

Effective Learning and Filtering of Faulty Heart-Beats for Advanced ECG Arrhythmia Detection using MIT-BIH Database

Vasileios Tsoutsouras, Dimitra Azariadi, Sotirios Xydis, Dimitrios Soudris
Institute of Communication and Computer Systems, Greece
Contact info: {billtsou, dimitra, sxydis, dsoudris}@microlab.ntua.gr

ABSTRACT

Electrocardiogram (ECG) signal has been established as one of the most fundamental bio-signals for monitoring and assessing the health status of a person. ECG analysis flow relies on the detection of points of interest on the signal with the QRS complex, located around an R peak of the heart beat, being the most commonly used. Using the MIT-BIH arrhythmia database, we evaluate the accuracy of various R peak detectors, showing a large number, i.e. several thousands, of falsely detected peaks. Considering the medical significance of the ECG analysis, we propose a machine learning based classifier to be incorporated in the ECG analysis flow aiming at identifying and discarding heart beats based on erroneously detected R peaks. Using Support Vector Machines (SVMs) and extensive exploration, we deliver a tuned classifier that i) successfully filters up to 75% of the false beats, ii) while keeping the correct beats mis-classified as false lower than 0.01% and iii) the computational overhead of the classifier sufficiently low.¹

Keywords

ECG analysis, Support-Vector-Machine (SVM), Machine Learning, False Heart-Beat Filtering

1. INTRODUCTION

Electrocardiogram (ECG) signal processing forms a valuable analysis and diagnostic tool in modern medicine. More specifically ECG signals or derived biomarkers have been utilized as a means of quantifying not only the physiology of the heart but also the activity of the entire body of a subject. Subsequently, the field of ECG analysis is extensive and spawns amongst various research fields thus requiring multidisciplinary approach to investigate it.

In medical research activities, the utilization of real ECG signals is of vital importance. As a result many medical research institutes have created initiatives and published data bases of ECG signals to be used for research purposes. One of the most utilized is the MIT-BIH database [1] resulted from a collaboration of Beth Israel Deaconess Medical Center and MIT. It is composed of 48 fully annotated half-hour two-lead ECG signals. In MIT-BIH, the provided

¹This work was partially supported by the H2020 project "Aegle", <http://www.aegle-uhealth.eu> and DAAD-funded project "Teacher".

ECG signal data are annotated by medical experts and a rich set of software tools [2] is provided to enable the structuring of ECG processing pipelines. Previous research works [3–5] examine several classification schemes to improve the characterization of ECG signals. However, they are focused on classifying ECG signals as normal or abnormal, without giving special care in the recognition of faulty detected ECG signals (see Section 2), thus compromising the quality of arrhythmia detection.

In this paper, we identify this inefficiency regarding faulty heart-beat detection, and propose an advanced ECG analysis flow that effectively learns and filters false heart-beats. Through extensive analysis over the entire MIT-BIH annotated database, we show that faulty heart-beats are an inherent characteristic of the ECG detectors provided in physionet toolkit. We then propose a mitigation strategy based on the new ECG analysis flow that incorporates an extra Support-Vector-Machine (SVM) classifier that learns the features of faulty ECG heart-beats, thus enabling their filtering prior to the final normal/abnormal classification. In order to deliver an optimized and cost effective ECG processing pipeline for arrhythmia detection, we perform extensive exploration over the ECG extracted features². We show that the proposed solution achieves filtering of up to 75% of false beats with a minor overhead of 0.01% correct beats erroneously classified as false and all these with acceptable computational cost.

2. PROBLEM STATEMENT: FAULTY HEART BEAT DETECTION

In ECG signal analysis, the most critical task is the determination of the location of certain points of interest, e.g. QRS peaks and P and T waves [6]. QRS peaks are essential to further analysing the morphology of the ECG signal as well as dictating ECG heart beat segmentation. Fig. 1 illustrates the measured amplitude of two heart-beats, with labels on the location of the aforementioned points of interest in the first beat.

