Short title: ECG Features Selection

In Search of Optimal Subset of ECG Features to Augment the Diagnosis of Acute Coronary Syndrome at the Emergency Department

By

Zeineb Bouzid, BS;[a] **Ziad Faramand**, MD;[e,h] **Richard E Gregg**, MS;[i]

Stephanie O. Frisch, PhD, RN;[c,e] **Christian Martin-Gill**, MD;[f,h] **Samir Saba**, MD;[g,h]

Clifton Callaway, MD, PhD;[f,h] **Ervin Sejdić**, PhD;[a,b,c,d] & **Salah Al-Zaiti**, RN, PhD[e,f,g]

From

(a) Department of Electrical & Computer Engineering and (b) Department of Bioengineering at Swanson School of Engineering; (c) Department of Biomedical Informatics at School of Medicine; (d) Intelligent Systems Program at School of Computing and Information; (e) Department of Acute & Tertiary Care Nursing; (f) Department of Emergency Medicine; and (g) Division of Cardiology at University of Pittsburgh, PA, USA; (h) University of Pittsburgh Medical Center (UPMC), Pittsburgh PA, USA; and (i) Advanced Algorithm Research Center, Philips Healthcare, Andover, MA, USA

Word count: ~3000 words

Tables: 3 tables

Figures: 4 figures

Funding: National Institute of Health grant # R01HL137761

Trial Registration: ClinicalTrials.gov # NCT04237688

Conflict of Interest: PCT patent application #WO2016168181A1

Corresponding Author: Salah Al-Zaiti, PhD, University of Pittsburgh, 3500 Victoria Street, 336 VB, Pittsburgh PA 15261, ssa33@pitt.edu

ABSTRACT

Background: Clinical practice primarily relies on classical ST amplitude measures during the initial evaluation of patients with suspected acute coronary syndrome (ACS). Machine learning, when driven by domain-specific knowledge, could help identify an optimal subset of ECG features to augment clinicians' decision during patient evaluation.

Methods: This was an observational study of consecutive patients evaluated at the emergency department for suspected ACS (Cohort 1 n=745, age 59±17, 42% Female, 15% ACS; Cohort 2 n=499, age 59±16, 49% Female, 18% ACS). A total of 554 temporal-spatial waveform features were extracted from baseline 12-lead ECGs. We identified a subset of 65 physiology-driven features that are mechanistically linked to myocardial ischemia, and compared their performance to a subset of 229 data-driven features selected by multiple machine learning algorithms. We then used random forest to select a subset of 73 most important ECG features that had both data- and physiology-driven basis to ACS prediction and compared their performance to clinical experts. Classifiers were evaluated using logistic regression (LR) and artificial neural network (ANN) with 10-fold cross-validation on cohort 1 followed by independent testing on cohort 2.

Results: Compared to physiology-driven features, classifiers based on data-driven features were superior during model training, but generalized poorly to testing data. LR classifiers based on the 73 hybrid features yielded a stable model that outperformed clinical experts in terms of predicting ACS and non-ST elevation ACS (net reclassification improvement 0.10 [-0.02–0.23] and 0.19 [0.04–0.33], respectively). For the latter, classical ST and T wave amplitudes had the least predictive importance, with metrics of non-dipolar electrical dispersion (i.e., circumferential ischemia), ventricular activation time (i.e., transmural conduction delays), QRS and T axes and angles (i.e., global remodeling), and PCA ratio of ECG waveforms (i.e., regional heterogeneity) playing a more important role.

Conclusions: We identified a subset of novel ECG features that would improve ACS detection. These features guided by domain-specific knowledge yielded stable LR classifiers highly adaptable to clinical decision support applications.

Key Words: machine learning, dimensionality reduction, acute coronary syndrome, electrocardiogram, ischemia

INTRODUCTION

The prompt identification of acute coronary syndrome (ACS) is a longstanding challenge in emergency practice.(1-3) The electrocardiogram (ECG) is readily available during initial patient evaluation, and sensitive ECG markers of acute myocardial ischemia can expedite the current time-consuming, biomarker-driven approach for ACS diagnosis.(4-6) The electrophysiological basis of acute myocardial ischemia has been thoroughly studied over the past few decades, (7) with many studies suggesting the abundance of hidden signatures of acute myocardial ischemia in the surface ECG signal (8, 9). Yet, current guidelines exclusively rely on the amplitude of ST segment and T wave for ACS detection (10), translating into a diagnostic sensitivity of approximately 40% for the standard 12-lead ECG (11). Given that ECG waveform is among the most extensively studied signals in cardiovascular medicine, existing computational algorithms can extract hundreds of features from a single 10-second 12-lead ECG. Thus, recent advances in pattern recognition and machine learning could help in identifying an optimal subset of features to augment clinicians' decision in detecting ACS during initial evaluation. (12)

Although it is being widely adopted in various clinical applications, machine learning is limited by the relatively small size of available clinical datasets and the difficulty of finding comparable external datasets for replication.(13) Accordingly, feature subset selection (FSS) plays a significant role in optimizing the accuracy of supervised classification systems, including improved understandability of the final classifier. In addition to available data-driven approaches of FSS, some studies suggest the need for domain-specific expertise to guide feature selection and model development during the learning process. (13) The electrophysiology of myocardial ischemia is well understood, and it is feasible to perform FSS based on cardiac physiology. However, there is a paucity of evidence regarding the effect of manual FSS on the performance of supervised classification systems. In fact, manual FSS is counter-intuitive to the premise of machine learning—*the discovery of hidden patterns in the data that might not be apparent to clinicians*. Accordingly, using two prospective clinical cohorts, we sought to (1) compare the accuracy of supervised classifiers in detecting ACS using ECG feature subsets selected based on either data-driven techniques or domain-specific knowledge; and (2) whether data-driven FSS techniques can identify ECG features indicative of ACS that were overlooked by domain-specific human experts.

METHODS

Design and Settings

This was a prospective observational cohort study of consecutive patients with chest pain transported by Emergency Medical Services to one of three tertiary care hospitals in the US between 2013 and 2015. The methods of this study were previously described in detail. (14) In short, we collected the prehospital 12-lead ECGs obtained by paramedics in the field and stored them for offline analysis. We then followed patients up to adjudicate study outcomes. Clinical data were obtained from medical charts by independent reviewers. Patients were recruited under a waiver of informed consent and the study was approved by the Institutional Review Board of University of Pittsburgh.

The primary study outcome was the presence of ACS (myocardial infarction or unstable angina) during the primary indexed admission, defined according to the 4th Universal Definition of myocardial infarction consensus statement as the presence of symptoms of ischemia (i.e. diffuse discomfort in the chest, upper extremity, jaw, or epigastric area for more than 20 minutes) with the presence of biomarker, nuclear, or angiographic evidence of myocardial ischemia and / or loss of viable myocardium. (10) Study outcomes were adjudicated by two independent physician reviewers and disagreement was resolved by a third physician reviewer. Patients discharged from the emergency department were classified as negative for ACS if they had no 30-day adverse events. Patients presenting ventricular tachycardia or fibrillation on prehospital ECG were excluded from this analysis.

