High Frequency Noise Detection and Handling in ECG Signals

Kjell Le,
Trygve Eftestøl,
and Kjersti Engan
Department of Electrical Engineering
and Computer Science
University of Stavanger
Email: kjell.le@uis.no

Stein Ørn*†
*Department of Cardiology
Stavanger University Hospital
†Department of Electrical Engineering
and Computer Science
University of Stavanger

Øyunn Kleiven
Department of Cardiology
Stavanger University Hospital

***Abstract*—After acquisition of new clinical electrocardiogram (ECG) signals the first step is often to preprocess and have a signal quality assessment to uncover noise. There might be restriction on the signal length and other issue that impose limitation where it is not possible to discard the whole signal if noise is present. Thus there is a great need to retain as much noise free regions as possible.**

A noise detection method is evaluated on a manually annotated subset (2146 leads) of a data base of 12-lead ECG recordings from 1006 bicycle race participants. The aim is to apply the noise detector on the unlabelled part of the data set before any further analysis is conducted. The proposed noise detector can be divided into 3 parts: 1) Select a high frequency signal as a base signal. 2) Apply a thresholding strategy on the base signal. 3) Use a noise detection strategy.

In this work receiver operating characteristic (ROC) curve and area under the curve (AUC) will be used to assess a high frequency noise detector designed for ECG signals. Even though ROC analysis is widely used to assess prediction models, it has its own limitation. However, it is a good starting point to assess discriminatory ability.

To generate the ROC curve the performance evaluation is based on sample-level. That is, each sample has a label whether it is noise or not. The threshold strategy and the chosen threshold will be the varying factor to generate ROC curves. The best model has an average AUC of 0.862, which shows a good detector to discriminate noise. This threshold strategy will be used for noise detection on the unlabelled part of the data set.

I. Introduction

A challenge with the handling of clinical electrocardiogram (ECG) signals is to identify noise regions that are present in parts of the signals. This is especially challenging with data sets of several thousand ECGs, which is the case in this work. Denoising, i.e. removing the noise that can be removed, and identify regions with noise that are impossible to remove must be done before any clinical interpretation can be drawn from the data material. Otherwise the noise may literally interfere with the veracity of the interpretation.

The importance of the resolution of the noise detector depends on which stage, from the acquirement to assessing the ECG signal, the noise identification is done. During acquisition, the signal quality assessment could use relative large segments of for example $10\,\mathrm{s}$ to evaluate the quality of the signal [1]–[4], i.e. the detection is done on segment-level. In this stage there is the luxury of remeasuring possibility if the signal quality is unacceptable. However, often another assessment is done after the entire data material is collected, and it is not possible to redo the ECG recording of a specific person. In this case, the noise detection should be much more localized, so to be sure not to discard signals that can be useful in clinical interpretations. A possible approach will be to develop and evaluate a noise detector on a manually annotated subset (2146 segments) of the complete data set. Furthermore, the best performing noise detector can be applied to the unlabelled part of the data set before any further analysis is conducted.

At this point it is desirable to remove noise to enhance the signal quality. Though, since there is always an overlapping in frequency bands between important information and noise, some information is expected to be lost.

Recommendation on preprocessing methods to remove baseline wander, powerline interference and high frequency noise can be found in [5].

One method to identify severe noise in the high frequency area is to extract a high frequency signal from the original signal and identify abnormality in the extracted signal. Abnormality could be the presence of large energy in the high frequency specter for a section of the signal compared to other sections. Various methods to extract the high frequency signal, not limited to these, are a highpass filter, stationary wavelet transform (SWT) [6], also known as *algorithme à trous*, and empirical mode decomposition (EMD) [2], [7]. The differences between the high frequency signals produced are relatively small when juxtaposed as shown in figure 2.

In the literature noise detection in ECG is usually done on a segment-level, where segments typically have a duration of $10\mathrm{s}$ [1]–[4]. For the purpose of this work the noise localization is defined at a sample-level, where each sample will be labelled as noise/not noise, permitting better exploitation of the data set.

