Journal of Human Centered Technology

https://www.humentech.utm.my E-ISSN: 2821-3467

Prediction of Myocardial Infarction Risks Using Support Vector Machine and Hyperparameter Optimization

Kai Wei Tay[1], Jia Ee Khor[1], Nur Shaqirah Mohd Shaharudin[1], Ruvanesh Prakash[1], Husam I.E. Abushar[1], Lukman Hakim Ismail[1]*

[1]Department of Biomedical Engineering and Health Sciences, Faculty of Electrical Engineering, Universiti Teknologi Malaysia, 81310 UTM Johor Bahru, Johor, Malaysia

*Corresponding Author lukman@utm.my

Cite: https://doi.org/10.11113/humentech.v5n2.136 Research Article

Abstract:
Cardiovascular diseases, notably myocardial infarction (MI) remain as the primary cause of global mortality, with millions of deaths each year, often attributed to delayed diagnosis and intervention. This study focused on predictive modelling for MI risks using Support Vector Machine (SVM) algorithm, with its hyperparameter tuned by both GridSearchCV and RandomizedSearchCV. Data extracted from UCI Heart Disease Dataset underwent comprehensive preprocessing pipeline including normalization and Chi-Square feature selection to identify key predictors. The predictive accuracy, precision, recall, F1-score, and Area Under the Receiver Operating Characteristic Curve (AUROC) were evaluated. Both the Grid Search and Randomized Search tuned models achieved a superior accuracy of 84.21%, with the Grid Search model yielding a slightly higher AUROC of 0.9137, confirming the effectiveness of hyperparameter tuning. This outcome highlights the superior discrimination and classification performance in predicting MI. This finding demonstrates the potential of machine learning to enable timely, accurate diagnosis of cardiovascular diseases and therefore improve patient care.

Keywords: Artificial intelligence; Hyperparameter tuning; Machine learning; Myocardial infarction; Support vector machine

1. INTRODUCTION

Cardiovascular diseases, specifically myocardial infarction (MI) remain as the primary cause of global mortality, with millions of deaths each year often attributed to delayed diagnosis and intervention (1). MI is more commonly known as heart attack, caused by an obstruction of blood flow to the heart. Coronary artery blockage is a critical condition which may lead to severe complications or fatality if not promptly treated (2). Traditional diagnostic methods are often limited by static thresholds and manual interpretations, which may result in delayed or inaccurate assessments. Leveraging machine learning (ML), especially Support Vector Machines (SVM), provides a promising solution for predicting MI risks by analyzing complex patient datasets (3). This study employs the UCI Heart Disease dataset to develop an SVM-based predictive model, enhanced through hyperparameter optimization techniques, to achieve high accuracy and reliability. This method shows how ML can revolutionize contemporary healthcare systems by supporting individualized treatment plans and enabling early detection (4).

1.1 Myocardial Infarction

Myocardial infarction, sometimes known as a heart attack, is a form of coronary heart disease which leads to either acute or chronic myocardial ischemia brought on by an imbalance between demand and supply of oxygen. Pathologically, it is defined as death of myocardial cell due to a sustained lack of blood flow. Initial ultrastructural alterations include reduced cellular glycogen, relaxed myofibrils, and sarcolemma disruption. These early changes are rapidly followed by abnormalities in the mitochondria (5).

Acute myocardial infarction (AMI) is a critical illness characterized by sudden blockage of coronary artery, typically caused by thrombus obstruction. This causes a severe cessation in blood flow to the myocardium, eventually resulting in heart failure and death (6). A person with acute myocardial infarction may exhibit a variety of symptoms, with chest discomfort being the most prevalent. At some acute incidents, some people do not report the usual symptoms which include chest pain or feeling of pressure on chest (7). The symptoms of MI are also influenced by age, gender, smoking, comorbidities and MI type. For instance, young patients frequently have intense, radiating chest pain, but elderly patients typically have intermittent chest pain that gradually gets worse. Women typically describe atypical symptoms, whilst men

Received 19 January 2026; Accepted 02 April 2026
Available online 06 August 2026

typically complain of conventional ones. Silent or asymptomatic MI is more common in patients with comorbidities, for example, diabetes mellitus, compared to people without such illnesses. (8).

Diagnosis of MI requires clinical evidence of ischemia, characteristic electrocardiogram (ECG) changes, imaging evidence of myocardial damage or confirmation of an intracoronary thrombus, together with a rise or fall in cardiac troponins exceeding the 99th-percentile upper reference limit. Because the ECG is the fastest and most informative initial test, current guidelines emphasize obtaining and interpreting ECG within the first 10 minutes of patient arrival. Immediate coronary angiography is recommended for patients demonstrating marked elevations of ST-segment or new left bundle branch block. For other patients, clinicians apply risk-based diagnostic and treatment strategies to ensure accurate risk stratification and timely management (9). Notably, ST-segment elevation criteria have been refined to consider differences in leads V2–V3 that are specific to age and sex. Specifically, ≥0.15 mV in females, ≥0.25 mV in males under 40, and ≥0.2 mV in males aged 40 or above while a rise of ≥0.1 mV in other leads remains diagnostic (9).

