Machine Learning for Prehospital Electrocardiogram Diagnosis of Acute Coronary Syndrome

Agnese Sbrollini[1], C. Cato ter Haar[2], Laura Burattini[1,*], Cees A. Swenne[1,3]

[1]Università Politecnica delle Marche, Ancona, Italy
[2]Amsterdam University Medical Center, Amsterdam, the Netherlands
[3]Leiden University Medical Center, Leiden, the Netherlands

Abstract

Identification of acute coronary syndrome (ACS) in patients with non-traumatic chest pain is essential for prehospital triage. This study evaluated machine-learning methods to detect ACS in prehospital electrocardiograms (ECGs). Data from 1,425 emergency medical service patients were obtained by the SUBTRACT study. For each patient, an ambulance ECG and a previous elective ECG were analyzed by extracting 18 ambulance ECG direct measurements (DM) and 29 serial features (SF, differences between the two ECGs). Eight classifiers (Linear Discriminant Analysis, Decision Tree, Support Vector Machine, Artificial Neural Network, Logistic Regression, Naive Bayes, K-Nearest Neighbors, Advanced Repeated Structuring & Learning Procedure - RS&LP) were trained and tested. Performance was assessed by the area under the receiver operating characteristic (AUC). The inclusion of SF significantly improved the performance of all classifiers. RS&LP and Support Vector Machine achieved the highest (AUC: 90 ± 12%) and the lowest (AUC: 64 ± 3%) performance, respectively. Overall, RS&LP demonstrated superior discriminative ability and generalizability, supporting its potential for clinical use. Future studies should validate these findings in larger multicenter cohorts to further improve ACS diagnosis and management directly in the ambulance.

1. Introduction

Rapid and accurate triage of patients with non-traumatic chest pain presenting to the emergency medical service (EMS) is crucial for patient prognosis, particularly for those with acute coronary syndrome (ACS). In a representative cohort of 1,425 EMS patients that we recently studied [1], 16.3% received a final hospital diagnosis of ACS. These patients should be identified as early as possible and transported directly to hospitals with facilities for primary percutaneous coronary intervention (PCI), enabling rapid restoration of coronary blood flow. Consequently, the ECG recorded at initial medical contact plays a central role in early diagnosis and patient triage.

In our previous work, we focused on detecting ischemia in prehospital ECGs, as ischemia is a hallmark of ACS; however, in a specific form of ACS, unstable angina, ischemic and non-ischemic episodes alternate. Also, ischemia may occur in patients with stable coronary artery disease. Moreover, ECG manifestations of ischemia are heterogeneous and require adjudication using clinical information obtained during hospital admission, introducing uncertainty into retrospective labeling [2].

In the present study, for the first time, we instead focus on the detection of ACS in prehospital ECGs. As the definitive diagnosis of ACS can only be established by integrating all clinical information collected during hospital admission, our objective is to predict this diagnosis directly from the ambulance ECG. This objective represents a broader pattern-recognition problem that is particularly well suited to artificial intelligence (AI) methods. A second objective is to compare the performance of AI algorithms using a clinically relevant cost function that quantifies the consequences of incorrect decisions. In particular, the current study serves as a benchmark for our recently developed deep-learning algorithm [3,4], specifically designed for relatively small datasets while maintaining good generalizability, and compares its performance with that of seven widely used AI algorithms. To our knowledge, this is the first study to compare multiple AI algorithms for ACS detection from prehospital ECGs while explicitly optimizing a clinically motivated decision-cost function.

2. Materials and Methods

2.1. Data

We studied data from the SUBTRACT study [1,2], a retrospective observational study conducted at the Leiden University Medical Center (LUMC) and the Amsterdam Medical Center (AMC), in collaboration with regional hospitals and emergency medical services. The SUBTRACT study aimed to assess the added value of

serial comparison of the acute ECG and an earlier recorded elective ECG in detecting ischemia/ACS in patients who were urgently transported by ambulance for non-traumatic chest pain. Of the 1425 patients included, 231 appeared to have ACS (cases); the other 1194 patients (controls) had other pathology (non-acute in 386). For each patient, two 10-second 12-lead ECGs were analyzed: an ambulance ECG (AECG) and a previously recorded non-acute elective ECG, serving as reference ECG (RECG). The AECGs were recorded using LIFEPAK 12 devices (Mason-Likar electrode configuration). The RECGs were retrieved from the ECG databases of the participating hospitals in the region; they had been recorded by various ECG recording systems, all using the standard electrode configuration.

