Media Statistika

p-ISSN 1979 – 3693 e-ISSN 2477 – 0647
MEDIA STATISTIKA 17(1) 2024: 69-80
http://ejournal.undip.ac.id/index.php/media_statistika

GENE MARKERS IDENTIFICATION OF ACUTE MYOCARDIAL INFARCTION DISEASE BASED ON GENOMIC PROFILING THROUGH EXTREME GRADIENT BOOSTING (XGBoost)

Rohmatul Fajriyah[1,2], Havidzah Asri Isnandar [2], Adhar Arifuddin[2]
[1] Department of Statistics, Universitas Islam Indonesia, Yogyakarta, Indonesia
[2] Master Program in Statistics, Universitas Islam Indonesia, Yogyakarta, Indonesia

e-mail: *966110101@uii.ac.id*

DOI: 10.14710/medstat.17.1.69-80

Article Info:
Received: 29 November 2023
Accepted: 13 October 2024
Available Online: 14 October 2024

Keywords:
Bioinformatics; Microarray; XGBoost; Pre-processing; Filtering; Heatmap

Abstract: One disease that can cause death is Acute Myocardial Infarction (AMI). AMI, also known as a heart attack, is a condition that causes permanent damage to heart muscle tissue due to prolonged ischemia or lack of blood flow that occurs due to blockage of the epicardial coronary arteries and results in blood clots and limiting blood supply to the myocardium. During the years the young AMI patients are increasing. One of the ways to diagnose early is providing information of biomarkers related to this disease by implementing the bioinformatics data analysis. The research was conducted to look at the genomic profile of patients suffering from AMI based on without recurrent events and normal control, using the XGBoost method, due to its scalability and efficiency. Based on the grid search of tuning hyperparameters, the XGBoost method gives a classification accuracy of 88.89%, AUC 90 and kappa 0.7805. These results indicate that the XGBoost method can classify patients suffering from AMI well. This research has identified three genes that contribute the most to classifying AMI patients, namely calponin 2, ribosomal protein S11 and myotropin. Based on the heatmap visualization, information was obtained that the three genes are class markers without recurrent events.

1. INTRODUCTION

In today's modern era, bioinformatics has experienced significant development and plays a vital role in biology and medicine. Bioinformatics is a computing-based approach that has substantially contributed to the organization, analysis, and interpretation of information related to biological macromolecules (Fajriyah, 2021). The development of bioinformatics has had a positive impact, especially in diagnosing, predicting, and treating various diseases, including cardiovascular diseases involving the heart system.

As we know, the heart is the central organ in the cardiovascular system, functioning as the primary regulator of blood flow, which carries oxygen and nutrients and manages carbon dioxide transport throughout the body. Therefore, integrating bioinformatics in medical practice related to heart disease will help increase in-depth understanding, make

more accurate diagnoses, and develop more effective therapies to maintain human cardiovascular health (National Heart, Lung, and Blood Institute, 2022).

Acute myocardial infarction (AMI) is a common reason of the admission to the intensive care unit (Shahu et al., 2024) and a pathological condition that causes permanent damage to heart muscle tissue as a result of ongoing ischemia, which ultimately results in obstruction of the epicardial coronary arteries (Mechanic, Gavin, and Grossman (2023)). This condition has potentially fatal complications, including heart attack. Delays in diagnosis and inappropriate management will determine failure on providing appropriate therapy.

AMI usually occurs in older people, but in years the trend of young people get AMI is increasing (Lv et al. (2021); Rizk and Blankstein (2021); Dimitrova (2023); Krittanawong et al. (2023)); Sood, Singh and Gadkari (2023); Pohle et al. (2024)). Wahyuningsih et al. (2023) reported that AMI incidence in South Asian including in Indonesia is increasing. According to their study, factors that influence young patient AMI can be divided into comorbid, psychological, lifestyle and gender factors. Men have higher risk than women.

In practice, AMI patients get treatment based on their clinical outcomes assessment (Shahu et al., 2024). Recent studies show that researchers have had integrating the bioinformatics technology in their research of AMI, for instance. Li, He, Li, Gong, and Liu (2019) identified of potential molecular targets of acute myocardial infarction to understand the disease mechanism. Cheng, An and Li (2022) identified genes associated with AMI. Kumar et al. (2022) identified gene signatures that involved in cardio protection and Wu et al (2022) has identified the potential diagnostic biomarkers of AMI and uncover the immune cell infiltration profile of AMI.

These genomics-related studies are useful approach to provide a better care for AMI patient, especially of young patient. As Wahyuningsih et al. (2023) have highlighted, for young patient, myocardial infarction is a very damaging and detrimental disease. Because of very significant morbidity, psychological and financial burden of patient and families. This disease also very impactful on them, particularly of unmarried one, due to the risk of not having children.

Therefore, in this paper we built a classification model for the early detection AMI based on the patient genomic profiling such that the model can be used to prevent the failure of providing the appropriate treatment for them. The second part of the paper describes the literature review of AMI and XGBoost methods. It is followed by material and method; the results and discussion and the last section is conclusion and remarks.