1.2 Support Vector Machine

The support vector machine is a nonlinear development of the support vector algorithm, which was developed in the 1960s as a generalized portrait approach. It has a strong foundation in statistical learning theory. It was initially developed for classification tasks but in recent years its application broadened to include regression problems as well. This method is recognized for its superior performance owing to its ability for mitigating overfitting and accurate response approximation (10).

In classification tasks, support vector machines aim to discover an optimal separating function that distinguishes between two classes based on the available training instances. Equation (1) and (2) formally define the linear and non-linear decision functions of an SVM classifier. In this context, w and b determine the optimal separating hyperplane and $K(a_i, x)$ represents the kernel function mapping the input into a higher-dimensional feature space. For linear SVM, b is the bias whereas w determines the orientation of the plane that separates from the origin. Non-linear SVM uses a function φ (x) to transfer the input vector x into the high dimensional feature space, z.

$$f(x) = w \bullet x + b \tag{1}$$

$$f(x) = \sum_{i=1}^{p} (a_i - a^*) K(a_i, x) + b \tag{2}$$

Support vector machine can be summarized through two core concepts: the sparse representation and the kernel methodology. In sparse representation, SVM requires access to the entire training dataset during model construction. Although all samples contribute to learning the decision boundary, only a portion of them, referred to as support vectors, are ultimately retained and used for prediction. These support vectors emerge from an optimization process involving Lagrangian relaxation and an objective function that balances classification error with margin maximization (11).

SVMs can handle data that is not linearly distinguishable in the original feature space attributable to this foundation for kernel methodology. A kernel function is used to indirectly transfer the input data into a higher-dimensional space prior to solving the underlying convex optimization problem. This transformation helps separate instances that overlap in the original space, allowing the SVM to construct an effective decision boundary even when no linear separator exists in the input domain (7).

1.3 Myocardial Infarction Prediction

Cardiovascular diseases remain as the primary cause of mortality in both the European Union and the United States, accounting for roughly 30% of all deaths. Developing predictive models for assessing AMI risk in patients with is therefore highly valuable. Common predictors include administrative and demographic characteristics, length of stay, diagnoses, and laboratory parameters. Based on recent literature, machine learning (ML) techniques, including Stochastic Gradient Descent (SGD), SVM, logistic regression and the data source from National Health Service (NHS) were utilized. Results show an accuracy of 80% from SGD while 81% for SVM (12).

Despite improvements in treating coronary heart disease (CHD), the risk of death during hospitalization remains high, highlighting the need for better prediction models to improve patient care using predictors of systolic blood pressure (SBP), heart rate, gender, age, height, and others. Types of machine learning used are multivariate logistic regression, stochastic gradient boosting methods and the data source are from Regional Vascular Centre of Primorsky Regional Clinical Hosp. Receiver operating characteristic (ROC) for IHM prognostic models demonstrate high accuracy with 0.85 for 1st scenario and 0.90 for 2nd scenario (13).

Predicting MI can also be done using SVM and to ensure its reliability, comparing with other machine earning technique is needed. Machine learning techniques used are plain Bayesian, linear discriminants and SVM. The predictors are laboratory test results and clinical outcomes with the data source from MIMIC-III database. Based on the findings, SVM has the highest accuracy of 92% (14). Based on the studies from Boudali, all models achieve accuracy above 73% and area under curve more than 80% (15).

As demonstrated by (16), the use of SVM to analyze differentially expressed genes (DEGs) highlights the potential of bioinformatics-driven approaches in MI diagnostics. The study identified critical risk genes like HES5 and VEGFA, achieving an impressive accuracy of 87%, underscoring the role of genetic markers in early MI prediction. This gene-focused approach can complement clinical models, bridging gaps in traditional risk assessments that rely heavily on physiological and demographic data.

Building on these findings, (17) emphasized the importance of incorporating comprehensive predictors, such as age, sex, and lab results, into Gradient Boosting Decision Tree (GBDT) models. With an AUROC of 0.91 and accuracy of 83%, the study demonstrated that GBDT models outperform traditional methods, providing clinicians with real-time insights for early intervention. These results align with (18), who highlighted the superiority of advanced ML systems like XGBoost and SVM in mortality prediction for AMI patients. By leveraging diverse predictors, including ECG patterns and demographic data, the study showcased the potential of ML to refine prognostic models and support decision-making.

Further advancing the scope, (19) investigated MI severity prediction using various ML techniques, identifying troponin levels as the most significant predictor. Their SVM model achieved an accuracy of 88.67%, underscoring critical value of incorporating key biomarkers into predictive frameworks. Similarly, (20) demonstrated that ML models, including Random Forests and Neural Networks, are capable of identifying complex, non-linear patterns among clinical variables. By outperforming traditional logistic regression, these models reveal the limitations of conventional statistical methods in addressing multifactorial nature of AMI.

A critical takeaway from these studies is the need for robust validation across diverse datasets and settings. As highlighted by (18), including longitudinal health data and testing across ethnic and regional variations can improve the generalizability of ML models. Combining insights from genetic, clinical, and demographic predictors offers a holistic approach to MI risk assessment, while advancements in computational techniques pave the path to scalable and customized healthcare solutions.

