International Journal of Innovative Computing, Information and Control
Volume **22**, Number **1**, February **2026**

 ISSN 1349-4198
pp. **145–159**

ENHANCED MACHINE LEARNING-BASED PREDICTION OF HEART FAILURE SURVIVAL USING FEATURE IMPROVEMENT

Minh Tuan Nguyen[1], Nhat Trang Linh Le[2] and Thang Nhat Le[3]

[1]Faculty of Telecommunications 1
[3]Postgraduate Studies Faculty
Posts and Telecommunications Institute of Technology
122 Hoang Quoc Viet Road, Cau Giay District, Hanoi 10000, Vietnam
{ nmtuan; thangln }@ptit.edu.vn

[2]UniSA Business School
The University of South Australia
101 Currie Street, Adelaide, South Australia 5000, Australia
leyny031@mymail.unisa.edu.au

Received August 2025; revised November 2025

Abstract. *Heart failure is a leading cause of death, which requires rapid identification of risk factors for accurate diagnosis to prevent mortality. In this work, we propose a novel algorithm for the detection of heart failure using a final combination of six features and a support vector machine model. The final combination is identified as the potential subset among 12 input features and the additional features. The potential subset is selected by the comparison of different feature-importance-based feature selection, including machine learning models and 5-fold cross-validation, which are then combined with the additional features computed by the fuzzy C-means clustering from the potential subset. The proposed algorithm addresses the support vector machine as a productive classifier by implementation of the grid search method in combination with a 5-fold cross-validation procedure from a set of four machine learning models. Compared with different intelligent models, the proposed model shows significant improvement of validated performance with an accuracy of 86.73%, precision of 87.67%, recall of 71.64%, and F1-score of 75.00%. The relatively high performance confirms the superiority of the proposed algorithm over the other models, which makes it become a well-fitted application in the clinic environment.*

Keywords: Machine learning, Feature selection, Heart failure, Survival prediction, Fuzzy C-means clustering

1. **Introduction.** Heart failure (HF) is a serious health problem that leads to a significant burden on morbidity, mortality, and healthcare systems worldwide with an estimated number of 55.5 million patients [1]. The HF is a consequence of the interruption of blood circulation caused by the malfunction of the heart muscle, which causes accumulation of blood. As a result, the lung fluid stagnates, weakening the human respiratory system. Certain heart disorders progressively weaken or stiffen the heart, hindering its ability to fill and pump blood correctly. The main symptoms of HF are shortness of breath during exercise or fatigue, swollen legs, and rapid heart rate when lying down [2]. Despite significant advances in medical science, the incidence of HF remains high and causes many deaths [3]. Therefore, early identification of mortality signs and prompt treatment with advice and drugs play a vital role in mortality reduction.

DOI: 10.24507/ijicic.22.01.145

Recently, machine learning (ML) and deep learning (DL) have emerged as advanced techniques to address the limitations of traditional approaches. A large number of studies in the field of medical diagnosis pay intense attention to the effective designs of ML models for the correct identification of potential markers which significantly improve the prediction of HF mortality [4]. Obviously, the utility of ML approaches produces better results in terms of time savings, high accuracy, and expert efforts to diagnose HF using relevant information and features extracted from a specific dataset [5, 6, 7]. In addition, the other powerful algorithm, DL, is also used in various existing works [8]. Unfortunately, these techniques require massive databases for effective learning processes [9]. In other words, problems involving small data volumes are insufficient for the development of DL designs. Hence, ML techniques are employed for the HF detection in this study.

Generally, the validation process has not been properly considered in previous studies, which results in unreliable performance in practical environments [10, 15]. Furthermore, due to the complex nature of HF [11], there are challenges in the early identification of features associated with mortality, which restrict detection performance. Therefore, a statistical valid manner, namely 5-fold cross-validation (CV) and effective feature engineering approaches are considered in this work to address the existing limitations of previous studies. Indeed, a novel algorithm that includes a set of risk factors called improved feature subset (IFS) and an SVM model is proposed. Firstly, feature importance-based feature selection methods using different ML models and 5-fold CV are adopted to select potential features (PFs) among all of input features (AIF), which are then combined with additional features (AFs) computed by the fuzzy C-means (FCM) clustering algorithm. Secondly, ML models are optimized by a grid search method with 5-fold CV procedure using the combinations of PFs and AFs (PFAFs) as the input. Finally, the selected ML model is estimated for its detection performance using the 5-fold CV procedure in the validation set. This framework is designed to improve the accuracy of the diagnosis of mortality in HF patients and support clinical decision-making processes.

The main contributions of this work are as follows: (i) The use of three feature ranking methods to identify the most important risk factors for the comparisons of method performance, and thereby selection of the potential factors for the proposed algorithm; (ii) Improvement of the selected risk factors by applying the FCM method to generating additional high quality features; (iii) Proposal of a novel algorithm for the detection of HF which is estimated by a 5-fold CV procedure based on a statistic method.

2. **Related Works.** Most of existing studies have focused on ML techniques in combination with effective solutions to the problem of imbalanced data to efficiently recognize the mortality of HF. Indeed, one of the most commonly used techniques is the synthetic minority over sampling technique (SMOTE) employed in [10, 11, 12]. Here, a thorough survival analysis and survival prediction are performed in [10] using the Kaplan-Meier estimation, Cox proportional hazard regression methods and ML models, respectively. Moreover, the above models are applied for the survival prediction using significant variables, which are identified from the survival analysis. Then, a performance comparison results in the highest accuracy of 83.33% for the Support Vector Machine (SVM) classifier, which is proposed as the final algorithm for the detection of HF mortality. The authors of [11] propose a reliable decision support system by integration of a sampling strategy and an ensemble learning framework using Random Forest (RF), AdaBoost (AB), K Nearest Neighbor (KNN), and SVM to assess the risk of mortality from HF. The highest accuracy of 76.25% is produced by the RF with 3 features selected as the final classifier for the detection of HF mortality. Another significant technique for the data imbalance is the SMOTE with the edited nearest neighbor as shown in [12]. Various ML algorithms

including DT, logistic regression (LR), Gaussian Naïve Bayes, RF, KNN and SVM are considered for the prediction of survival from HF. The highly experimental result generated by the RF model with an accuracy of 90% shows the superior performance among ML models. In addition, a multi-objective stacked enable hybrid model (MO-SEHM), which combines a multi-objective feature selection method, namely non-dominated sorting genetic algorithm II (NSGA-II) and SEHM as a classification method is employed in [3] using SMOTE and Tomek links to balance the input data. The proposed algorithm using nine informative features produces an accuracy of 94.87%, which is the highest performance in comparison with that of five other ML models.