The proposed noise detection is a 3 parts dissection of a method proposed by Satija et.al. [8]. That is: 1) Extract a high frequency signal, base signal, to use as an input. 2) A threshold

ISBN 978-90-827970-1-5

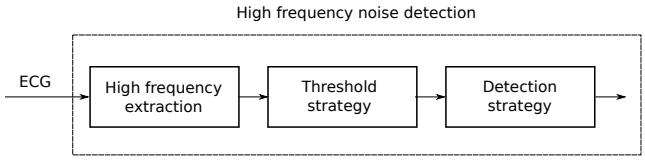

Fig. 1. Block diagram for noise detection. ECG is fed into a high frequency extractor to extract a base signal. Then a threshold strategy is applied and in the end a detection strategy is used. Noise regions are the output.

strategy on the base signal. 3) Using a detection strategy on the thresholded signal. The similarity of both methods is the first step, i.e extract a base signal. In this case the authors use a composition of the 3 first intrinsic mode functions (IMFs) of a complete ensemble empirical mode decomposition (EMD) [9] to detect high frequency noise. In the two other steps the authors have intertwined the threshold- and detection strategy. With the unraveling of the steps it should be clearer what is happening under the hood.

The proposed method will be assessed with receiver operating characteristic (ROC) and area under the curve (AUC) parameter [10]. The performance evaluation used to generate ROC curve is based on sample-level and not segment-level.

II. Dataset

The data set used is from The North Sea Race Endurance Exercise Study (NEEDED), ClinicalTrials.gov identifier: NCT02166216. It contains data measured from 1006 participants in a bicycle race from Egersund to Sandnes in Norway. The length of the race is $91\,\mathrm{km}$, and is a unique study with the largest number of collected ECG samples from assumed healthy individuals recorded.

For each participant in the study 3 measurements of 12-lead ECG were taken, 1 day before the race, 3 hours after reaching the finish line, and 24 hours after the race. The duration of ECG is $10\,\mathrm{s}$ for each episode. The sampling frequency is $600\,\mathrm{Hz}$. And the conversion factor from raw data to mV is $3.75/1000\,\mathrm{mV}$.

Because there exist only one recording of each day, and the main motivation of the NEEDED study is to study the ECGs across the days to discover potentially changes in the morphology, it is crucial to not reject the entire $10\,\mathrm{s}$ segment when only some regions contain severe noise.

This work uses a subset of the data, and there are 3 different noise classes depending on the severity of the noise. The annotations of 2146 ECG leads were done by two clinicians. Before a final consensus each clinician worked separately.

III. Proposed Method

In this section the main step of the proposed method will be explained. Figure 1 shows a block diagram of the proposed method. The main difference between Satija's method [8], [11] and the proposed method is the unraveling of the intertwined of threshold- and detection strategy in Satija's.

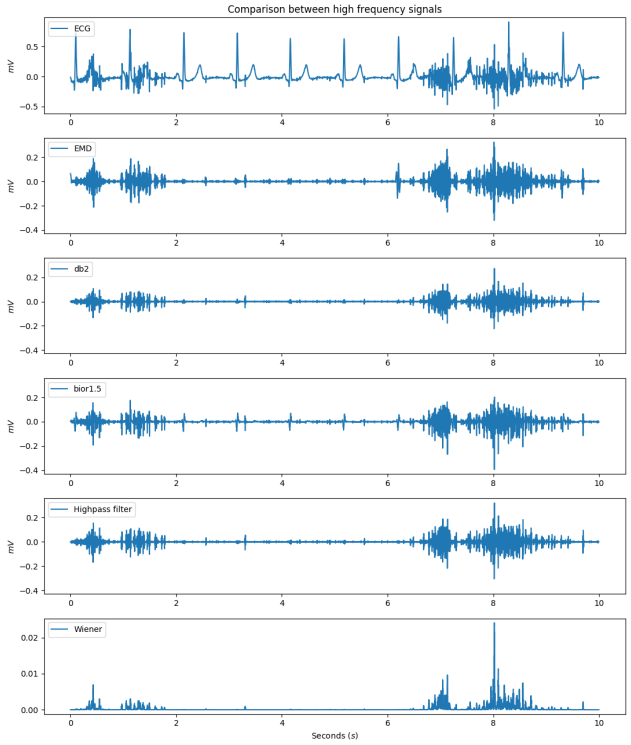

Fig. 2. Comparison between high frequency signal. From top to bottom shows the source ECG signal, first intrinsic mode function, level 1 detail coefficient of SWT with "db2" wavelet, level 1 detail coefficient of SWT with "bior1.5" wavelet, an IIR order 2 butterworth highpass filter, $f_c = 75\,\mathrm{Hz}$, and finally Wiener (variance) base signal.