ECG Preprocessing and Feature Extraction

Each ECG was manually over-read by an independent reviewer. ECGs with excessive noise or artifact (n=24, 2%) were substituted by the next serial ECG obtained during emergency evaluation. ECGs with ventricular tachycardia or fibrillation were excluded from this analysis (n=7, 0.5%). All other available ECGs, including those with secondary repolarization changes (i.e., pacing, BBB, coarse atrial fibrillation, or LVH with strain, n=178, 14%) were included in the analysis. We decided to keep these ECGs because their removal had no effect on the performance of subsequent predictive models. Besides, the ability to classify these challenging ECGs would have huge clinical utility during emergency care.

Then, 10-second, 12-lead ECG signals (500 s/s, HeartStart MRx, Philips Healthcare) were pre-processed at Philips Healthcare Advanced Algorithm Research Center (Andover, MA). Raw ECG signals were decompressed to extract individual ECG leads. Noise, artifact, and ectopic beats were then removed, and representative average beats were computed for each ECG lead to eliminate residual baseline noise and artifacts. This technique yields high signal-to-noise ratio and stable average waveform signal for each of the 12 leads.

Next, fiducial points from these representative beats were identified and corresponding ECG features were extracted. The details of feature extraction from this dataset was previously described in detail. (12) In short, a total of 554 features were extracted from each 12-lead ECG. First, duration, amplitude, and area of various waveform deflections were calculated from each of the 12 leads, yielding 444 temporal ECG features (Figure 1A). Second, the 12 representative beats were superimposed, and global intervals and subintervals were computed, yielding 6 more temporal ECG features (Figure 1B). Third, principal component analysis (PCA) on time-voltage data was performed on orthogonal leads I, II, V1–V6 to compute PCA ratios of the eigenvalues of various ECG waveforms, yielding 13 spatial ECG features (Figure 1C). Finally, axes, angles, loops, and gradients of QRS and T vectors from xy, xz, yz, and xyz planes were computed, yielding 91 more spatial ECG features (Figure 1D).

All extracted ECG features were then z-score normalized. Missing data, representing less than 0.2% of the total features' values available in our dataset, were imputed using the mean or the mode of the corresponding feature.

FSS using Domain-Specific Human Expertise

Two research scientists trained in cardiac electrophysiology reviewed the 554 extracted ECG features and agreed on a reduced set of 65 features that had strong physiological basis as plausible markers of acute myocardial ischemia, including 24 classical measures (amplitude of J+80 point and T wave from each of the 12 leads), and 41 supplemental features that may correlate with acute cardiac ischemia: depolarization and repolarization times (i.e., QRS duration, JT_{end}, JT_{peak}, $T_{peak\text{-}end}$, and QT interval, k=6); depolarization and repolarization vectors (QRS and T axes and angles, k=8); repolarization velocity (i.e., T wave peak inflection, amplitude, and slope, k=5); global electrical dispersion (PCA ratios between QRS, STT, J, and T eigenvalues, k=13); repolarization characteristics (i.e., T wave morphology and T loop features, k=7); and high frequency signal noise values (k=2). The selection of these candidate features was based on review of literature (15) and our previous work. (8, 16, 17)

FSS using Data-Driven Algorithms

We used three different data-driven algorithms to identify a list of features most important for optimizing the performance of the classification algorithm. First, we used **Cohen's d effect size**, which compares how distinguishable ACS vs. non-ACS distributions of a given feature are in terms of the distance between the means. All distributions were evaluated for normality of distributions and homogeneity of variances. Features corresponding to an effect size lower than 0.35 are assumed to fail to differentiate between the two populations and were excluded from our dataset. Using this cutoff value, only 23 features out of 554 remained (4%). Second, we used **recursive features elimination** as part of logistic regression. We evaluated 20 features per iteration and used F1 scores to evaluate model performance. F1 scores provides the best tradeoff between precision and recall using imbalanced datasets like ours, which had a 6:1 ratio of non-ACS to ACS subgroups. The selection of the optimal set of features went through a 10-fold cross-validation process. Using this technique, 156 features out of 554 (28%) were selected. Finally, we used **LASSO regression** to select the most important features with non-zero coefficients. We used the L1 norm method to penalize the least square error between the outcome and an affine function of the input variables. The regularization parameter alpha was set by the means of a 10-fold cross-validation. Using this technique, 96 features out of 554 (17%) were selected.

Next, given that the three FSS techniques described above use complementary, non-competing approaches, we identified the features that received at least one vote (i.e., appeared in at least one FSS algorithm). This yielded a total of 229 features. We used these data-driven features in subsequent training and testing of machine learning classifiers in order to compare against the domain-specific manually selected features. It is noteworthy that this step-by-step process for FSS was selected after a comprehensive evaluation of our dataset. This is important to note because the performance of machine learning algorithms is dependent on the inherent properties of the dataset used. Several studies have used multiple FSS procedures to tackle one specific disease diagnosis.(18)

FSS using a hybrid data- and physiology-driven approach

To identify any important ECG features that were missed by domain-specific experts, we mapped the 229 data-driven features against the major components of the 12-lead ECG signal, identifying the overlap between the data-driven features and the ones selected by domain-specific experts. We identified pertinent data-driven features that could be mechanistically linked to ischemia and yet missed by human experts. This yielded a total of 100 hybrid features that are both data-driven and judged by clinicians as presumably contributing as signatures of myocardial ischemia. To reduce the apparent redundancy in these features, we used random forest to identify and keep the important features for the task of ACS detection. This yielded a final novel subset of 73 features that we used in subsequent tuning of ML classifiers.

Machine Learning Methods

Logistic regression (LR) and artificial neural networks (ANN) have been preferentially used in previous studies focusing on ECG-based prediction of ACS. (19-21) Considering the size of our dataset and the expected reduction of model complexity achieved through FSS, we started with LR as the machine-learning classifier of choice to address the aims of our study. We then used ANN to explore whether FSS approaches would have a similar effect on more sophisticated, non-linear machine learning classifiers.

Our LR and ANN classifiers were trained using a 10-fold cross-validation on Cohort 1 and, afterwards, tested on an independent Cohort 2 being completely blinded to its outcomes. We started with all 556 available features (554 ECG features with age and sex) without any FSS (i.e., LR_{554} and ANN_{554}). Next, we built models using the 65 manual features selected by domain-specific human experts (i.e., LR_{65} and ANN_{65}), the 229 data-driven features (i.e., LR_{229} and ANN_{229}), and the 73 hybrid data- and physiology-driven features (i.e., LR_{73} and ANN_{73}). The algorithms were trained using 10-fold cross-validation and then evaluated on an independent testing set that was blinded to the outputs.

The classification performance of each classifier was evaluated using the area under the receiver operating characteristic (AUROC) curve. This tool is powerful because it reflects the ability of binary classifiers to distinguish between two populations. We used DeLong's test to compare the difference between the mean AUC of two correlated ROC curves of different classifiers (22), and we opted for pairwise comparisons. We set alpha at $p<0.05$ for two tailed hypothesis testing.