2. LITERATURE REVIEW

Research about AMI and XGBoost have been conducted extensively, globally and nationally. AMI is a severe disease with elevated morbidity and mortality rate worldwide. It has remained a common reason for admission to the intensive care unit (ICU). In Indonesia, some studies of AMI have been conducted by Fahmi et al. (2023); and Hafni, Yaswir and Desywar (2024).

Classifying the AMI patient can be based on the clinical outcome or genomically. On the clinical outcome -based, the XGBoost classification and research related to AMI have been carried out by many researchers and it showed that XGBoost has a very good performance. For example, in Moore and Bell (2022), Li et al. (2023), Doudesis et al. (2023), and Kim et al. (2024).

The genomically-based research, where they tried to find the markers for AMI disease, have been carried out by Wang et al. (2021), Hu et al. (2022), Xue et al. (2022), Kang et al. (2023), Miao et al. (2023), Li et al. (2023), Sun et al. (2023), Yang et al. (2023) and You and Dong (2023). They implemented the AI-Machine learning related based methods such as random forest, xgboost. support vector machine, support vector machine - recursive feature elimination, lasso, decision tree, and network analysis.

In this paper, we used the XGBoost method to classify patients of acute myocardial infarction (AMI) based on their genomic profiling and then retrieved the most important genes in classifying the patients. These important genes are the gene markers for the AMI. As Hafni, Yaswir and Desywar (2024) mention that one of the criteria to diagnosis of AMI is increased cardiac biomarkers. Therefore, it is very important to find the markers of the AMI in order preventing the young AMI patient untreated appropriately.

3. MATERIAL AND METHOD

3.1. Data

The data used in this research is secondary data taken from NCBI, sourced from research by Suresh et al. (2014), which has been updated in 2019. The study used blood volume samples obtained from patients who experienced their first myocardial infarction (AMI) within 48 hours after the myocardial infarction, as well as control patients who had regular echocardiograms and no history of heart disease.

The samples consisted of blood samples collected from 52 subjects, of which 31 subjects experienced AMI and 21 were controls. The data generated from this research was then stored in the NCBI Gene Expression Omnibus (GEO) database with access number GSE48060. Patients were grouped into three classes, namely Patients with recurrent events (R), without recurrent events (WR), and Normal control (NC), with 54675 genes.

3.2. Pre-processing and Filtering

After the microarray data is obtained, before it is analyzed further, the data must go through the pre-processing and filtering stages (Gentleman et al. (2024) and Fajriyah (2021)). The pre-processing stage is crucial (Fajriyah, 2021) such that the data used in the analysis is representative and the error is minimized. There are four steps in the pre-processing, namely background correction, normalization, probe correction and summarization. In this paper we use RMA method for background correction, quantile method for normalization, no probe correction and median polish method for summarization (Baans, Jambek and Said (2019), and Federico et al. (2022))**.**

The filtering stage is a stage to reduce variables that are irrelevant and do not contribute significantly so that the classification will be more effective. There are two steps in filtering, namely non-specific and specific filterings. The non-specific filtering is filtering the genes which have low expression and variation, control probes and probe ID. The specific filtering is filtering the genes which have no different expression across class/treatment.

3.3. Analysis Method

The XGBoost classification method was used in data analysis. XGBoost is a machine learning method that combines boosting with gradient boosting (Mienye and Sun (2022)). Gradient boosting (GB) is one of several methods often used, namely the gradient descent boosting approach. The development of this algorithm includes Extreme Gradient Boosting

(XGBoost). This method is a tree ensembles algorithm with several classification and regression trees (CART) and were proven to be more effective in predicting the results of various clinical problems. In XGBoost, the boosting method is used to build the model, and when building a new model, the gradient descent method is used to reduce errors.

For a given data set with n observation of y response variable (class of object) and m feature x, suppose $f(x)$ is the function to predict y based on x. The XGBoost minimize the lost function $\hat{F}(x)$ to determine the most suitable feature. The loss function is defined as

$$\hat{F}(x) = \arg\min_f E_{x,y}\left[L(y, \hat{f}(x))\right] \quad (1)$$

where $L(y, \hat{f}(x))$ is the residual of the sample. In building the XGBoost model, the parameters used will influence how good the built model to predict (Elgeldawi, 2021). Therefore, it is crucial to select the right combination of parameters such that the accuracy will be high. The hyperparametres in the XGBoost are eta, gamma, max depth, min child weight, subsample and colsample by tree.