The above solutions for the data imbalance have been proven their effectiveness for the improvement of the model performance. However, the bias is also associated with the use of such methods due to the generation of synthetic data, which certainly are non-existing in the actual data distributions collected from the practical environments. Therefore, these techniques are omitted in [13, 14, 15]. Indeed, the author of [13] introduces two prediction models including survival analysis related to death events using time as the target features and classification. Moreover, four feature selection methods, namely analysis of variance, Chi-square, mutual information (MI), and recursive feature elimination are implemented to select the optimal features, which are serum creatinine, ejection fraction, and sex. As a result, Gradient Boosting model for survival analysis and RF for classification are chosen with a concordance index of 0.714 and balanced accuracy of 0.74, respectively. In [14], the essential risk factors are addressed by the utility of feature ranking-based feature selection using the feature importance. Consequently, serum creatinine and ejection fraction are found to be significant risk factors for the HF mortality. Additionally, the DT model is selected as the final classifier which produces the highest accuracy of 81% compared to the others. In [15], a filter method using correlation coefficients is used to investigate the features, which are positively or negatively correlated with all-cause mortality. Then, various ML models are considered for the development of a framework to diagnose hospital HF mortality, which includes the RF model achieving a highest accuracy of 88% among seven ML models.

Feature selection plays a crucial role in identification of the most potential factors with respect to the model performance improvement. More specifically, feature ranking-based feature selection has been widely used in various studies [16, 17, 18]. In [16], different feature ranking techniques, namely MI, analysis of variance (ANOVA), and Chi-Square are evaluated by the predictive performance of the ML and DL models using a dataset of clinical indicators for heart diseases. The authors of [17] employ the ML models in combination with permutation importance (PI) and partial dependence plots to investigate the risk of 3-year all-cause mortality. Furthermore, three categories of filter, wrapper, and evolutionary algorithms including 16 feature selection techniques are implemented in [18], which result in significant improvement of the model performance using the selected feature subsets. It is noteworthy that the filter-based feature selection generates the highest quality feature subset compared with that of the other algorithms in terms of ML model performance. In [19], the convolutional neural networks (CNNs) are applied to identifying papillary thyroid cancer from histopathological images, achieving robust classification accuracy and demonstrating the growing reliability of CNN-based diagnostic models.

To improve the quality of the subset selected by the feature selection algorithms, the selected features are reinforced by a transformation technique such as FCM method. Precisely, a hybrid method of genetic algorithm (GA) and FCM in which the former plays a role of feature selection to select the most informative features, while the latter computes an additional features to make an expanded subset with better quality representing successfully the significant characteristics of the input data [20]. Moreover, the authors

of [21] propose a framework of FCM and deep neural network for the diagnosis of coronary artery disease, which achieves an accuracy of 99.91% on cardiac magnetic resonance imaging dataset.

3. **Method.** Figure 1 illustrates the proposed method, which consists of four main steps including data preprocessing, feature construction, model optimization, and model validation. In the first step, data normalization and outlier removal are implemented to improve data quality. Then, the data are divided into 50% for training and 50% for testing. Due to the small sample size, the entire dataset is applied for model validation using a 5-fold CV procedure. In the second step, different feature selection algorithms using MI based feature ranking, tree-based (TB) method, and PI in combination with ML models and 5-fold CV procedure are performed to select the most relevant PFs among AIFs, which are then put into the FCM algorithm for the generation of AFs. In the third step, a grid search with 5-fold CV is employed to identify the optimal ML models using the PFAFs which are the combination of PFs and AFs. Finally, the selected ML models are estimated of their detection performance using 5-fold CV procedure on the validation set.

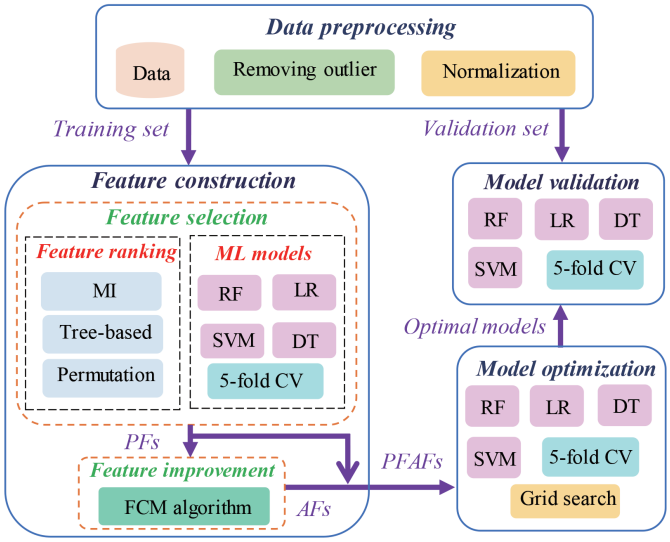


FIGURE 1. Flow of method

3.1. **Data.** We utilize a dataset of HF survival events in this work [22], including a total of 299 subjects in which 96 patients are dead and remaining are survived. The average follow-up time is 130 days, with a minimum of 4 days and a maximum of 285 days. The dataset is collected at the Faisalabad Institute of Cardiology and at the University Allied Hospital in Faisalabad (Punjab, Pakistan). Table 1 provides a descriptive comparison of clinical and demographic variables between survival and death events in patients with HF.

3.2. **Feature construction.** To identify PFAFs, the proposed framework comprises 3 stages. First, the importance of 12 input features known as AIFs is calculated by 3 methods including TB, PI, and MI. Then, the features are ranked from the highest to lowest values of the importance to construct 12 feature combination of each feature selection algorithm. Second, 4 ML models in combination with 5-fold CV procedure are utilized to evaluate these combinations to identify the PFs. In this process, the dataset is divided into five equal parts, of which four parts are used for training and one part for validation in each iteration. This process is repeated five times such that each part is used once as

TABLE 1. Comparison of variables according to HF events outcome

		Survival (N=203)	Death (N=96)	p-value
Age, mean(std)		58.76 (10.64)	65.22 (13.21)	0.0000
Anaemia, N(%)	0	120 (59.11%)	50 (52.08%)	0.3073
	1	83 (40.89%)	46 (47.92%)	0.3073
CPK, mean(std)		540.05 (753.80)	670.20 (1316.58)	0.3692
Diabetes, N(%)	0	118 (58.13%)	56 (58.33%)	1.0000
	1	85 (41.87%)	40 (41.67%)	1.0000
EF, mean(std)		40.27 (10.86)	33.47 (12.53)	0.0000
HBP, N(%)	0	137 (67.49%)	57 (59.38%)	0.2141
	1	66 (32.51%)	39 (40.62%)	0.2141
Platelets, mean(std)		266657.49 (97531.20)	256381.04 (98525.68)	0.3993
SCR, mean(std)		1.18 (0.65)	1.84 (1.47)	0.0001
SS, mean(std)		137.22 (3.98)	135.38 (5.00)	0.0019
Sex, N(%)	Female	71 (34.98%)	34 (35.42%)	1.0000
	Male	132 (65.02%)	62 (64.58%)	1.0000
Smoking, N(%)	0	137 (67.49%)	66 (68.75%)	0.9318
	1	66 (32.51%)	30 (31.25%)	0.9318
time, mean(std)		158.34 (67.74)	70.89 (62.38)	0.0000