ECG Reference Standards

We compared the performance of the final LR_{73} classifier against two current ECG reference standards: (1) clinical experts' interpretation and (2) commercial interpretation software. To get these annotations, each 12-lead ECG was over-read by two experienced clinicians. Each reviewer classified each ECG according to the likelihood of underlying ACS (yes / no) taking into account diagnostic ST-T changes as per the fourth Universal Definition of Myocardial Infarction consensus statement, (10) and the presence of other suspicious ECG findings (i.e., contiguous territorial involvement, evidence of reciprocal changes, changes beyond those caused by secondary repolarization, and lack of ECG evidence of non-ischemic chest pain etiologies). Disagreements were resolved by a board-certified cardiologist. Next, we used Philips diagnostic 12/16 lead ECG analysis program (Philips Healthcare, Andover, MA) for automated ECG interpretation. This software is commercially available and is used in practice to denote the diagnostic likelihood of ACS on the ECG printout (i.e., "***Acute MI***").

We computed and compared the sensitivity, specificity, and positive and negative predictive values for the final ML classifier and the reference standards. We also computed the net reclassification improvement (NRI) index for our final ML classifier against each reference standard. Finally, in subsequent sensitivity analyses, we re-evaluated the diagnostic performance of our final ML classifier in detecting patients with non-ST elevation ACS (NSTE-ACS) after excluding patients with confirmed STEMI on their prehospital ECG and who were sent to the catheterization lab emergently.

RESULTS

Baseline Characteristics

Our sample consisted of 1,244 patients from two study cohorts: a training cohort (n=745, age 59±17, 42% Female, 40% Black) and a testing cohort (n=499, age 59±16, 49% Female, 40% Black). Most patients were evaluated for chest pain (90%) or shortness of breathing (39%); most patients presented in normal sinus rhythm (88%) or atrial fibrillation (9%); and the rate of 30-day cardiovascular death was 4.6%. Table 1 summarizes the baseline characteristics of each cohort. The two cohorts were comparable in terms of demographics, past medical history, chief complaint, baseline ECG characteristics, and clinical outcomes.

Performance of ML classifiers

The primary study outcome was ACS, which occurred in 114 out of 745 patients (15.3%) in the training cohort and 92 out of 499 patients (18.4%) in the testing cohort. Figure 2 compares the AUROC curves of the different LR and ANN classifiers considered in this study. On training set (Fig. 2A, left panel), both manual FS and data-driven FSS techniques had better performance compared to no-FSS, with the best performance (lowest bias) achieved using the data-driven approach. However, on independent testing (Fig. 2A, right panel), data-driven FS approach generalized poorly (high variance). Manual FSS, on the other hand, generalized well to the testing set, suggesting a better bias-variance tradeoff. Comparing the area under ROC curve of manual FSS and data-driven FSS yielded a statistically significant difference for the Logistic Regression model with a p-value equal to 0.0105. The same trend was observed using ANN. The data-driven FSS approach performed best on the training set (Fig. 2B, left panel), but generalized poorly to the testing set (Fig. 2B, right panel), again suggesting more overfitting compared to manual FSS approach, with a p-value equal to 0.0411.

Overlap in Features between FSS Approaches

Among the 229 data-driven features, 31 features (14%) were among the ones manually selected by human experts. These data-driven features with physiological plausibility for ACS classification included (1) lead-specific ST and T wave amplitudes; (2) T peak–Tend interval; (3) frontal and horizontal QRS and T axes; (4) spatial QRS-T angle and total-cosine R-to-T angle; (5) T loop morphology dispersion; (6) PCA ratio of QRST waveform, STT waveform, and T wave; and (7) the non-dipolar component of J wave. Among these features, T peak–T end was specifically selected by all three data-driven FSS algorithms, and was also ranked by LR classifiers as the most important feature among the ones selected by human experts. Finally, to discern which data-driven features contributed to noise vs. contributed to true prognostic value in ACS prediction, we mapped the 229 data-driven features against the major components of the 12-lead ECG signal (Table 2). This table highlights a potential subset of features that data-driven algorithms ranked as important for the task of ACS detection but were not selected by domain-specific experts.

Performance of Hybrid Subset of Novel Features

The final hybrid subset included 73 features that had both data- and physiology-driven basis. Figure 3A compares the AUROC curves of the three LR classifiers based on data-driven basis alone, domain-expertise alone, and hybrid data- and physiology-driven basis. As seen in this panel, the hybrid features model generalized well to the testing set, outperforming the other two models. Similar trends were seen with ANN algorithms, but without any additional gain compared to LR algorithms (LR_{73} 0.79 vs. ANN_{73} 0.76). Thus, compared the diagnostic accuracy of the final LR_{73} against the reference standards (Table 3). As seen in this table, our LR classifier had higher sensitivity compared to expert clinicians and the commercial software while maintaining higher negative predictive value (i.e., superior rule out performance). Although the LR classifier had lower specificity than other reference standards, it achieved positive overall net reclassification improvement (0.10 [-0.02–0.23] and 0.21 [0.10–0.32], respectively).

Finally, in our sensitivity analyses, we re-evaluated the diagnostic performance of our final ML classifier in detecting patients with NSTE-ACS. Figure 3B and Table 3 show the AUROC of LR_{73} and its corresponding diagnostic accuracy values as compared to the reference standards. Similar to previous results, our classifier had higher sensitivity compared to expert clinicians and the commercial software while maintaining higher negative predictive value (i.e., superior rule out performance), achieving positive overall net reclassification improvement for NSTE-ACS detection (0.19 [0.04–0.33] and 0.29 [0.15–0.42], respectively). Figure 4 displays the importance ranking of the novel ECG features for the task of NSTE-ACS detection. Intriguingly, classical ST and T wave amplitudes had the least predictive importance, with metrics of non-dipolar electrical dispersion, ventricular activation time, QRS and T axes and angles, and PCA ratio of ECG waveforms playing a more important role.

DISCUSSION

This study evaluated the effect of two FSS techniques on the accuracy of machine learning classifiers in augmenting the ECG detection of ACS. Using two prospective clinical cohorts, our data show that machine learning classifiers have better bias-variance tradeoff when built based on features manually selected by human experts as compared to no FSS or using data-driven techniques alone. On independent testing, our data show that using a hybrid subset of 73 novel ECG features based on data- and physiology-driven approaches yields not only more powerful and interpretable model, but also outperforms clinical experts and commercial rule-based software in detecting any ACS event, as well as NSTE-ACS events. More interestingly, feature importance ranking demonstrates the presence of novel and plausible markers of ischemia that are highly adaptable to clinical decision support applications.

