Statistical analysis of cross-correlation index for identifying abnormal ECG signals

1st Joana Florescano Olguín
School of Sciences of Engineering and Technology
Autonomous University of Baja California
Tijuana, México
joana.florescano.olguin@uabc.edu.mx

2nd Daniela M. Martínez
School of Sciences of Engineering and Technology
Autonomous University of Baja California
Tijuana, México
daniela.martinez@uabc.edu.mx

3rd Norma A. Barboza Tello
School of Sciences of Engineering and Technology
Autonomous University of Baja California
Tijuana, México
norma.barboza@uabc.edu.mx

4th Irma Uriarte Ramírez
School of Sciences of Engineering and Technology
Autonomous University of Baja California
Tijuana, México
irma_ur@uabc.edu.mx

***Abstract*—In this study, we propose conducting a statistical analysis on the cross-correlation index of a given electrocardiogram (ECG) signal to swiftly ascertain whether the signal corresponds to a healthy heart. This information serves as a rapid triage for patients presenting with chest pain, facilitating the prompt exclusion of the possibility of a heart disease. The results illustrate that the proposed algorithm adeptly classified as much as 80% of the normal signals alongside up to 85% of the abnormal signals under specific parameter configurations. This indicates its effectiveness in detecting the presence of abnormal signals.**

***Index Terms*—Electrocardiogram signal, signal classification, cross correlation, statistical analysis**

Paper to be considered in the CSCI-RTHI

I. Introduction

As per the United States Department of Health and Human Services, chest pain stands as the second most common reason for adults seeking care at the emergency department (ED). When assessing a patient with chest pain, medical providers must initially determine whether the pain stems from a life-threatening cause or other non-urgent factors. According to these statistics, among the life-threatening causes, 31% of cases align with acute coronary syndrome, a potentially fatal heart condition. Conversely, nearly 60% of adults arriving at the ED with chest pain receive a diagnosis related to non-life-threatening causes, such as gastrointestinal reflux disease or musculoskeletal issues [1].

Furthermore, it is widely acknowledged that EDs frequently experience congestion due to the diverse range of individuals presenting with complaints of varying severity levels [2]. To ensure that patients seeking emergency medical services (EMS) receive timely and appropriate medical attention, there is a pressing need to enhance triage systems with automated and precise technological solutions. These solutions aim to facilitate swift and reliable diagnoses, enhancing the overall efficiency of emergency healthcare services.

Multiple efforts have been made in developing technological solutions to improve triage in EMS. In [3] the authors delve into various technology-driven solutions aimed at assessing stroke and and furnishing tools to aid in triage decisions. The authors of [4] offer a comprehensive review of machine learning-based proposals designed to prioritize patients in telemedicine triage scenarios.

This study aims to introduce an algorithm designed to detect abnormal ECG signals by focusing on two key characteristics: the presence of ectopic rhythm and morphological features of the ECG waveform. Ectopic rhythm determination involves assessing the heart rate and the amplitude of the QRS interval, while morphological features are evaluated through statistical analysis of the cross-correlation index. The key advantages of the proposed algorithm lie in its ease of implementation, high sensitivity, and the absence of a requirement for training sequences to make decisions. As a result, this algorithm holds promise for further development as a responsive and resource-efficient technological solution, potentially streamlining the triage process in EMS.

Based on the simulation results, the presented algorithm successfully classified as much as 80% of normal and up to 85% of abnormal ECG signals under specific parameter configurations. However, the analysis of these outcomes suggests the need for further considerations to enhance algorithm accuracy. One such aspect for improvement is the assessment of peak amplitude, indicating a potential avenue for refining the algorithm's performance.

II. Electrocardiogram signals

The Electrocardiogram (ECG) is the final outcome of a complex series of physiological and technological processes aimed at representing the electrical activity of the heart originated by the contractile activity of the cardiac muscle, and is one of the

This investigation is supported by the project 351/3446 registered by the Autonomous University of Baja California.

most recognized and used biomedical signals in the field of medicine.

An ECG signal can be acquired by using electrodes attached to the skin surface. ECG signal visualization changes depending on the electrodes placement and these different representations are given the name of leads. The most common leads are that formed by the electrical potential difference between the left leg and the right arm. [7].