A. Base Signal

Extracting a high frequency base signal is the first step in the noise detection algorithm.

The main job of the base signal is to suppress characteristic waves, such as QRS complexes, P- and T waves, in ECG and enhancing noise. Ideally this would be a highpass signal. However, since QRS complexes also contain high frequency it is not possible to fully suppress QRS complexes.

The different base signals used in this work are: 1) first intrinsic mode from EMD, 2) first level detail coefficient of SWT, 3) a highpass signal from a highpass filter and 4) a base signal derived from a Wiener denoising scheme which will be described in the next subsection. The base signals are shown in figure 2.

B. Threshold Strategy

As can be seen from figure 2, the base signals contain more energy, and have higher amplitudes in noisy parts of the ECG. Thus a thresholding strategy on the base signal can be used to discriminate the severe noise areas. The choice of the appropriate threshold can be viewed as a choice of signal to noise (SNR) ratio considered to be acceptable.

For a ROC analysis, an exact threshold is not important. What is important is what type of threshold is used. A strategy is to use a fixed threshold, t_{fix} based on physical properties [8], [11]. Thus to generate the ROC curve an additive offset

ISBN 978-90-827970-1-5 © EURASIP 2018

could be used. Such that the threshold, t_{th}, used in the method becomes $t_{th} = t_{fix} + \Delta$ where Δ is the offset.

Though, one could also use an adaptive threshold based on the signal properties. One such adaptive approach based on wavelet shrinkage (denoising method) is the universal threshold [12] which is defined as

$$t_{uni} = \sigma_0 \sqrt{2 \ln(N)} \tag{1}$$

where N is the length of segment and σ_0 is the standard deviation of the (Gaussian zero mean) noise. σ_0 is often unknown, but can be estimated:

$$\hat{\sigma}_0 \approx \frac{\text{median}(|x(n)|)}{\sqrt{2}\text{erf}^{-1}(\frac{1}{2})} \tag{2}$$

where erf is the Gauss error function and $x(n)$ is the detail coefficients. However, $x(n)$ can also be any highpass signal.

To generate the ROC curve with this threshold a multiplicative offset could be used. This is because $\sqrt{2 \ln(N)}$ is basically a multiplicative factor. Thus the resultant threshold becomes $t_{th} = t_{uni}\Delta$.

With the two thresholds described above a hard thresholding rule [12] is applied on the base signal $x(n)$ such that the new signal $x^*(n)$ is defined as

$$x^*(n) = \begin{cases} 0, & \text{if } |x(n)| < \text{threshold} \\ x(n), & \text{else} \end{cases} \tag{3}$$

The idea with wavelet shrinkage is to filter out details which has low amplitude. Details with high amplitude are untouched. Therefore these are the high frequency which should be detected. Hench everything that is not in the signal $x^*(n)$ is declared as noise.

Since wavelet shrinkage is a denoising method, another view on the matter is to determine how much denoising should be allowed. This can be done by defining a function, $w(n)$, which controls how much the detailed parts should be attenuated,

$$x^*(n) = x(n)w(n) \tag{4}$$

Thereafter a threshold is applied on $w(n)$ instead of $x(n)$. $w(n)$ can be regarded as the new base signal. $w(n)$ can be defined in different ways. A possible strategy, applied in this work is to calculate the variance of a sliding window. This is because high variance is related to noise. This approach will be similar to Wiener denoising method [13]–[16].