Despite substantial progress in machine learning-based MI predictions, many existing studies rely on limited preprocessing pipelines, complex feature sets, and narrow hyperparameter exploration. This work advances the field by integrating a robust, comprehensive preprocessing pipeline. The pipeline combines missing data handling, outliers' removal, feature normalization, feature selection with a systematic comparison of hyperparameter optimization methods applied to SVM. This study provides a direct, controlled evaluation of two widely adopted optimization approaches (GridSearchCV and RandomizedSearchCV) using a publicly accessible dataset (UCI Heart Disease).

2. METHODOLOGY

Figure 1 illustrates the overall methodological workflow for MI predictions. This study aims to predict myocardial infarction (MI) risks using a Support Vector Machine (SVM) model optimized with hyperparameter tuning based on the following steps: a) Data preparation includes loading the dataset, target labeling, handling missing values, removing outliers using the IQR method, and performing exploratory data analysis; b) Data Transformation which involves feature normalization and feature selection using the Chi-Square test to identify the most relevant features; c) Training of the SVM model with initial parameters, followed by hyperparameter fine-tuning using GridSearchCV and RandomizedSearchCV; and d) Performance evaluation on the test dataset, comparing metrics like accuracy, precision, recall, F1-score, as well as AUROC.

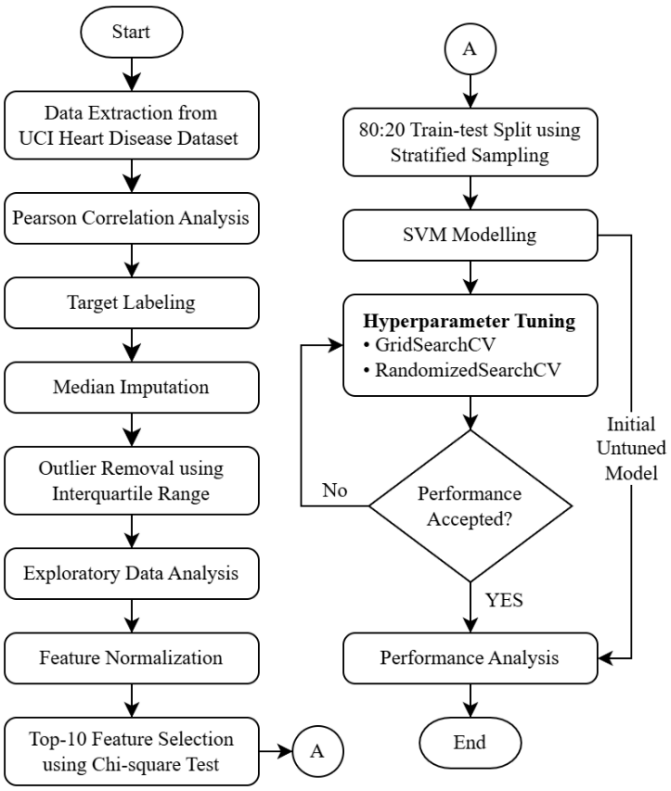

Figure 1. Overall methodological workflow of this study.

2.1 Description of Clinical Dataset

This study's dataset came from the UCI Machine Learning Repository which can be assessed by the public through online resources (21, 22). The dataset comprises 13 features critical health metrics which are age, gender (sex), chest pain type (cp), resting blood pressure (trestbps), serum cholesterol (chol), fasting blood sugar (fbs), resting ECG results (restecg), maximum heart rate achieved (thalach), exercise induced angina (exang), ST depression induced by exercise relative to rest (oldpeak), slope of the peak exercise ST segment (slope), number of major vessels colored by flourosopy (ca) and thalassemia types (thal: normal, fixed defect, reversible defect). Patients with no heart disease (value 0) and those with increasing degrees of heart disease (values 1-4) are categorized by the target label. The dataset includes approximately 284 patient records, with a mix of continuous and categorical data. These features reflect key clinical observations that are commonly analyzed by cardiologists to assess cardiovascular health and predict the risk of heart disease.

2.2 Correlation Analysis Between Features

Initially, the dataset was loaded into Google Colab environment. The magnitude and direction of linear correlations between pairs of features and label were then assessed using a correlation matrix. For this study, Pearson's correlation coefficient was employed. Coefficient closer to value 1 represents greater positive linear correlation whereas coefficient closer to value −1 represents greater negative linear correlation. In contrast, value closer to 0 reflects lower linear correlation. Based on Figure 2 which shows the correlation between features and target label, features such as ca, thal and exang show strong correlations with the target_binary, indicating high predictive value for myocardial infarction (MI). For instance, exang (0.43), ca (0.46) and thal (0.53) have strong positive correlations making these features critical for model training. In contrast, features like age, chol, fbs and restecg exhibit weak correlations with the target (~0.04–0.22), suggesting limited predictive power. This analysis is important to avoid adding redundant features as they may add noise to the model.

