Journal of Electrocardiology 51 (2018) 683–686

Contents lists available at ScienceDirect

Journal of Electrocardiology

journal homepage: www.jecgonline.com

A model for predicting emergency physician opinion of electrocardiogram tracing data quality

Mat Goebel [a,*], Luke Busico [b], Greg Snow [c], Joseph Bledsoe [d]

[a] UC San Diego School of Medicine, San Diego, CA, United States
[b] Intermountain Medical Center, EKG Department, Murray, UT, United States
[c] Intermountain Office of Research, Murray, UT, United States
[d] Intermountain Medical Center, Emergency Department, Murray, UT, United States

ARTICLE INFO

Available online xxxx

Keywords:
STEMI
Emergency department
Data quality
Regression modeling

ABSTRACT

Background: Limited work has established an objective measure of ECG quality that correlates with physician opinion of the study. We seek to establish a threshold of acceptable ECG data quality for the purpose of ruling out STEMI derived from emergency physician opinion.
Methods: A panel of three emergency physicians rated 240 12-Lead ECGs as being acceptable or unacceptable data quality. Each lead of the ECG had the following measurements recorded: baseline wander, QRS signal amplitude, and artifact amplitude. A lasso regression technique was used to create the model.
Results: The area under the curve for the model using all 36 elements is 1.0, indicating a perfect fit. A simplified model using 22 terms has an area under the curve of 0.994.
Conclusions: This study demonstrated that emergency physician opinion of ECG quality for the purpose of ruling out STEMI can be predicted through a regression model.

Introduction

The electrocardiogram (ECG) is an essential component to a timely diagnosis of ST-segment elevated myocardial infarction (STEMI), and many prehospital systems and emergency departments rely on a computerized diagnosis of STEMI to trigger a physician overread [1–3]. While it is widely recognized that the technical quality of an ECG tracing affects its diagnostic utility, limited work has established an objective measure of ECG data quality that correlates with physician opinion of the study. Previous research used both physicians and non-physicians with varying levels of experience with ECGs [4–18]. None of the raters were representative of the emergency department, or tailored to any one particular clinical use over another. These studies also created complex computer algorithms that relied on having the ECG in a digital format that can be passed to a computer (e.g., XML). Image files (e.g., jpeg or PDF) cannot be used for this purpose, even though hospitals may access all of their ECGs in this format. No research has ventured to create a system that establishes a data quality cutoff where an emergency physician is or is not willing to interpret a test for ruling out the diagnosis of STEMI. Though acceptable quality may vary based on factors like infarct location and amount of overall ST-segment elevation, or may be more accurately be expressed as a confidence interval, this study seeks to identify some threshold of acceptable ECG data quality for ruling out the diagnosis of STEMI derived from the opinion of emergency physicians.

Methods

A priori calculations indicated a needed sample size of 240 ECGs in a 2:1 diagnostic to non-diagnostic ratio to achieve 80% power in our regression model. 160 ECGs were randomly selected from all ECGs obtained in an urban academic emergency department in 2011 by query of the MUSE ECG database (GE Healthcare Inc., Chicago, IL USA). 80 ECGs were randomly selected that had been electronically confirmed by a physician with a statement indicating the ECG was of poor technical quality. Both groups were random sampling of all ED findings, and none were associated with an actual diagnosis of STEMI. We intentionally excluded ECGs diagnostic of STEMI in order to force physicians to look at ECGs with the highest level of scrutiny, thus identifying what is likely the upper bound of acceptable quality. All ECGs were completely anonymized, and the computerized interpretation was removed prior to utilization in the study – i.e., raters only saw a page of waveforms. A privacy waiver was granted by our institutional review board for collecting and analyzing the data. Funding for the statistical analysis was provided by a grant from the Intermountain Research and Medical Foundation.

* Corresponding author at: 3148 Lighthouse Ridge Lane, San Diego, CA 92110, United States.
E-mail address: mrgoebel@ucsd.edu (M. Goebel).

https://doi.org/10.1016/j.jelectrocard.2018.05.001
0022-0736/

Three emergency physicians were randomly selected from a group of 113 emergency physicians staffing 5 hospitals with a combined annual ED census of ~230,000 patients in an urban/suburban region with an American College of Surgeons designated Level 1 trauma center serving as the tertiary referral center for the 23 hospital integrated healthcare system. The panel of three emergency physicians were asked to "rate each of the following ECGs as being acceptable or unacceptable data quality for the purpose of ruling out STEMI, given the patient has a clinical presentation consistent with acute coronary syndrome." Consensus of the ECG quality was obtained when at least 2 of the 3 physicians agreed.

Each lead of the 12-lead ECG had the following measurements recorded: baseline wander, QRS signal amplitude, and artifact amplitude. Measurements were manually taken with ECG calipers, measured peak-to-peak, and recorded in mm (see Fig. 1). Each ECG was independently measured by two hospital ECG technicians randomly assigned from a pool of four. Each pair of measurements was averaged to obtain a single value that was passed to the statistician for analysis.