The normal waveform morphology of an ECG consists of a P wave, that represents the atrial depolarization; a QRS complex, that represents the ventricular depolarization; and a T wave, that represents the ventricular repolarization. The time interval between waves of two heartbeats depends on the heart rate and rhythm [8]. A graphical representation of the waves comprising a heartbeat is presented in Figure 1.

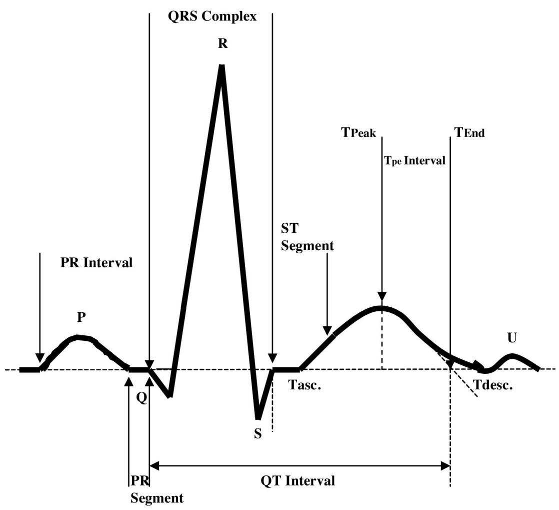

Fig. 1. P, QRS, and T waves of a heartbeat. Image source: Ref. [7]

The characteristics of a normal ECG signal that presents sinus rhythm are:

- Heart rate ranging from 60 to 100 beats per minute (bpm).
- Periodic occurrence of P, QRS and T waves without any significant changes in their typical amplitudes or intervals.
- QRS complex amplitude ranging from 0.5 - 2.5mV.

Any heart rhythm that is different from sinus rhythm is called ectopic rhythm. An ectopic heart rhythm could be caused by several heart conditions including chronic cardiovascular diseases and other abnormalities such as cardiac arrhythmia. Cardiac arrhythmias are anomalies in the heart, manifested either by irregularity in the heartbeats or by abnormally fast or slow heart rates, known as tachycardia and bradycardia, respectively. [9]

III. Cross-correlation based ECG classification

A way to identify a potentially abnormal ECG signal is through the analysis of the temporal and morphological characteristics of the electrical waveform. In this work, we propose a cross-correlation based technique to examine the level of similarity between a given ECG waveform and a healthy baseline ECG signal. Additionally, rhythm is evaluated by means of a peak detection routine that allows to determine if the heart rate and QRS amplitude of the ECG signal corresponds to that of a sinus rhythm. Figure 2. depicts the methodology for evaluating the proposed algorithm.

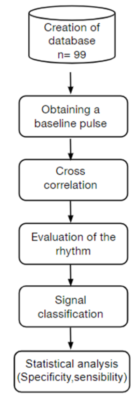

Fig. 2. Flowchart of the methodology used

A. ECG signal database

In order to construct the baseline ECG signal, a set of 16 filtered and preprocessed normal ECG records were considered. The data was obtained from ECG-ID Database from PhysioNet [10] which is an online and open access database that contains 310 ECG recordings obtained from 90 persons. Additionally, in order to evaluate the proposed algorithm, a second set consisting of 50 ECG records was considered. For this second set, ECG signals presenting an ectopic rhythm were selected from the ECG Fragment Database for the exploration of dangerous arrhythmia [11] from PhysioNET. This latter database contains a total of 1016 ECG recordings of the following arrhythmias: Ventricular flutter and fibrillation, ventricular arrhythmias and arrhythmias.

Table I presents the acquisition parameters of the ECG signals considered for this work:

TABLE I
ACQUISITION PARAMETERS FOR EACH ECG RECORD CONSIDERED FOR BASELINE ECG SIGNAL CONSTRUCTION AND ALGORITHM EVALUATION

Parameter	Range or value
Lead	Type II
Sample frequency	500 Hz
Record time	20 s
Number of samples	10,000
Number of waveforms	10 - 20

B. Healthy baseline ECG signal construction

The healthy baseline ECG signal will function as a pattern for identifying sinus (healthy), and ectopic (abnormal) ECG signals. This baseline signal is basically a heartbeat pattern built from 16 healthy ECG records, as they were labeled at the

ECG-ID database. Only one heartbeat of each record is used for building the baseline pulse. This is illustrated in Figure 3.

