Original Article

Healthc Inform Res. 2023 April;29(2):120-131.
https://doi.org/10.4258/hir.2023.29.2.120
pISSN 2093-3681 • eISSN 2093-369X

HIR
Healthcare Informatics Research

Data Modeling Using Vital Sign Dynamics for In-hospital Mortality Classification in Patients with Acute Coronary Syndrome

Sarawuth Limprasert[1,2], Ajchara Phu-ang[2]

[1]Division of Cardiology, Department of Medicine, Phramongkutklao Hospital, Bangkok, Thailand
[2]College of Innovation, Thammasat University, Bangkok, Thailand

Objectives: This study compared feature selection by machine learning or expert recommendation in the performance of classification models for in-hospital mortality among patients with acute coronary syndrome (ACS) who underwent percutaneous coronary intervention (PCI). **Methods:** A dataset of 1,123 patients with ACS who underwent PCI was analyzed. After assigning 80% of instances to the training set through random splitting, we performed feature scaling and resampling with the synthetic minority over-sampling technique and Tomek link method. We compared two feature selection methods: recursive feature elimination with cross-validation (RFECV) and selection by interventional cardiologists. We used five simple models: support vector machine (SVM), random forest, decision tree, logistic regression, and artificial neural network. The performance metrics were accuracy, recall, and the false-negative rate, measured with 10-fold cross-validation in the training set and validated in the test set. **Results:** Patients' mean age was 66.22 ± 12.88 years, and 33.63% had ST-elevation ACS. Fifteen of 34 features were selected as important with the RFECV method, while the experts chose 11 features. All models with feature selection by RFECV had higher accuracy than the models with expert-chosen features. In the training set, the random forest model had the highest accuracy (0.96 ± 0.01) and recall (0.97 ± 0.02). After validation in the test set, the SVM model displayed the highest accuracy (0.81) and a recall of 0.61. **Conclusions:** Models with feature selection by RFECV had higher accuracy than those with feature selection by experts in identifying patients with ACS at high risk for in-hospital mortality.

Keywords: Acute Coronary Syndrome, Machine Learning, Mortality, Vital Sign, Data Mining

Submitted: April 21, 2022
Revised: 1st, October 11, 2022; 2nd, January 13, 2023
Accepted: February 15, 2023

Corresponding Author
Sarawuth Limprasert
Division of Cardiology, Department of Medicine, Phramongkutklao Hospital, 315 Ratchvidhi Rd. Thung Phayathai, Ratchathewi, Bangkok 10400. Thailand. Tel: +66814159199, E-mail: sarawuthlim@pmk.ac.th (https://orcid.org/0000-0002-9506-6525)

I. Introduction

Cardiovascular disease is a leading cause of death worldwide. According to a report from the World Health Organization, 17.9 million people (32% of all deaths) died from cardiovascular disease in 2019 [1]. A report on non-communicable diseases in Thailand described cardiovascular disease as the second most common cause of death, with a crude number of 32.84 deaths per 100,000 population [2]. These reports also revealed increased cardiovascular deaths in the past decades, impacting the healthcare system. Acute coronary syndrome (ACS) is one of the most common cardiovascular diseases, but the associated mortality rate remains high even

with the use of advanced therapies, particularly percutaneous coronary intervention (PCI). A mortality rate of 5% has been documented in the Thai ACS Registry [3].

One effective strategy to prevent death in patients with ACS is close monitoring and prompt treatment. Hence, identifying patients at high risk for mortality and observing them in the intensive care unit (ICU) or coronary care unit can mitigate the risk of death. However, due to the high workloads of healthcare professionals, correctly classifying patients can be challenging. Patients classified as low risk may be admitted to a regular ward, with less monitoring and delayed proper management. Thus, the misclassification of patient risk can lead to death. In each inpatient, vital signs are routinely recorded. A change in vital signs resulting from pathophysiological alteration can precede a patient's deterioration. Therefore, many early warning systems using vital signs have been developed to help physicians classify patients according to their risk [4-6]. Nevertheless, the application of these warning systems in clinical practice may be unreliable in improving clinical outcomes, perhaps due to the incorporation of few variables and infrequent recording of vital signs. In a regular ward, vital signs may be measured only twice daily. Thus, some patients must be admitted to the ICU or coronary care unit for more frequent vital sign measurements. In this context, considerable data can be retrieved for analysis, and data models have been introduced to identify patients with high risk and predict mortality. Despite the high performance of these models, deploying them in clinical practice has limitations due to issues such as the understandability of the data and metric evaluation [7]. Hence, research is required on data mining with physician guidance and model evaluation from physicians' perspectives. In this study, we compared model performance between expert assessment and machine learning techniques for feature selection.

1. Related Works

The use of vital signs as features for predictive models has been of great interest in medicine over the past decade. Bloch et al. [8] obtained electronic medical records from patients with sepsis who had been admitted to the ICU of Rabin Medical Center. Vital signs, including blood pressure, heart rate, respiratory rate, and body temperature, were recorded every 10 minutes. Using these features, the researchers developed models of the probability of sepsis within the subsequent 4 hours and compared predictive performance. A support vector machine (SVM) with a radial basis function model showed the best performance, with an area under the curve (AUC) of 88.83%. Other data mining studies incorporated factors in addition to vital signs. Kim et al. [9] included patient age and body weight in models for various machine learning algorithms to predict all-cause mortality in pediatric ICU patients. Their convolutional neural network model yielded the highest AUC (0.89–0.97) for mortality prediction between 6 and 60 hours before death. Rojas et al. [10] added patient laboratory results and treatment data to models for predicting ICU readmission. In that study, the gradient-boosting algorithm demonstrated the highest performance. The model had AUCs of 0.73 and 0.72 in the prediction of ICU readmission within 72 hours and after 72 hours, respectively. Another advanced machine learning technique was tested by Kwon et al. [11], who developed a deep learning-based early warning system using time-series data inputs. Their design consisted of three recurrent neural network layers with a long short-term memory unit. This model had a high sensitivity (24.3%) and a low false alarm rate (41.6%) for detecting patients with in-hospital cardiac arrest. Furthermore, many researchers have introduced models for predicting in-hospital mortality in patients with acute heart failure. Radhachandran et al. [12] developed several machine learning models for predicting 7-day mortality in these patients, employing features including age, sex, vital signs, and laboratory results. Their best-performing model had an area under the receiver operating characteristic curve (AUROC) of 0.84. In summary, machine learning models have been used in many healthcare contexts to optimize patient care. However, no model has been constructed to classify patients with ACS based on risk of in-hospital mortality.

