → miRNA	Incorrect noun number	Correctness
307.	~~specific~~ → particular	Word choice	Engagement
308.	*is up-regulated*	Passive voice misuse	Clarity
309.	~~in comparison to~~ → than in	Wordy sentences	Clarity

310.	*To be specific, miR-186 is the only differentially expressed miRNA between CAD and MI, with a clear up-regulation in CAD, indicating its main role in the promotion of atherosclerosis.*	Unclear sentences	Clarity
311.	~~a clear~~ → an apparent, an evident	Word choice	Engagement
312.	~~main~~ → central, leading, prominent, primary	Word choice	Engagement
313.	*was also up-regulated*	Passive voice misuse	Clarity
314.	~~that of~~ → in	Wordy sentences	Clarity
315.	*is thought*	Passive voice misuse	Clarity
316.	*It is thought that the up-regulation of miRNA-21 in PBMCs is a compensatory reaction to reduce Treg lymphocyte number in response to the reduction of TGFβ1 secretion into the plasma through a TGFβ1/smad-independent pathway.*	Unclear sentences	Clarity
317.	the coronary	Determiner use (a/an/the/this, etc.)	Correctness
318.	*Recent data have supported the elevation of miR-32 in CAD samples with the calcification of coronary artery.*	Unclear sentences	Clarity
319.	some reports are	Wordy sentences	Clarity
320.	~~remained~~ → remains	Faulty tense sequence	Correctness
321.	*be elucidated*	Passive voice misuse	Clarity
322.	~~approachs~~ → approaches, approach	Misspelled words	Correctness
323.	, mir-29a	Punctuation in compound/complex sentences	Correctness

324.	. Data	Improper formatting	Correctness
325.	*is significantly up-regulated*	Passive voice misuse	Clarity
326.	~~in comparison~~ → compared	Wordy sentences	Clarity
327.	~~completely~~ → thoroughly	Word choice	Engagement
328.	*been completely examined*	Passive voice misuse	Clarity
329.		Tone suggestions	Delivery
330.	~~suppressing~~ → stopping	Word choice	Engagement
331.	~~with~~	Wrong or missing prepositions	Correctness
332.	~~with~~	Wordy sentences	Clarity
333.	*it can be concluded*	Passive voice misuse	Clarity
334.	*been identified*	Passive voice misuse	Clarity
335.	~~key~~ → critical	Word choice	Engagement
336.	, including	Punctuation in compound/complex sentences	Correctness
337.	, 2018	Punctuation in compound/complex sentences	Correctness
338.	study,	Punctuation in compound/complex sentences	Correctness
339.	*In a study it has been shown that miR-296 expression levels are significantly increased in the PBMCs of CAD patients compared to healthy controls, suggesting its involvement in regulating pro-inflammatory cytokines like IL-6 and TNF-a (Fard et al. 2020).*	Unclear sentences	Clarity

340.	*These findings suggest that miR-296 may have a significant impact on the pathogenesis of atherosclerosis and could potentially serve as a diagnostic biomarker for CAD or MI.*	Unclear sentences	Clarity
341.		Tone suggestions	Delivery
342.	~~samples~~ → pieces	Word choice	Engagement
343.	performed better	Wordy sentences	Clarity
344.	~~a accuracy~~ → an accuracy	Determiner use (a/an/the/this, etc.)	Correctness
345.	*This*	Intricate text	Clarity