The procedure carried out in XGBoost is: 1) Divide the data into two, namely train data and test data, with a ratio of 80%: 20%. 2) Build an XGBoost classification model using training data and calculate classification criteria. 3) Test the model with test data and calculate classification criteria. The XGBoost works as following

1. Create a single leaf tree, which is initializing prediction probability $P_i^0, i = 1,2, \dots, n$
2. For the first tree, compute the average of response variable as prediction and calculate the residuals using the desired loss function $Residual_i^t = Y_i - P_i^{t-1}$. The subsequent trees' residuals is computed from prediction based on the previous tree.
3. Compute the similarity score $Similarity\ score = \frac{Residual^2}{Cover + \lambda}$ where, Gradient^2 is squared sum of residuals $= \sum_{i=1}^{n} Residual_i^t$, Cover is sum of previous probability *(1-previous probability) $= \sum_{i=1}^{n} P_i^t(1 - P_i^t)$ and λ is Gamma in hyperparameter.
4. Based on similarity score, select the right node and calculate information gain. $Information\ Gain = Left\ Similarity + Right\ Similarity - Similarity\ for\ Root$
5. Compute the leaf output value $Output_i = \frac{\sum_{i=1}^{n} Residual_i^t}{\sum_{i=1}^{n} P_i^t(1-P_i^t)+\lambda}$
6. Compute log odds value $Log\ odds_i^t = log\left(\frac{P_i^t}{1-P_i^t}\right)$
7. Compute predicted value = new probability $Predicted\ Value = Log\ odds_i^t + \eta\ Output_i$ where η is Eta, the learning rate.
8. Compute the probability of predicted value as $P_i^{t+1} = \frac{\exp(Predicted\ Value)}{1+\exp(Predicted\ Value)}$
9. Go back to step 1 and repeat the process for all the trees.

In this paper, we use the R packages xgboost to do the xgboost classification for the data set.

3.4. Evaluation Criteria

The model performance is assessed through confusion matrix, where the metrics such as accuracy, precision, recall/sensitivity, and specificity can be computed. Precision is a metric to measure the true positive predictions for the overall results predicted. Recall is the ratio to compare true positive predictions with the sum of true positives and false negatives. To measure the proportion of correctly predicted toward all data then the accuracy metric is implemented. In this paper we use the F1-Score to evaluate the XGBoost classification model.

Table 1. Confusion Matrix

Actual	Prediction	
	Class 1	Class 2
Class 1	TP	FN
Class 2	FP	TN

Based on confusion matrix in Table 1, then some metrics can be computed such as $\text{Accuracy} = \frac{\text{TP+TN}}{TP+TN+FP+FN} x100\%$, $\text{Precision} = \frac{\text{TP}}{TP+FP} x100\%$, $\text{Recall (Sensitivity)} = \frac{\text{TP}}{TP+FN} x100\% = AUC$, $\text{Specificity} = \frac{\text{TN}}{TN+FP} x100\%$, and $F1 - Score = 2x \frac{\text{Precision x Recall}}{\text{Precision+ Recall}}$

4. RESULTS AND DISCUSSION

4.1. Pre-processing and Filtering

In the pre-processing stage, the affyPLM package from Bioconductor is implemented. Pre-processing is carried out to minimize non-biological errors in the data, namely errors caused by technical or experimental factors, reduce noise in the data, and reduce the impact of these variables. The pre-processing carried out is background correction, normalization, and summarization. The RMA method is selected in the background correction step. Meanwhile, the methods for normalization and summarization are the quantile and median Polish methods, respectively.

Before pre-processing process, the genes expression among samples are showing distributed differently, see Figure 1. This different can be biologically different or due to some non-biological error measurement. Because of this then the pre-processing is needed to be implemented, where the results can be seen in Figure 1. The next step is filtering, at the filtering stage, the Bioconductor genefilter package is used through the nsFilter function. In this case, the selected filtering criteria is to filter genes with low IQR and variability, which are included in the control probe and have ID duplications.

Before filtering we have 54675 genes and after filtering process we obtained 10412 genes only. In this research we only have two classes, namely AMI and normal patients. Due to these two classes, it is possible that some genes (among 10412 genes) will have the same expression between two classes. Then this led to the feature selection process. Feature selection was done using the t-test, and the p-value threshold used was 0.00001. The results of the feature selection process provide 14 genes. These 14 genes used as independent variables (features) to build an XGBoost classification model.