II. Methods

The data mining processes in this study are summarized in Figure 1.

1. Data Collection

The Institutional Review Board of the Royal Thai Army Medical Department approved this study protocol on March 26, 2022 (No. IRBRTA 0409/2022). This cross-sectional study initially incorporated data from all patients who underwent vessel intervention in Pharmongkutklao Hospital between August 2014 and August 2021. That dataset included 3,109 patients. For the present study, 1,986 patients were excluded. The reasons for exclusion were scheduled elective coronary intervention in 1,817 patients, data loss from the procedure in 154 patients, and non-coronary artery intervention in 15 patients. Hence, we analyzed data from 1,123 patients with ACS who underwent PCI. Patient enrollment

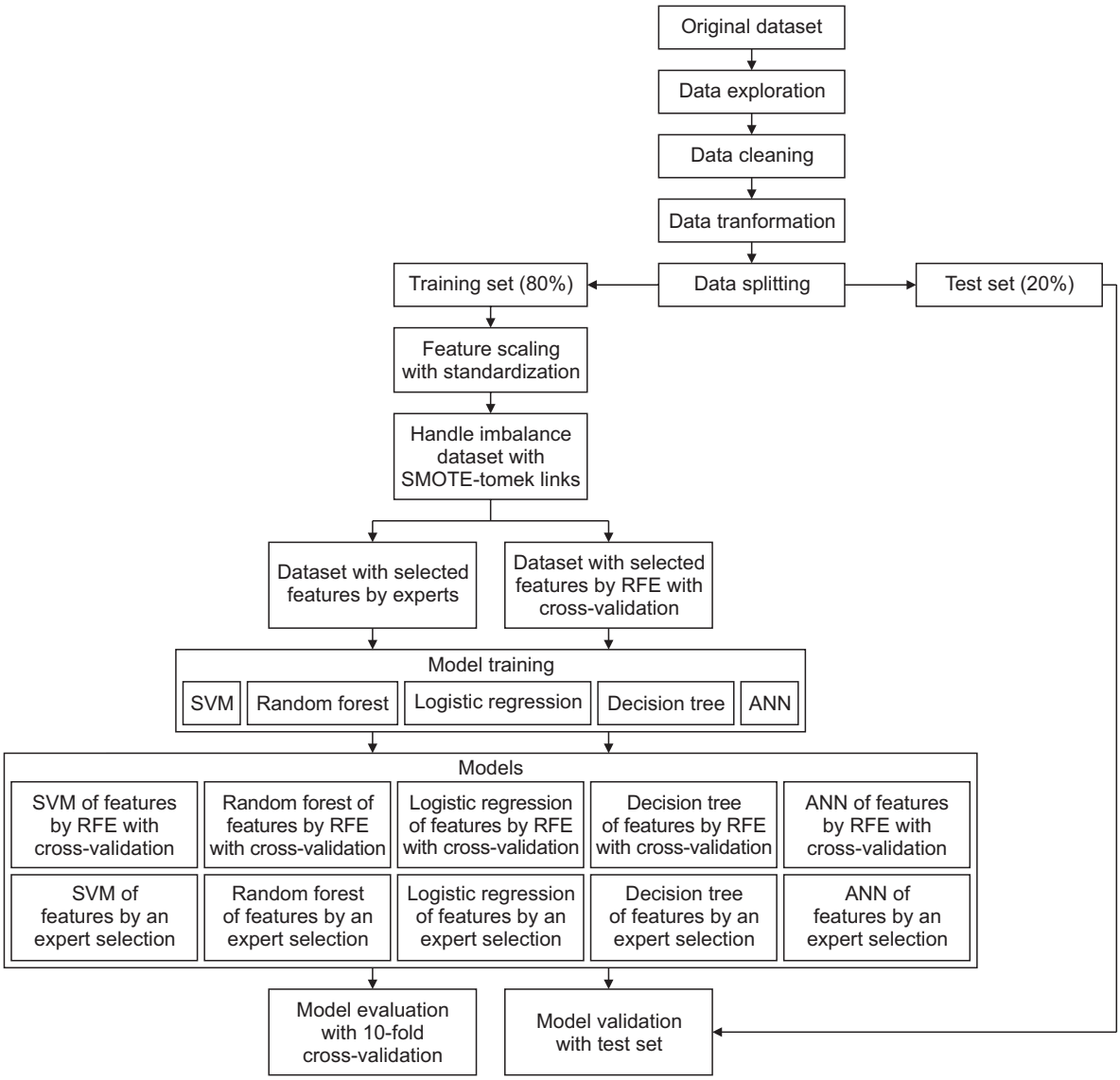

Figure 1. Summary of data mining processes. RFE: recursive feature elimination, SVM: support vector machine, ANN: artificial neural network.

is summarized in Figure 2.

2. Data Description

The data included a total of 34 features and one label. The features comprised the patients' clinical profile, creatinine level as an indicator of renal function, inotropic agent or vasopressor use, and vital signs. Patient characteristics included the diagnostic type of ACS (ST-elevation or non-ST-elevation), sex, age, and comorbidities (diabetes mellitus, hypertension, dyslipidemia, and atrial fibrillation). Vital signs included systolic blood pressure, diastolic blood pressure, pulse rate, respiratory rate, and oxygen saturation, each with five measurement periods: before starting the procedure and between minutes 0–15, 15–30, 30–45, and 45–60. The class label was binary (in-hospital death or discharge).

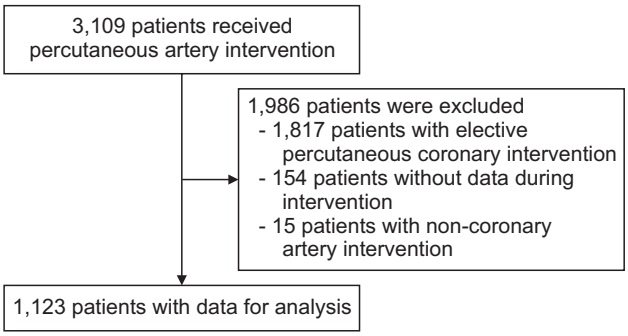

Figure 2. Patient enrollment in the dataset.