Extensive research activities have been performed on heart-beat detection algorithms [7]. The current research trend has shifted from detecting these points to using them as a basis to retrieve more complex information about a heart beat in an effort to assess its medical status. Deriving mathematical formulations to capture ECG's correlation has proved to be very complex and thus researchers have turned to the use of machine learning (ML) approaches to achieve this goal. ML approaches are focused on determining the best set of ECG signal features that will drive the training of an optimized heart-beat classifier [4, 5]. A subset of the MIT-BIH database is used by picking a number of indicative beats to use as training data set. The training set is formed utilizing the ECG annotations provided by the medical experts.

²Differing feature sets generate differing SVM configurations.

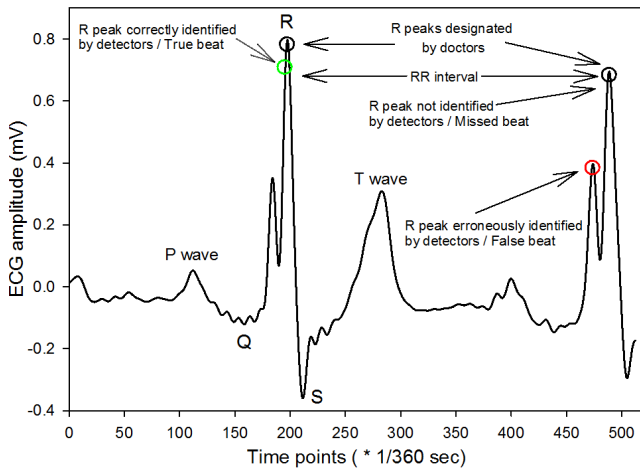


Figure 1: ECG signal: Correct vs. faulty QRS detection

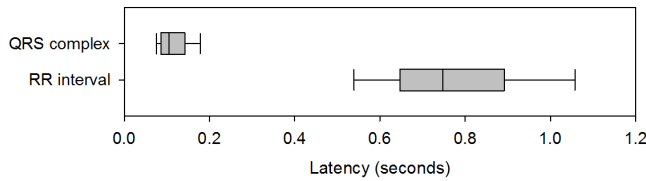


Figure 2: Latency values of heart beats

In reality, QRS detectors fail to correctly identify all points of interest [7]. An ECG signal in the form of a time series is provided as input to the detector, which generates a set of points in time corresponding to the R peaks. However comparing the detected points with the respective annotations, there is a mismatch not only in the number of identified R peaks but also in their exact time value. In Fig. 1, black circles indicate the R peaks defined by medical experts. The rest of the circles indicate points as they have been identified by an R peak detector. According to the proximity of these points compared to the annotated R peaks, the detected R peaks can be classified as (i) True, (ii) False or (iii) Missed. A True R peak is one close to an R peak annotation (green circle in Fig. 1). A False R peak is one which is far from the corresponding R peak annotation and is erroneously identified by the detector (red circle in Fig. 1). In case that the detector has failed to identify the second R peak of Fig. 1, it is considered to be a missed one.

In order to derive the class of a detected R peak, we use the distance from the actual R peak. Thus, there is the need to define a threshold T to characterize the heart-beat class. A statistical analysis of the ECG signals provided by the MIT-BIH database was conducted and focused on deriving a robust estimation about the time lapse between two consecutive heart beats and the duration of the QRS complex. The results regarding these durations are summarized in Fig. 2. The values are in seconds and they are measured using the entire MIT-BIH database and the corresponding annotations.

It is evident that the value which exhibits the most robust behaviour is the QRS complex duration and therefore its median value which is approximately 0.11 seconds will be used to determine the threshold T . With an ECG's sampling frequency $F_s = 360Hz$, QRS complex is approximately 40 data samples wide. We define the threshold T as half the duration of the QRS complex, ergo 20 samples.