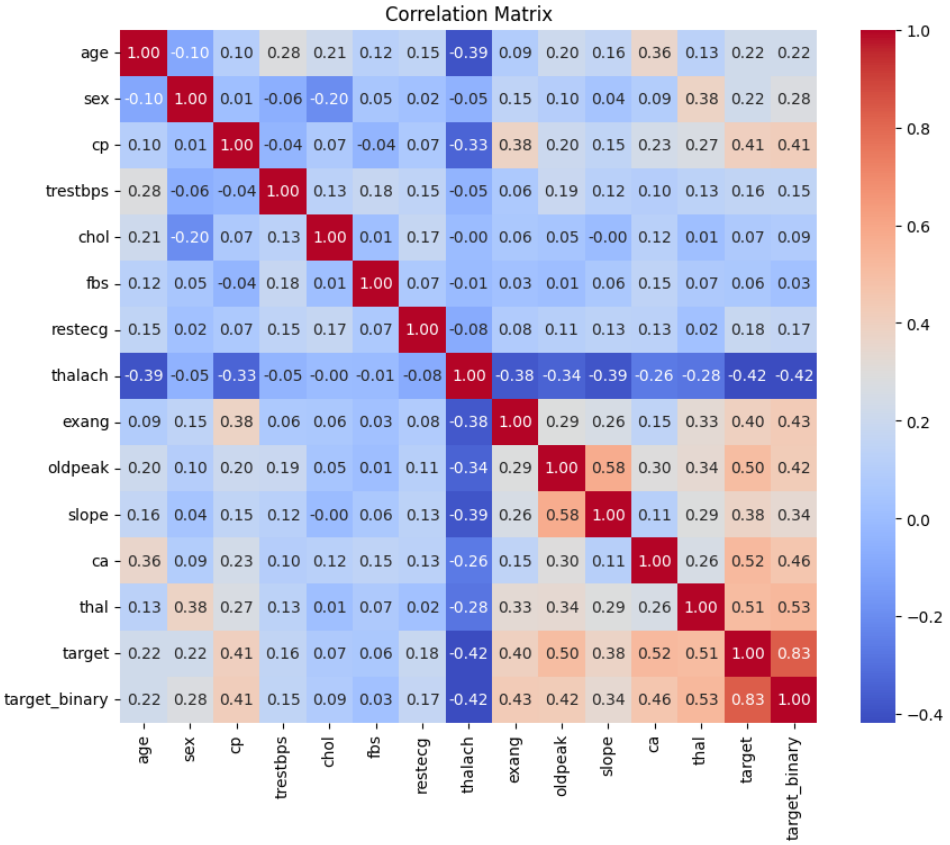

Figure 2. Correlation matrix of features and label.

2.3 Data Preprocessing

The target label originally ranged from 0 to 4, where a value of 0 indicated absence of heart disease whereas values 1 through 4 represented increasing levels of disease severity. To simplify the classification task, this label was binarized into two classes, where 0 denoted no heart disease and 1 indicated its presence.

Missing values were mitigated using median imputation, as median is less sensitive to extreme values compared to mean, thereby preserving data integrity without excluding any records. Outliers were then identified and removed using the Interquartile Range (IQR) approach (23), which finds values outside the interval [Q1−1.5*IQR, Q3+1.5*IQR] where Q1 and Q3 are the first and third quartiles, respectively. Removing outliers improves the model's ability to generalize.

2.4 Exploratory Data Analysis

The dataset includes 284 patient records, with a mix of continuous and categorical data as shown in Table 1 and Table 2. Continuous features like age, resting blood pressure, cholesterol, maximum heart rate, and ST depression provide critical physiological measurements, while categorical features such as cp, sex, thal, and ca capture discrete patient characteristics. These features were preprocessed through outlier removal, normalization, and feature selection to ensure cleaned and optimized dataset for predictive modelling and subsequent analysis.

Table 1. Exploratory data analysis of continuous features.

Description	Age	Resting Blood Pressure	Cholesterol	Maximum Heart Rate	ST Depression
Count	284	284	284	284	284
Mean	54.07	129.98	242.54	150	0.95
Standard deviation	9.12	15.45	44.73	22.66	1.03
Minimum	29.00	94.00	126.00	88.00	0.00
Quartile 1 (Q1)	47.00	120.00	210.75	133.75	0.00
Median	55.00	130.00	239.50	153.00	0.60
Quartile 3 (Q3)	60.00	140.00	271.00	168.00	1.60
Maximum	77.00	170.00	360.00	202.00	4.00

Table 2. Exploratory data analysis of categorical features.

Feature	Count (Percentage)
Sex	0 (Female): 86 (30%)
	1 (Male): 198 (70%)
Chest pain type	1 (Typical angina): 22 (8%)
	2 (Atypical angina): 49 (17%)
	3 (Non-anginal pain): 82 (29%)
	4 (Asymptomatic): 131 (46%)
Fasting blood sugar	0 (False): 244 (86%)
	1 (True (fasting blood sugar > 120 mg/dl)): 40 (14%)
Resting ECG results	0 (Normal): 144 (51%)
	1 (ST-T wave abnormality): 2 (1%)
	2 (Probable left ventricular hypertrophy): 138 (48%)
Exercise induced angina	0 (No): 194 (68%)
	1 (Yes (exercise-induced angina)): 90 (32%)
Slope of peak exercise ST segment	1 (Upsloping): 138 (48%)
	2 (Flat): 130 (46%)
	3 (Downsloping): 16 (6%)
Number of major vessels colored by flourosopy	0 (0 vessels): 170 (60%)
	1 (1 vessel): 63 (22%)
	2 (2 vessels): 35 (12%)
	3 (3 vessels): 16 (8%)
Thalassemia type	3 (Normal): 162 (57%)
	6 (Fixed defect): 17 (6%)
	7 (Reversible defect): 105 (37%)
Label	0 (No heart disease): 158 (56%)
	1 (Heart disease): 126 (44%)