3. Libraries for Data Mining

All data mining processes and statistical analyses were executed using Google Colaboratory with Python code in the

browser. Data management was done using the Scikit-learn version 1.0.2, Pandas version 1.3.5, and NumPy version 1.21.6 libraries. Statistical analysis was performed using the SciPy version 1.7.3 library. The imbalanced-learn version 0.8.1 library was employed to handle the imbalanced labels of the dataset. The Scikit-learn library provided recursive feature elimination with cross-validation and trained conventional models. The artificial neural network model was trained using the TensorFlow version 2.9.2 library. The Matplotlib version 3.22 library was utilized to plot data visualization. Finally, the Scikit-learn library was used to evaluate the performance of all models. All hyperparameter settings were default values without tuning.

4. Data Exploration

1) Missing values

We detected and imputed the missing values in each instance. If the missing value was a patient characteristic, we replaced it with the value documented in the medical records. If the missing value was a vital sign, we assigned it the average of the values from the other measurement periods.

2) Outliers

Human error can impact data from medical records. In this study, outliers were defined as vital sign measurements that were clinically impossible. Systolic and diastolic blood pressures outside the range of 30 to 300 mmHg, pulse rates outside the range of 10 to 300 beats per minute (bpm), respiratory rates outside the range of 3 to 60 breaths per minute, and oxygen saturation levels outside the range of 60% to 100% were corrected and replaced with the values documented in the medical records.

5. Data Preparation

1) Data transformation

In machine learning, all values must be transformed into numbers before training the data. In our dataset, some features were recorded as Boolean values. For these features, we assigned a value of 1 if true and 0 if false. The transformed value of sex was 1 for male and 0 for female participants.

2) Data splitting into training and test sets

We randomly split the dataset into a training set (comprising 80% of all patients) and a test set (containing the remaining 20%). The training set was used in the model training processes, while the test set was prepared for final validation. The similarity between the training and test sets was analyzed with the Mann-Whitney U test and the chi-square test for continuous and categorical features, respectively. Table 1 details the comparison between the training set and test set.

3) Feature scaling

Feature scaling, or the calculation of distances separating data, is crucial in machine learning algorithms. To avoid any feature disproportionately influencing the model due to its large magnitude, we needed to adjust the values of all features to exist on the same scale. The medical dataset contained inherent differences in unit measurement and a wide range of values for each feature. Due to the non-normal distribution, we used the standardization method to scale the values.

4) Imbalanced dataset handling

Since a minority of patients had in-hospital mortality, our dataset was imbalanced. Model training with an imbalanced dataset can lead to unreliable performance, particularly for predicting the minority class. Thus, we utilized a combination of the synthetic minority over-sampling technique (SMOTE) and the Tomek link method to avoid poor predictive performance of the models.

5) Feature selection

We compared two methods of feature selection. In one method, features were selected based on the consensus of two interventional cardiologists specializing in ACS and PCI. The other method was recursive feature elimination with 10-fold cross-validation (RFECV). The SVM algorithm was employed in the RFECV process. This estimator was trained, and features were selected via the coefficients. The least important features were removed. This process was repeated recursively until the optimal number of features was obtained to achieve maximal recall.

6. Data Modeling

We compared the performance of five simple models: SVM, decision tree, logistic regression, random forest, and artificial neural network.

7. Performance Evaluation

We assessed three performance metrics for each model: accuracy, recall, and false-negative rate. Because we were focused on classifying patients with in-hospital mortality, the priority was recall, representing the ratio of patients predicted to have in-hospital mortality to the actual number of patients who died in-hospital. For the training set, 10-fold cross-validation was also applied to calculate the average of

Table 1. Comparison of features between the training and test sets

Feature	Training set	Test set	p-value[a]
Age (yr)	66.28 ± 12.94	65.99 ± 12.65	0.61
Sex, male	623 (69.3)	147 (65.33)	0.29
STE-ACS	603 (67.07)	143 (63.56)	0.36
Diabetes	533 (59.29)	126 (56.0)	0.41
Hypertension	654 (72.75)	155 (68.89)	0.28
Dyslipidemia	651 (72.41)	153 (68.0)	0.22
Atrial fibrillation	767 (85.32)	202 (89.78)	0.10
Inotropic agent or vasopressor use	720 (80.09)	183 (81.33)	0.74
Creatinine level	1.77 ± 1.97	1.76 ± 2.1	0.52
Vital signs at starting procedure			
SBP	149.96 ± 32.3	152.37 ± 32.22	0.58
DBP	80.59 ± 17.83	82.58 ± 17.2	0.31
PR	78.22 ± 18.7	78.89 ± 17.24	0.49
RR	17.41 ± 4.44	17.65 ± 4.07	0.40
Oxygen saturation	98.48 ± 2.76	98.58 ± 2.36	0.97
Vital signs between 0–15 min of procedure			
SBP	135.30 ± 30.22	138.00 ± 32.1	0.37
DBP	68.56 ± 15.99	68.77 ± 15.85	0.92
PR	78.32 ± 18.57	78.28 ± 16.3	0.77
RR	17.61 ± 4.15	17.77 ± 4.09	0.66
Oxygen saturation	98.41 ± 2.74	98.47 ± 2.87	0.54
Vital signs between 15–30 min of procedure			
SBP	132.22 ± 28.83	133.77 ± 30.36	0.70
DBP	67.74 ± 15.59	70.34 ± 16.16	0.05
PR	78.70 ± 18.72	77.26 ± 16.45	0.65
RR	17.55 ± 4.08	17.53 ± 4.24	0.91
Oxygen saturation	98.43 ± 2.63	98.54 ± 2.47	0.93
Vital signs between 30–45 min of procedure			
SBP	128.46 ± 28.55	130.79 ± 31.12	0.54
DBP	67.02 ± 15.48	69.28 ± 15.75	0.06
PR	78.83 ± 18.11	77.94 ± 17.44	0.66
RR	17.74 ± 3.95	17.72 ± 4.27	0.92
Oxygen saturation	98.57 ± 2.39	98.56 ± 2.82	0.76
Vital signs between 45–30 min of procedure			
SBP	128.64 ± 27.54	131.11 ± 28.0	0.39
DBP	67.54 ± 15.3	68.76 ± 14.31	0.30
PR	79.14 ± 18.65	78.48 ± 16.28	1.00
RR	17.81 ± 4.08	17.86 ± 4.05	0.85
Oxygen saturation	98.63 ± 2.20	98.38 ± 3.79	0.60

Values are presented as mean ± standard deviation or number (%).