Algorithm 1 Feature importance using MI for feature selection

1: $F = \{F_1, F_2, \ldots, F_n\}$ is set of input features; y is target; $FSimp_n$ is feature score; S_n is the combined subset of features, $n = 12$
2: **function** COMPUTEMI(F, y)
3: **for** $i \leq n$ **do**
4: $FSimp[i] \leftarrow$ Calculating the MI score for feature F_i relative to target y.
5: **return** $FSimp$
6: Feature scores: $FSimp \leftarrow$ COMPUTEMI(F, y)
7: Arrangement of the feature set F_n based on the feature score $FSimp_n$ ordered from the highest to lowest
8: $S_1 \leftarrow F_1$
9: **for** $i < n$ **do**
10: $S_{i+1} \leftarrow \{S_{(i)}, F_{i+1}\}$
11: Separation of training set into 5 folds of recodes $V(j)$
12: **for** $i \leq 5$ **do**
13: **for** $k \leq n$ **do**
14: Training ML models (RF, LG, SVM, DT) with $V(t)$, $t \neq i$ for the combined feature sets S_k
15: Calculation of accuracy of the models on $V(i)$ for the combined feature sets S_k
16: Calculation of mean accuracy
17: Identification of PFs by selecting the combination of features with the highest mean accuracy corresponding to the ML models

the validation set, ensuring that all samples are used for both training and validation, and therefore eliminating the overfitting problem. Finally, the FCM algorithm is applied for the PFs to generating 2 AFs. The PFs and AFs are then combined as the improved feature subsets used as the input of different ML models. Algorithm 1 shows the feature

selection algorithm using the MI based feature importance. The remaining 2 methods are performed similarly.

3.2.1. *Feature importance.* There are 3 different methods adopted in this work to calculate the importance of features for which the features are ranked from the highest to lowest values of the importance.

a) Mutual information (MI): This method [23] is used to quantify the dependency between the feature variables and the target variable. It calculates the extent to which the knowledge of a feature variable decreases the uncertainty about the target variable. In this research, MI is employed to evaluate and rank features according to their individual contributions to the accurate prediction of the target event. A more substantial MI score indicates a more pronounced dependency between a feature and the target outcome.

b) Tree-based (TB): The importance of features in TB models [24] is determined based on how effectively each feature splits the data to create purer nodes. In this study, we use RF to calculate feature importance, which is determined by measuring the extent to which a feature contributes to reducing impurity or increasing the amount of information gained in the model's predictions. RF calculates this importance by averaging the impurity reduction across all decision trees in the forest.

c) Permutation importance (PI): The method [25] is a model inspection technique that offers a valuable perspective on the relationship between features and the predictive power of a model. Permutation methods rely on measuring the change in value or accuracy when values of a feature are exchanged by shuffling versions of itself to determine feature relevance. In the permutation method to evaluate the importance of features, DT model is used to determine the influence of each feature on the prediction result.

3.2.2. *Feature selection.* A total of 36 feature combinations are generated by the utility of 3 feature importance based feature ranking methods, which are then fed into different ML models to estimate their classification performance in terms of HF mortality. The feature subsets used as the input of individual ML models, which produce the highest accuracy corresponding to specific feature ranking method, are selected as the PFs for the further improvement.

3.2.3. *Feature improvement.* Fuzzy C-means (FCM) clustering [26] is a soft clustering technique in which each data point can belong to clusters with multiple degree values of membership, unlike traditional hard clustering methods that assign each point to a single cluster. In this paper, the FCM algorithm is used to generate AFs from the PFs selected by different feature selection algorithms and from the AIF. The FCM algorithm utilizes these features to partition the data into two clusters. Consequently, a specific AF value ranges from 0 to 1, which represents the probability of a data point assigned to a cluster. The feature subsets are improved by the combinations of PFs, AIF and corresponded AFs known as PFAFs and AIFAFs.

3.3. **Model optimization.** A grid search-based method combined with 5-fold CV procedure to optimize ML models (RF [27], LR [28], SVM [29], and DT [30]) using the input of AIFs, PFs, AFs, PFAFs, and AIFAFs is an effective hyperparameter optimization algorithm, which searches for optimal values of the model parameters. As a result, the detection performance is improved for the selected models, which also helps avoid successfully overfitting problems.

3.4. **Model validation.** The optimal models are evaluated using all of PFAFs, AIFAFs, PFs, AFs, and AIF of the previous step and the 5-fold CV-based method on the validation set. Here, the validation set is randomly divided into 5 folds for which a fold is for testing

and others are used for training. Every fold needs to be the testing data for the completion of this process to compute the mean and standard deviation of the model performance. One of the feature combinations used as the input of an ML model, which produces the highest detection accuracy, is proposed as the final algorithm for the recognition of mortality HF.

4. Simulation Results.

4.1. Measurement.

The performances of ML models are evaluated by different measured parameters, namely accuracy (Ac), precision (Pr), recall (Re), F1-score (F1), and an area under the curve (AUC). Ac measures the percentage of patients who are correctly diagnosed. Pr and Re show the percentage of patients predicted as mortality who are truly dead out of the total predicted as HF and the fraction of patients predicted as mortality out of the total of actual mortality, respectively. F1 is calculated by the harmonic mean of Pr and Re, which provides an overall view of the performance of the models in predicting HF events. AUC is an essential evaluation statistic that evaluates the performance of a binary classification model, specifically the area under the receiver operating characteristic (ROC) curve, indicating the model's ability to distinguish between survival and mortality classes.

$$\mathrm{Ac} = \frac{\mathrm{TP} + \mathrm{TN}}{\mathrm{TP} + \mathrm{TN} + \mathrm{FP} + \mathrm{FN}} \tag{1}$$

$$\mathrm{Pr} = \frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FP}} \tag{2}$$

$$\mathrm{Re} = \frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FN}} \tag{3}$$

$$\mathrm{F1} = \frac{2 \times \mathrm{Pr} \times \mathrm{Re}}{\mathrm{Pr} + \mathrm{Re}} \tag{4}$$

where true positive (TP) and false negative (FN) represent the number of deaths, which are identified correctly and incorrectly as survivals. True negative (TN) and false positive (FP) show the number of survivals, which are addressed correctly and incorrectly as mortality, respectively.