2.2. ECG Processing, Feature Extraction

After converting the AECGs to standard 12-lead ECGs by using the Leiden matrix [5], all ECGs were analyzed by the LEADS program [6]. LEADS computes a vectorcardiogram (VCG) and determines an averaged dominant QRST complex with its onset-QRS, J-point, and end-T landmarks. All processing steps are semi-automated (*i.e.*, they include expert review and, if necessary, correction procedures). Eventually, LEADS outputs a wide range of ECG/VCG variables, from which we selected the variables used in the current study.

We characterized the AECG of each subject by 18 direct measurements (DM), comprising sex, age, and 16 ECG features: QRS duration, maximal QRS vector magnitude, QRS-integral vector magnitude, QRS complexity, J-point vector magnitude, sum of the absolute values of the J-point amplitudes of the 8 independent leads I/II/V1-V6, sum of the absolute values of the J-point amplitudes in all leads, maximal T-vector magnitude, T-integral vector magnitude, T-wave complexity, T-wave symmetry, number of leads with positive T waves, QT interval, ventricular-gradient magnitude, QRS-T spatial angle, and heart rate. Moreover, for each subject, we computed 29 serial ECG features (SF) by subtracting RECG from AECG features: age difference, signed and absolute QRS-duration differences, signed and absolute maximal QRS-vector magnitude differences, signed and absolute QRS-integral vector-magnitude differences, signed and absolute QRS-complexity differences, J-point vector-magnitude difference, summed absolute J-point amplitude differences in the independent leads, summed absolute J-point amplitude differences in all leads, signed and absolute maximal T-vector magnitude differences, signed and absolute T-integral vector magnitude differences, signed and absolute T-wave complexity differences, signed and absolute T-wave symmetry differences, differences between the number of leads with positive T waves, number of leads with changed T-wave polarity, signed and absolute QT-interval differences, ventricular gradient magnitude difference, signed and absolute QRS-T spatial angle differences, and signed and absolute heart rate differences.

2.3. Machine-Learning Classifiers

The features were used to train eight machine-learning classifiers: Linear Discriminant Analysis (LDA), Decision Tree (DT), Support Vector Machine (SVM), Artificial Neural Network (ANN), Logistic Regression (LR), Naive Bayes Classifier (NBC), K-Nearest Neighbors (KNN) and, finally, a neural network developed using the Advanced Repeated Structuring and Learning Procedure (RS&LP) [1,5] (see Table 1 for the hyperparameters). Each classifier was trained and evaluated in 200 trials. In the first 100 trials, only direct measurements (DM) were used as input features; in the remaining 100 trials, the complete feature set (DM + SF) was used. The trials differed in the random partitioning of the dataset: 70% of the data were used for training and the remaining 30% for testing, while preserving the case/control prevalence in both subsets.

2.4. Statistical Analysis

Classifier performance was assessed using the receiver operating characteristic (ROC) and its area under the curve (AUC). Each threshold value on the ROC separates the subjects (n_{CS} cases; n_{CT} controls) into true positives (TP; correctly identified cases), true negatives (TN; correctly identified controls), false positives (FP; incorrectly identified controls), and false negatives (FN; incorrectly identified cases). The operating point on the ROC was found by minimizing the cost function:

$$Cost = FP + \frac{n_{CT}}{n_{CS}} FN \quad (1)$$

Sensitivity (SE) and specificity (SP) were computed at the selected operating point. Normality of the distributions was assessed using the Lilliefors test. Normal distributions were expressed as mean and standard deviation (μ±σ). Comparisons between classifiers trained using only DM and those trained using the complete feature set (DM + SF) were performed using the paired Student's T-test. Using the same statistical approach, the performance of the RS&LP classifier was compared to that of all other classifiers. Statistical significance was set at 5% ($P < 0.05$).

3. Results

Table 2 lists the classifier performances and operating points. Inclusion of SF significantly improved all classifier performances. The RS&LP classifier performed significantly better than the other classifiers ($P < 0.05$). Figure 1 shows the average costs for all classifiers; RS&LP had the lowest costs in the test sets, indicating superior generalizability.

Table 1. Machine-learning classifiers and their hyperparameter settings.

Classifier	Hyperparameters
LDA	Solver: Singular Value Decomposition
DT	Splitting criterion: Gini impurity; Splitter: Best; Minimum samples/node: 2; Minimum samples/leaf: 1
SVM	Kernel: Radial Basis Function; Regularization parameter: 1.0; Polynomial degree: 3; Tolerance: 10^{-3}
ANN	Hidden layer size: 100 neurons; Activation function: Rectified Linear Unit; Optimizer: Adam; Regularization term: 10^{-4}; Initial learning rate: 0.001; Maximum iterations: 200
LR	Regularization term: 1.0; Solver: quasi-Newton; Maximum iterations: 100; Tolerance: 10^{-4}
NBC	Variance smoothing: 10^{-9}
KNN	Number of neighbors: 5; Weighting scheme: Uniform; Distance metric: Minkowski; Leaf size: 30
RS&LP	Architecture and learning parameters defined according to the original implementation [3]

Table 2. Performance of the machine-learning classifiers.