STE-ACS: ST-elevation acute coronary syndrome, DBP: diastolic blood pressure, PR: pulse rate, RR: respiratory rate, SBP: systolic blood pressure.

[a]Using the Mann-Whitney U test for continuous data and the chi-square test for categorical features.

these metrics. Cross-validation is a common model validation technique that is used to estimate a model's performance on unseen data. It is a resampling procedure used to evaluate models with a limited data sample. As our study involved 10-fold cross-validation, the data were split into 10 groups. One group was considered the test set, and the remaining groups were defined as the training set. Then, a model was fit on the training set and evaluated on the test set. Finally, the test set was rotated until all 10 groups had been used as the test set.

III. Results

1. Baseline Characteristics

The baseline characteristics of the patients with ACS who underwent PCI are summarized in Table 2. Most patients were men (68.51%), and the average patient age was 66.22 ± 12.88 years. The sample included 378 patients (33.63%) with ST-elevation ACS. Overall, 465 (41.37%), 809 (71.98%), 804 (71.53%), and 155 (71.53%) patients had diabetes, hypertension, dyslipidemia, and atrial fibrillation, respectively. The average creatinine level (indicating renal function) was 1.77 ± 2.0 mg/dL. Inotropic agents or vasopressors were administered to 221 patients (19.66%). The total mortality among all patients was 10.4%.

2. Feature Selection

1) Features selected by experts

The experts selected 11 features, including ACS type, diabetes, atrial fibrillation, creatinine level, each systolic blood pressure level throughout the procedure, the pulse rate before the procedure, and oxygen saturation before the procedure. The selected features were ranked as summarized in Table 3.

Table 2. Baseline characteristics (n = 1,123)

Characteristic	Value
Age (yr)	66.22 ± 12.88
Sex, male	770 (68.51)
STE-ACS	378 (33.63)
Diabetes	465 (41.37)
Hypertension	809 (71.98)
Dyslipidemia	804 (71.53)
Atrial fibrillation	155 (71.53)
Inotropic or vasopressor use	221 (19.66)
Creatinine (mg/dL)	1.77 ± 2.0

Values are presented as mean ± standard deviation or number (%). STE-ACS: ST segment elevation acute coronary syndrome.

2) Features selected by RFECV

The SVM model with 15 features demonstrated the highest average recall in the RFECV process. The first-ranked features were the most important and had the maximum influence on the prediction model, while the other ranked features were less important. The 15 important features were derived from this technique. The features related to clinical characteristics were age, diabetes mellitus, hypertension, dyslipidemia, the use of inotropic agents or vasopressors, and creatinine level. Among the vital signs, the important features included systolic blood pressure before the procedure; pulse rate and respiratory rate within 0–15 minutes of the procedure; systolic blood pressure, diastolic blood pres-

Table 3. Selection and ranking of features by the RFECV method and expert assessment

	Rank	
RFECV	1	PR during 0–15 min of procedure
		SBP during 15–30 min of procedure
		DBP during 15–30 min of procedure
		PR during 15–30 min of procedure
		RR during 15–30 min of procedure
		SBP during 30–45 min of procedure
		DBP during 30–45 min of procedure
		SBP at the start of procedure
		Creatinine level
		Inotropic agent or vasopressor use
		Dyslipidemia
		Hypertension
		Diabetes
		Age
		RR during 0–15 min of procedure
Experts	1	Type of ACS
	2	SBP at the start of procedure
		SBP during 0–15 min of procedure
		SBP during 15–30 min of procedure
		SBP during 30–45 min of procedure
		SBP during 45–60 min of procedure
	3	Oxygen saturation at the start of procedure
	4	PR at the start of procedure
	5	Creatinine level
	6	Diabetes
	7	Atrial fibrillation

RFECV: recursive feature elimination with cross-validation, PR: pulse rate, RR: respiratory rate, SBP: systolic blood pressure, DBP: diastolic blood pressure, ACS: acute coronary syndrome.

sure, pulse rate, and respiratory rate within 15–30 minutes of the procedure; and systolic blood pressure and diastolic blood pressure within 30–45 minutes of the procedure.

Figure 3 demonstrates the recall of the model with various selected features in each iteration of 10-fold cross-validation. Table 3 lists selected features and their ranking using the RFECV algorithm and expert selection.

3. Model Performance: Training Set

Table 4 shows the performance metrics for each model for the training set. The models with features selected by RFECV had higher accuracy and recall than the models with features selected by experts. The random forest model with features selected by RFECV had the highest accuracy (0.96 ± 0.01) and recall (0.97 ± 0.02). Figure 4 shows a comparison of the performance metrics among models for the training set.

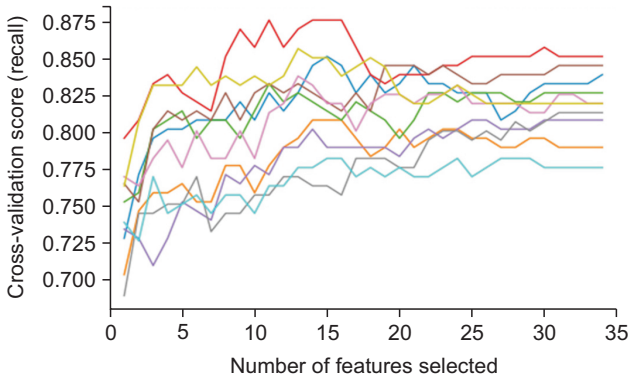

Figure 3. Recall values in recursive feature elimination with 10-fold cross-validation. Each line represents each iteration of the 10-fold cross-validation.