2.5 Feature Normalization and Selection

The continuous features were scaled to a range of 0 to 1 to ensure all features are treated equally by the model since some might have larger values than others. Chi-Square test shown in Table 3 was used to measure the relationship between each feature and the target variable, from which the ten most significant features are chosen. This is to reduce

the data's dimensionality, avoiding potential overfitting issues. Based on Table 3 and Figure 2, both analyses show that calcium level, thalassemia and exercise-induced angina ranked among the most significant predictors of MI.

Table 3. Chi-Square score of features.

Ranking	Features	Score
1	Number of major vessels colored by flourosopy	74.5143
2	Thalassemia types	63.5311
3	Exercies induced angina	35.4703
4	ST depression induced by exercise relative to rest	14.4417
5	Chest pain type	13.8409
6	Resting ECG results	9.5972
7	Sex	8.3120
8	Slope of the peak exercise ST segment	6.6845
9	Maximum heart rate achieved	3.6187
10	Age	0.9841
11	Resting blood pressure	0.3669
12	Cholesterol	0.2538
13	Fasting blood sugar	0.1592

2.6 Model Development and Validation

To maintain the original class distribution (44% heart disease vs. 56% non-disease) while assessing the model's performance, the dataset was split into an 80% training set and a 20% testing set using stratified sampling. The initial model was trained using randomly chosen hyperparameters to establish baseline performance. To optimize the model, hyperparameter fine-tuning was performed using GridSearchCV and RandomizedSearchCV, which systematically explored combinations of hyperparameters to find the best configuration. The optimal parameters identified by GridSearchCV were C=6, degree=2, gamma=0.001, and kernel='linear'. Similarly, RandomizedSearchCV yielded the best parameters as C=8, degree=4, gamma=0.001, and kernel='linear'. These fine-tuning methods ensured improved model performance by selecting the most suitable parameter combinations for the support vector machine (SVM).

The performance of models trained with GridSearchCV and RandomizedSearchCV are compared using metrics like accuracy, precision, recall, F1-score, Receiver Operating Characteristic (ROC), and Area Under the ROC Curve (AUROC) (24). Accuracy is the ratio of correctly predicted instances to the total instances in the dataset using:

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{3}$$

Recall estimates a model's capacity to find true positives in the dataset. The model's ability to capture positive examples is shown by a higher recall. It is calculated as:

$$Recall = \frac{TP}{TP + FN} \tag{4}$$

F1 Score strikes a balance between precision and recall by taking their harmonic means. When there is an unequal distribution of classes, it is highly beneficial. The F1 Score is calculated as:

$$F1\ Score\ = \frac{Precision \times Recall}{Precision + Recall} \tag{5}$$

where Precision is defined as:

$$Precision\ = \frac{TP}{TP + FP} \tag{6}$$

The trade-off between True Positive Rate (Recall) and False Positive Rate (FPR) under varying limit values is shown graphically by ROC. The ROC curve aids in displaying a binary classifier's performance. AUROC measures a classifier's overall performance across all thresholds. AUROC ranges from 0 to 1, where value closer to 1 indicates excellent discriminatory ability whereas value of 0.5 reflects performance no better than random chance. Thus, a higher AUROC signifies a more effective and reliable predictive model (23, 24).

3. RESULTS AND DISCUSSION

This study has developed 3 MI predictive models based on different hyperparameter tuning methods. They include the initial untuned model, Grid Search tuned model and Randomized Search tuned model. In this section, their performance and the importance of features will be analyzed and presented.

The predictive capacity of the SVM model for myocardial infarction (MI) risk was rigorously examined using the held-out test set following hyperparameter optimization. Both the GridSearchCV and RandomizedSearchCV tuning processes yielded a peak accuracy of 84.21%. The superior performance was further evidenced by the high AUROC, which was recorded at 0.9137 for the model tuned with Grid Search.

The performance summary of the developed models is presented in Table 4. The table shows that the hyperparameter tuned models, which are Grid Search tuned model and Randomized Search tuned model outperformed the initial untuned model in all metrics. Based on the AUROC and ROC graph in Figure 3, Grid Search tuned model achieved a slightly higher AUROC than the Randomized Search tuned model, indicating a marginally stronger ability to separate classes. However, during the searching process, this study found that Grid Search method consumes longer processing time than Randomized Search. Despite the slightly better AUROC, the other metrics of both tuning methods are comparable. These results confirm the model's strong discriminatory power, demonstrating a robust balance between true positive and true negative predictions in identifying patients at risk of MI.