Let $\sigma_L^2(n)$ be the (sample) variance of a window with length L at sample n:

$$\sigma_L^2(n) = \frac{1}{L} \sum_{k=-(L-1)/2}^{(L-1)/2} (x(n-k) - m_L)^2 \tag{5}$$

m_L is the sample mean of the window. Where the window length is odd. Defining now $\sigma^2(n)$ as the minimum variance of $\sigma_L^2(n)$ with different $L = 3, 5, 7, 9$, $w(n)$ can be found as

$$w(n) = \frac{\sigma^2(n)}{\sigma^2(n) + \sigma_0^2} \tag{6}$$

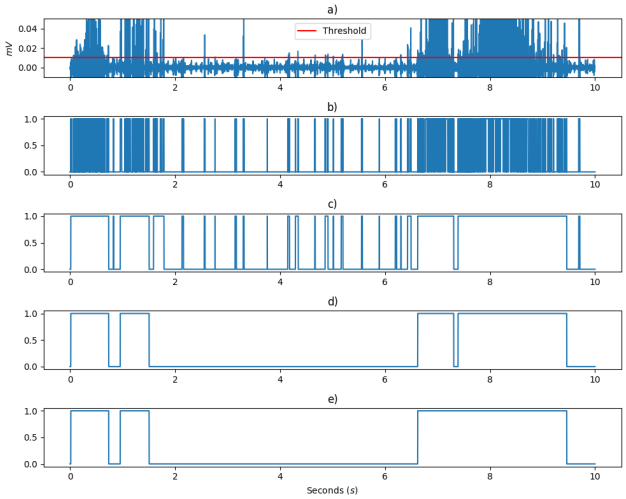

Fig. 3. Visualization of noise detection. a) A zoom base signal (Highpass filter). b) Mask of noise region after thresholding (Fixed thresholding). c) Merge adjacent noise region in a neighborhood $\epsilon_{th,1}$. d) Remove noise regions where the duration is less than d_{th}. e) Merge adjacent noise region in a neighborhood $\epsilon_{th,2}$.

where σ_0^2 is the variance of (unknown) noise [14]–[16]. σ_0^2 can also be viewed as an acceptance noise variance factor.

A fixed threshold or the universal threshold with the hard thresholding rule could now be applied to the base signal $w(n)$.

C. Detection Strategy

The detection strategy is defined as the steps which describe how to detect noise from the thresholded base signal $x^*(n)$. A continuous non-zero amplitude of $x^*(n)$ is defined to be noise.

The proposed detection strategy is to 1) Merge adjacent region in a neighborhood of $\epsilon_{th,1}$. 2) Remove regions where the duration is less than d_{th}. 3) Merge adjacent region in a neighborhood of $\epsilon_{th,2}$. Step 2 reduces the chance of detecting QRS complexes as noise. A visualization of the intermediate results of the noise detection strategy is shown in figure 3.

Equivalent rules to Satija's method of thresholding [8], [11] is to merge adjacent region in a neighborhood of $50\,\text{ms}$ and extends regions with $50\,\text{ms}$ each on both sides, then remove regions where the duration is less than $350\,\text{ms}$.

D. Receiver Operating Characteristic (ROC)

For ROC curve generating an offset is added or multiplied to the threshold to calculate sensitivity/true positive rate (TPR) and specificity/true negative rate (TNR). The ROC curve is the plot between 1 - TNR vs. TPR. A good ROC curve is a curve closest to the point $(0, 1)$, the upper left corner [17]. A bad ROC curve is a curve that is close to the line $y = x$ (line of equality) or beyond.

IV. Experiment and Result

In the experiment the EMD base signal is based on complete ensemble EMD [9] with 10 ensembles, Gaussian noise standard deviation is 0.001, and the stopping criterion uses

ISBN 978-90-827970-1-5

TABLE I
AREA UNDER THE CURVE, GENERATED FROM FIXED THRESHOLD WITH ADDITIVE OFFSET.

subset	0	1	2	3	4	5
db2	0.821	0.878	0.835	0.851	0.856	0.850
bior1.5	0.821	0.904	0.856	0.878	0.862	0.851
HP	0.820	0.875	0.832	0.852	0.850	0.850
EMD	0.826	0.886	0.835	0.856	0.862	0.862
Wiener	0.822	0.870	0.831	0.848	0.856	0.848

Huang's standard deviation test [7] where the standard deviation is 0.001. Only the first IMF is used. Two SWT base signals will be used. The wavelets used in decomposition are "db2" and "bior1.5". For the highpass filter the cutoff frequency is $f_c = 75\,\mathrm{Hz}$. In the Wiener base signal an acceptance noise variance is set to $\sigma_0^2 = 1$. Detection strategy parameters used are $\epsilon_{th,1} = 75\,\mathrm{ms}$, $\epsilon_{th,2} = 125\,\mathrm{ms}$, and $d_{th} = 350\,\mathrm{ms}$.