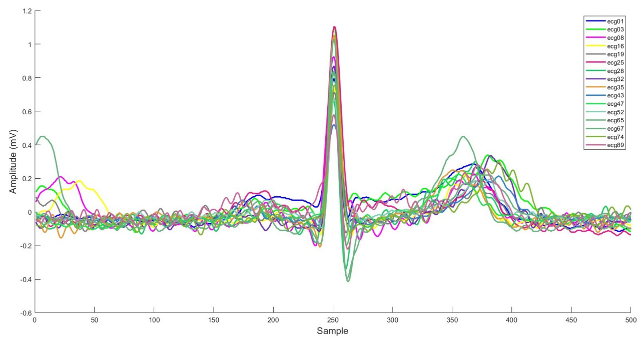

Fig. 3. Family of time series used to construct the ECG reference signal

After syncing the 16 selected pulses, the baseline pulse is obtained by averaging the signal samples as in:

$$x_{bl}\left[1,j\right] = \frac{1}{m}\sum_{i=1}^{m} x\left[i,j\right] \tag{1}$$

For $j = 1 \ldots n$ and n stands for the number of samples considered for each pulse and m = 16 stands for the number of pulses considered for averaging. In this work $n = 500$ samples were considered for each pulse. The resulting baseline pulse is depicted in Figure 4

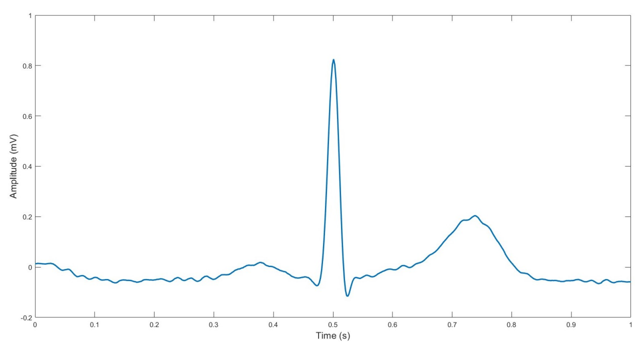

Fig. 4. Baseline ECG pulse used as the reference signal for cross-correlation

C. ECG classification algorithm

The proposed algorithm is divided into two main processes and final decision is made following an AND rule taking as inputs the outputs of each process.

The first process consists of performing cross correlation between the baseline pulse and the received ECG signal. Assuming that $y\left[n\right]$ stands for the discrete representation of the received ECG signal, the discrete cross correlation function can be calculated as follows [12]:

$$r\left[k\right] = \frac{1}{2n-1}\sum_{i=1}^{n} x_{bl}\left[i+k\right] y\left[i\right] \tag{2}$$

Where k corresponds to the discrete time shift applied to $x_{bl}\left[n\right]$. A graphical representation of the cross correlation process is presented in Figure 5.

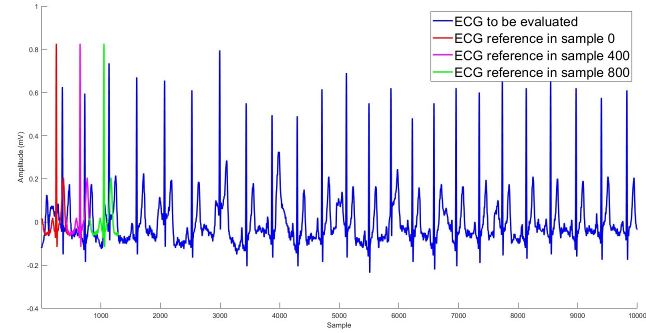

Fig. 5. The blue signal represents the patient's ECG signal, the red, magenta and green signal represents the healthy ECG reference signal at different points.