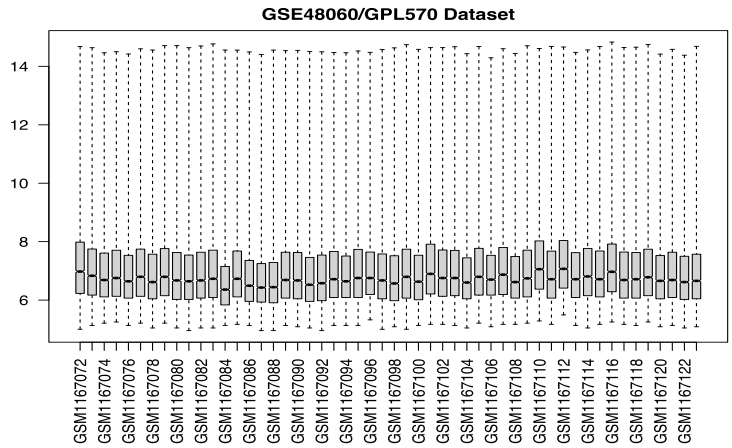

Figure 1. Gene Expression Before Pre-Processing

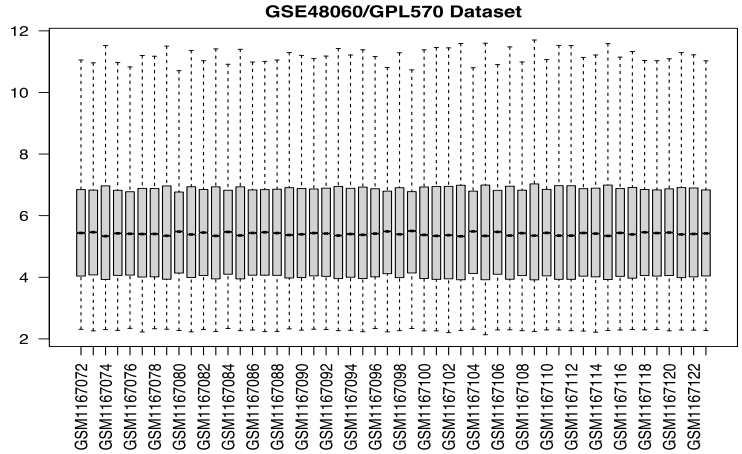

Figure 2. Gene Expression After Pre-Processing

4.2. Model Building

The hyperparameter tuning process uses the grid search method to obtain parameters that optimize the XGBoost classification model. The best parameters values are nrounds=500, eta= 0.5, max depth=4, gamma=1, subsample=0.5, min child weight = 2 and colsample by tree=1. The model produces accuracy, precision, recall, specificity, F1-Score and AUC are respectively 88.89, 100, 80, 100, 88.89 and 90. These values indicate that the model is very good to classify the AMI patients based on their genomics profile.

The model provides 16 trees where the variables X201605_x_at plays prominent role. One of the trees from the model can be seen in Figure 3. Figure 3 show that the initial leaf (root node) in the classification results of AMI patient X201605_x_at, the calponin 2 gene, and X223925_s_at, the myotrophin gene. Myotrophin is a more powerful predictor of major adverse cardiac events following acute coronary syndrome (Khan et al. (2007)), which is significantly related to AMI.

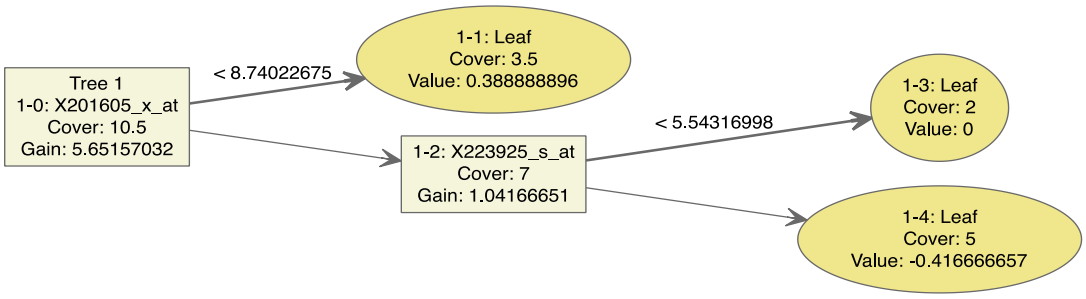

Figure 3. The XGBoost Tree Model

4.3. Feature Importance

In classifying the patients, understanding which genes play important role into it, is part of investigating which genes are considered as markers for the disease. Table 2 shows ten most important genes (features), based on the gain criterion, and their full description.

The research results in Tables 2 inline with the results obtained by Ahammad (2018). Table 2 also shows that, MPTN, also known as the myotropin gene, is related to cardiac hypertrophy and is considered a strong predictor of dangerous cardiovascular events (Khan

et al. (2007) and Ahammad, (2018)). The CNN2, according to Yang et al. (2023) has been linked to HXP to treat AMI.

Table 2. Importance Variable Annotation

Probe Id	Gain	Symbol	Entrezid	Genename
201605_x_at	0.284	CNN2	1265	calponin 2
213350_at	0.225	RPS11	6205	ribosomal protein S11
223925_s_at	0.156	MTPN	136319	myotrophin
208052_x_at	0.079	CEACAM3	1084	CEA cell adhesion molecule 3
244889_at	0.068	LOC400499	400499	putative uncharacterized protein LOC400499
220646_s_at	0.060	KLRF1	51348	killer cell lectin like receptor F1
210865_at	0.049	FASLG	356	Fas ligand
230742_at	0.049	RBM5	10181	RNA binding motif protein 5
215342_s_at	0.031	RABGAP1L	9910	RAB GTPase activating protein 1 like

Study from Qiu and Liu (2019) found that LOC400499 is related to AMI incidents and FASLG is related to recurrent AMI. According to recent studies of AMI, Jia et al.(2024) proposed that FASLG is a possible biomarker for the early diagnosis for AMI. This finding is very important since early diagnosis and treatment is very crucial.