			LDA	DT	SVM	ANN	LR	NBC	KNN	RS&LP
DM	Training	AUC	80§±3	97§±3	69§±6	76§±4	76§±3	89§±6	81§±4	86±9
		SE	72±9	89§±9	50§±7	74±12	68§±10	88§±13	76±12	74±13
		SP	75§±8	93§±8	88§±3	66§±12	74§±9	76§±11	71§±11	85±11
	Validation	AUC	78§±6	88§±7	64§±9	70§±6	72§±7	89§±7	77§±8	83±8
		SE	72§±14	86±14	47§±15	70§±19	67§±17	88§±14	63§±19	82±14
		SP	75±15	79§±14	87§±10	67§±19	73±17	80§±13	80§±17	72±13
	Testing	AUC	78§±2	81§±6	64§±5	70§±4	72§±2	81§±6	77§±6	81±9
		SE	69§±10	80§±15	42§±9	66§±15	61§±10	85§±14	60§±15	74±18
		SP	74±10	84§±10	87§±10	63§±15	73±10	78±10	78±13	75±14
DM+SF	Training	AUC	87*,§±3	96*±4	76*,§±3	77§±10	82*,§±3	94*±4	80§±4	93*±12
		SE	79*,§±8	85*±10	60*,§±5	67*,§±19	72*,§±8	91*,§±7	76§±12	86*±19
		SP	79*,§±8	96*,§±6	89*±1	79*,§±14	78*,§±7	86*,§±12	70§±10	92*±12
	Validation	AUC	83*,§±5	83*,§±9	65§±9	75*,§±12	74*,§±6	93*,§±5	76§±8	91*±12
		SE	77*,§±14	71*,§±20	50§±16	66§±21	65§±15	92*±10	61§±19	93*±12
		SP	78±13	83±12	84§±13	82*±16	78*±16	87*,§±10	81±16	80*±15
	Testing	AUC	83*,§±3	83*,§±8	64§±3	75*,§±11	74*,§±2	83*,§±4	76§±6	90*±12
		SE	74*,§±10	68*,§±19	42§±6	62§±19	59§±10	88±9	58§±18	89*±15
		SP	76§±9	85§±9	87§±5	82*±11	77*±10	89*,§±10	79±15	80*±15

* P<0.05 when comparing performance obtained by DM vs those obtained by DM+SF in the same machine-learning method:
§ P<0.05 when comparing performance obtained by a machine learning method vs those obtained by the RS&LP

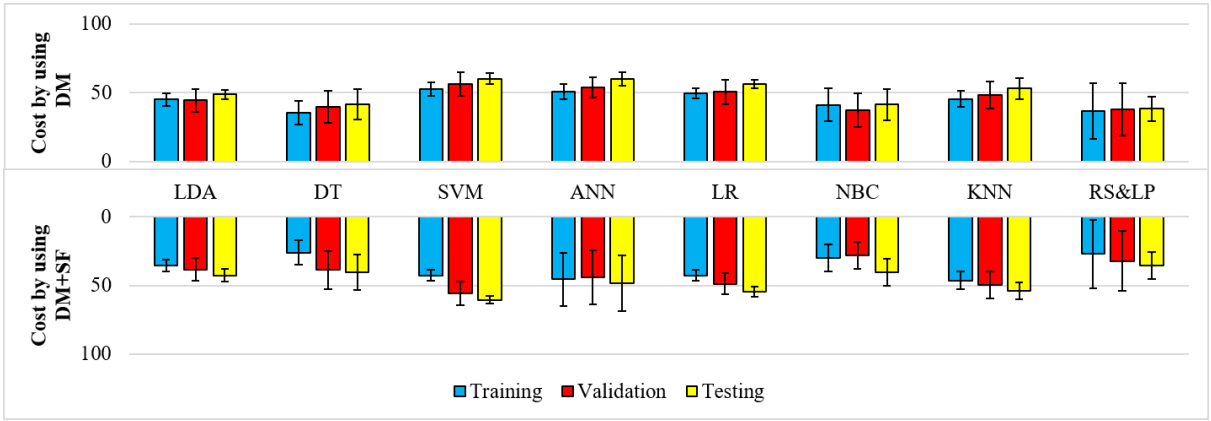

Figure 1. Mean ± standard deviations of training, validation, and testing costs in all machine learning methods.

4. Discussion and Conclusion

Our present study compared the ability of eight machine-learning algorithms to detect ACS directly from prehospital ECG data. As expected, the addition of serial ECG features significantly improved performance in all evaluated machine-learning methods. This finding underscores the usefulness of comparing an acute ECG with an earlier elective ECG of the same patient, as recommended in the ACS guidelines [7], where changes may be indicative of an acute process. With the increasing availability of cloud storage for ECGs, such serial comparisons are becoming a logistic reality.

