

A METHOD FOR DETECTING CORONARY ARTERY DISEASE USING NOISY ULTRASHORT ELECTROCARDIOGRAM RECORDINGS

Orestis Apostolou*

Vasileios Charisis*

Georgios Apostolidis*

Leontios J. Hadjileontiadis^{†,*}

[†]Department of Biomedical Eng., Khalifa University, PO Box 127788, UAE

* Department of Electrical & Computer Eng., Aristotle University of Thessaloniki, GR-54124, Greece

ABSTRACT

The current study aims at creating an algorithm able to detect Coronary Artery Disease (CAD), using ultrashort (duration of 30 seconds) one-lead ECG recordings. The presented method is designed to allow both electrode and noisy recordings (deriving from a smartwatch) as input. This is achieved by using an autoencoder neural network, which inspects the quality of each recording. The algorithm's core is a Support Vector Machine (SVM) model, which evaluates each patient's recordings and predicts whether they indicate CAD. Using statistics and combining the models mentioned above, a light, reliable, easy to use predicting system is created, suitable for deployment in a mobile application, which uses a smartwatch as its recording tool.

Index Terms— coronary artery disease, autoencoder, SVM, electrocardiogram, smartwatch

1. INTRODUCTION

During the last decade, there has been an exponential rise in the applications which use Artificial Intelligence (AI). More problems, which seemed unsolvable in the previous century, are now resolved with remarkable effectiveness using AI systems. The capabilities created by utilizing the tool abundance a scientist has in hands are limitless. The results of this scientific flourishing are effortless to spot in everyday life. The current study suggests a diagnostic method for CAD [1]. CAD is the narrowing or the blockage of the heart arteries caused by the buildup of fat, cholesterol, and other substances, a situation known as atherosclerosis. If not detected timely, it can lead to heart failure or death. In 2015, CAD was responsible for 15.6% of the total deaths in the year, making it the most common death factor globally.

During the last years, many AI algorithms have been created trying to detect CAD at early stages. Most of them use long-duration ECG recordings (24-hour or 15-minute)

as input, and with the aid of either classic machine learning or deep learning methods, they predict the existence of CAD. Some of those techniques are very effective, approaching the accuracy of 100% in their predictions. However, those techniques are challenging to use in daily life due to people's limited access to accurate digital ECGs. Even when that access is feasible, it is impractical to utilize this technique for everyday use.

In the current study, we approach the issue from a different perspective. There has been an effort to create a light algorithm that uses ECG recordings of 30 seconds duration, suitable for coping with either noiseless waveforms deriving from electrodes or noisy waveforms coming from smartwatches. Initially, ECGs are separated into R-R intervals, and afterward, each interval is imported into an autoencoder network. The network tries to reconstruct the initial R-R waveform and estimates the current noise level. If this level exceeds a certain threshold, then the waveform is characterized as noisy. After evaluating each R-R interval, it is adjudicated if the 30-second sample contains a significant amount of noise. If so, the waveform is rejected as noisy. If not, it proceeds in the following stages of the procedure. During these stages, feature extraction is executed, which is followed by a Principal Component Analysis. Then, the modified features are imported into an SVM model to be classified as having CAD or not. The last stage of the algorithm is the evaluation of more recordings coming from the same person so that the final result can be more accurate.

Regarding the structure of the current study, the preprocessing algorithm stages are described in Section 2. In Section 3, all details of the SVM model are presented. The statistical methods used for the evaluation are included in Section 4. In Section 5, the method and its individual stages are evaluated. In Section 6, comparisons to other published studies are presented, and Section 7 concludes the paper.

2. PREPROCESSING

The preprocessing stages presented here were applied both for the training and testing data of the SVM. This way, objective evaluation of the procedure was ensured.

This research project was funded by the Abu Dhabi Department of Education and Knowledge (ADEK), UAE, under the Award for Research Excellence (AARE) 2018, ref. no: 29934

2.1. Data acquisition

The data used in the application fall into two categories: data of people having CAD and data of people without CAD. The separation was designed so that the final classification problem would be whether a person has CAD or not. The databases used were MIT-BIH Normal Sinus Rhythm Database [2], Long Term ST Database [3], and the Combined measurement of ECG, Breathing, and Seismocardiograms [4]. Those three databases contain long-duration ECGs. Therefore the data was separated into samples of 30 seconds. There were available recordings of 56 people with CAD (35 men, 21 women) and 39 healthy people (17 men, 22 women) in total.