Having defined the threshold T , a rule based procedure was real-

Table 1: Beat categories for different detectors

Detector	False beats	Missed beats	True beats
wqrs	3823	3571	100293
sqrs	3231	4312	99552
gqrs	3473	3419	100445

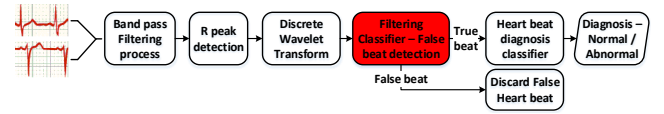


Figure 3: Proposed ECG analysis flow

ized to quantify the detected R peak by a detector as (i) True, (ii) False or (iii) Missed, using as reference input the respective annotated R peaks. The two discrete time series, i.e. \mathbf{D} :detected and \mathbf{A} :annotated R peaks, are traversed in parallel and an index in each one (i_D and i_A , respectively) indicates which R peaks are compared. If the detected and annotated R peaks under examination are at most T time units away, $|\mathbf{D}[i_D] - \mathbf{A}[i_A]| \leq T$, then the detected is considered True and both indexes are increased. If the detected R peak follows after more than T points, $\mathbf{D}[i_D] - \mathbf{A}[i_A] > T$, then the annotated R peak is considered a missed one and i_A is increased. Conversely, if the annotated R peak precedes for more than T points, $\mathbf{A}[i_A] - \mathbf{D}[i_D] > T$, the detected one is considered False and i_D is increased. This process takes place at the ECG analysis flow at *design time* and in **no case** are doctor annotated files utilized in ECG analysis flow at *run-time*.

We evaluate the accuracy of three R peak detection algorithms, i.e. (i) WQRS, (ii) SQRS and (iii) GQRS, found in MIT-BIH's physionet tool-suite. In Table 1, it is shown that the false and missed beats are significantly lower than true detected beats. However, in an effort to create a successful ECG analysis flow this erroneous behaviour should be minimized. In addition, in the evaluation of typical ECG pipelines, i.e. with classifiers trained to determine Normal or Abnormal beats using as reference the annotated R peaks, under false heart-beats testing scenarios, we reported about 86% of the false beats classified as abnormal, which severely affects the diagnostic abilities offered by the ECG analysis flow.

3. MACHINE LEARNING FOR FALSE BEAT FILTERING

In this section, we propose a machine-learning based classifier that extends the typical ECG analysis flow with the responsibility of distinguishing whether an R peak determined by one of the aforementioned detectors is True or False.

In a typical ECG analysis pipeline [3, 8], the two leads of the ECG signal are filtered by a band pass filter for noise removal. The filtered signals are directed to R peak detection phase which in turn produces a possible R peak. Using this R peak as a center point, a heart beat is derived and it is used as an input to a feature extraction mechanism in order to make its description more compact and accurate. As a part of this work we use Discrete Wavelet Transform (DWT) [9] as a feature extraction mechanism since it has been proven to produce very accurate results [4]. Finally, the feature vector of the heart beat is directed to a classifier which concludes about the nature of this heart beat i.e normal or abnormal.

Fig. 3 shows the proposed ECG pipeline. The proposed false heart-beat classifier (red box in Fig. 3) extends the ECG pipeline by detecting and discarding heart beats which are based on False R beat detection. After the feature vector of the heart beat has been derived, it is filtered through the classifier. This is an very important

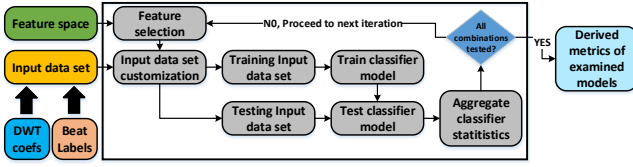


Figure 4: Design space exploration framework

aspect of the flow since it implies that it is based on a feature extraction stage which has to be computed for the diagnosis classification and it does not burden the flow with an extra feature extraction process.