Equation 2 implies that as $x_{bl}\left[n\right]$ is shifting, the normalized correlation index is calculated on each shift, giving place to a total of $K = N - n + 1$ correlation indexes, where N corresponds to the length of $y\left[n\right]$. Each correlation index indicates the level of similarity between $x_{bl}\left[n\right]$ and $y\left[n\right]$ for the specific sample interval defined by $\left[k+1, n+k\right]$.

In order to determine if the cross correlation process outputs a positive result for an ectopic ECG signal, all the K correlation indexes are compared to a threshold (α) and if the number of $r\left[k\right] > \alpha$ for $k = 1, 2, \ldots, K$ is lower than β, then the ECG is classified as abnormal. α and β can be considered as design parameters for the algorithm, representing the level of similarity expected between two normal ECG pulses and the number of pulses classified as normal expected in the entire record of a healthy ECG signal, respectively.

The second process implied in the algorithm is the assessment of the heart rate and the amplitude of the QRS complex. Heart rate can be calculated by the following expression:

$$\text{bpm} = 60\left(\frac{R_{peaks}}{t_{R-R}}\right) \tag{3}$$

Where R_{peaks} stands for the number of R waves present during the recording and t_{R-R} stands for the time interval between two consecutive R waves. In this proposal, R_{peaks} is calculated by applying a peak finder routine and, considering T as the total recording time, t_{R-R} is estimated by:

$$t_{R-R} = \frac{T}{R_{peaks}} \tag{4}$$

Additionally, the maximum amplitude of the QRS complex is estimated by the maximum value of the ECG recording, this is:

$$V_{max} = \max(y\left[n\right]) \tag{5}$$

Finally, the decision is made based on the AND rule as described in the algorithm 1. The output $class = 0$ determines that the received ECG signal can be considered as normal; while the output $class = 1$ means that the ECG signal is abnormal.

Algorithm 1 ECG classification algorithm

Require: **r** α β bpm V_{max}

$P_n \leftarrow 0$
for $k \leftarrow 1$ to K **do**
  if $r[k] \geq \alpha$ **then**
    $P_n \leftarrow P_n + 1$
  end if
end for
$class \leftarrow 1$
if $P_n \geq \beta$ **then**
  if bpm ≥ 60 & bpm ≤ 100 **then**
    if $V_{max} \geq 0.5$ **then**
      $class \leftarrow 0$
    end if
  end if
end if

D. Performance metrics

The reliability of the algorithm is evaluated using two metrics: specificity and sensitivity. The sensitivity refers to the ability of correctly classifying an ECG signal presenting ectopic rhythm as abnormal. Sensitivity (S) is defined by the conditional probability:

$$S = P(class = 1 \mid \text{ectopic rhytmh}) \tag{6}$$

On the other hand, the specificity refers to the ability of correctly classifying as normal an ECG signal that presents sinus rhythm. Specificity (E) is determined by the following conditional probability:

$$E = P(class = 0 \mid \text{sinus rhythm}) \tag{7}$$

IV. Analysis of Results

In order to evaluate the proposed algorithm, an evaluation set up was developed in Matlab, supported mainly by the Signal Processing Toolbox [13].

For the evaluation, a total of 99 ECG signals from the Physionet database were considered. 49 corresponds to healthy ECG signals and 50 to abnormal ECG signals. Evaluation consists in inputting each of the ECG signals into the algorithm and record the output in order to quantify S and E. The evaluation parameters considered during the evaluation are listed in Table II.

TABLE II
EVALUATION PARAMETERS

Parameter	Value or range
α	0.5 - 1
β	53 and 59
heart rate	60 - 100 bpm
QRS amplitude in mV	≥ 0.5

The α and β values were determined after extensive simulations and statistical analysis. It is well known that the normalized correlation index for two identical signals is equal to 1. Therefore α was chosen to ensure that the similarity between the baseline pulse and each pulse of the received ECG signal was strong enough without compromising significantly the specificity. On the other hand, the values determined for β corresponds to the second decile and the average of the set $\mathbf{R_n}$ defined as the set of correlation indexes that complied with $r[k] > \alpha$ for $k = 1, 2, \ldots, K$ in the 16 ECG normal signals used for building the baseline ECG pulse.