4. Model Performance: Test Set

Table 5 shows the performance metrics for each model for the test set. The models with features selected by the RFECV technique also demonstrated higher accuracy than those with features selected by experts. Only the SVM and logistic regression algorithms had recall values of more than 0.5. However, the SVM algorithm yielded the highest accuracy (0.81) and a recall of 0.61. Even the logistic regression algorithm showed high accuracy, although this accuracy was at most 0.80 with the lasso, ridge, or elastic net penalties for regularization. Figure 5 shows a comparison of the performance metrics among models for the test set.

IV. Discussion

This study was the first implementation of basic machine learning models for classifying the risk of in-hospital mortality of patients with ACS who underwent PCI. Two factors were important in this effort: the selection of the features for optimal model performance and the assessment of algorithm performance for patient classification.

Feature selection is a crucial process of data mining. Feature selection by a machine learning algorithm primarily differs from hand-picking by experts in how features are prioritized. While experts use their knowledge and experience to suggest clinically important features, the recursive feature elimination technique involves training a model repeatedly on a smaller subset of features and removing the least important one from the dataset. Cross-validation combined with this process can be used to assess the model's performance with different subsets of features. In general, RFECV

Table 4. Model performance metrics for the training set with 10-fold cross-validation

Model	Feature selection	Accuracy	Recall	FNR
SVM	RFECV	0.81 ± 0.02	0.81 ± 0.02	0.19 ± 0.02
	Expert selection	0.74 ± 0.04	0.74 ± 0.04	0.26 ± 0.04
Random forest	RFECV	0.96 ± 0.01	0.97 ± 0.02	0.04 ± 0.03
	Expert selection	0.94 ± 0.01	0.95 ± 0.03	0.04 ± 0.03
Decision tree	RFECV	0.91 ± 0.03	0.93 ± 0.05	0.07 ± 0.05
	Expert selection	0.88 ± 0.02	0.91 ± 0.03	0.26 ± 0.05
Logistic regression	RFECV	0.81 ± 0.03	0.81 ± 0.03	0.19 ± 0.03
	Expert selection	0.75 ± 0.04	0.74 ± 0.05	0.26 ± 0.05
ANN	RFECV	0.92 ± 0.02	0.96 ± 0.03	0.04 ± 0.03
	Expert Selection	0.85 ± 0.04	0.88 ± 0.05	0.12 ± 0.05

Values are presented as mean ± standard deviation.

FNR: false negative rate, SVM: support vector machine, ANN: artificial neural network, RFECV: recursive feature elimination with cross-validation.

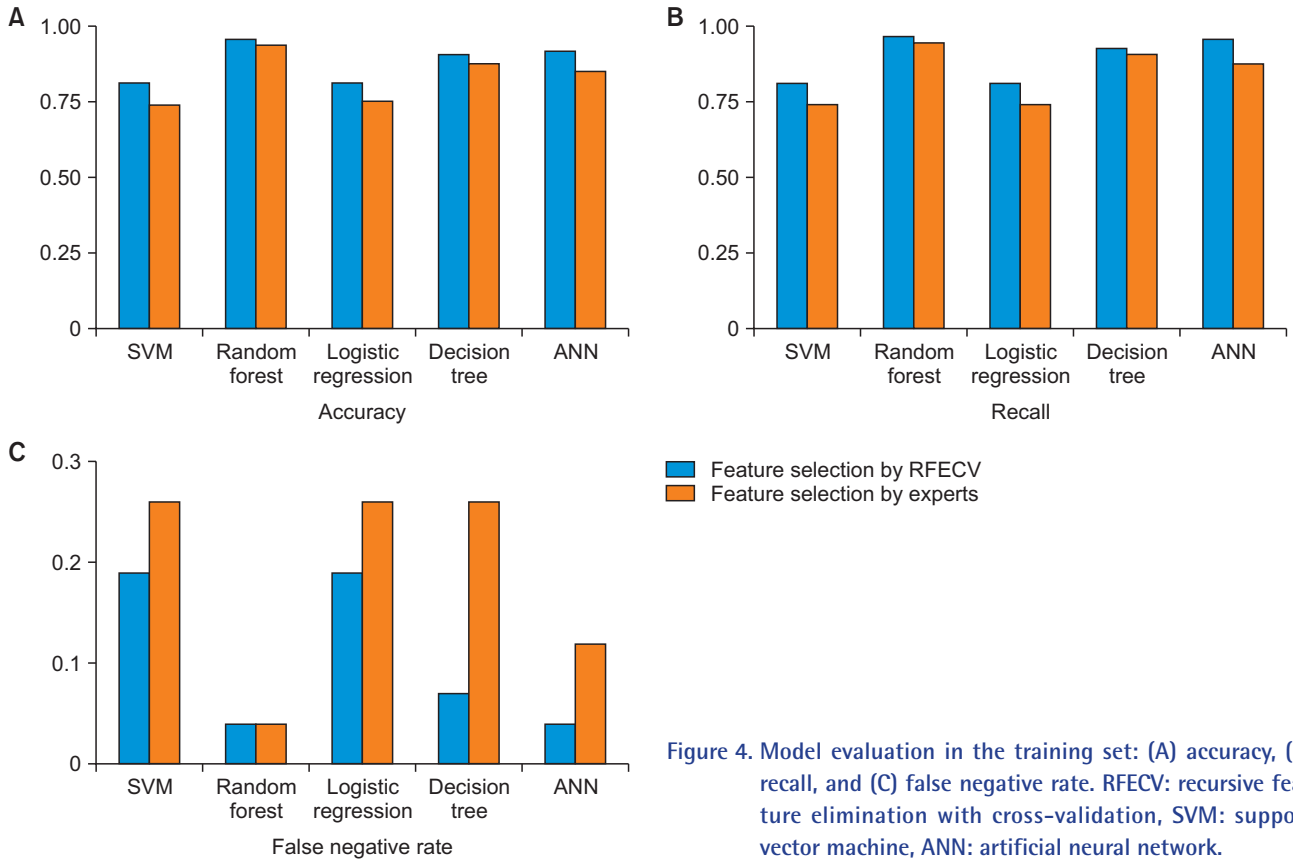

Figure 4. Model evaluation in the training set: (A) accuracy, (B) recall, and (C) false negative rate. RFECV: recursive feature elimination with cross-validation, SVM: support vector machine, ANN: artificial neural network.