4.2. Feature construction.

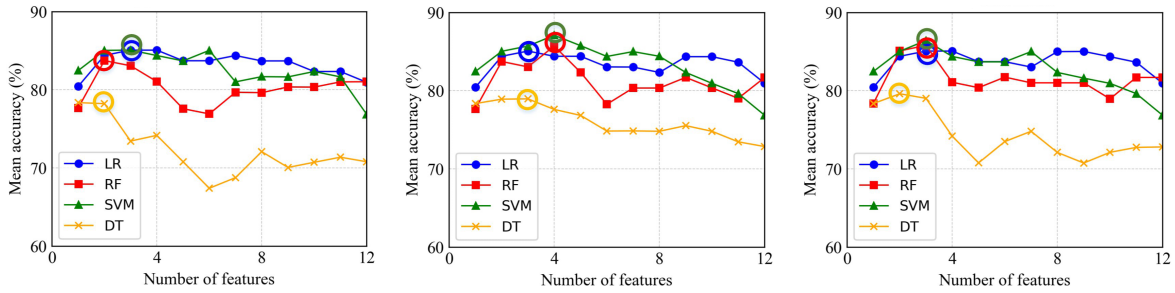
4.2.1. *Feature ranking.* Table 2 shows the features ranked by the importance values, which are generated by three algorithms including TB, PI, and MI.

4.2.2. *Feature selection.* Figure 2 illustrates the performance of various ML models using 36 feature combinations. The highest accuracy of different ML models corresponding to the input combinations, which are then named the PFs, is given as follows.

- The PFs of Time, EF, and age are selected by the LR and SVM models while remaining models identify Time and EF as the PFs using MI based feature ranking method as shown in Figure 2(a).
- A number of 3 PFs such as Time, EF, and SS selected by the PI based feature ranking method in combination with LR and DT models which is smaller than that of 4 PFs, namely Time, EF, SS, and Platelets identified by the above feature ranking combined with SVM and RF models as represented in Figure 2(b).
- TB method combined with DT model selects only 2 PFs of Time and EF while that in combination with LR, RF, and SVM models identify 3 PFs of Time, EF, and SCR as given in Figure 2(c).

TABLE 2. Feature importance based feature ranking using 3 methods including MI, PI and TB

	MI		PI		TB	
	Feature	Importance	Feature	Importance	Feature	Importance
1	time	0.227572	time	0.235374	time	0.289324
2	EF	0.121555	EF	0.221769	EF	0.170525
3	Age	0.086256	SS	0.072109	SCR	0.136987
4	SCR	0.080220	Platelets	0.058503	Age	0.099002
5	SS	0.027645	CPK	0.057143	CPK	0.081010
6	Diabetes	0.026874	Age	0.053061	Platelets	0.072576
7	Sex	0.018008	SCR	0.021769	SS	0.069798
8	Platelets	0.003035	Smoking	0.020408	Anaemia	0.031541
9	Smoking	0.000000	Anaemia	0.012245	Sex	0.013869
10	HBP	0.000000	Sex	0.000000	Smoking	0.013111
11	CPK	0.000000	HBP	0.000000	HBP	0.011345
12	Anaemia	0.000000	Diabetes	0.000000	Diabetes	0.010911



(a) Mutual information (b) Permutation importance (c) Tree-based

FIGURE 2. Evaluation of 36 feature combinations based on average accuracy of different ML models

We define the feature sets as follows: PFs1 (Time, EF); PFs2 (Time, EF, Age); PFs3 (Time, EF, SS); PFs4 (Time, EF, SS, Platelets); and PFs5 (Time, EF, SCR). Moreover, the additional features are named as AFs1, AFs2, AFs3, AFs4, and AFs5 corresponding to the above feature sets, respectively, while AFs6 is for the AIF.

4.2.3. *Feature improvement.* A total of 5 PFs as the outcomes of feature selection step and an AIF are used to compute 6 sets of AFs using the FCM algorithm. Obviously, there are 5 PFAFs and an AIFAF, which are then adopted for the model optimization.

4.3. **Model optimization.** There are 24 optimal models of DT, RF, LR and SVM, which are selected by the grid search-based method in combination with 5-fold CV procedure using 5 PFAFs and an AIFAF. The optimal parameters of the selected models are given in Table 5. We also implement a similar procedure to address 24 models, which use the input of 5 PFs and an AIF; 24 other models using the input of 6 AFs for further performance comparisons as shown in Table 3 and Table 4.

4.4. **Model validation.** Table 3, Table 4 and Table 5 show the validation performance of various optimal models using PFs, AIF, AFs, AIFAFs, and PFAFs. The highest accuracy of 85.38% is generated by the RF model using 3 PFs of time, EF, and SS as given in Table 3, SVM model using AFs4 as shown in Table 4, while that of the SVM model using