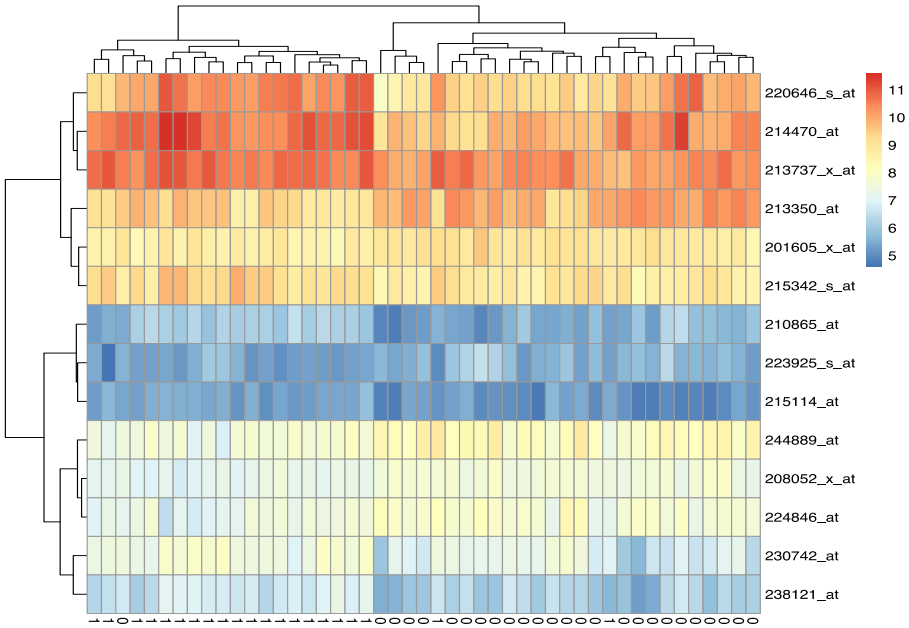

Figure 4. Heatmap of Importance Variables, 0: WR, 1: NC

Further analysis of these importance genes, which is using the differentially expressed genes (DEG) and visualized them using the heatmap visualization can be seen in Figure 4. Figure 4 shows the up and down regulated genes of AMI. In total, we have 14 genes AMI related. The up-regulated genes are 208052_x_at, 213350_at, 244889_at, 224846_at, 223925_s_at, and 201605_x_at. Meanwhile, the down-regulated genes are 214470_at,215342_s_at, 220646_s_at, 213737_x_at, 230742_at, 238121_at, 215114_at, and 210865_at. The term of up and down regulated mean that the expression of this gen is higher

and lower according to statistical interpretation if we compare them based on the two patient treatment or condition.

4.4. Remarks

Acute Myocardial Infarction (AMI) or heart attack is a condition that occurs when blood flow to the heart muscle suddenly stops, usually due to blockage of the coronary arteries. In this study, the research focus is on AMI without recurrent events. It refers to cases where the patient does not experience a repeat heart attack after the first. This research was conducted to identify marker genes associated with Acute Myocardial Infarction (AMI), including without recurrent events. Some significant marker genes were found, namely 208052_x_at, 213350_at, 244889_at, 224846_at, 223925_s_at, and 201605_x_at for AMI and for normal control class is characterized by the marker genes 214470_at,215342_s_at, 220646_s_at, 213737_x_at, 230742_at, 238121_at, 215114_at, and 210865_at.

In the study, marker genes show differences in expression between patients with AMI without recurrent events and the normal control group. They provide information about disease mechanisms and help diagnose, predict, and treat. Some genes marker has identified based on the study. However, further research is needed to identify further the specific function or involvement of these genes in AMI without recurrent events.

The control (normal) class is characterized by the genes marker 214470_at,215342_s_at, 220646_s_at, 213737_x_at, 230742_at, 238121_at, 215114_at, and 210865_at. Similar to the previous marker genes, there is no information about the specific function or involvement of these genes in normal conditions or AMI without recurrent events. These findings indicate differences in gene expression between patients with AMI without recurrent events and the control group (normal). However, further research is needed to understand deeply the role and mechanisms of these genes in AMI without recurrent events.

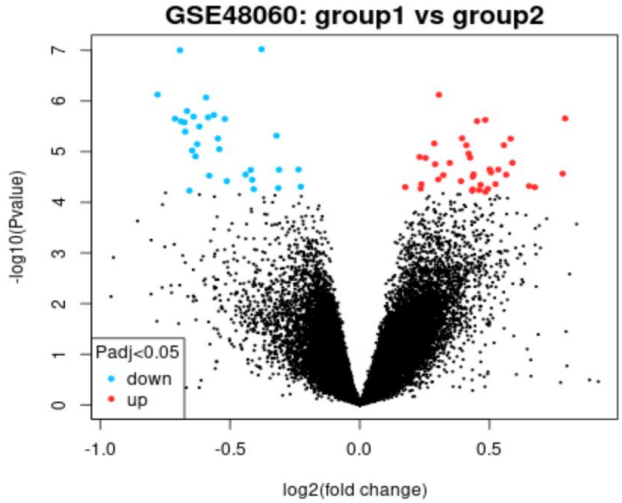

Figure 5. Down and Up Regulated Genes

NCBI has provided an online analysis for the available data sets in its website, GEO2R. The analysis is limited to the t.test and or one-way anova. In comparison toward our research and GEO2R, we can highlight some information as following. The GEO2R for 2 conditions, WR and Normal AMI patients, provided 51 differentially expressed genes where 13 are down regulated genes (Normal) and 38 are up regulated genes (AMI). The genes in this research are part of them. Volcano plot where group 1 is Normal patient and group 2 is AMI patient, visualizing the down and up regulated genes, as in Figure 5.