Among the CAD patients, some subjects also had other medical diagnoses. To achieve a total classification focus in CAD detection, we used only samples from people who suffered exclusively from CAD or were completely healthy in the model training procedure. Thus, there were 25 subjects diagnosed with CAD and 18 healthy used for training. The rest of the data was used for testing purposes.

2.2. R-R intervals separation

It is widely known that the movement of the heart muscles is periodic, and it can be separated into segments. The most characteristic section in a circadian circle is the R peak, representing the heartbeat. The detection of the R peaks is a critical stage of the preprocessing of the application. The problem of peak detection has been encountered with numerous different approaches and, many times, with exceptional accuracy. The current application uses a variation of the classic Pan-Tompkins algorithm [5, 6]. This variation is not real-time; hence, it is not restricted to processing causal signals. This property helps the algorithm correct the mispredictions created at the beginning of the waveform caused by the initialization of the filters. Also, it allows the algorithm to rectify other mispredictions as well.

2.3. Autoencoder

Taking into consideration that the recordings used in the application may originate from smartwatches, they may include significant noise levels. The current application tries to detect recordings containing significant noise levels and reject them so that they do not lead to mispredictions. For that reason, an autoencoder neural network was constructed aimed to detect abnormalities in the R-R samples. This network consists of one encoder and one symmetrical decoder, each one having three hidden layers. The neural network intends to reconstruct the initial R-R sample. Random sample intervals from the healthy dataset were used (some of the non-healthy people display heart episodes that cause non-regular heartbeats) to train the neural network. Af-

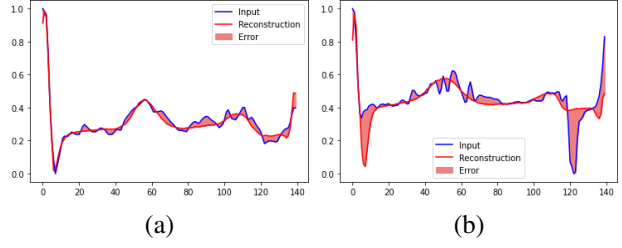


Figure 1: Example of R-R interval reconstruction created by the autoencoder network. (a) Noiseless and (b) Noisy reconstruction.

ter completing the network training, all the reconstructions' mean absolute error is calculated by comparing the initial waveform and its reconstruction. If it is found to exceed the selected threshold, the R-R interval is labeled as noisy. This threshold was chosen to be equal to $E\{MAE\} + 2.5 \cdot \sigma\{MAE\}$. Finally, if the percentage of the noisy R-R intervals is greater than 50%, the whole 30-second sample is being labeled as noisy and gets rejected from the rest of the procedure.

2.4. Features extraction

For the construction of the machine learning model, numerous features were extracted in the time and frequency domains, as well as some non-linear features. Subsequently, we will analyze some of their characteristics.

2.4.1. Time domain features

Most of the time-domain features are related to Heart Rate Variability (HRV) [7]. This metric is affected by the multiple processes inside the human body, and, usually, its fluctuation can not be fully understood. The features presented here are divided into two categories: statistic metrics, like SDNN, and geometric metrics, like HTI. Features referencing R-R intervals are affected from each R-R interval inside the initial 30-second waveform. In contrast, features referencing N-N intervals are affected only by regular R-R intervals, meaning the intervals whose duration does not differ majorly to their adjacent intervals.

2.4.2. Frequency domain features

Fast Fourier Transformation (FFT) was used to transform our signal from time to frequency domain. Even though there are more available techniques for executing this transformation, we have chosen to use only this algorithm since its computational cost is negligible and the information gain we had from implementing different algorithms was not significant.

Table 1: List of all extracted features

Name	Domain	Description
SDNN	Time	Standard deviation of N-N intervals
SDRR	Time	Standard deviation of R-R intervals
AHR	Time	Average heart rate
SDSD	Time	Standard deviation of successive diff.
pNN50	Time	Percentage of adjacent N-N intervals with more difference than 50 ms
HTI	Time	Heart rate variability triangular index
HRMM	Time	Heart rate maximum minus minimum
LFBE	Freq.	Low frequency band energy
LFBEF	Freq.	Low frequency band energy percentage
LFP	Freq.	Low frequency peak
HFBE	Freq.	High frequency band energy
HFBEF	Freq.	High freq. band energy percentage
HFP	Freq.	High frequency peak
LF/HF	Freq.	Low freq. to high freq. energy ratio
PSD1	NonLin.	Poincaré primary standard deviation
PSD2	NonLin.	Poincaré secondary standard deviation
PR	NonLin.	Poincaré Ratio
PA	NonLin.	Poincaré ellipsis area
AEM	NonLin.	Mean approximate entropy
AES	NonLin.	Std. of approximate entropy
SEM	NonLin.	Mean sample entropy
SES	NonLin.	Std. of sample entropy

2.4.3. Non-linear features

Since the nature of the problem of detecting CAD is non-linear, approaching its solution using only linear features would be inaccurate. In this section, we will discuss some of the non-linear features that were extracted, along with some comments on their programming implementation based on the computational limitations this problem imposes.