In addition, Figure 4 shows the confusion matrix of the Grid Search tuned model and the Randomized Search tuned model. The results obtained showed that both models generate the same confusion matrix. The models exhibited a good predictive ability based on the matrix, predicting 20 actual positives, 28 actual negatives, 4 false alarms, and 5 false negatives. However, recognizing the restricted number of samples and feature limitations of the UCI dataset, this convergence suggests the SVM models have reached a performance ceiling. Therefore, additional tuning may yield diminishing returns unless new predictors, larger datasets, or alternative machine learning algorithms are introduced.

Table 4. Performance of the developed SVM models.

Metric	Before Tuning	Grid Search	Randomized Search
Accuracy	0.7895	**0.8421**	0.8421
Precision	0.7600	**0.8333**	0.8333
Recall	0.7600	**0.8000**	0.8000
F1-score	0.7600	**0.8163**	0.8163
AUROC	0.8825	**0.9137**	0.9026

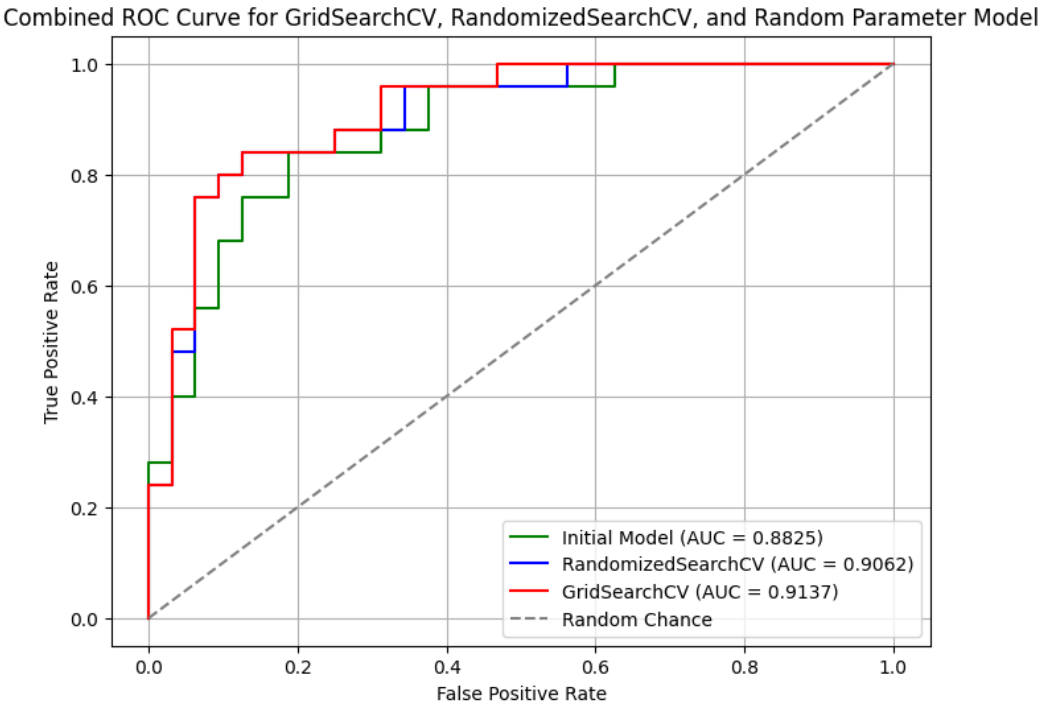

Figure 3. ROC graph of the developed SVM models.

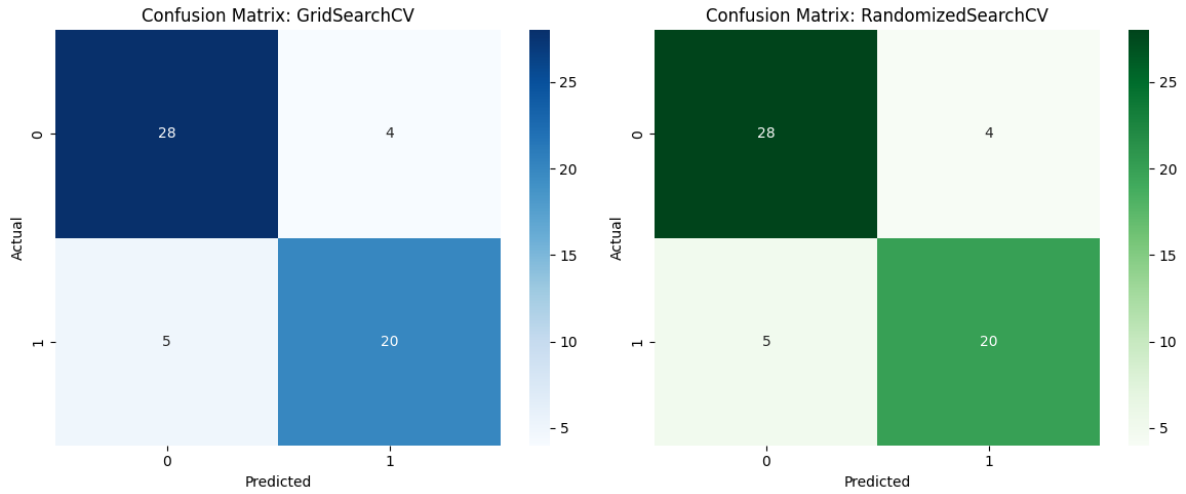

Figure 4. Confusion matrix of developed SVM models.