TABLE 3. Performance of ML models using PFs and AIF on the validation set

Model	Parameter	Feature	Ac (%)	Pr (%)	Re (%)	F1 (%)
LR	iter=100, C=2	PFs1	82.31±3.97	80.09±14.06	59.05±11.27	66.74±10.10
	iter=50, C=3	PFs2	82.31±3.13	76.95±12.20	62.51±6.07	68.47±7.44
	iter=50, C=3	PFs3	83.00±3.84	80.48±13.75	60.07±8.71	68.25±9.43
	iter=100, C=1	PFs4	83.33±3.66	84.44±13.82	56.76±10.44	67.12±10.45
	iter=100, C=1	PFs5	83.33±1.68	83.74±9.01	58.70±8.55	68.22±5.62
	iter=50, C=3	AIF	83.70±4.82	76.59±12.23	67.72±11.61	71.56±10.77
RF	leaf=4, tree=177	PFs1	83.69±6.09	77.32±15.67	70.40±9.44	72.80±9.38
	leaf=1, tree=154	PFs2	84.70±2.99	78.55±9.84	71.00±8.73	74.00±6.15
	leaf=4, tree=76	**PFs3**	**85.38±4.84**	**83.48±12.29**	**66.53±9.55**	**73.59±8.83**
	leaf=5, tree=171	PFs4	83.69±5.39	79.25±16.91	67.97±7.25	72.15±8.86
	leaf=4, tree=76	PFs5	84.70±4.53	76.78±8.81	73.05±9.37	74.52±7.52
	leaf=2, tree=219	AIF	83.68±3.94	77.88±7.62	66.67±8.06	71.50±6.05
SVM	C=2, kernel='rbf'	PFs1	83.68±6.10	80.70±16.14	63.48±9.58	70.57±10.72
	C=1, kernel='rbf'	PFs2	84.37±4.31	78.23±9.67	67.84±9.95	72.47±8.99
	C=2, kernel='rbf'	PFs3	84.70±6.50	82.28±15.60	65.43±11.57	72.34±11.38
	C=2, kernel='rbf'	PFs4	84.71±6.40	81.42±15.26	66.61±11.13	72.83±11.15
	C=2, kernel='rbf'	PFs5	82.32±4.21	74.53±10.60	66.36±11.76	69.45±8.55
	C=1, kernel='linear'	AIF	82.67±3.23	75.79±11.10	62.94±8.96	68.60±9.09
DT	leaf=5, max_depth=5	PFs1	77.22±6.00	65.39±16.44	61.04±11.51	62.21±10.87
	leaf=2, max_depth=50	PFs2	80.61±6.91	72.44±16.66	66.67±12.36	67.99±9.64
	leaf=2, max_depth=5	PFs3	84.35±4.72	79.32±14.10	69.15±6.05	73.28±7.55
	leaf=1, max_depth=10	PFs4	77.89±3.22	65.18±11.23	64.71±8.57	64.25±7.41
	leaf=5, max_depth=10	PFs5	79.94±3.57	65.84±6.52	73.66±9.03	69.29±6.54
	leaf=2, max_depth=5	AIF	79.93±4.35	69.45±11.51	59.52±10.83	63.80±10.47

TABLE 4. Performance of ML models using AFs on the validation set

Model	Parameters	Features	Ac (%)	Pr (%)	Re (%)	F1 (%)
LR	iter=50, C=0.1	AFs1	65.29±7.48	47.15±9.93	81.88±6.33	59.57±9.90
	iter=50, C=0.5	AFs2	66.65±6.70	48.25±9.81	78.72±5.79	59.55±9.15
	iter=50, C=0.1	AFs3	65.29±7.48	47.15±9.93	81.88±6.33	59.57±9.90
	iter=50, C=0.1	AFs4	64.95±7.40	46.81±9.91	80.77±6.47	58.99±9.87
	iter=50, C=1	AFs5	65.97±7.75	47.79±10.19	81.88±6.33	60.06±10.11
	iter=50, C=0.1	AFs6	46.29±6.77	28.88±6.65	49.10±10.37	36.12±7.54
RF	leaf=9, tree=17	AFs1	74.15±4.34	60.15±8.00	51.73±10.25	55.05±7.76
	leaf=3, tree=217	AFs2	72.09±2.51	56.27±8.26	48.63±7.59	51.67±6.30
	leaf=12, tree=44	AFs3	74.14±3.86	60.09±3.93	52.54±4.18	55.93±3.08
	leaf=12, tree=44	AFs4	73.45±4.39	60.44±11.23	45.45±4.83	51.64±6.75
	leaf=9, tree=232	AFs5	75.14±5.20	63.59±10.13	50.10±8.28	55.54±7.69
	leaf=13, tree=282	AFs6	65.96±5.59	35.50±36.62	35.50±36.62	10.44±10.26
SVM	C=100, kernel='rbf'	AFs1	71.41±5.59	55.16±11.97	56.60±6.23	55.28±8.67
	C=7.7, kernel='rbf'	AFs2	74.81±3.78	60.08±10.55	59.51±4.62	59.45±6.81
	C=100, kernel='rbf'	AFs3	75.15±4.88	61.48±8.50	56.91±7.79	58.70±6.76
	C=18, kernel='rbf'	**AFs4**	**75.29±3.98**	**61.16±7.45**	**59.82±5.29**	**60.27±5.42**
	C=24, kernel='rbf'	AFs5	71.41±4.09	55.60±9.85	60.41±7.25	56.72±4.12
	C=0.003, kernel='rbf'	AFs6	58.19±17.31	4.75±9.49	20.00±40.00	7.67±15.34
DT	leaf=8, max_depth=5	AFs1	71.74±5.06	55.50±14.10	45.57±11.04	49.69±11.63
	leaf=2, max_depth=20	AFs2	75.15±3.34	60.34±5.30	53.63±17.48	55.68±11.92
	leaf=2, max_depth=10	AFs3	73.47±3.17	61.51±15.30	53.97±6.22	55.79±2.42
	leaf=8, max_depth=5	AFs4	72.43±4.39	56.14±11.51	53.81±8.24	53.81±8.24
	leaf=4, max_depth=5	AFs5	73.44±5.69	57.39±11.02	63.29±10.85	59.42±9.01
	leaf=3, max_depth=5	AFs6	68.71±4.61	52.46±15.70	14.53±5.99	21.88±7.54

TABLE 5. Performance of ML models using various combinations of PFAFs and AIFAFs on the validation set

Model	Parameter	Feature	Ac (%)	Pr (%)	Re (%)	F1 (%)
LR	C=3, iter=30	PFs1, AFs1	82.33±4.45	77.87±11.16	60.12±9.42	67.53±8.93
	C=4, iter=30	PFs2, AFs2	83.31±3.73	80.94±8.58	61.19±8.73	69.27±7.15
	C=3, iter=100	PFs3, AFs3	84.00±4.47	80.21±9.62	64.03±13.78	70.31±11.28
	C=1, iter=50	PFs4, AFs4	84.00±4.66	80.55±11.93	64.03±13.78	70.27±11.61
	C=1, iter=50	PFs5, AFs5	84.69±3.39	81.76±7.13	65.19±10.75	71.97±8.46
	C=3, iter=50	AIF, AFs6	84.38±4.16	78.41±10.18	67.72±11.61	72.33±9.94
RF	leaf=4, tree=23	PFs1, AFs1	84.71±3.82	82.74±10.14	66.44±8.06	72.90±5.17
	leaf=2, tree=55	PFs2, AFs2	85.70±3.88	82.92±7.48	68.68±9.94	74.64±7.57
	leaf=4, tree=76	PFs3, AFs3	86.40±3.54	83.00±8.39	70.63±13.89	75.31±9.83
	leaf=1, tree=171	PFs4, AFs4	85.72±3.48	83.14±9.38	67.10±11.65	73.57±9.69
	leaf=2, tree=164	PFs5, AFs5	86.72±4.77	81.10±6.06	73.43±16.23	76.17±11.39
	leaf=2, tree=219	AIF, AFs6	84.68±2.49	78.28±8.38	68.48±7.18	72.96±7.24
SVM	C=2, kernel='rbf'	PFs1, AFs1	86.07±2.19	86.42±10.30	65.39±4.17	74.18±5.17
	C=1, kernel='rbf'	PFs2, AFs2	84.68±3.29	80.54±9.97	67.26±8.88	72.72±7.55
	C=2, kernel='rbf'	PFs3, AFs3	86.38±4.47	86.24±6.40	66.09±12.06	74.44±10.09
	C=2, kernel='rbf'	**PFs4, AFs4**	**86.73±4.10**	**87.67±7.85**	**71.64±11.05**	**75.00±9.61**
	C=2, kernel='rbf'	PFs5, AFs5	85.37±4.13	83.71±6.27	65.14±11.79	72.79±9.46
	C=2, kernel='linear'	AIF, AFs6	83.00±2.33	76.95±9.38	62.86±10.02	68.84±8.29
DT	leaf=3, max_depth=5	PFs1, AFs1	84.03±4.33	82.87±11.63	64.18±11.76	70.97±7.19
	leaf=5, max_depth=5	PFs2, AFs2	81.63±4.60	73.86±7.40	64.76±11.51	68.25±8.13
	leaf=2, max_depth=5	PFs3, AFs3	84.33±4.05	78.33±6.67	70.31±11.48	73.44±6.54
	leaf=5, max_depth=10	PFs4, AFs4	83.34±3.42	76.57±10.04	72.70±12.09	73.05±3.36
	leaf=2, max_depth=5	PFs5, AFs5	83.68±0.77	75.72±7.15	70.37±7.57	72.41±4.80
	leaf=2, max_depth=10	AIF, AFs6	80.61±2.98	71.07±9.75	63.70±5.69	66.75±5.82