The Poincaré plot [8] is the first group of non-linear extracted features. This plot depicts the relationship between the duration of each R-R interval duration in comparison to the duration of the next interval. It contains crucial statistical information about the HRV of the subject, and by using it, we extract the standard deviation on the primary and secondary axes, the area of the ellipsis created from these two vectors, and the ratio between the two values.

Approximate entropy is another group of extracted features we analyze in this preprocessing step. Approximate entropy measures the regularity and predictability of a signal [9]. An irregular and unpredictable waveform has a greater approximate entropy.

Since the computational complexity of this feature is $O(N^2)$, we could not calculate it directly. Instead, we separate our signal in 5-seconds intervals with 50% overlapping. Afterward, we undersample the signals at a frequency of 64 Hz and then calculate the approximate entropy of each individual. Finally, we calculate the mean value and the standard deviation of these intervals and use them as our

extracted features. This process is followed so that this feature can be calculated quickly, even in mobile processors.

Sample entropy is the last group of features we extract. Similarly to approximate entropy, it measures the regularity and predictability of a signal and has a complexity of $O(N^2)$ [10]. Thus, we follow the same procedure with approximate entropy to calculate its mean value and standard deviation.

2.5. Principal component analysis

A crucial step in designing machine learning models is to implement dimensionality reduction appropriately. It is expected that there is a high level of information redundancy inside our extracted features. Using algorithms to reduce this redundancy is necessary to avoid creating an overfitting model. In the current research, we utilized the Principal Component Analysis (PCA) algorithm [11]. Initially, we had 23 extracted features. After implementing the algorithm, we had 12 components and a percentage of 94.66% variance compared to the previous state.

3. SUPPORT VECTOR MACHINE MODEL

Having finished the preprocessing tasks, the next step is to design a machine learning model to classify our samples. Among the available models, we decided to use a Support Vector Machine (SVM) [12]. SVM is a supervised machine learning model suitable for classification problems. We constructed an SVM model using the reduced extracted features mentioned in the previous chapter. The kernel of the model was chosen to be radial, as it seemed to produce better results and had a shorter training time. Gamma value was selected to be equal to 0.005.

Even though the subjects selected to participate in the training process diagnosed with CAD were significantly more than the healthy ones (25 against 18), we decided to feed the model with the same number of samples for each class during the training. This has been done so as not to impose another factor of imbalance on the model. Thus, the training was completed using 40000 30-second samples from each class.

4. DECISION ON SUBJECT LEVEL

In the proposed method, the subject is allowed to use more than one 30-second ECG to detect CAD. It is understandable that a sample of 30 seconds can not contain the same amount of information as a 24-hour recording would. Thus, by using more than one recording, subjects can enhance the accuracy of their predictions. The algorithm calculates the ratio of the positive to total predictions (excluding the ones that are rejected as noisy). If this ratio exceeds a certain

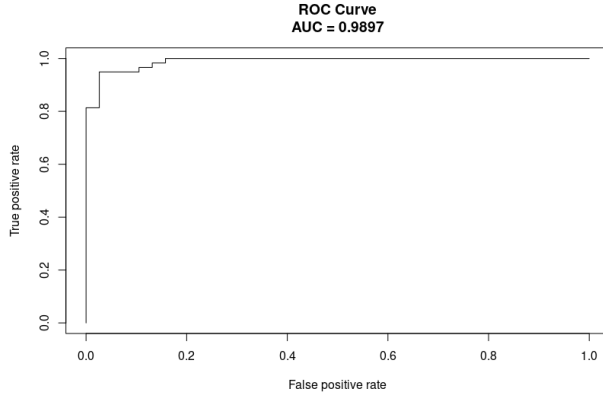


Figure 2: ROC Curve for different evaluation thresholds.

threshold, the subject is characterized as having CAD. Else, he is characterized as not having CAD. A minimum number of 30-second samples for a reliable result was empirically determined to be 10, as a trade-off between practicality and effectiveness.