Robust performance of the developed SVM is significantly attributable to the structured preprocessing pipeline. Normalization ensured that no feature disproportionately influenced the SVM training process, while the Chi-Square feature selection effectively reduced the feature space. This method isolated the top 10 most relevant clinical and demographic indicators, confirming the importance of factors such as number of major vessels colored by flourosopy (Ca), Thalassemia type (Thal), and the presence of exercise-induced angina (Exang) as dominant predictors of MI risk within this dataset. The exclusion of less informative features mitigated the risk of overfitting and enhanced the model's generalization capability.

4. CONCLUSION

This study has successfully demonstrated the use of Support Vector Machine, with comprehensive hyperparameter tuning via GridSearchCV and RandomizedSearchCV to develop a high-performance predictive model for myocardial infarction risk. The integrated methodology, which included rigorous data normalization and the application of Chi-Square feature selection to focus on the most salient predictors (Ca, Thal, Exang), proved highly effective. The resulting model performed excellently, achieving accuracy of 84.21% and AUROC of 0.9137 on the test set derived from the UCI Heart Disease Dataset.

These results affirm that optimized machine learning algorithms can process complex clinical data and generate accurate, reliable risk assessments. The developed model has substantial potential to support the non-invasive, early diagnosis of cardiovascular disease and serves as a crucial tool for proposing highly individualized and timely intervention strategies. Furthermore, the model is deployable as a clinical decision-support system, serving as an additional layer of evaluation for clinicians. Because the required input variables, including heart rate, ECG features, chest pain type, and basic demographics are typically collected at the point of admission, the model is well-positioned to provide rapid risk stratification that complements clinical judgement.

Future work should focus on evaluating this model on a more extensive, multi-center dataset to examine its robustness and generalizability across diverse patient populations. Furthermore, exploring advanced machine learning algorithms (e.g., deep learning architectures and ensemble methods like Random Forest and Gradient Boosting) and integrating time-series data could offer avenues for further enhancing predictive ability.

AUTHORSHIP CONTRIBUTION STATEMENT

Kai Wei Tay: conceptualization, methodology, investigation, data curation, validation, formal analysis, visualization, writing – original draft; Jia Ee Khor: conceptualization, validation, formal analysis, writing – original draft; Nur Shaqirah Mohd Shaharudin: conceptualization, investigation, validation, visualization, writing – original draft; Ruvanesh Prakash: conceptualization, validation, writing – original draft; Husam I.E. Abushar: conceptualization, investigation, validation, writing – original draft; Lukman Hakim Ismail: conceptualization, validation, resources, supervision, writing – review and editing.

DATA AVAILABILITY

The original data are openly available in a public repository at https://doi.org/10.24432/C52P4X. The secondary data produced by this study is not available.

DECLARATION OF COMPETING INTEREST

The authors have no conflict of interest relevant to the content of this study to declare.

DECLARATION OF GENERATIVE AI

ChatGPT and Gemini were used during the preparation of this manuscript for grammar checking and language clarity. All content was carefully reviewed and edited by the authors, who take full responsibility for the integrity and accuracy of the final manuscript.

ACKNOWLEDGMENT

The authors would like to express their sincere gratitude to the creators of UCI Heart Disease dataset for their development and open access provision.