Table 5. Model performance metrics for the test set

Model	Feature selection	Accuracy	Recall	FNR
SVM	RFECV	0.81	0.61	0.39
	Expert selection	0.76	0.65	0.35
Random forest	RFECV	0.88	0.30	0.70
	Expert selection	0.86	0.39	0.61
Decision tree	RFECV	0.85	0.35	0.65
	Expert selection	0.77	0.39	0.61
Logistic regression	RFECV	0.80	0.61	0.39
	Expert selection	0.77	0.65	0.35
ANN	RFECV	0.86	0.30	0.69
	Expert Selection	0.81	0.47	0.52

FNR: false negative rate, SVM: support vector machine, ANN: artificial neural network, RFECV: recursive feature elimination with cross-validation.

is a powerful technique for identifying the most informative subset of features. In contrast, features selected based on the experience and clinical knowledge of experts may vary to a greater extent due to background differences.

Many features have not been explicitly studied in clinical trials. Therefore, experts might assume that these features could affect in-hospital mortality based on similar clinical studies. However, most of the selected features are independent predictors of in-hospital mortality. Systolic blood pressure of less than 100 mmHg was reported to be a crucial predictor for elderly patients with ACS in a developing country, with a hazard ratio of 2.75 [13]. A decrease in systolic blood pressure by less than 20 mmHg from baseline has also been associated with in-hospital mortality [14]. In the present study, the experts selected systolic blood pressure in all periods from the beginning through the first hour of the procedure. Systolic blood pressure was also selected with the RFECV method, but for only certain time periods. The results align with the reasoning that systolic blood pressure in the first 15 minutes of the procedure can be influenced by multiple factors, especially the patient's initial stress level. After 45 minutes, coronary revascularization will have been performed successfully in almost all patients. In addition, sedation medication may have taken effect. Therefore, systolic blood pressures during these two intervals are less relevant and were not selected by the RFECV method. The diastolic blood pressures between minutes 15 and 30 and between minutes 30 and 45 of the procedure were selected as important features by the RFECV method. However, no clinical study has demonstrated a correlation between dia-

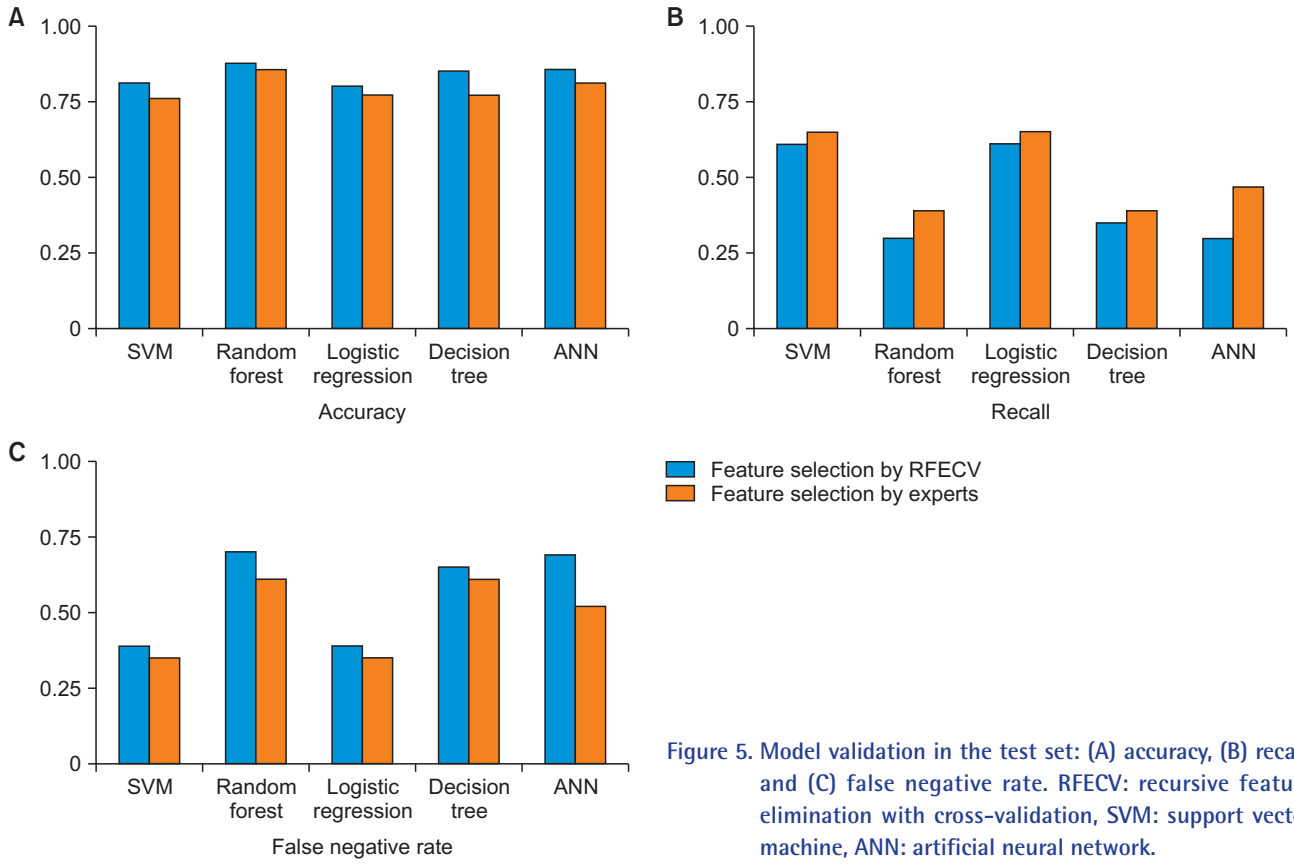

Figure 5. Model validation in the test set: (A) accuracy, (B) recall, and (C) false negative rate. RFECV: recursive feature elimination with cross-validation, SVM: support vector machine, ANN: artificial neural network.

stolic blood pressure and mortality in patients with ACS. A larger dataset of patients is needed to confirm the clinical importance of diastolic blood pressure in predicting mortality with machine learning. For pulse rate, experts selected the rate at the beginning of the procedure as important. In contrast, the RFECV method categorized pulse rates within the first 15 minutes and during minutes 15–30 of the procedure as important features. An increase in pulse rate is a crucial predictor of in-hospital mortality [14], but the exact time period and details of the dynamic change must be clinically evaluated. Furthermore, pulse rate is the most sensitive of the vital signs and can change rapidly depending on the patient's status. Respiratory rate is the vital sign that is most frequently ignored due to the lack of evidence of clinical outcomes. However, this study indicated that the respiratory rates during the first 0–15 minutes and 15–30 minutes of the procedure were important features in predicting in-hospital mortality.