TABLE 6. Performance of the selected ML models corresponding to various PFs and AIF with or without AFs on testing set

Model	Feature	Ac (%)	Pr (%)	Re (%)	F1 (%)
LR (iter=50, C=3)	AIF	77.55	68.57	52.17	59.26
RF (leaf=4, tree=76)	Time, EF, SS	85.03	80.00	69.57	74.42
DT (leaf=2, max_depth=5)		80.27	72.97	58.79	65.06
SVM (C=2, kernel='rbf')	Time, EF, SS, Platelets	82.31	76.32	63.04	69.05
LR (C=1, iter=50)	Time, EF, SCR, AFs	81.63	73.17	65.22	68.97
RF (leaf=2, tree=164)		87.07	82.93	73.91	78.16
SVM (C=2, kernel='rbf')	**Time, EF, SS, Platelets, AFs**	**87.76**	**83.33**	**76.09**	**79.55**
DT (leaf=2, max_depth=5)	Time, EF, SS, AFs	81.63	78.79	56.52	65.82

4 PFs of time, EF, SS, Platelets and 2 AFs is up to 86.73% as presented in Table 5. In addition, Table 6 presents the performance of ML models using PFs and AIF with or without AFs on testing set. It is noteworthy that AFs are not considered for the model estimation on the testing set because they are computed from original PFs. Figure 3 shows the ROC curves and AUC values for these models on the testing set. These results confirm that incorporating AFs improves classification performance across most models. Among them, SVM model achieves the highest AUC of 92%.

5. **Discussion.** The early diagnosis of HF plays an essential role in the delivery of curable treatments to improve the chance of survival. Today, the increasing costs associated

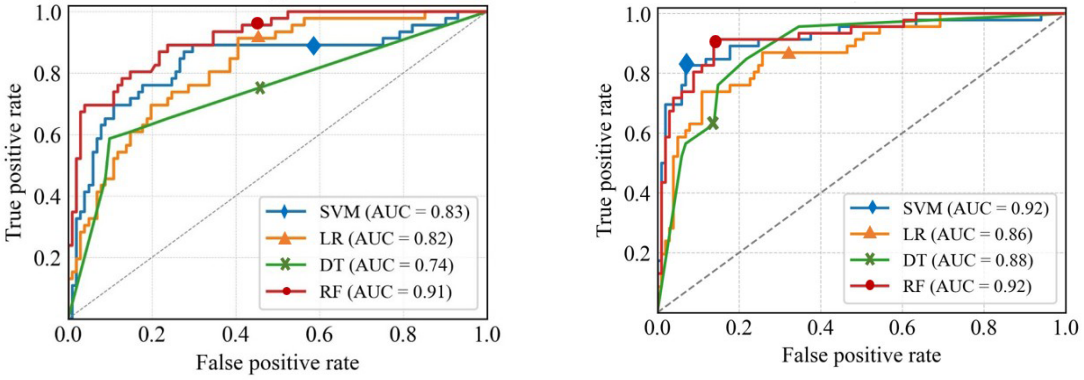


FIGURE 3. ROC curves of the models using PFs and (a) without AFs, (b) with AFs

with conventional medical treatments for the diagnosis of HF have highlighted the importance of developing diagnostic systems using ML methods. In addition, the high quality healthcare services also require the application of cutting-edge technologies such as artificial intelligence for the healthcare industry. Certainly, a large number of processes such as detection, diagnosis, treatment, and monitoring are efficiently improved by these above techniques. The aim of this work is to develop a novel mortality detection algorithm for HF patients using ML methods, which can be applied in clinic environments.

The feature selection has been widely used in numerous studies of biomedical area due to significant contributions to the performance improvement. Clearly, the irrelevant features are successfully eliminated from the input feature space, which definitely improves the quality of the remaining feature subset. Moreover, the feature selection only considers the most independent features, which represent a large fraction of information related to medical classification tasks. Another significant characteristic of the feature selection is that the ML models maintain relatively high performance using a small number of features selected from the input features, which reduces dramatically complexity for the practical applications. In this work, we use 3 feature importance-based feature ranking methods, namely MI, PI, and TB, for opportunity improvement to certainly select the most informative features. Here, a comprehensive view of feature contributions is exposed and compared using multiple feature selection approaches. Indeed, the results of the comparative analysis of the feature importance of 3 feature selection methods are shown in Table 2, which proves the most relevance of Time and EF features. Moreover, the third and fourth rank features of Age, SS, SCR and SCR, Platelets, Age are addressed by the feature selection algorithm using MI, PI, and TB, respectively, imply the significance of those features for the detection of HF mortality. The rational behind the use of 3 feature importance-based feature ranking methods is to investigate expanded feature combinations. Consequently, it provides a better opportunity for the selection of a small number of high quality features, which definitely contributes to the improvement of the HF diagnosis using the ML model.