Effect of FSS Approach on Classifiers Performance

Our data show that, compared to no-FSS, physiology-driven features optimized our LR classifier and yielded a generalizable model. This finding is expected given that using domain-specific knowledge not only tremendously reduced the dimensionality (65 out of 556 features), but also intuitively reduced the redundancy in the data, both of which are compatible with linear classifiers. On the other hand, our data show that the initial gain observed by using data-selected features generalized poorly to an independent unseen cohort. Our training set results are similar to the ones reported by Green et al. (2006). In their work, they built the model based on 16 ECG features chosen using the Principal Component Analysis (PCA) approach. Their cohort consisted of a comparable sample size (634 patients) and ACS prevalence (130 ACS patients i.e. ≈ 20.5%). (20) However, Green et al. did not have an independent testing set for validation. In our data, we showed that data-driven FSS lacked generalizability on a new test example, indicating overfitting of training data coupled with a substantial variability of classifier performance. Although this finding was surprising, the small dataset size as well as the inclusion of patients with confounders in our datasets could provide a simple rationale for this unexpected finding. Besides, some strict requirements about data nature, such as the homogeneity of variances for the Cohen's d effect size algorithm, were not satisfied which may jeopardize the predictive performance, including its generalizability.

We observed similar trends in results when we applied ANN as a non-linear classifier. These findings are a little bit counterintuitive given that ANN is expected to better capture the underlying characteristics of the dataset when fed with more features. This divergence can be attribute to the small sample size, especially for training data, which is incompatible with learning a complex model without increasing the risk of overfitting. (23) This was observed as a significant reduction in ANN classifiers performance using all available features (k=554) or the data-selected ones (k=229). Again, we speculate the reduced dimensionality and data redundancy when using physiology driven features reduced the complexity of the ANN classifiers, yielding a more generalizable model.

Finally, it is worth noting that using ANN classifiers consistently yielded higher classification accuracy when compared to LR classifiers, with or without any FSS (Figure 2). However, this gain in accuracy was negligible when using the physiology-driven features (ANN_{65} = 0.77 vs. LR_{65} = 0.76 [for test set]). Given that LR classifiers are easily interpretable, our results suggest that using an LR_{65} classifier with physiology-driven features can yield a fully understandable decision support tool for clinical use.

Overlap between Data- and Physiology-Driven Features

The secondary aim of this study was to explore whether data-driven FSS techniques might identify ECG features indicative of ACS that were overlooked by domain-specific human experts. Table 2 mapped the 229 data-driven features against the major components of the 12-lead ECG signal, identifying the overlap between the data-driven features and the ones selected by domain-specific expertise. More interestingly, this table summarizes the cluster of data-driven features that were overlooked by human-experts. Some of these overlooked data-driven features are contextually understandable, like ST slope, ST deviation morphology, and T wave attributes, but some other features were more challenging to classify. Upon careful annotation, we classified the overlooked data-driven features in one of these three broad categories: (1) noise attributed to existing comorbidities or patient medications (i.e., lead-specific P duration, P amplitude, and PR interval); (2) redundant information quantified by simultaneous ECG features (i.e., lead-specific Q, R, and S wave attributes that are redundant with scar size, and lead-specific QRS duration and QT interval that are redundant with principal component analysis); and (3) features that could be mechanistically linked to myocardial ischemia and can serve as plausible features of ACS (i.e., presence of fragmented QRS and lead-specific ventricular activation time).

<u>Novel ECG Features of Ischemia</u>

The novel features identified in this study as plausible markers of ACS that are potentially mechanistically linked to myocardial ischemia bring a valuable addition to clinical knowledge. Intriguingly, although the classical ST and T wave amplitude measures were among the predictive features, they ranked as the least important when compared to the contribution of other novel features (Figure 4). Some of the observed patterns and clusters of the most important features can be summarized in the following major categories:

1. Features of the non-dipolar voltage, which quantifies the spatial electrical dispersion in the fourth to eighth eigenvalues. In the context of ST, T, and J components, the non-dipolar voltage would indicate the magnitude of diffusion or widespread global changes,(24) a probable measure of circumferential ischemia in ACS.
2. Ventricular activation time, which quantifies the time from Q onset to R peak. Whereas depolarization of the whole ventricular myocytes is assessed through global QRS duration, localized regional depolarization can be assessed using individual leads facing that myocardial region. Thus, ventricular activation time measured from anterior and inferior leads would primarily indicate transmural conduction delays in the left ventricle and apex,(25) a probable consequence of localized ischemia in these regions.
3. QRS and T axes and corresponding angles, which characterize the propagation direction of depolarization and repolarization signals and, hence, global electrical dispersion. In the context of ACS, these features can reflect the altered electromechanical forces in the ventricular myocardium and probably the resulting global remodeling after myocardial injury.(26)
4. Waveform eigenvalues and corresponding ratios, which quantifies the principal components of ECG signal in perpendicular space. The altered signal propagation speed and velocity between healthy and ischemic myocardium leads to spatial

heterogeneity and significantly impacts these features.(9) Thus, in the context of ACS, these eigenvalues would resemble regional myocardial ischemia (or injury vectors).(8)

5. Other T wave metrics that quantify duration (e.g., T peak T end), amplitude (e.g., relative R-to-T), area (e.g., JTpeak area), morphology (e.g., T asymmetry), and loop characteristics (e.g., loop dispersion). Some studies have demonstrated that such simple T wave metrics may better predict early ischemia as compared to ST segment,(27) a finding that is supported by our current results.
6. Residual high frequency noise in the signal. Although this might be a simple incidental finding reflective of acuity of illness at the time of ECG acquisition, we previously demonstrate that such noise highly correlates with beat-to-beat repolarization lability.(16) This lability can resemble the alternans of intracellular Ca^{+2} transient in adjacent cells during acute myocardial ischemia.

Clinical Implications

Unlike the majority of previous studies that primarily used the limited, open-source MIT-PTB diagnostic ECG database, our results are based on two large clinical cohorts with real-world ECG data. Thus, our study has some immediate clinical implications. Our machine learning algorithms are fully interpretable and can be easily incorporated into existing ECG software or embedded into ECG interpretation platforms for decision support. These algorithms can help clinicians in identifying NSTE-ACS events in real-time, which constitutes a long-lasting challenge in clinical practice. Given that our algorithm has higher sensitivity and negative predictive value compared to experienced clinicians, our models are well-suited as an initial screening tool (i.e., rule out). This has the potential to better allocate hospital resources by avoiding prolonged observations, unnecessary admissions, or invasive testing. With an average net reclassification improvement of 20%, our approach can positively impact the initial triage of 1.4 out of the 7 million Americans evaluated at the emergency department for chest pain every year. This is inclusive of the challenging group of patients whom baseline ECGs are typically deemed un-interpretable for ischemia (e.g., pacing, BBB, LVH, etc.). Finally, given that our machine learning model are less dependent on classical ST and T wave amplitude measures, they can be used to augment (rather than replace) commercial rule-based ECG software that follow published recommendations by AHA/ACC guidelines.

Study Limitations

Strengths of our current study include the quality of our prehospital ECG dataset, using two independent training and validation sets, the selection of features mechanistically linked to ischemia, the emphasis on the interpretability and clinical relevance, and the comparison against a reference standard. Yet, our study had some limitations. Even though the data were collected from multiple healthcare centers, both training and testing sets were still restricted to one region. Thus, the study may be biased by disparities inherent to sex, race and other factors' distributions in the community. Our algorithms need to be tested on a more diverse population including data from more geographically distant healthcare centers. Besides, the patient to feature ratio, which reaches almost 1:1 value for one of the classifiers, is low. This fact, aggregated with the unbalanced dataset presenting only 15.3% prevalence of outcome, would considerably influence the performance of the classifiers, especially ANN. Future research needs to include more patients in the study while ensuring the collection of similar proportions of diseased and healthy patients with respect to the primary outcome of the study.