In this work we focus on using Support Vector Machine (SVM) based classifier for the proposed false heart-beat filtering component [10], mainly due to its ability to support non-linear classification with efficient accuracy and computation cost. The main challenge of the target classification problem is that false beats are much less in number compared to true beats (Section 2). This creates a very imbalanced training data set and if no action is taken to take this into account during training phase then the classifier fails at acceptably classifying false beats. In other words, since the goal of the training phase is to adequately train the classifier according to the input dataset, if a classifier which can successfully identifies only true beats is produced, then the overall classification accuracy of the model is high since true beats vastly dominate the training data set. To alleviate this phenomenon we train our classifier using different penalty weights for the misclassification of false beats, while the default is for these penalty weights to be equal for each misclassification. This is a feature which is offered by SVMs and more precisely by their implementation of libSVM [11].

3.1 Design space exploration for SVM tuning

The design goal is to create an SVM based classifier which given a heart beat to be classified in a binary manner aims at:

- $Sensitivity = \frac{Num. of correctly classified Positive points}{Total number of Positive points}$ to be maximized i.e. the best possible recognition of True beats is achieved.
- $Specificity = \frac{Num. of correctly classified Negative points}{Total number of Negative points}$ to be maximized i.e. the best possible recognition of False beats is achieved.
- $Accuracy = \frac{Number of correctly classified points}{Total number of points}$ to be maximized.
- Minimizing the computational overhead inflicted by the extra SVM false beat classifier on the overall ECG analysis flow to sustain real-time operation.

To achieve these design goals, we performed a design space exploration on the feature vectors produced by the DWT feature extraction phase. The wavelet base for the DWT is Daubechies of order 2 (db2) [9] and we perform 4 levels of decomposition as proposed in [4]. The 4 levels produce 8 sets of coefficients each one for 4 levels of detailed and 4 levels of approximate coefficients. Since we utilize both leads of the ECG signal, i.e. two time series of ECG these 8 sets are multiplied by 2 thus resulting in 16 sets of DWT coefficients. As a feature vector we use combinations of these 16 sets.

The implemented exploration framework is illustrated in Fig. 4. To reduce exploration execution we input to it the data set accompanied by the extracted sets of DWT coefficients for each heart beat. DWT is calculated only once, thus saving a great amount of time.

Table 2: Comparison of different diagnosis classifiers

Classifier	Accuracy (%)	Sensitivity (%)	Specificity (%)
Utilized classifier	97.87	99.40	99.40
Classifier proposed in [4]	95.35	99.15	69.91

From that point on, there is an iteration over the combinations of coefficients that the designer would like to evaluate as a feature vector for the description of heart beat. In more detail we refer to combinations of coefficients as picking any of the 16 sets of coefficients and testing them as the selected features for the building of the classifier. Each combination results in a different classifier both in accuracy and number of support vectors. The product of number of support vectors and size of feature vector is a reliable metric for inducing the computational requirements of the produced classifier since it is the number of multiplications required for a new heart beat to be classified [10].

4. EXPERIMENTAL RESULTS: IMPACT OF DESIGN ALTERNATIVES

To effectively quantify the impact of the false beat filtering classifier there was the need to utilize a diagnosis classifier. We performed an extensive search over a variety of input feature vectors for both leads of the incoming ECG signal and the best results we acquired for a feature vector based on the coefficients of the 4th level of DWT decomposition using data from the entire MIT-BIH without choosing specific, representative beats as in [4]. We compared our classifier against a state-of-the-art similar classification model proposed in [4]. The results are summarized in Table 2 and show that the produced classifier is better in all metrics compared to the one presented in [4] which is justified on the argument that the latter had been trained and tested based on a subset of the MIT-BIH database.

Regarding the exploration conducted to determine the best classifier for the pre-diagnosis filtering phase, the first comment is that design alternatives for all combinations of the 16 sets of DWT coefficients create a design space large enough to render examination all alternatives a rather time consuming task. Fig. 5 presents a box plot of the values of computational requirements (Section 3.1) for combinations of 1 to 6 sets from the coefficients. The trend shows that increasing the numbers of combined sets, results to increase in necessary computations for a new heart beat to be classified. Having achieved classification accuracy of up to 99.4%, we did not examine a higher number of combined sets in order to keep these computations as low as possible, which is one design requirement of the filtering classifier.