For the regression, we started by fitting a logistic regression model using the physician rating variable as the response and the full set of 36 predictor variables. Since the goal was a predictive model, the lasso was applied to push the coefficient estimates towards zero [19]. The lasso is a technique that finds the best fitting model subject to the constraint that the sum of the absolute values of the standardized slopes be less than or equal to a specified bound [19–21]. This biases the estimates of the coefficients towards zero, but gives a more stable model with smaller standard errors. We fit the lasso model with several bounds and then examined the generalized cross-validation score to select the bound for use in the final model.

Results

Interrater agreement was analyzed using Fleiss' kappa, and in pairs using Cohen's kappa. Physician raters exhibited moderate to good interrater reliability, with a Fleiss' kappa of 0.598 and Cohen's kappa values ranging from 0.521–0.662 (see Table 1). Agreement amongst raters was also analyzed by interclass correlation, which was 0.872. Measurement discrepancy between ECG technician measurements ranged from −1 to 1 mm, mean 0.04 mm, (SD = 0.48 mm).

During the regression process, two of the coefficients went to zero indicating they could be dropped from the model. Details of the final coefficients are found in Table 2. The prediction from this model can be converted to a probability using the inverse logit transform exp.(x)/(1 + exp.(x)).

Table 1
Physician interrater reliability.

Pair	Kappa
MD1 vs MD2	0.625 ($p < 0.001$)
MD1 vs MD3	0.521 ($p < 0.001$)
MD2 vs MD3	0.662 ($p < 0.001$)

When the predictions from this model are applied to the original data there is complete separation. All the data points with a predicted physician rating of zero (unacceptable quality) had a predicted probability <0.5 and all the data points with a predicted physician rating of one (acceptable quality) had a predicted probability >0.5 (see Fig. A.1), i.e. there was no overlap in the predictions between groups, so there was a no gray area in classification. The area under the curve is 1, indicating a perfect fit (see Fig. 2). The coefficients for a simplified model with 22 rather than 37 terms are found in Table A.1. This simplified model has an area under the curve of 0.994 (see Fig. 3).

Discussion

We have described a method for modeling ECG data quality for the purpose of diagnosing STEMI derived from emergency physician consensus. Our approach is novel, both in using emergency physician opinion as the gold standard, and in designing our method to be compatible with image formats as well as other formats of ECG data. When working with outside organizations, such as EMS, ECG data may only be available as images; eg. PDF, JPEG, etc. Our model was extremely accurate in predicting emergency physician opinion, with an area under the curve of 1. Most emergency departments rely on the treating emergency physician to activate the cardiac catheterization lab, therefore utilizing emergency physician opinion adds to the validity of this study.

Performing quality assurance and improvement for ECG acquisition is impossible without knowing how to differentiate between diagnostic and non-diagnostic ECGs. This test could allow auditing of ECG quality between facilities or emergency medical service agencies, identifying if there is a consistent source of low quality ECGs. When problems with quality are objectively identified, resources to address those deficiencies can be justified, improvement can be tracked, and the cost effectiveness of training to improve ECGs can be studied.

The differences between the full and simplified models may illustrate which measurements are more important to emergency physicians in determining ECG quality. The simplified model eliminated the baseline wander measurements in the augmented voltage limb leads, the signal measurement in most of the precordial leads, and some of

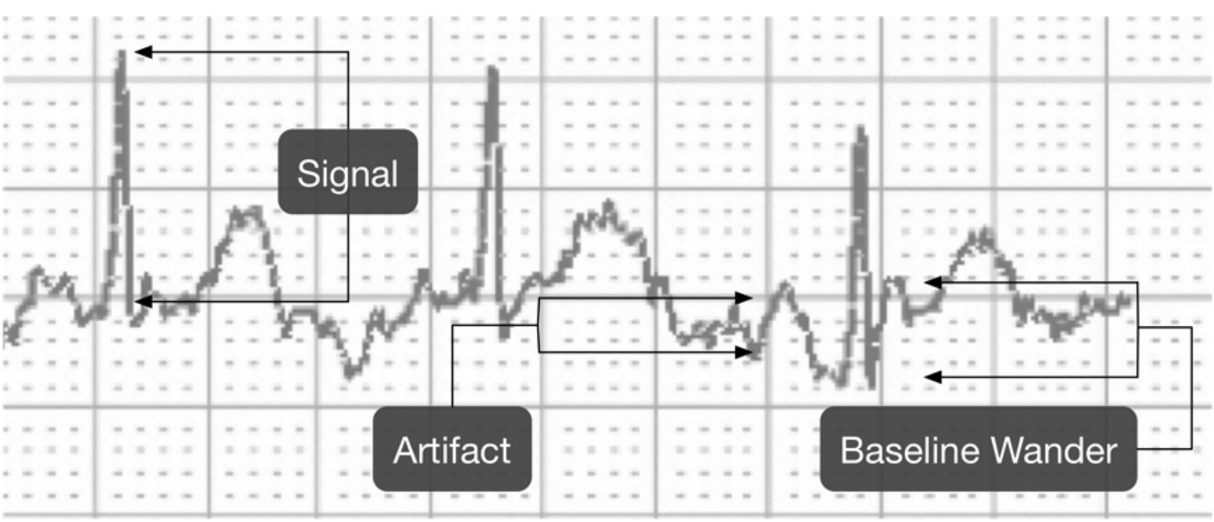

Fig. 1. Sample ECG with calipers showing baseline wander, artifact amplitude, and QRS amplitude measurements.