CONCLUSION

In this prospective analysis, we explored the value of different algorithms to identify an optimal subset of ECG features that can augment the diagnosis of ACS at the Emergency Department. In this context, we arrived at the conclusion that LR classifiers guided with domain-specific expertise yield the most reliable classification performance and are consequently more adapted to developing clinically relevant decision support tools. However, data-driven classifiers identified a subset of novel ECG features that would improve ACS detection by providing important insights for developing cardiac electrical biomarkers that are mechanistically linked to ischemia and can be clinically relevant.

Table 1: Baseline Study Characteristics

	Cohort 1 (N=745) (Training Set)	*Cohort 2 (N=499) (Testing Set)*
Demographics		
Age in years	59 ± 17	59 ± 16
Sex (Female)	317 (42%)	243 (49%)
Race (Black)	301 (40%)	202 (40%)
Past Medical History		
Hypertension	519 (69%)	329 (66%)
Diabetes mellitus	196 (26%)	132 (26%)
Old myocardial infarction	205 (27%)	122 (24%)
Known CAD	248 (33%)	179 (36%)
Known heart failure	130 (17%)	74 (15%)
Prior PCI / CABG	207 (28%)	124 (25%)
Clinical Presentation		
Chest Pain	665 (89%)	454 (91%)
Shortness of Breathing	250 (34%)	234 (47%)
Normal Sinus Rhythm	648 (87%)	442 (88%)
Atrial Fibrillation	71 (9%)	46 (9%)
Course of Hospitalization		
Length of Stay (median [IQR])	2.3 [1.0–3.0]	1.2 [0.6-2.5]
Confirmed ACS (all events)	114 (15.3%)	92 (18.4%)
NSTE-ACS	83 (11.1%)	74 (14.8%)
Treated by Primary PCI / CABG	74 (10%)	65 (13%)
30-Day CV Death	33 (4.4%)	24 (4.8%)

Table 2: Overlap in Features Between Data-driven and Human-Expert Techniques

12-Lead ECG Component	***Number of Features Selected***		***Comparison between techniques***	
	Human Expert	**Data-Driven**	**Overlap in Features**	**Features Overlooked by Clinicians**
ECG normalization (k=2)	2	2	Age and sex	-
P duration, amplitude, or area (k=72)	0	25	-	Lead-specific P duration & amplitude
PR interval metrics (k=26)	1	11	Global PR interval	Lead-specific PR interval
Q duration or amplitude (k=24)	0	10	-	Lead-specific Q wave presence
R duration or amplitude (k=48)	0	23	-	Lead-specific R amplitude
S duration or amplitude (k=48)	0	16	-	S amplitude in precordial leads
Other QRS complex metrics (k=74)	1	31	Global QRS duration	QRS notch; ventricular activation time; lead-specific QRS duration or area
Selvester Score (k=19)	1	0	Total scar size	-
ST amplitude, duration, or slope (k=72)	12	31	Lead-specific ST amplitude	Lead-specific ST duration and slope
ST deviation morphology (k=14)	0	7	-	Presence of concaved ST deviation
T duration, amplitude, or area (k=76)	14	33	Lead-specific T amplitude, T-to-R relative amplitude	Lead-specific T duration or area; presence of notched T wave
QT interval and subintervals (k=23)	4	12	Global QTc, T peak−T end	Lead-specific QT interval
QRS axis (k=12)	1	7	Frontal plane QRS axis	Horizontal and spatial QRS axis
T axis (k=11)	4	6	T axis in frontal, horizontal, and spatial planes	-
QRS and T vector angles (k=5)	2	3	QRS-T angle and TCRT	-
T loop morphology (k=6)	4	4	T asymmetry & dispersion	-
Principal Components Analysis (k=16)	16	6	PCA ration of J, T, and STT	-
Noise signal (k=8)	3	2	Noise & baseline wander	-

Table 3: Diagnostic Accuracy Measures of Machine-Learning Classifiers against Gold Standard Reference on the Testing Set (n=499)

	Clinical Experts Interpretation	Commercial Software Read	Novel ECG Features (LR_{73})
Predicting Any ACS Event			
Sensitivity	0.40 (0.30–0.51)	0.25 (0.17–0.35)	0.72 (0.61–0.81)
Specificity	0.94 (0.92–0.96)	0.98 (0.96–0.99)	0.73 (0.68–0.77)
Positive Predictive Value	0.63 (0.51–0.73)	0.79 (0.62–0.90)	0.38 (0.33–0.42)
Negative Predictive Value	0.88 (0.86–0.89)	0.85 (0.83–0.87)	0.92 (0.89–0.94)
NRI Index	Reference –	– Reference	0.10 (-0.02–0.23) 0.21 (0.10–0.32)
Predicting NSTE–ACS Event			
Sensitivity	0.26 (0.16-0.37)	0.12 (0.06–0.22)	0.72 (0.60–0.82)
Specificity	0.94 (0.92–0.97)	0.98 (0.96–0.99)	0.68 (0.63–0.72)
Positive Predictive Value	0.46 (0.33–0.60)	0.60 (0.35–0.80)	0.29 (0.25–0.33)
Negative Predictive Value	0.87 (0.85–0.89)	0.85 (0.84–0.87)	0.93 (0.90–0.95)
NRI Index	Reference –	– Reference	0.19 (0.04–0.33) 0.29 (0.15–0.42)

Figure Legends:

Figure 1: Computation of ECG Features

This figure shows the computation of 554 features from each 12-lead ECG. (a) Duration, amplitude, and area of various waveform deflections are calculated from the median beat of each of the 12 leads. (b) The 12 median beats are superimposed, and global intervals and subintervals are computed. (c) Principal component analysis (PCA) on time-voltage data is performed on the orthogonal leads I, II, V1–V6 to compute PCA ratios of the eigenvalues of various ECG waveforms. (d) Axes, angles, loops, and gradients of QRS and T vectors from xy, xz, yz, and xyz planes are computed.

Figure 2: ***Classification Performance using LR and ANN classifiers***

These plots show the performance of logistic regression (LR) and artificial neural network (ANN) classifiers on training data (Cohort 1) and testing data (Cohort 2) using all available ECG features (k=554), data-driven subset of ECG features (k=229), or physiology-driven subset of ECG features (k=65). P values are based on non-parametric method by Delong.

Figure 3: ***Classification Performance using different subsets of novel ECG features***

These plots show the performance of logistic regression (LR) classifiers on testing data (Cohort 2) for predicting (A) acute coronary syndrome (ACS) and (B) non-ST elevation acute coronary syndrome (NSTE-ACS) using data-driven subset of ECG features (k=229), physiology-driven subset of ECG features (k=65), or hybrid subset with novel features (k=73).