Fig. 6 presents all the results regarding sensitivity, specificity and computational requirements of the aforementioned combinations of DWT coefficients. We focus on these metrics and not accuracy since the goal of the figure is to quantify the ability of each different classifier to meet the design goals presented in Section 3. We can see that in all cases sensitivity is above 99.80% or in other words the True beats misclassified as False are relatively low. Furthermore, specificity i.e. the False beats successfully filtered is above 72% and reaching up to 88%.

In an effort to determine the best of the design alternatives, we highlighted the region of Pareto optimal design points which maximize sensitivity and specificity while minimizing computational requirements of the trained classifier. The most medically reasonable choice is to decide the design alternative which maximizes sensitivity in order to reduce the True beats classified as False as much as possible. Inevitably, this impacts on the classifier's ability to filter false beats. Consequently if ones takes into account the tu-

Table 3: Baseline vs extended diagnosis flows

Analysis flow	Accuracy (%)	Sensitivity (%)	Specificity (%)	Number of False alarms
Baseline	97.87	99.40	99.40	2175
Extended	97.86	99.35	99.46	550

ple (sensitivity, specificity, computations), this results in (99.992%, 65.56%, 272900). However, we notice that a small reduction in sensitivity leads to a design alternative with (99.808%, 87.678%, 95410) which means that an increase 22% in specificity was achieved by a classifier which is 3 times less computationally demanding by sacrificing only 0.184% in sensitivity. In absolute numbers, this configuration has misclassified only 96 True beats out of testing data set of 50116.

Finally, Table 3 summarizes comparative results between the typical and proposed ECG analysis flows. Taking a closer inspection on the results, we can see that accuracy of the extended diagnosis flow is 0.01 less compared to the baseline diagnosis flow. To explain that, we will focus on the other metrics. Regarding sensitivity, the extended diagnosis flow lacks 0.05% compared to the baseline one. This is because, there is a small number of True beats who are erroneously filtered as False and therefore sensitivity of the complete flow decreases. On the contrary, due to the fact that the filtering classifier discards a large number of False beats, the specificity of the extended diagnosis flow is increased compared to the baseline one. We do not observe a steep rise due to the fact that most of the discarded False beats are successfully classified as Abnormal in the baseline flow. Apart from the provided metrics regarding the behaviour of the two diagnosis flows, the greatest advantage of the filtering classifier is that, supposing that the diagnosis flow raises an alarm whenever a heart beat is considered to be Abnormal, there is about 75% reduction in these alarms owned to False beats, i.e. False alarms. The trade-off is that the filtering classifier imposes an average 27% in the required execution time of the complete diagnosis flow.

5. CONCLUSION

In this paper we analysed the ability of R peak detectors to successfully identify R peaks in an ECG signal. Based on a significant number of erroneously identified peaks we proposed an Support Vector Machine (SVM) based classifier to be incorporated in the ECG analysis flow in order to identify and discard these false R peaks. Design space exploration conducted on different feature vectors describing a heart beat, resulted in a variety of classifiers with different accuracy and computational characteristics. The pro-