5. RESULTS

In the current section, we will present the testing procedure results, both for each recording individually and for each subject in total. To keep the testing procedure unbiased, we used a variation of the Leave-one-out cross-validation. Each time we tested the algorithm results for either a 30-second recording or a subject, all of the samples belonging to the current subject were excluded from the training process.

For the individual evaluation of the recordings, the precision of the model was 96.04%, the recall was 83.16%, and the f-score was 89.14%, having a total accuracy of 83.46%. The significant difference between precision and recall was caused firstly by the imbalance of testing samples and, secondly, by the imbalance of the subjects participating in the training process.

Regarding the subjects' evaluation, we had to inspect the ideal threshold of their CAD predictions ratio. For this purpose, we created the ROC curve shown in Figure 2. The Area Under Curve (AUC) was calculated to be 0.9897%. We decided to use 0.5 as a final threshold, meaning that subjects whose ratio of CAD to total predictions is greater than 0.5 will be characterized as having CAD; else, they will be characterized as not having CAD. With the use of this threshold, the total precision was 92.1%, recall was 98.3%, f-score was 95.1%, and accuracy was 93.8%. These results were evaluated after testing all the available recordings of each subject.

6. COMPARISON AND DISCUSSION

The problem of automatic detection of CAD has been an active field of research for a long time. Various approaches have been proposed in the past two decades. Subsequently, we will refer to some of the most popular ones, compare them to our work, and comment on their differences.

In 2013, Giri et al. [13] used 15-minute ECGs and an SVM model to achieve accuracy between 85% to 90%. In 2017, Dolatabi et al. [14] proposed a method using one-lead 24-hour ECGs and an SVM, which reached an accuracy of 99.2%. Also, Patidar et al. [15] used 15-min ECGs and an LS-SVM model to achieve 99.7% accuracy. Almost all of the previous models achieved better accuracy than our current work; however, the length of the ECG recording that these models need as input would be prohibiting for our present application. Since we try to create a tool that people can use in their everyday life, demanding long ECG recordings as input could not be practical.

As for the models that use shorter time intervals, there have been published numerous papers as well. Altan et al. [16] used 15-second ECGs and a deep belief network and achieved 98.05% accuracy with a 10-fold validation. Acharya et al. [17] had an accuracy of 94.95% using 5-second ECGs, a CNN, and 10-fold validation as well, while Lih et al. [18] used 2-second ECGs, a CNN-LSTM network, and 10-fold validation and had the accuracy of 98.51% for a 4-class problem. All of the methods mentioned above have higher accuracy and a smaller ECG duration as input. However, there is a not negligible probability that cross-validation results in an overestimation of the performance scores, since different subsets may include samples deriving from the same person, allowing the algorithms to perceive impeccable signs deriving from each heart's physiology. Thus, it can not be corroborated in real-life settings.

Finally, Tan et al. [19] used 5-second one-lead ECG recordings and a CNN and LS network. They also used a subject-specific validation which led to a 99.14% accuracy on non-CAD subjects and 76.56% accuracy on CAD subjects. However, the CAD dataset in that research consisted of only seven subjects.

7. CONCLUSION

The method presented in this paper proposes a non-invasive, effective, reliable, and lightweight algorithm for detecting CAD using 30-second ECG recordings. It delivers a subject-specific accuracy of 93.8% through a blind testing procedure. Also, it is designed to accept noisy ECGs derived from smartwatches, and, moreover, it is light enough to be deployed on a regular mobile phone. All of the factors mentioned above make it suitable for an everyday application that can contribute to the early detection of CAD.