Likewise, additional interpretable clinical machine learning studies should be conducted. Interestingly, oxygen saturation at the beginning of the procedure was selected by experts but not by the RFECV method. Heart failure is one of the most common complications in patients with ACS, and low oxygen saturation can be detected in patients with heart failure. Therefore, experts might assume that oxygen saturation could predict in-hospital mortality in patients with ACS. However, the fact that the RFECV method did not identify this feature may not necessarily imply that oxygen saturation level is not associated with patient severity, but rather that it is less relevant than other features.

In addition to vital signs, certain patient characteristics are important predictors of in-hospital mortality in patients with ACS. Both the experts and the RFECV algorithm selected diabetes comorbidity and renal function as important features. Multiple studies have demonstrated that impaired renal function is associated with in-hospital mortality and major bleeding [14-16], while diabetes is a traditional cardiovascular risk factor [14]. Diabetes not only causes ACS, but also increases the risk of PCI-related complications [17]. Regarding ACS type, in one large cohort of patients admitted with ACS, the mortality rate was lower among patients with ST-elevation ACS than among those with non-ST-elevation ACS [18]. Experts are generally familiar with this result. Therefore, the experts in this study selected ACS type as an important feature. Interestingly, the type of ACS was considered less relevant in the RFECV feature selection method. Again, this does not necessarily undermine the importance of the ACS type, but rather suggests that other features were more

https://doi.org/10.4258/hir.2023.29.2.120

relevant in our dataset. Patients with atrial fibrillation who are hospitalized for ACS have significantly increased risk of in-hospital mortality [19]. This explains why the experts included atrial fibrillation as an important feature, although it was unexpectedly not listed by the RFECV method. Patient age, hypertension, dyslipidemia, and use of inotropic agents or vasopressors were selected as important features by the RFECV method. However, these features were not chosen by the experts, perhaps due to the lack of a known impact on clinical outcomes. Moreover, the use of an inotropic agent or vasopressor can imply low systolic blood pressure, which the experts could have viewed as redundant with the systolic blood pressure measurements.

Based on the performance of the models, our study showed that feature selection by a machine learning algorithm using RFECV was more accurate for classification than expert assessment and was even comparable in recall when validated with the test set. The superior performance achieved with feature selection using the RFECV technique may be driven by its selection of the most important variables and eradication of the redundant and irrelevant ones, which could improve the predictive performance. This represents a slight increment in accuracy compared to models with expert-selected features but is consistent with previously developed models. Seib et al. [20] applied the super learner algorithm, which included ensemble-forward stepwise feature selection, to predict complications following thyroidectomy. This provided a modest improvement in outcome prediction as well. Another study, by Alam et al. [21], involved the use of feature-ranking-based ensemble classifiers to predict survivability among ICU patients. The implications of feature ranking can improve model performance in all datasets as well as all algorithms. Hence, we confirm that feature selection can boost classification model performance. However, the optimal method for feature selection depends on the characteristics of the dataset.

In this study, the SVM algorithm was the classification model with the highest accuracy (0.81). Furthermore, the SVM model showed slightly higher accuracy than logistic regression. A few factors may explain why an SVM classification model might perform better than a logistic regression model on this dataset. First, overfitting was not an issue for the training dataset because it was already handled by preprocessing or feature engineering methods. Therefore, regularization would not make a significant difference in performance. Moreover, our data may be sparse or include some noise, which might not affect the performance of the SVM algorithm. Lastly, logistic regression assumes a linear decision boundary and independence of features, which may not be true for some datasets. SVM does not rely on these assumptions and captures the complex boundary between classes. Logistic regression is focused on maximizing likelihood, while SVM is focused on maximizing the margin between the classes. In the present study, even basic models could produce high accuracy for classification. Nevertheless, the recall rates (the highest of which was 0.61) should be considered, as they illustrate that even that model could misclassify patients as low risk in 4 out of 10 instances. This would be a relatively high rate if the model were applied to clinical practice; nearly half of the patients at high risk for in-hospital mortality could be transferred to the regular ward, which has less intensive monitoring than the ICU. In the binary classification task of machine learning, recall is synonymous with sensitivity in diagnostic tests in clinical research. In the context of classification models for mortality prediction in critical patients, our models yielded slightly lower recall values than some previous studies. In an in-hospital cardiac arrest prediction model by Kwon et al. [11], the highest sensitivity (0.76) was obtained with the logistic regression model. Liu et al. [22] examined another predictive model for mortality in critical patients with acute kidney injury. They utilized the least absolute shrinkage and selection operator regression method for feature selection, obtaining a recall of 99.4% with the XGBoost model. The high granularity of data—including vital signs, care plans, nurse charting, disease severity, laboratory results, diagnosis, and treatment information—in their dataset likely also enhanced their model performance. Thus, extensive granularity of features, hyperparameter tuning, and more complex algorithms should be applied to each step of data mining to improve the recall associated with classification models.

In conclusion, we implemented simple machine learning algorithms with feature selection by the RFECV technique, which yielded higher accuracy than algorithms with feature selection by experts. Despite lower recall due to limitations in feature dimensionality and non-complex algorithms, further studies should be conducted, incorporating more clinical features and more complex algorithms to develop an optimal model for classifying patients with ACS.

Conflict of Interest

No potential conflict of interest relevant to this article was reported.

Acknowledgments

Our gratitude goes to all the patients and staff who participated in this study.