Generally, feature selection can be considered as the first layer of feature engineering, which includes feature improvement as the second layer. Here, we reinforce the PFs by the use of FCM algorithm, which calculates two AFs. The combination of PFs and AFs generates a powerful PFAF, which are then used as the input of different ML models for the optimization and validation related to identification of the best models proposed as the final algorithm for HF detection. Obviously, the classification performance of the

models using PFAFs in Table 5 is higher than that of the models using only PFs in Table 3 or AFs in Table 4. The performance of various ML models using the AFs as the input is dramatically low with the highest accuracy of 75.29% as shown in Table 4, which is inappropriate to apply for HF survival detection. Clearly, the AFs significantly contribute to improving the quality of the PFAFs and the final detection performance of the proposed algorithm. The reinforced PFAFs consist of both selected risk factors identified by the feature selection method and additional information computed by FCM from the above selected factors. Consequently, the PFAFs emphasize significantly the characteristics of HF disease in terms of survivals and deaths. The optimal model, which is addressed by a grid search-based algorithm using the PFAFs in combination with 5-fold CV procedure, generates definitely high diagnosis performance.

Overfitting is a vital problem, which needs to be avoided for the improvement of learning and classification processes. Therefore, we implement the 5-fold CV-based statistical method for all steps in our method to select the feature combinations and optimal ML models by comparing the statistical performance results of various ML models, which ensure a high reliability of the proposed algorithm for the application in clinic environments. Here, the validation set is separated into 5 parts for which a part is for testing and the others are for training. This procedure is repeated 5 times to ensure every part being testing data. Then, the performance mean and standard deviation of the ML models are calculated to identify the most productive and stable model proposed as the final algorithm for the HF mortality detection. Moreover, we also split the input data into training and testing sets, and each of them accounts for 50% of the total. Then, the entire data are considered for the validation set. The rational behind the use of total input data as the validation set is the small input data, which may cause the overfitting problem. Hence, we use a validation set including a separated training set, which has an amount of half of the validation set to minimize the overlapped samples for feature selection using ML model optimization and validation. Another method to avoid the overfitting problem is model optimization with 5-fold CV procedure for the identification of the optimal parameters. Indeed, we employ the above method to select 72 optimal models using PFs, AIF, AFs, PFAFs, AIFAFs as shown in Table 3, Table 4 and Table 5, which certainly remove the overfitting problem.

The SVM model using an IFS of Time, EF, SS, Platelets, and AFs generated by the FCM algorithm shows the highest validated accuracy of 86.73% among the other models as shown in Table 5, which is proposed as the final algorithm for HF mortality detection. Moreover, the proposed algorithm produces the highest accuracy on the testing set as given in Table 6, which represents the classification performance of the optimal models with the highest validated performance in Table 3 and Table 5. It is noteworthy that this IFS is selected by the PI-based feature ranking in combination with the RF model, which results in the effectiveness of PI compared to the MI and TB methods. The Time feature represents the follow-up time of HF patients while EF measures the blood pumped from the left ventricle with each shrinkage. There are 2 conditions associated with EF: (i) preserved EF, also known as diastolic HF in which heart muscle contracts normally, but the ventricles do not relax as they normally do during ventricular filling; (ii) reduced EF, also known as systolic HF in which the heart muscle does not contract effectively, causing less oxygen-rich blood to be pumped. The SS feature presents a level of sodium in the blood, which needs to be maintained stably for proper muscle and nerve function. Hence, a low sodium level is a key factor in the prognosis of HF. Blood clotting is highly dependent on Platelets feature for which the abnormal platelet counts are associated with complications of HF. In addition, platelet levels outside the normal range contribute to an increased risk of thromboembolism. Table 7 shows the performance comparisons of

TABLE 7. Performance comparisons of the proposed algorithm with existing works

References	Data balancing method	Validation method	Model	Ac (%)	Pr (%)	Re (%)	F1 (%)
[11], 2021	SMOTE	5-fold CV	RF	76.25	–	89.97	–
[10], 2022	SMOTE	85%-15%	RF	83.33	86.36	90.48	88.37
[13], 2023	–	5-fold CV	RF	78.00	66.00	64.00	65.00
[14], 2025	–	80%-20%	DT	78.33	–	–	–
Proposed	–	50%-50%	SVM	87.76	83.33	76.09	79.55
algorithm	–	5-fold CV	SVM	86.73	87.67	71.64	75.00

the proposed algorithm to the existing methods, which confirms the superiority of the proposed algorithm.

The first limitation of this work is small and single data used for method development. Training and validation are employed as parts of this dataset, which results in no additional data for external estimation of the proposed algorithm. Furthermore, the limited size of the input data possibly leads to bias during the model training process. Clearly, the utility of a single dataset definitely reduces the generalization of the proposed algorithm. The second limitation is a time-consuming process due to the implementation of different ranking-based feature selection methods in combination with 5-fold CV procedure. To address these limits, multiple and independent HF datasets should be used in future works to obtain reliable estimation, generalization improvement of the final diagnosis algorithm corresponding to optimal risk factors selected by an effective feature ranking-based feature selection algorithm, which definitely decreases prediction time while maintaining relatively high detection performance.

6. **Conclusion.** Accurate diagnosis of HF mortality definitely improves the curable opportunities for patients. Hence, identification of the important risk factors, which contribute to the improvement of the HF detection is vital to immediately prevent the risk of death. In this paper, we proposed a novel and simple algorithm for the HF mortality detection, which consists of 4 risk factors and 2 additional features generated by the FCM algorithm and an SVM model. Here, 3 feature-ranking-based feature selection methods in combination with different ML models were used to investigate the performance of extended sets of risk factors. Consequently, an efficient search for risk factors was performed for the most informative subset of risk factors in terms of the final HF mortality detection performance. Furthermore, the FCM algorithm employed for generation of reinforced features, which were then combined with the selected risk factor subset used as input to the ML model, confirmed a superiority of the PI-based feature ranking method among the others. Indeed, the proposed algorithm produced the statistically validated performance with an accuracy of 86.73%, precision of 87.67%, recall of 71.64%, and F1-score of 75.00%, which is certainly suitable for application in clinic environments.

Acknowledgment. This work was funded by Vietnam Ministry of Science and Technology (MST) under Grant number ĐT.67/25.

REFERENCES

[1] Q.-F. Chen et al., Global burden of heart failure and its underlying causes in 204 countries and territories, 1990-2021, *European Heart Journal – Quality of Care and Clinical Outcomes*, vol.11, no.4, pp.493-509, 2025.

[2] NHLBI, *What Is Heart Failure?*, https://www.nhlbi.nih.gov/health-topics/heart-failure/, 2022.