After all the 99 signals were processed by the algorithm, each of the 99 signal classifications were registered as one of the following results type:

1) Normal ECG signal classified as normal, i.e. true negative (TN)
2) Normal ECG signal classified as abnormal, i.e false positive (FP).
3) Abnormal ECG signal classified as normal, i.e. false negative (FN).
4) Abnormal ECG signal classified as abnormal, i.e. true positive (TP).

Finally, S and E were estimated by empirical probabilities according to equations 8 and 9:

$$\hat{S} = \frac{nTP}{nFN + nTP} \tag{8}$$

$$\hat{E} = \frac{nTN}{nFP + nTN} \tag{9}$$

nTP, nTN, nFP and nFN stands for the number of TP, TN, FP and FN results, respectively.

Figure 6 presents the sensitivity and specificity obtained for all the set of configuration parameters considered in II. From this results we can observe that for values of $\alpha \geq 0.85$ the specificity starts to decrease dramatically, while the sensitivity remains mostly constant. This behavior remains similar for both β = second decile and β = 59 (average). However, specificity is overall greater for β = 53 (second decile).

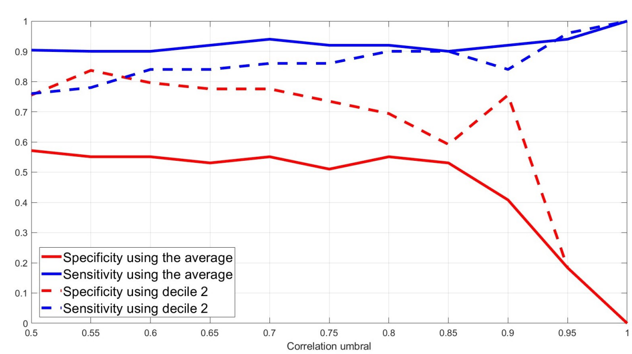

Fig. 6. The solid lines represents the sensitivity and specificity evaluated with the average and the dotted lines represents the sensitivity and specificity evaluated with the decile 2.

According to the results observed in figure 6 the value of β has a significant impact on the sensitivity and the specificity. In this sense, another rounds of simulations were carried out to analyze the performance in terms of different values of β. For this evaluations, the values defined for β are the corresponding to the ten deciles of the set $\mathbf{R_n}$. The corresponding $\hat{S}$ and $\hat{E}$ are shown in figure 7. From figure 7 can be observed how the specificity decreases almost linearly as the decile number is increased, while the sensitivity exhibits only a slightly increase. This results were expected since as we increase the decile number, the amount of required pulses that exhibit a strong similarity with the baseline ECG pulse is also increased, thus giving place to a higher probability of classifying normal signals as abnormal.

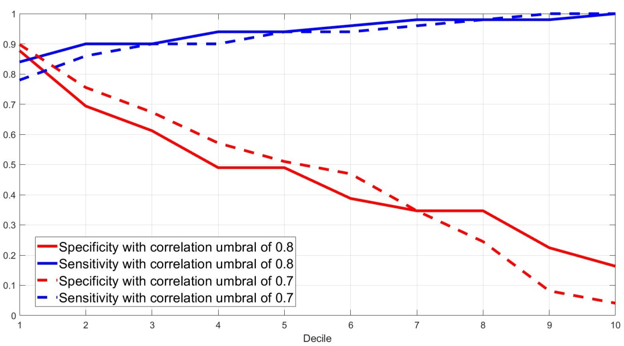

Fig. 7. The solid lines represents the sensitivity and specificity evaluated with a correlation threshold of 0.8 and the dotted lines represents the sensitivity and specificity evaluated with a correlation threshold of 0.7.

V. Conclusions and Future Work

In this work, a cross correlation based algorithm for classifying normal and abnormal ECG signals is proposed and evaluated by means of statistical analysis. From the obtained results we observed that the behavior of sensitivity when evaluated using deciles remains constant but the specificity is dramatically decreased from decile 3. In this sense, the most appropriate value β is the value located in the second decile, achieving a sensitivity of 90% and specificity of 70% for $\alpha = 0.8$. On the other hand, for $\alpha = 0.7$ and second decile, the sensitivity value is 86% and specificity value is 75%, which can be considered a better trade off for the two metrics.