5. CONCLUSION

Based on the analysis that has been carried out, with optimal grid search hyperparameter tuning values, namely nround = 500, eta = 0.5, max_depth = 4, gamma = 1, subsample = 0.5, min_child_weight = 2, and colsample_bytree = 1, the XGBoost method provides a very high classification accuracy of 100, in classifying patients into without recurrent event or normal control based on their genomic profiling. Some gene markers related to determining patients, including those without recurrent events, were identified. The genes marker of AMI are 208052_x_at, 213350_at, 244889_at, 224846_at, 223925_s_at, and 201605_x_at. The normal control class is characterized by the genes marker 214470_at,215342_s_at, 220646_s_at, 213737_x_at, 230742_at, 238121_at, 215114_at, and 210865_at. In this research, the built model did not consider imbalance classes. That is, it only used two classes out of 3 classes in the original data. Future research is expected to include all classes by considering imbalance classes in the modeling. We can also use genomics data on acute myocardial infarction patients from Indonesia.

REFERENCES

Ahammad, I. (2018). Identification of Key Proteins Associated with Myocardial Infarction using Bioinformatics and Systems Biology. Preprints.org. https://doi.org/10.20944/preprints201804.0339.v1..

Azab, A. & Elsayed, A. (2017). Acute Myocardial Infarction Risk Factors and Correlation of its Markers with Serum Lipids. *Applied Biotechnology & Bioengineering*, 3(4), 00075.

Baans, O., Jambek, A., & Said, K. A. M. (2019). Analysis of Normalization Method for DNA Microarray Data. *Asia Pacific Journal of Molecular Biology and Biotechnology*. 27(4), 30-37. https://doi.org/10.35118/apjmbb.2019.027.4.04

Cheng, M., An, S., & Li J. (2017). Identifying Key Genes Associated with Acute Myocardial Infarction. *Medicine (Baltimore)*. 96(42), e7741. https://doi.org/10.1097/MD.0000000000007741. PMID: 29049183. PMCID: PMC5662349.

Dimitrova, I. N. (2023). Acute Myocardial Infarction in Young Individuals: Demographic and Risk Factor Profile, Clinical Features, Angiographic Findings and In-Hospital Outcome. *Cureus,* 15(9), e45803. https://doi.org/10.7759/cureus.45803

Doudesis, D., Lee, K. K., Boeddinghaus, J., Bularga, A., Ferry, A. V., Tuck, C., Lowry, M. T. H., Lopez-Ayala, P., Nestelberger, T., Koechlin, L., Bernabeu, M. O., Neubeck, L., Anand, A., Schulz, K., Apple, F. S., Parsonage, W., Greenslade, J. H., Cullen, L., Pickering, J. W., Than, M. P., Gray, A., Mueller, C., & Mills, N. L. (2023). CoDE-ACS Investigators. Machine Learning for Diagnosis of Myocardial Infarction using Cardiac Troponin Concentrations. *Nat Med*. 29(5), 1201-1210. https://doi.org/10.1038/s41591-023-02325-4. Epub 2023 May 11. PMID: 37169863. PMCID:PMC10202804.

Elgeldawi, A.S. E. (2021). Hyperparameter Tuning for Machine Learning Algorithms. *Arabic Sentiment Analysis. Informatics*, 8(4), 79. https://doi.org/10.3390/informatics8040079

Fajriyah, R. (2021). Paper Review: An Overview of Microarray Technologies. *Bulletin of Applied Mathematics and Mathematics Education*, 1(1), 21. https://doi.org/10.12928/bamme.v1i1.3854

Federico, A., Saarimäki, L. A., Serra, A., Del Giudice, G,. Kinaret, P. A. S., Scala, G., & Greco, D. (2022). Microarray Data Pre-processing: From Experimental Design to Differential Analysis. Methods in Molecular Biology, 2401, 79-100. https://doi.org/10.1007/978-1-0716-1839-4_7. PMID: 34902124.

Gentleman, R., Carey, V., Huber, W., & Hahne, F. (2024). Genefilter: Methods for Filtering Genes From High-Throughput Experiments. *R package version 1.86.0*

Hafni, N., Yaswir, R., & Desywar. (2024). Overview Profile of High Sensitive Troponin I (hsTnI) Levels Based on Infarction Location in Acute Myocardial Infarction Patients: A Single Center Observational Study at Dr. M. Djamil General Hospital, Padang, Indonesia. *Bioscientia Medicina: Journal of Biomedicine and Translational Research*, 8(8), 4820-4828. https://doi.org/10.37275/bsm.v8i8.1058

Hu, Z., Liu, R., Hu, H., Ding, X., Ji, Y., Li, G., Wang, Y., Xie, S., Liu, X., & Ding, Z., 2022, Potential Biomarkers of Acute Myocardial Infarction Based on Co-Expression Network Analysis. *Experimental and Therapeutic Medicine*, 23, 162. https://doi.org/10.3892/etm.2021.11085

Jia, H. M., An, F. X., Zhang, Y., Yan, M. Z., Zhou, Y., & Bian, H. J. (2024). FASLG as a Key Member of Necroptosis Participats in Acute Myocardial Infarction by Regulating Immune Infiltration. *Cardiol Res.* 15(4), 262-274. https://doi.org/10.14740/cr1652. Epub 2024 Jul 30. PMID: 39205966; PMCID: PMC11349138.