Comparison of the AUCs shows clear differences in discriminative performance between the evaluated algorithms. With only DM features, RS&LP achieved the highest testing AUCs (81%), significantly outperforming all competing methods. Superiority of the RS&LP performance becomes more evident after inclusion of the serial ECG features, with a testing AUC increasing to 90%. Among the conventional machine-learning algorithms, NBC and DT showed the best overall performance, whereas SVM consistently yielded the lowest AUC values. These findings support the rationale behind the RS&LP algorithm design, which was specifically developed for small datasets where robust generalization is often more important than maximizing training performance.

The use of a cost function rather than a conventional threshold was motivated by the clinical context of ACS detection. In emergency cardiovascular care, false-negative classifications are substantially more harmful than false-positive classifications because they may delay referral to a PCI-capable hospital and postpone appropriate treatment. The adopted cost function therefore weights false negatives by the ratio between controls and cases, effectively compensating for class imbalance while emphasizing the clinical consequences of missed ACS diagnoses. This approach aligns the optimization process with real-world decision-making and provides a performance metric that is more clinically meaningful than accuracy alone. Comparing cost values across training, validation, and test datasets provides important information about model generalization. In principle, one would expect training costs to be lower than validation and testing costs, since the classifiers are optimized on the training data. Such a pattern is indeed observed for several algorithms, particularly DT and NBC, whose excellent training performance is accompanied by higher costs on validation and testing datasets, suggesting overfitting. Of note, the RS&LP classifier also caused higher test costs than training costs, but still lower than those of all other classifiers, confirming its maintained generalization.

Several limitations should be acknowledged. First, external validation in independent cohorts is necessary before widespread clinical implementation is realistic. Second, although the dataset is relatively large for a prehospital ECG study, controls was substantially less than cases, potentially influencing classifier training despite the use of a cost function specifically designed to address class imbalance. On the other hand, it must be recognized that the composition of our database is not different from that of other EMS populations, and that the control patients are far from normal [2], which makes prehospital triage a considerable challenge. Third, because ECG features were considered, additional information contained in the raw ECG signals may not have been fully exploited [8]. Also, repeated ECGs made during the ambulance ride might offer additional information.

Future studies should validate serial ECG analysis in every patient for whom a previous ECG is available and accessible in larger datasets and assess its integration with clinical data and decision-support systems to improve ACS diagnosis and management.

References

[1] C. C. Ter Haar et al., "An initial exploration of subtraction electrocardiography to detect myocardial ischemia in the prehospital setting." *Ann. Noninvas. Electrocard.*, vol. 25, no. 3, May 2020.

[2] C. C. Ter Haar, and C. A. Swenne, "Post hoc labeling an acute ECG as ischemic or non-ischemic based on clinical data: A necessary challenge." *J. Electrocardiol.*, vol. 81, pp. 75-79, Nov.-Dec. 2020.

[3] A. Sbrollini et al., "Advanced repeated structuring and learning procedure to detect acute myocardial ischemia in serial 12-lead ECGs." *Physiol. Meas.*, vol. 44, no.8, Aug. 2023.

[4] A. Sbrollini et al. "Serial electrocardiography to detect newly emerging or aggravating cardiac pathology: a deep-learning approach." *Biomed. Eng. Online*, vol. 18, no. 1, Feb. 2019.

[5] S.C. Man et al. "Reconstruction of standard 12-lead electrocardiograms from 12-lead electrocardiograms recorded with the Mason-Likar electrode configuration." *J. Electrocardiol.*, vol. 41, no. 3, pp. 211-219, May-Jun 2008.

[6] H.H.M. Draisma et al. "LEADS: an interactive research-oriented ECG/VCG analysis system." *Comput. Cardiol.*, vol. 32, pp. 515-518, Sep. 2005.

[7] S.V. Rao et al. "2025 ACC/AHA/ACEP/NAEMSP/SCAI guideline for the management of patients with acute coronary syndromes: a report of the American College of Cardiology/American Heart Association Joint Committee on Clinical Practice Guidelines." *Circulation*, vol. 151, no. 13, pp. e771-e862, Apr. 2025.

[8] R.Q. Ji et al. "Novel machine learning fusion architectures integrating electrocardiogram representations: applications to acute coronary event detection." *Eur. Heart J. - Digit. Health*, vol. 7, no. 5, June 2026

Address for correspondence:

Kees Swenne,
Cardiology Department,
Leiden University Medical Center,
Albinusdreef 2, 2333 ZA Leiden, the Netherlands.
E-mail address: c.a.swenne@lumc.nl