REFERENCES

(1) Fox KAA, Dabbous OH, Goldberg RJ, Pieper KS, Eagle KA, Van de Werf F, Avezum A, Goodman SG, Flather MD, Anderson Jr FA, Granger CB. Prediction of risk of death and myocardial infarction in the six months after presentation with acute coronary syndrome: Prospective multinational observational study (GRACE). BMJ 2006; 333:1091. https://doi.org/10.1136/bmj.38985.646481.55.
(2) Thygesen K, Alpert JS, Jaffe AS, Chaitman BR, Bax JJ, Morrow DA, White HD. Fourth universal definition of myocardial infarction. J Am Coll Cardiol 2018; 72:2231–2264. https://doi.org/10.1016/j.jacc.2018.08.1038.
(3) Azmi J, Arif M, Nafis MT, Alam MA, Tanweer S, Wang G. A systematic review on machine learning approaches for cardiovascular disease prediction using medical big data. Med Eng Phys 2022; 105:103825. https://doi.org/10.1016/j.medengphy.2022.103825.
(4) Bergstra J, Bengio Y. Random search for hyper-parameter optimization. J Mach Learn Res. 2012; 13:281–305.
(5) Srivastava R, Mahawar P, Banerjee M, Bagade S. Research trends in science and technology Volume I, 1st ed. Khalasa Kolhapur: Bhumi Publishing; 2023.
(6) Zhang Q, Wang L, Wang S, Cheng H, Xu L, Pei G, Wang Y, Fu C, Jiang Y, He C, Wei Q. Signaling pathways and targeted therapy for myocardial infarction. Signal Transduct Target Ther. 2022; 7(1):78. https://doi.org/10.1038/s41392-022-00925-z.
(7) Schmitz T, Harmel E, Raake P, Freuer D, Kirchberger I, Heier M, Peters A, Linseisen J, Meisinger C. Association between acute myocardial infarction symptoms and short- and long-term mortality after the event. Canadian J Cardiol. 2024; 40:1355–1366. https://doi.org/10.1016/j.cjca.2024.01.019.
(8) Abed MA, Ali RMA, Abu Ras MM, Hamdallah FO, Khalil AA, Moser DK. Symptoms of acute myocardial infarction: A correlational study of the discrepancy between patients' expectations and experiences. Int J Nurs Stud. 2015; 52:1591–1599. https://doi.org/10.1016/j.ijnurstu.2015.06.003.
(9) Vafaie M. State-of-the-art diagnosis of myocardial infarction. Diagnosis 2016; 3:137–142. https://doi.org/10.1515/dx-2016-0024.
(10) Roy A, Chakraborty S. Support vector machine in structural reliability analysis: A review. Reliab Eng Syst Saf. 2023; 233:109126. https://doi.org/10.1016/j.ress.2023.109126.
(11) Awad M, Khanna R. Support vector machines for classification. In: Awad M, Khanna R, editors. Efficient Learning Machines. Apress; Berkeley CA; 2015, p. 39–66. https://doi.org/10.1007/978-1-4302-5990-9_3.
(12) Oliveira M, Seringa J, Pinto FJ, Henriques R, Magalhães T. Machine learning prediction of mortality in acute myocardial infarction. BMC Med Inform Decis Mak. 2023; 23:70. https://doi.org/10.1186/s12911-023-02168-6.
(13) Shakhgeldyan KI, Kuksin NS, Domzhalov IG, Rublev VY, Geltser BI. Interpretable machine learning for in-hospital mortality risk prediction in patients with ST-elevation myocardial infarction after percutaneous coronary interventions. Comput Biol Med. 2024; 170:107953. https://doi.org/10.1016/j.compbiomed.2024.107953.
(14) Zhou X, Li X, Zhang Z, Han Q, Deng H, Jiang Y, Tang C, Yang L. Support vector machine deep mining of electronic medical records to predict the prognosis of severe acute myocardial infarction. Front Physiol. 2022; 13:991990. https://doi.org/10.3389/fphys.2022.991990.
(15) Boudali I, Chebaane S, Zitouni Y. A predictive approach for myocardial infarction risk assessment using machine learning and big clinical data. Healthcare Analytics. 2024; 5:100319. https://doi.org/10.1016/j.health.2024.100319.
(16) Song X, Zheng Y, Xue W, Li L, Shen Z, Ding X, Zhai Y, Zhao J. Identification of risk genes related to myocardial infarction and the construction of early SVM diagnostic model. Int J Cardiol. 2021; 328:182–190. https://doi.org/10.1016/j.ijcard.2020.12.007.
(17) Liu R, Wang M, Zheng T, Zhang R, Li N, Chen Z, Yan H, Shi Q. An artificial intelligence-based risk prediction model of myocardial infarction. BMC Bioinformatics. 2022; 23:217. https://doi.org/10.1186/s12859-022-04761-4.
(18) Wang S, Li J, Sun L, Cai J, Wang S, Zeng L, Sun S. Application of machine learning to predict the occurrence of arrhythmia after acute myocardial infarction. BMC Med Inform Decis Mak. 2021; 21:301. https://doi.org/10.1186/s12911-021-01667-8.
(19) Chen Z, Shi J, Pommier T, Cottin Y, Salomon M, Decourselle T, Lalande A, Couturier R. Prediction of myocardial infarction from patient features with machine learning. Front Cardiovasc Med. 2022; 9:754609. https://doi.org/10.3389/fcvm.2022.754609.
(20) Khera R, Haimovich J, Hurley NC, McNamara R, Spertus JA, Desai N, Rumsfeld JS, Masoudi FA, Huang C, Normand SL, Mortazavi BJ, Krumholz HM. Use of machine learning models to predict death after acute myocardial infarction. JAMA Cardiol. 2021; 6:633. https://doi.org/10.1001/jamacardio.2021.0122.
(21) Janosi A, Steinbrunn W, Pfisterer M, Detrano R. Heart Disease [Dataset]. UCI Machine Learning Repository 1989. https://doi.org/10.24432/C52P4X.

(22) Detrano R, Janosi A, Steinbrunn W, Pfisterer M, Schmid JJ, Sandhu S, Guppy KH, Lee S, Froelicher V. International application of a new probability algorithm for the diagnosis of coronary artery disease. Am J Cardiol 1989; 64:304–310. https://doi.org/10.1016/0002-9149(89)90524-9.
(23) Tay KW. Trade-off analysis between temporal window durations and performance of patient deterioration predictive models based on machine learning. Proceedings of International Exchange and Innovation Conference on Engineering & Sciences (IEICES) 2025; 11:131–136. https://doi.org/10.5109/7395508.
(24) Vujovic ŽĐ. Classification model evaluation metrics. Int J Adv Comput Sci Appl. 2021; 12:599–606. https://doi.org/10.14569/IJACSA.2021.0120670.