Figure 4: ***Importance Rank of subset of novel ECG features for the task of NSTE-ACS detection***

This plot shows the feature importance ranking obtained using a Random Forest model on a hybrid dataset including novel ECG features with prehospital ECG data after excluding STEMI patients.

Figure 1: ***Computation of ECG Features***

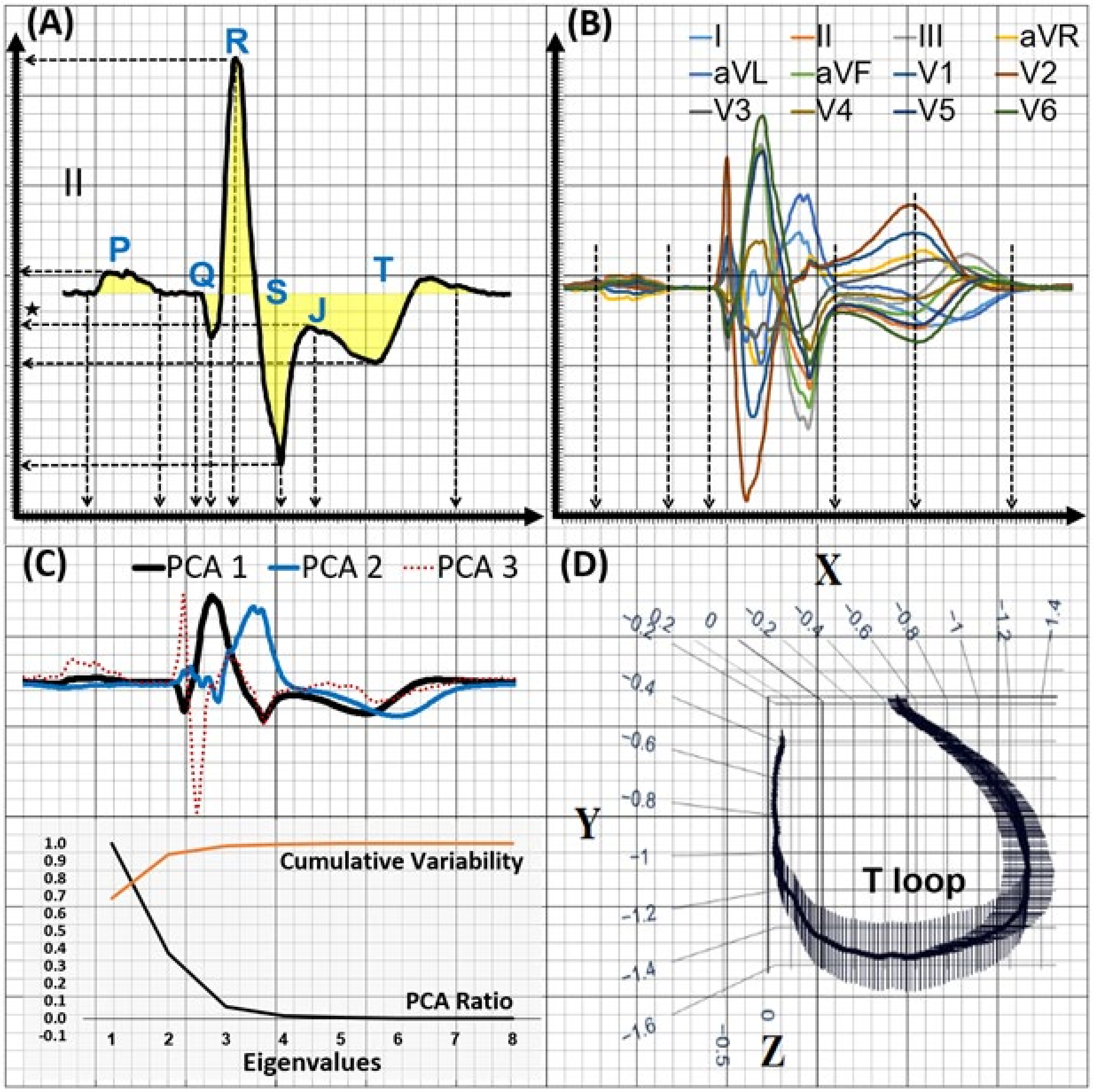