Kang, L., Zhao, Q., Jiang, K., Yu, X., Chao, H., Yin, L., & Wang, Y. (2023). Uncovering Potential Diagnostic Biomarkers of Acute Myocardial Infarction Based on Machine Learning and Analyzing Its Relationship with Immune Cells. *BMC Cardiovasc Disord*, 23(2). https://doi.org/10.1186/s12872-022-02999-7

Khan, S. Q. Kelly,D., Quinn, P., Favies, J.E., & Ng, L.L. (2007). Myotrophin is a More Powerful Predictor of Major Adverse Cardiac Events Following Acute Coronary Syndrome Than N-Terminal Pro-B-Type Natriuretic Peptide. *Clinical Science*, 112(4), 251–256. https://doi.org/10.1042/CS20060191.

Kim, M., Kang, D., Kim, M.S., Choe, J.C., Lee, S. Ahn, J.H., Oh, J., Choi, J.H., Lee, H.C., Cha, K.S., Jang,K., Bong, W.I., Song, G., & Lee, H. (2024). Acute Myocardial Infarction Prognosis Prediction with Reliable and Interpretable Artificial Intelligence System. *Journal of the American Medical Informatics Association*, 31(7), 1540–1550. https://doi.org/10.1093/jamia/ocae114.

Krittanawong, C., Khawaja, M., Tamis-Holland, J. E., Girotra, S., & Rao, S.V. (2023). Acute Myocardial Infarction: Etiologies and Mimickers in Young Patients. *Journal of the American Heart Association*, 12(18), e029971. https://doi.org/10.1161/JAHA.123.029971.

Kumar, S., Shih, C. M. , Tsai, L. W., Dubey, R., Gupta, D., Chakraborty, T., Sharma, N., Singh, A. V., Swarup, V., & Singh, H. N. (2022). Transcriptomic Profiling Unravels Novel Deregulated Gene Signatures Associated with Acute Myocardial Infarction: A Bioinformatics Approach. *Genes (Basel).* 13(12), 2321. https://doi.org/10.3390/genes13122321. PMID: 36553589; PMCID: PMC9777571.

Li, H., Sun, X., Li, Z., Zhao, R., Li. M. & Hu, T. (2023). Machine Learning-Based Integration Develops Biomarkers Initial the Crosstalk Between Inflammation and Immune in Acute Myocardial Infarction Patients. *Front. Cardiovasc. Med.* 9, 1059543. https://doi.org/10.3389/fcvm.2022.1059543.

Li, Y., He, X. N., Li, C., Gong, L., & Liu, M. (2019). Identification of Candidate Genes and MicroRNAs for Acute Myocardial Infarction by Weighted Gene Coexpression Network Analysis. *Biomed Res Int.*, 11, 5742608. https://doi.org/10.1155/2019/5742608. PMID: 30886860; PMCID: PMC6388335.

Li, Y. W., Yang, X. D., Cai, Y. C., & Kong, X. Y. (2023). Artificial Intelligence-based Prediction of Acute Myocardial Infarction Mortality Risk, *2nd International Conference on Health Big Data and Intelligent Healthcare (ICHIH)*, 164-169. https://doi.org/10.1109/ICHIH60370.2023.10396320.

Lv, J., Ni, L., Liu, K., Gao, X., Yang, J., Zhang, X., Ye, Y., Dong, Q., Fu, R., Sun, H., Yan, X., Zhao, Y., Wang, Y., Yang, Y., & Xu, H. (2021). Clinical Characteristics, Prognosis, and Gender Disparities in Young Patients With Acute Myocardial Infarction. *Frontiers in Cardiovascular Medicine*, 8, 720378. https://doi.org/10.3389/fcvm.2021.720378

Mechanic. O. J., Gavin, M., & Grossman, S. A. (2023). Acute Myocardial Infarction. 2023 Sep 3. Treasure Island (FL): StatPearls Publishing. PMID: 29083808.

Miao, M., Cao, S., Tian, Y. , Liu, D., Chen, L., Chai, Q., Wei, M., Sun, S., Wang, L., Xin, S., Liu, G., & Zheng, M. (2023). Potential Diagnostic Biomarkers: 6 Cuproptosis- and Ferroptosis-Related Genes Linking Immune Infiltration in Acute Myocardial Infarction. *Genes Immun*, 24, 159–170. https://doi.org/10.1038/s41435-023-00209-8

Moore, A. & Bell, M. (2022). XGBoost, A Novel Explainable AI Technique, in the Prediction of Myocardial Infarction: A UK Biobank Cohort Study. *Clin Med Insights Cardiol*., 8(16), 11795468221133611. https://doi.org/10.1177/11795468221133611. PMID: 36386405; PMCID: PMC9647306.