[3] F M J. M. Shamrat et al., An explainable multi-objective hybrid machine learning model for reducing heart failure mortality, *PeerJ Computer Science*, vol.11, e2682, 2025.
[4] A. A. Ahmad and H. Polat, Prediction of heart disease based on machine learning using jellyfish optimization algorithm, *Diagnostics*, vol.13, no.14, 2392, 2023.
[5] X. Li et al., Clinical applications of machine learning in heart failure, *State of the Art in Neural Networks and Their Applications*, pp.217-233, 2023.
[6] M. A. Mohammad, Advancing heart failure research using machine learning, *The Lancet Digital Health*, vol.5, no.6, pp.e331-e332, 2023.
[7] A. Banerjee et al., Identifying subtypes of heart failure from three electronic health record sources with machine learning: An external, prognostic, and genetic validation study, *The Lancet Digital Health*, vol.5, no.6, pp.e370-e379, 2023.
[8] M.-N. Kim et al., Deep learning for predicting rehospitalization in acute heart failure: Model foundation and external validation, *ESC Heart Failure*, vol.11, no.6, pp.3702-3712, 2024.
[9] S. M. Kasongo and Y. Sun, A deep learning method with filter based feature engineering for wireless intrusion detection system, *IEEE Access*, vol.7, pp.38597-38607, 2019.
[10] S. Mishra, A comparative study for time-to-event analysis and survival prediction for heart failure condition using machine learning techniques, *Journal of Electronics, Electromedical Engineering, and Medical Informatics*, vol.4, no.3, pp.115-134, 2022.
[11] A. Newaz, N. Ahmed and F. S. Haq, Survival prediction of heart failure patients using machine learning techniques, *Informatics in Medicine Unlocked*, vol.26, 100772, 2021.
[12] M. M. Nishat et al., A comprehensive investigation of the performances of different machine learning classifiers with SMOTE-ENN oversampling technique and hyperparameter optimization for imbalanced heart failure dataset, *Scientific Programming*, 3649406, 2022.
[13] P. A. Moreno-Sánchez, Improvement of a prediction model for heart failure survival through explainable artificial intelligence, *Frontiers in Cardiovascular Medicine*, vol.10, 1219586, 2023.
[14] M. A. Insel et al., Predicting survival of patients with heart failure by optimizing the hyperparameters, *International Journal of Knowledge-Based and Intelligent Engineering Systems*, vol.29, no.2, pp.262-277, 2025.
[15] D. Mpanya et al., Predicting in-hospital all-cause mortality in heart failure using machine learning, *Frontiers in Cardiovascular Medicine*, vol.9, 1032524, 2023.
[16] B. Ahmad, J. Chen and H. Chen, Feature selection strategies for optimized heart disease diagnosis using ML and DL models, *arXiv Preprint*, arXiv: 2503.16577, 2025.
[17] K. Wang et al., Interpretable prediction of 3-year all-cause mortality in patients with heart failure caused by coronary heart disease based on machine learning and SHAP, *Computers in Biology and Medicine*, vol.137, 104813, 2021.
[18] Z. Noroozi, A. Orooji and L. Erfannia, Analyzing the impact of feature selection methods on machine learning algorithms for heart disease prediction, *Scientific Reports*, vol.13, no.1, 22588, 2023.
[19] N. H. Shabrina, D. Gunawan and A. S. Harahap, Convolutional neural networks for identifying papillary thyroid cancer histopathological image, *International Journal of Innovative Computing, Information and Control*, vol.21, no.2, pp.565-576, 2025.
[20] I. Cherian et al., Early detection of Alzheimer's disease using fuzzy C-means clustering and genetic algorithm-based feature selection from PET scans, *International Journal of Intelligent Systems and Applications in Engineering*, vol.12, no.3s, pp.452-463, 2023.
[21] J. H. Joloudari et al., FCM-DNN: Diagnosing coronary artery disease by deep accuracy fuzzy C-means clustering model, *arXiv Preprint*, arXiv: 2202.04645, 2022.
[22] T. Ahmad et al., Survival analysis of heart failure patients: A case study, *PloS One*, vol.12, no.7, e0181001, 2017.
[23] P. E. Latham and Y. Roudi, Mutual information, *Scholarpedia*, vol.4, no.1, 1658, 2009.
[24] Z. Zhou and G. Hooker, Unbiased measurement of feature importance in tree-based methods, *ACM Transactions on Knowledge Discovery from Data (TKDD)*, vol.15, no.2, pp.1-21, 2021.
[25] S. Nirmalraj et al., Permutation feature importance-based fusion techniques for diabetes prediction, *Soft Computing*, pp.1-12, 2023.
[26] S. Miyamoto et al., *Algorithms for Fuzzy Clustering*, Springer, Heidelberg, 2008.
[27] G. Biau and E. Scornet, A random forest guided tour, *Test*, vol.25, no.2, pp.197-227, 2016.
[28] M. P. LaValley, Logistic regression, *Circulation*, vol.117, no.18, pp.2395-2399, 2008.
[29] W. S. Noble, What is a support vector machine?, *Nature Biotechnology*, vol.24, no.12, pp.1565-1567, 2006.

[30] C. Kingsford and S. L. Salzberg, What are decision trees?, *Nature Biotechnology*, vol.26, no.9, pp.1011-1013, 2008.

Author Biography

Minh Tuan Nguyen received the B.S. degree from the Post and Telecommunications Institute of Technology, Vietnam, in 2004, the M.S. degree from Hanoi University of Science and Technology, Vietnam, in 2008, both in Electronics and Telecommunications Engineering, and the Ph.D. degree at the Gwangju Institute of Science and Technology, Korea, in 2018. He is now with Posts and Telecommunications Institute of Technology. His research interests include network security, Internet of Things, biomedical signal processing, gene analysis, sentiment analysis, brain computer interface, machine learning, deep learning, optimization, and biomedical application design.

Nhat Trang Linh Le received the Bachelor of Business from UniSA Business School, The University of South Australia (UniSA), Australia in 2025, with research interests in data analytics, artificial intelligence, machine learning, deep learning, optimization.

Thang Nhat Le received the B.Eng. degree in Radio-Electronics and Communication from Hanoi University of Science and Technology (HUST), Vietnam, in 1995, the M.Eng. degree in Telecommunications from Asian Institute of Technology (AIT), Thailand, in 2000 and Ph.D. degree in Information and Communication Technology (ICT) from the Department of Computer Science and Telecommunications (DIT), University of Trento, Italy, in 2006. He is currently an Assoc. Professor and the Dean of Postgraduate Studies Faculty, Posts and Telecommunications Institute of Technology (PTIT), Vietnam. His research interests now are performance analysis, modeling and simulations, wireless communications systems, physical layer security, queueing theory and applications, machine learning, deep learning, optimization, and biomedical applications.