Figure 2. Classification Performance using LR and ANN classifiers

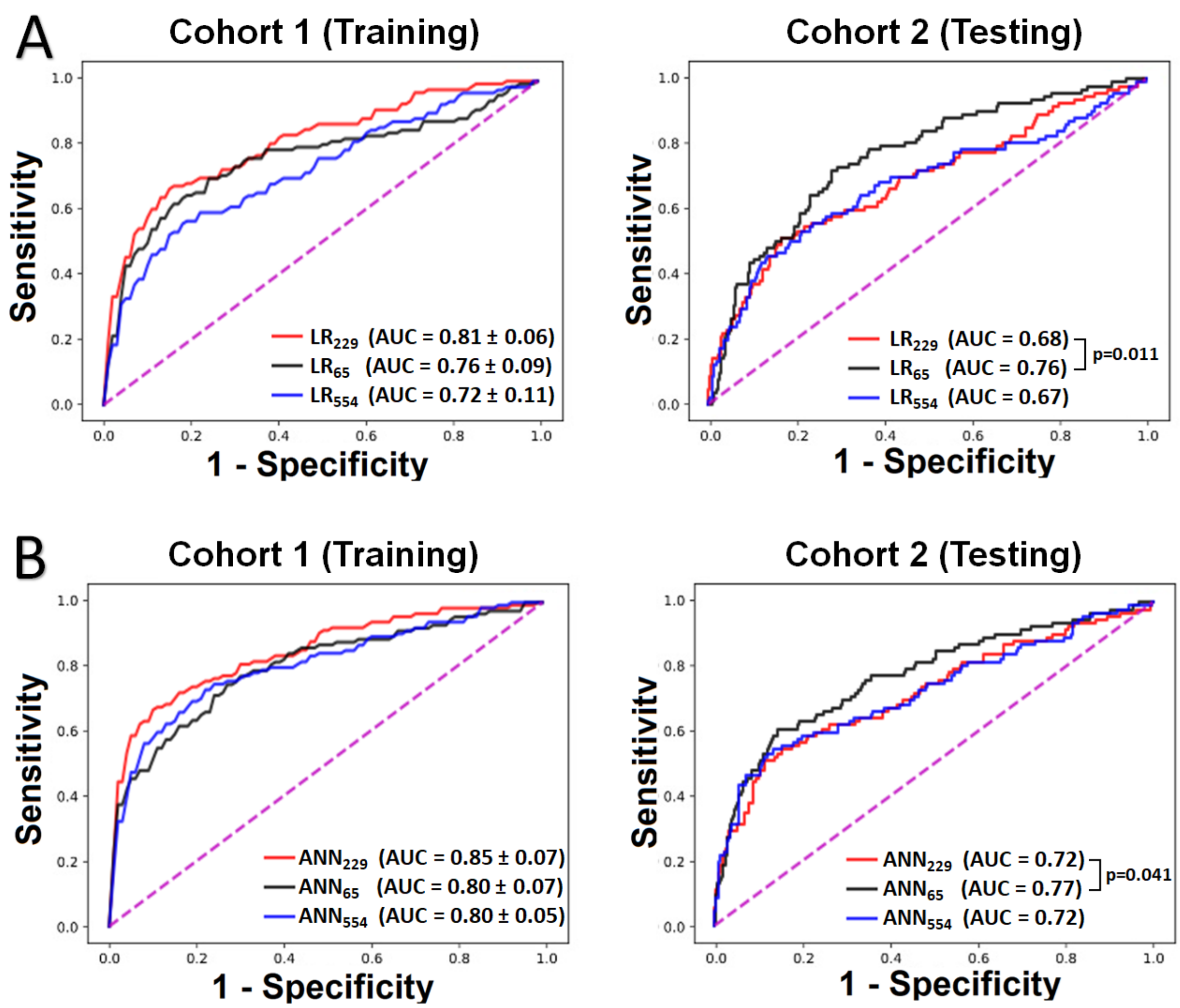


Figure 3. Classification Performance using different subsets of novel ECG features

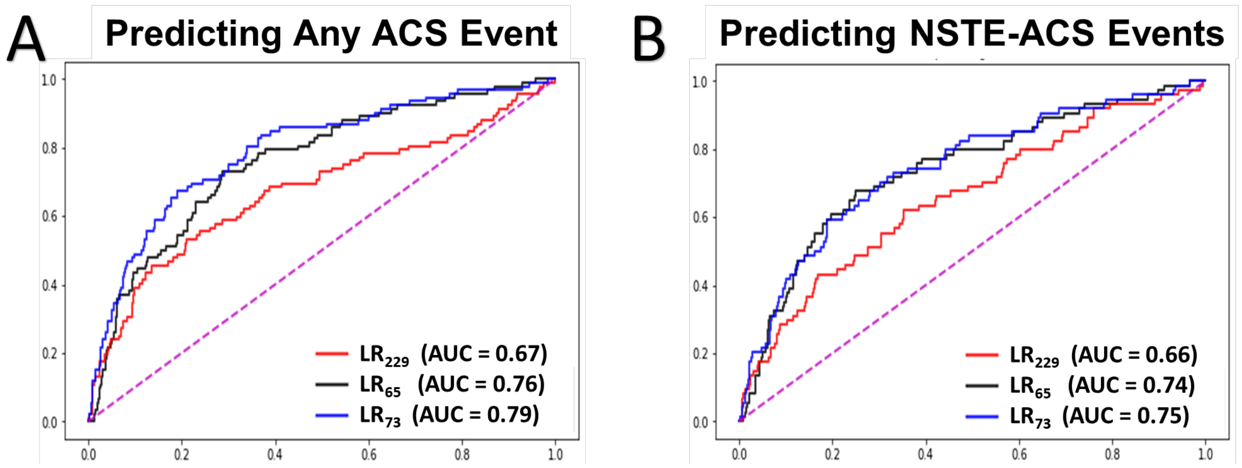
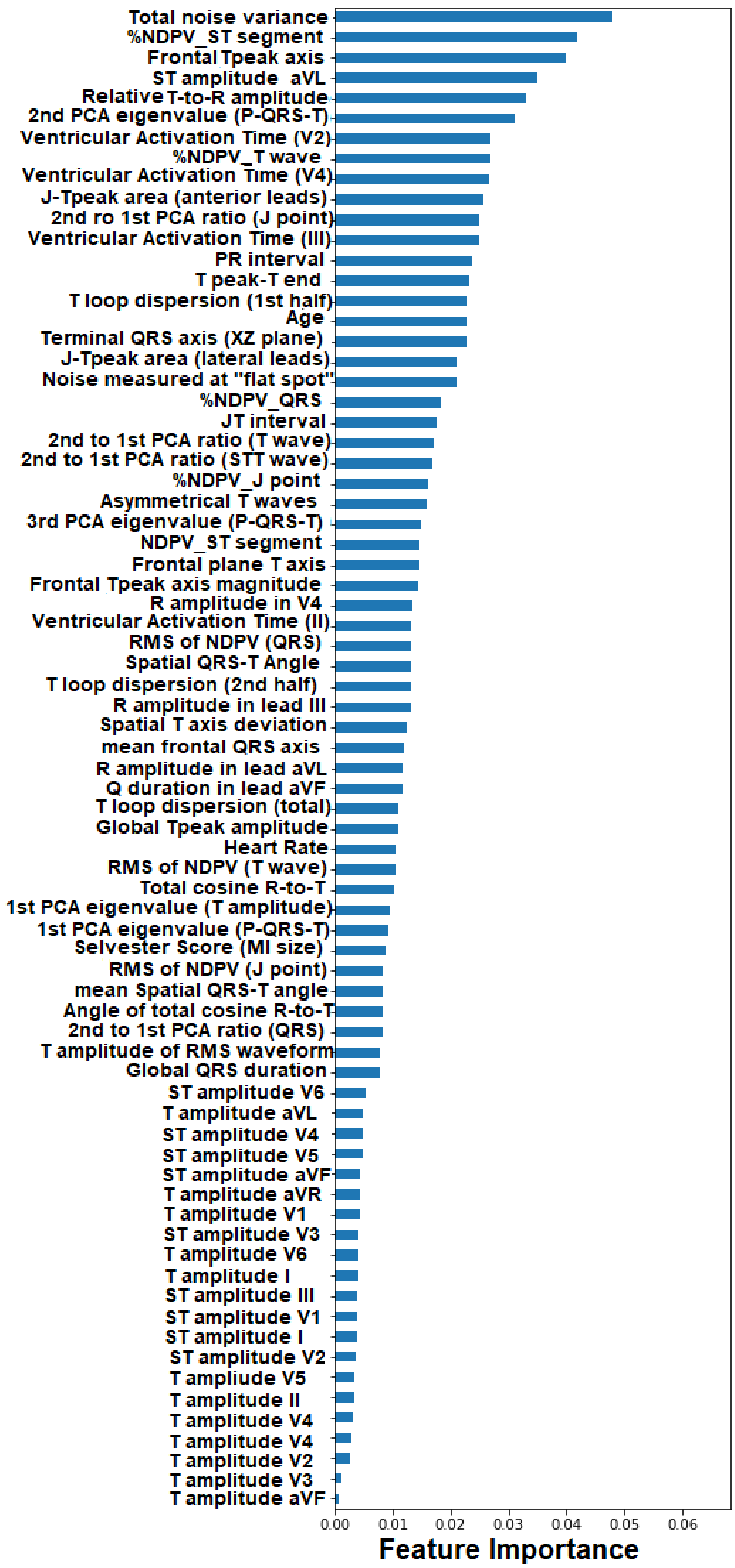


Figure 4. Importance Rank of novel ECG features for the task of NSTE-ACS detection