ORCID

Sarawuth Limprasert (https://orcid.org/0000-0002-9506-6525)
Ajchara Phu-ang (https://orcid.org/0000-0002-9498-3890)

References

1. World Health Organization. Cardiovascular diseases (CVDs) [Internet]. Geneva, Switzerland: World Health Organization; 2021 [cited at 2023 Mar 30]. Available from: https://www.who.int/news-room/fact-sheets/detail/cardiovascular-diseases-(cvds).
2. Srivanichakorn S. Morbidity and mortality situation of non-communicable diseases (diabetes type 2 and cardiovascular diseases) in Thailand during 2010-2014. Dis Control J 2017;43(4):379-90. https://doi.org/10.14456/dcj.2017.4
3. Srimahachota S, Boonyaratavej S, Kanjanavanit R, Sritara P, Krittayaphong R, Kunjara-Na-ayudhya R, et al. Thai Registry in Acute Coronary Syndrome (TRACS): an extension of Thai Acute Coronary Syndrome registry (TACS) group: lower in-hospital but still high mortality at one-year. J Med Assoc Thai 2012;95(4):508-18.
4. Cei M, Bartolomei C, Mumoli N. In-hospital mortality and morbidity of elderly medical patients can be predicted at admission by the modified early warning score: a prospective study. Int J Clin Pract 2009;63(4):591-5. https://doi.org/10.1111/j.1742-1241.2008.01986.x
5. Goldhill DR, McNarry AF, Mandersloot G, McGinley A. A physiologically-based early warning score for ward patients: the association between score and outcome. Anaesthesia 2005;60(6):547-53. https://doi.org/10.1111/j.1365-2044.2005.04186.x
6. Groarke JD, Gallagher J, Stack J, Aftab A, Dwyer C, McGovern R, et al. Use of an admission early warning score to predict patient morbidity and mortality and treatment success. Emerg Med J 2008;25(12):803-6. https://doi.org/10.1136/emj.2007.051425
7. Alam N, Hobbelink EL, van Tienhoven AJ, van de Ven PM, Jansma EP, Nanayakkara PW. The impact of the use of the Early Warning Score (EWS) on patient outcomes: a systematic review. Resuscitation 2014;85:587-94. https://doi.org/10.1016/j.resuscitation.2014.01.013
8. Bloch E, Rotem T, Cohen J, Singer P, Aperstein Y. Machine learning models for analysis of vital signs dynamics: a case for sepsis onset prediction. J Healthc Eng 2019;2019:5930379. https://doi.org/10.1155/2019/5930379
9. Kim SY, Kim S, Cho J, Kim YS, Sol IS, Sung Y, et al. A deep learning model for real-time mortality prediction in critically ill children. Crit Care 2019;23(1):279. https://doi.org/10.1186/s13054-019-2561-z
10. Rojas JC, Carey KA, Edelson DP, Venable LR, Howell MD, Churpek MM. Predicting intensive care unit readmission with machine learning using electronic health record data. Ann Am Thorac Soc 2018;15(7):846-53. https://doi.org/10.1513/AnnalsATS.201710-787OC
11. Kwon JM, Lee Y, Lee Y, Lee S, Park J. An algorithm based on deep learning for predicting in-hospital cardiac arrest. J Am Heart Assoc 2018;7(13):e008678. https://doi.org/10.1161/JAHA.118.008678
12. Radhachandran A, Garikipati A, Zelin NS, Pellegrini E, Ghandian S, Calvert J, et al. Prediction of short-term mortality in acute heart failure patients using minimal electronic health record data. BioData Min 2021;14(1):23. https://doi.org/10.1186/s13040-021-00255-w
13. Chehab O, Qannus AS, Eldirani M, Hassan H, Tamim H, Dakik HA. Predictors of in-hospital mortality in patients admitted with acute myocardial infarction in a developing country. Cardiol Res 2018;9(5):293-9. https://doi.org/10.14740/cr772w
14. Granger CB, Goldberg RJ, Dabbous O, Pieper KS, Eagle KA, Cannon CP, et al. Predictors of hospital mortality in the global registry of acute coronary events. Arch Intern Med 2003;163(19):2345-53. https://doi.org/10.1001/archinte.163.19.2345
15. Pitsavos C, Kourlaba G, Panagiotakos DB, Kogias Y, Mantas Y, Chrysohoou C, et al. Association of creatinine clearance and in-hospital mortality in patients with acute coronary syndromes: the GREECS study. Circ J 2007;71(1):9-14. https://doi.org/10.1253/circj.71.9
16. Santopinto JJ, Fox KA, Goldberg RJ, Budaj A, Pinero G, Avezum A, et al. Creatinine clearance and adverse hospital outcomes in patients with acute coronary syndromes: findings from the global registry of acute coronary events (GRACE). Heart 2003;89(9):1003-8. https://doi.org/10.1136/heart.89.9.1003
17. Bittl JA. Percutaneous coronary interventions in the diabetic patient: where do we stand? Circ Cardiovasc Interv 2015;8(4):e001944. https://doi.org/10.1161/CIRCINTERVENTIONS.114.001944
18. Terkelsen CJ, Lassen JF, Norgaard BL, Gerdes JC, Jensen

T, Gotzsche LB, et al. Mortality rates in patients with ST-elevation vs. non-ST-elevation acute myocardial infarction: observations from an unselected cohort. Eur Heart J 2005;26(1):18-26. https://doi.org/10.1093/eurheartj/ehi002

19. Dai Y, Yang J, Gao Z, Xu H, Sun Y, Wu Y, et al. Atrial fibrillation in patients hospitalized with acute myocardial infarction: analysis of the china acute myocardial infarction (CAMI) registry. BMC Cardiovasc Disord 2017;17(1):2. https://doi.org/10.1186/s12872-016-0442-9
20. Seib CD, Roose JP, Hubbard AE, Suh I. Ensemble machine learning for the prediction of patient-level outcomes following thyroidectomy. Am J Surg 2021;222(2):347-53. https://doi.org/10.1016/j.amjsurg.2020.11.055
21. Alam MZ, Masud MM, Rahman MS, Cheratta M, Nayeem MA, Rahman MS. Feature-ranking-based ensemble classifiers for survivability prediction of intensive care unit patients using lab test data. Inform Med Unlocked 2021;22:100495. https://doi.org/10.1016/j.imu.2020.100495
22. Liu J, Wu J, Liu S, Li M, Hu K, Li K. Predicting mortality of patients with acute kidney injury in the ICU using XGBoost model. PLoS One 2021;16(2):e0246306. https://doi.org/10.1371/journal.pone.0246306