8. REFERENCES

- [1] Andrew Cassar, David R. Holmes, Charanjit S. Rihal, and Bernard J. Gersh, "Chronic Coronary Artery Disease: Diagnosis and Management," *Mayo Clin Proc.*, vol. 84, pp. 1130–1146, 2009.
- [2] Ary L Goldberger, Luis AN Amaral, Leon Glass, Jeffrey M Hausdorff, Plamen Ch Ivanov, Roger G Mark, Joseph E Mietus, George B Moody, Chung-Kang Peng, and H Eugene Stanley, "Physiobank, physiobank, and physionet: components of a new research resource for complex physiologic signals," *circulation*, vol. 101, no. 23, pp. e215–e220, 2000.
- [3] Franc Jager, Alessandro Taddei, George B Moody, Michele Emdin, G Antolić, Roman Dorn, Ales Smrdel, Carlo Marchesi, and Roger G Mark, "Long-term st database: a reference for the development and evaluation of automated ischaemia detectors and for the study of the dynamics of myocardial ischaemia," *Medical and Biological Engineering and Computing*, vol. 41, no. 2, pp. 172–182, 2003.
- [4] Miguel A García-González, Ariadna Argelagós-Palau, Mireya Fernández-Chimeno, and Juan Ramos-Castro, "A comparison of heartbeat detectors for the seismocardiogram," in *Computing in Cardiology 2013*. IEEE, 2013, pp. 461–464.
- [5] Jiapu Pan and Willis J. Tompkins, "A real time QRS detection algorithm," *IEEE Transactions on Biomedical Engineering*, vol. BME-32, no. 3, pp. 230–236, 1985.
- [6] Hooman Sedghamiz, "Biosigkit: a matlab toolbox and interface for analysis of biosignals," *Journal of Open Source Software*, vol. 3, no. 30, pp. 671, 2018.
- [7] Fred Shaffer and J. P. Ginsberg, "An overview of heart rate variability metrics and norms," *Frontiers in Public Health*, vol. 5, no. 258, 2017.
- [8] Reem Satti, Noor-Ul-Hoda Abid, Matteo Bottaro, Michele De Rui, Maria Garrido, Mohammad R. Raoufy, Sara Montagnese, and Ali R. Mani, "The application of the extended poincaré plot in the analysis of physiological variabilities," *Frontiers Physiology*, vol. 10, pp. 116, 2019.
- [9] Steve M. Pincus, "Approximate entropy as a measure of system complexity," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 88, pp. 2297–2301, 1991.
- [10] Joshua S. Richman, Douglas E. Lake, and J. Randall Moorman, "Sample Entropy," *Methods in Enzymology*, vol. 384, pp. 172–184, 2004.
- [11] Lingjun Li, Shigang Liu, Yali Peng, and Zengguo Sun, "Overview of principal component analysis algorithm," *Optik*, vol. 127, pp. 3935–3944, 2016.
- [12] Jair Cervantes, Farid Garcia-Lamont, Lisbeth Rodríguez-Mazahua, and Asdrubal Lopez, "A comprehensive survey on support vector machine classification: Applications, challenges and trends," *Neurocomputing*, vol. 408, pp. 189–215, 2020.
- [13] Donna Giri, U. Rajendra Acharya, Roshan Joy Martis, S. Vinitha Sree, Teik-Cheng Lim, Thajudin Ahamed, and Jasjit S. Suri, "Automated diagnosis of coronary artery disease affected patients using LDA, PCA, ICA and discrete wavelet transform," *Knowledge-based systems*, vol. 37, pp. 274–282, 2013.
- [14] Azam Davari Dolatabadi, Siamak Esmael Zadeh Khadem, and Babak Mohammadzadeh Asl, "Noninvasive detection and classification of coronary artery occlusions using wavelet analysis of heart sounds with neural networks," *Computer methods and programs in biomedicine*, vol. 138, pp. 117–126, 2017.
- [15] Shivnarayan Patidar, Ram Bilas Pachori, and U. Rajendra Acharya, "Automated diagnosis of coronary artery disease using tunable-Q wavelet transform applied on heart rate signals," *Knowledge-based systems*, vol. 82, pp. 1–10, 2015.
- [16] Gokhan Altan, Novruz Allahverdi, and Yakup Kutlu, "Diagnosis of coronary artery disease using deep belief networks," *European Journal of Engineering and Natural Sciences*, vol. 2, pp. 29–36, 2017.
- [17] U. Rajendra Acharya, Hamido Fujita, Oh Shu Lih, Muhammad Adam, Jen Hong Tan, and Chua Kuang Chua, "Automated detection of coronary artery disease using different durations of ecg segments with convolutional neural network," *Knowledge-Based System*, vol. 132, pp. 62–71, 2017.
- [18] Oh Shu Lih, V. Jahmunah, Tan Ru San, Toshitaka Yamakawa Edward J. Ciaccio, Masayuki Tanabe, Makiko Kobayashi, Oliver Faust, and U. Rajendra Acharya, "Comprehensive electrocardiographic diagnosis based on deep learning," *Artificial Intelligence In Medicine*, vol. 103, pp. 1–8, 2020.
- [19] Jen Hong Tan, Yuki Hagiwara, Winnie Pang, Ivy Lim, Shu Lih Oh, Muhammad Adam, Ru San Tan, Ming Chen, and U. Rajendra Acharya, "Application of stacked convolutional and long short-term memory network for accurate identification of cad ecg signals," *Computers in Biology and Medicine*, vol. 94, pp. 19–26, 2018.