REFERENCES

1. Body R, Cook G, Burrows G, Carley S, Lewis PS. Can emergency physicians ‘rule in’and ‘rule out’acute myocardial infarction with clinical judgement? Emergency Medicine Journal. 2014;31(11):872-6.
2. Hess EP, Agarwal D, Chandra S, Murad MH, Erwin PJ, Hollander JE, et al. Diagnostic accuracy of the TIMI risk score in patients with chest pain in the emergency department: a meta-analysis. Cmaj. 2010;182(10):1039-44.
3. Hess EP, Brison RJ, Perry JJ, Calder LA, Thiruganasambandamoorthy V, Agarwal D, et al. Development of a clinical prediction rule for 30-day cardiac events in emergency department patients with chest pain and possible acute coronary syndrome. Annals of emergency medicine. 2012;59(2):115-25. e1.
4. Al-Zaiti SS, Shusterman V, Carey MG. Novel technical solutions for wireless ECG transmission & analysis in the age of the internet cloud. Journal of electrocardiology. 2013;46(6):540-5.
5. Birnbaum Y, Wilson JM, Fiol M, de Luna AB, Eskola M, Nikus K. ECG diagnosis and classification of acute coronary syndromes. Annals of Noninvasive Electrocardiology. 2014;19(1):4-14.
6. Quinn T, Johnsen S, Gale CP, Snooks H, McLean S, Woollard M, et al. Effects of prehospital 12-lead ECG on processes of care and mortality in acute coronary syndrome: a linked cohort study from the Myocardial Ischaemia National Audit Project. Heart. 2014;100(12):944-50.
7. Wagner GS, Macfarlane P, Wellens H, Josephson M, Gorgels A, Mirvis DM, et al. AHA/ACCF/HRS recommendations for the standardization and interpretation of the electrocardiogram: part VI: acute ischemia/infarction a scientific statement from the American Heart Association Electrocardiography and Arrhythmias Committee, Council

on Clinical Cardiology; the American College of Cardiology Foundation; and the Heart Rhythm Society Endorsed by the International Society for Computerized Electrocardiology. Journal of the American College of Cardiology. 2009;53(11):1003-11.

8. Al-Zaiti SS, Callaway CW, Kozik TM, Carey MG, Pelter MM. Clinical utility of ventricular repolarization dispersion for real-time detection of non-ST elevation myocardial infarction in emergency departments. Journal of the American Heart Association. 2015;4(7):e002057.

9. Lux RL. Non-ST-Segment Elevation Myocardial Infarction: A Novel and Robust Approach for Early Detection of Patients at Risk. Am Heart Assoc; 2015.

10. Thygesen K, Alpert JS, Jaffe AS, Chaitman BR, Bax JJ, Morrow DA, et al. Fourth universal definition of myocardial infarction (2018). European heart journal. 2019;40(3):237-69.

11. Leisy PJ, Coeytaux RR, Wagner GS, Chung EH, McBroom AJ, Green CL, et al. ECG-based signal analysis technologies for evaluating patients with acute coronary syndrome: A systematic review. Journal of electrocardiology. 2013;46(2):92-7.

12. Al-Zaiti SS BL, Bouzid Z, Faramand Z, Frisch S, Martin-Gill C, Gregg R, Saba S, Callaway C, & Sejdić E. Machine Learning-Based Prediction of Acute Coronary Syndrome Using Only the Pre-Hospital 12-Lead Electrocardiogram. Nature Communications. 2020.

13. Deo RC. Machine learning in medicine. Circulation. 2015;132(20):1920-30.

14. Al-Zaiti SS, Martin-Gill C, Sejdić E, Alrawashdeh M, Callaway C. rationale, development, and implementation of the electrocardiographic methods for the prehospital identification of non-st Elevation Myocardial Infarction Events (EMPIRE). Journal of electrocardiology. 2015;48(6):921-6.

15. Arini PD, Baglivo FH, Martínez JP, Laguna P. Evaluation of ventricular repolarization dispersion during acute myocardial ischemia: spatial and temporal ECG indices. Medical & biological engineering & computing. 2014;52(4):375-91.

16. Al-Zaiti SS, Alrawashdeh M, Martin-Gill C, Callaway C, Mortara D, Nemec J. Evaluation of beat-to-beat ventricular repolarization lability from standard 12-lead ECG during acute myocardial ischemia. Journal of electrocardiology. 2017;50(6):717-24.

17. Al-Zaiti S, Sejdić E, Nemec J, Callaway C, Soman P, Lux R. Spatial indices of repolarization correlate with non-ST elevation myocardial ischemia in patients with chest pain. Medical & biological engineering & computing. 2018;56(1):1-12.

18. Xie J, Wang C. Using support vector machines with a novel hybrid feature selection method for diagnosis of erythemato-squamous diseases. Expert Systems with Applications. 2011;38(5):5809-15.

19. Forberg JL, Green M, Björk J, Ohlsson M, Edenbrandt L, Öhlin H, et al. In search of the best method to predict acute coronary syndrome using only the electrocardiogram from the emergency department. Journal of electrocardiology. 2009;42(1):58-63.

20. Green M, Björk J, Forberg J, Ekelund U, Edenbrandt L, Ohlsson M. Comparison between neural networks and multiple logistic regression to predict acute coronary syndrome in the emergency room. Artificial intelligence in medicine. 2006;38(3):305-18.

21. Wu C-C, Hsu W-D, Islam MM, Poly TN, Yang H-C, Nguyen P-AA, et al. An artificial intelligence approach to early predict non-ST-elevation myocardial infarction patients with chest pain. Computer methods and programs in biomedicine. 2019;173:109-17.

22. DeLong ER, DeLong DM, Clarke-Pearson DL. Comparing the areas under two or more correlated receiver operating characteristic curves: a nonparametric approach. Biometrics. 1988:837-45.

23. Myers PD, Scirica BM, Stultz CM. Machine learning improves risk stratification after acute coronary syndrome. Scientific reports. 2017;7(1):1-12.

24. Abächerli R, Twerenbold R, Boeddinghaus J, Nestelberger T, Mächler P, Sassi R, et al. Diagnostic and prognostic values of the V-index, a novel ECG marker quantifying spatial heterogeneity of ventricular repolarization, in patients with symptoms suggestive of non-ST-elevation myocardial infarction. International Journal of Cardiology. 2017;236:23-9.

25. Alhamaydeh M, Gregg R, Ahmad A, Faramand Z, Saba S, Al-Zaiti S. Identifying the most important ECG predictors of reduced ejection fraction in patients with suspected acute coronary syndrome. Journal of Electrocardiology. 2020;61(4):81-5.

26. Strebel I, Twerenbold R, Wussler D, Boeddinghaus J, Nestelberger T, de Lavallaz JdF, et al. Incremental diagnostic and prognostic value of the QRS-T angle, a 12-lead ECG marker quantifying heterogeneity of depolarization and repolarization, in patients with suspected non-ST-elevation myocardial infarction. International journal of cardiology. 2019;277:8-15.

27. Lines GT, de Oliveira BL, Skavhaug O, Maleckar MM. Simple t-wave metrics may better predict early ischemia as compared to st segment. IEEE Transactions on Biomedical Engineering. 2016;64(6):1305-9.