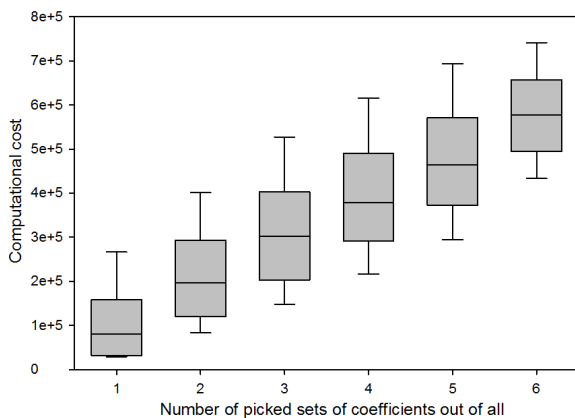


Figure 5: Comp. effort for the classification of a heart beat

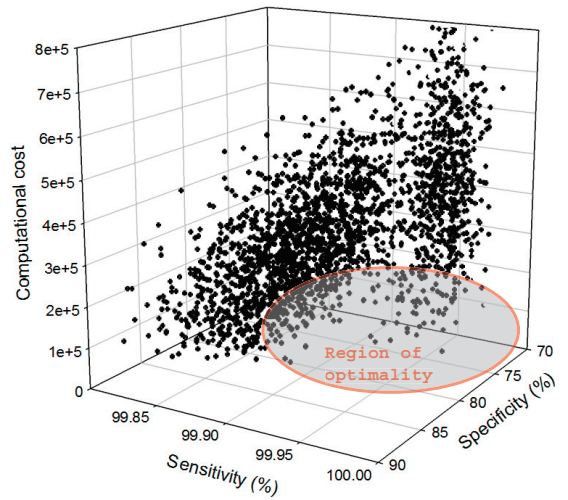


Figure 6: Design space of pre-diagnosis filtering classifier

posed false beat filtering classifier is able to filter up to 75% of false R peaks while keeping the number of True beats misclassified as False less than 0.01%. The required high accuracy creates a complex classifier but exploration enabled us to keep its computational requirements low resulting of 27% in the execution time of the ECG analysis flow.

6. REFERENCES

- [1] George B Moody and Roger G Mark. The impact of the mit-bih arrhythmia database. *Engineering in Medicine and Biology Magazine, IEEE*, 20(3):45–50, 2001.
- [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, physiotoolkit, and physionet components of a new research resource for complex physiologic signals. *Circulation*, 101(23):e215–e220, 2000.
- [3] Philip De Chazal, Maria O Dwyer, and Richard B Reilly. Automatic classification of heartbeats using ecg morphology and heartbeat interval features. *Biomedical Engineering, IEEE Transactions on*, 51(7):1196–1206, 2004.
- [4] Elif Derya Übeyli. Ecg beats classification using multiclass support vector machines with error correcting output codes. *Digital Signal Processing*, 17(3):675–684, 2007.
- [5] Mohammad R Homaeinezhad, SA Atyabi, E Tavakkoli, Hamid Najjaran Toosi, Ali Ghaffari, and Reza Ebrahimpour. Ecg arrhythmia recognition via a neuro-svm-knn hybrid classifier with virtual qrs image-based geometrical features. *Expert Systems with Applications*, 39(2):2047–2058, 2012.
- [6] Rangaraja M Rangayyan and Narendra P Reddy. Biomedical signal analysis: a case-study approach. *Annals of Biomedical Engineering*, 30(7):983–983, 2002.
- [7] Mohamed Elgendi, Björn Eskofier, Socrates Dokos, and Derek Abbott. Revisiting qrs detection methodologies for portable, wearable, battery-operated, and wireless ecg systems. *PloS one*, 9(1), 2014.
- [8] Mohammed Shoaib, Niraj K Jha, and Naveen Verma. Algorithm-driven architectural design space exploration of domain-specific medical-sensor processors. *Very Large Scale Integration (VLSI) Systems, IEEE Transactions on*, 21(10):1849–1862, 2013.
- [9] Ingrid Daubechies. The wavelet transform, time-frequency localization and signal analysis. *Information Theory, IEEE Transactions on*, 36(5):961–1005, 1990.
- [10] Corinna Cortes and Vladimir Vapnik. Support-vector networks. *Machine learning*, 20(3):273–297, 1995.
- [11] Chih-Chung Chang and Chih-Jen Lin. Libsvm: A library for support vector machines. *ACM Transactions on Intelligent Systems and Technology (TIST)*, 2(3):27, 2011.