Mienye, I. D. & Sun, Y. (2022). A Survey of Ensemble Learning: Concepts, Algorithms, Applications, and Prospects. *IEEA Access*, 29-49.

National Heart, Lung, and Blood Institute. (2022). *How The Heart Works*: https://www.nhlbi.nih.gov/health/heart

Pohle, C. L., Ferreira, J. S., Coelho, R. A., Quintal, J., Bernardes, P., Esteves, A. F., Sa, C., & Seixo, F. (2024). Young Patients with Acute Myocardial Infarction: Early Screening is Warranted. *European Journal of Preventive Cardiology*, 31(1), zwae175.091. https://doi.org/10.1093/eurjpc/zwae175.091.

Qiu, L. & Liu, X. (2019). Identification of Key Genes Involved in Myocardial Infarction. *Eur J Med Res.* 24(1), 22. https://doi.org/10.1186/s40001-019-0381-x. PMID: 31269974; PMCID: PMC6607516.

Rizk T. & Blankstein R. (2021). Not All Heart Attacks are Created Equal: Thinking Differently About Acute Myocardial Infarction in the Young. *Methodist DeBakey Cardiovasc Journal*, 17(4), 60-67. https://doi.org/10.14797/mdcvj.345

Shahu, A., Namburar, S., Banna, S., Harris, A., Schenck, C., Trejo-Paredes, C., Thomas, A., Ali, T., Carnicelli, A. P., Barnett, C. F., Solomon, M. A., & Miller, P. E. (2024). Outcomes for Mechanically Ventilated Patients With Acute Myocardial Infarction

Admitted to Medical vs Cardiac Intensive Care Units. *JACC: Advances*, 3(9), Part 1, 101199. https://doi.org/10.1016/j.jacadv.2024.101199.

Sood, A., Singh, A., & Gadkari C. (2023). Myocardial Infarction in Young Individuals: A Review Article. *Cureus*, 5(4), e37102. https://doi.org/10.7759/cureus.37102. PMID: 37168155; PMCID: PMC10166330.

Sun, Y., Zhong, N., Zhu, X., Fan, Q., Li, K., Chen, Y., Wan, X., He, Q., & Xu, Y. (2023). Identification of Important Genes Associated with Acute Myocardial Infarction using Multiple Cell Death Patterns. *Cellular Signalling*, 112. https://doi.org/10.1016/j.cellsig.2023.110921.

Suresh, R., Li, X., Chiriac, A., Goel, K., Terzic, A., Perez-Terzic, C., & Nelson, T. J. (2014). Transcriptome from Circulating Cells Suggests Dysregulated Pathways Associated with Long-Term Recurrent Events Following First-Time Myocardial Infarction. *NCBI*, 13-21.

Wahyuningsih, I., Deni, W. M. I., Rahayu, H.T., & Pratiwi, I.D. (2023). Risk Factors for Acute Myocardial Infarction in Young Adults:Literature Review. *2nd International Conference on Medica lHealth Science, KnEMedicine*, 106–111. https://doi.org/10.18502/kme.v3i2.13042

Wang, Y., Zhang, X., Duan, M., Zhang, C., Wang, K., Feng, L., Song, L., Wu, S., & Chen, X. (2021), Identification of Potential Biomarkers Associated with Acute Myocardial Infarction by Weighted Gene Coexpression Network Analysis, *Oxidative Medicine and Cellular Longevity*, 5553811, 11 pages, https://doi.org/10.1155/2021/5553811.

Wu, Y., Jiang, T., Hua, J., Xiong, Z., Chen, H., Li, L., Peng, J., & Xiong, W. (2022). Integrated Bioinformatics-Based Analysis of Hub Genes and the Mechanism of Immune Infiltration Associated With Acute Myocardial Infarction, Frontiers in Cardiovascular Medicine, Sec. Cardiovascular Genetics and Systems Medicine, 9. https://doi.org/10.3389/fcvm.2022.831605.

Xue, J., Chen, L., Cheng, H., Song, X., Shi, Y., Li, L., Xu, R., Qin, Q., Ma, J., & Ge, J. (2022). The Identification and Validation of Hub Genes Associated with Acute Myocardial Infarction Using Weighted Gene Co-Expression Network Analysis. *J Cardiovasc Dev Dis.*, 9(1), 30. https://doi.org/10.3390/jcdd9010030. PMID: 35050240; PMCID: PMC8778825.

Yang, Q., Ding, J., Luo, Z., & Hu, P. (2023). A Bioinformatics Approach to The Identification of Hub Genes of Huo Xin Pill (HXP) for the Treatment of Acute Myocardial Infarction. *Tropical Journal of Pharmaceutical Research*, 21, 2651-2658. https://doi.org/10.4314/tjpr.v21i12.21.

You, H. & Dong, M. (2023). Identification of Immuno-Inflammation-Related Biomarkers for Acute Myocardial Infarction Based on Bioinformatics. *J Inflamm Res.*, 7(16), 3283-3302. https://doi.org/10.2147/JIR.S421196. PMID: 37576155; PMCID: PMC10417757.