The performance in terms of both metrics is also analyzed with respect to α. Results show that for $0.5 \leq \alpha \leq 0.85$ both sensitivity and specificity remains fairly constant. Therefore, it can be concludes that β has a greater impact on the performance of the algorithm.

It is important to prioritize the value of sensitivity with respect to specificity, since it has a greater impact on the patient's health. A low level of sensitivity means a higher probability of diagnosing a signal as normal when, in fact, it presents an anomaly. However, if the specificity is significantly reduced it could mean that unnecessary resources would be spent in running further clinical tests for diagnosing a healthy heart. Therefore, a trade off between the two metrics must be defined.

This work corresponds to the first stage of a more complex system for classifying cardiac arrhythmias from ECG signals. The objective of this classification system is to analyze the morphology and characteristics of every arrhythmia to automatically adjusts its parameters by means of artificial intelligence or machine learning to ensure reliable classification.

References

[1] J. Tozer and K. Shaw, "Chest Pain," Manual of Austere and Prehospital Ultrasound, pp. 49–73, Jan. 2023, doi: 10.1007/978-3-030-64287-7_7.

[2] S. L. Dong and M. Bullard, "Emergency Department Triage," Evidence-Based Emergency Medicine, pp. 58–65, Aug. 2022, doi: 10.1002/9781444303674.ch7.

[3] J. C. Martinez-Gutierrez, R. V. Chandra, J. A. Hirsch, and T. Leslie-Mazwi, "Technological innovation for prehospital stroke triage: ripe for disruption," J Neurointerv Surg, vol. 11, no. 11, pp. 1085–1090, Nov. 2019, doi: 10.1136/NEURINTSURG-2019-014902.

[4] O. H. Salman, Z. Taha, M. Q. Alsabah, Y. S. Hussein, A. S. Mohammed, and M. Aal-Nouman, "A review on utilizing machine learning technology in the fields of electronic emergency triage and patient priority systems in telemedicine: Coherent taxonomy, motivations, open research challenges and recommendations for intelligent future work," Comput Methods Programs Biomed, vol. 209, Sep. 2021, doi: 10.1016/J.CMPB.2021.106357.

[5] D Mirvis. Goldberger A. L. "Electrocardiography", Heart disease, vol. 1, p. 82-128, 2001.

[6] J. Smith et al., "Placeholder Text: A Study," Citation Styles, vol. 3, Jul. 2021, doi: 10.10/X

[7] E. J. da S. Luz, W. R. Schwartz, G. Cámara-Chávez, and D. Menotti, "ECG-based heartbeat classification for arrhythmia detection: A survey," Comput Methods Programs Biomed, vol. 127, pp. 144–164, Apr. 2016, doi: 10.1016/j.cmpb.2015.12.008.

[8] A. Isin and S. Ozdalili, "Cardiac arrhythmia detection using deep learning," in Procedia Computer Science, Elsevier B.V., 2017, pp. 268–275. doi: 10.1016/j.procs.2017.11.238.

[9] J. A. Kastor and M. M. Saba, "Cardiac Arrhythmias," eLS, Dec. 2009, doi: 10.1002/9780470015902.A0002112.PUB2.

[10] V. Batchvarov, "The Morphology of the Electrocardiogram." Accessed: Oct. 23, 2023. [Online]. Available: https://www.academia.edu/6490566/The_Morphology_of_the_Electrocardiogram

[11] L. A. Manilo, A. P. Nemirko, E. G. Evdakova, and A. A. Tatarinova, "ECG Database for Evaluating the Efficiency of Recognizing Dangerous Arrhythmias," Proceedings - 2021 IEEE Ural-Siberian Conference on Computational Technologies in Cognitive Science, Genomics and Biomedicine, CSGB 2021, pp. 120–123, May 2021, doi: 10.1109/CSGB53040.2021.9496029.

[12] J. Semmlow, "Circuits, Signals and Systems for Bioengineers (Third Edition)," 2018.

[13] "Signal Processing Toolbox Documentation - MathWorks América Latina." Accessed: Oct. 03, 2023. [Online]. Available: https://la.mathworks.com/help/signal/