To generate ROC curves a subset of 2146 annotated ECG segments are used. Further on this subset is divided into 6 disjointed (almost equal sized) subsets. The 3 noise classes are merged into one class. The elements in the subsets are randomized without balancing with respect to the 3 classes nor the merged class. The subsets are not restricted to mutally exclusive contain data from one person within a day or for all 3 days, i.e. different subsets may contain different leads within a day or for all 3 days from the same person.

Area under the curve is calculated using the composite trapezoidal rule. Figure 4 and 5 shows average ROC curves with fixed and universal threshold respectively. Table I and II show the area under curve for each subset.

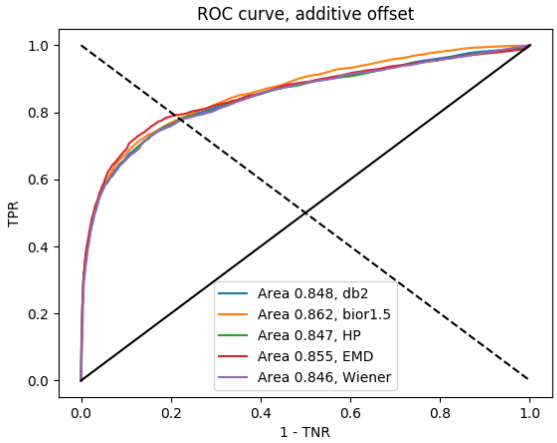

Fig. 4. Average ROC curve, generated from fixed threshold with additive offset.

V. DISCUSSION

From the fixed threshold with additive offset, the data subset 1 has the highest AUC, as shown in table I. The highest achievable AUC of 0.904 is attained with the "bior1.5" base signal. Subset 0 has lowest AUC of 0.820 with the highpass base signal. Table I shows that the variation between the AUC within a subset between the base signals is minimal. The variation between AUC, between the subsets are some what consistent.

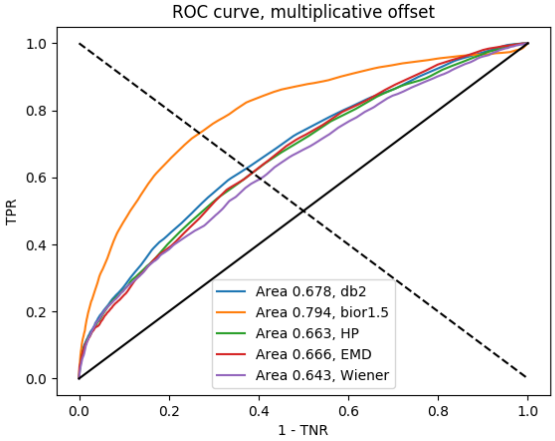

Fig. 5. Average ROC curve, generated from universal threshold with multiplicative offset.

TABLE II
AREA UNDER THE CURVE, GENERATED FROM UNIVERSAL THRESHOLD WITH MULTIPLICATIVE OFFSET.

subset	0	1	2	3	4	5
db2	0.668	0.643	0.603	0.648	0.674	0.834
bior1.5	0.797	0.794	0.736	0.787	0.779	0.870
HP	0.652	0.604	0.589	0.634	0.670	0.829
EMD	0.633	0.601	0.612	0.634	0.681	0.837
Wiener	0.615	0.603	0.570	0.600	0.642	0.829

For the universal threshold with multiplicative offset as shown in table II, the subset 5 has the highest AUC of 0.870. This is attained with the base signal "bior1.5". The lowest AUC can be found in subset 2 with a value of 0.589, from the highpass base signal. Variation of AUC between the base signals within a subset appears to be minimal, with the exception of "bior1.5" base signal. This base signal shows a much greater AUC as reflected in figure 5 compared to the other base signal. The variation between AUC, between the subsets are not consistent. Subset 5 seems to have much greater AUC than the rest.

Juxtaposing the tables and figures of the result, it is easily to identify that the fixed threshold is a viable choice. An AUC between 0.8 and 0.89 [17] shows a good model to detect noise. The fixed threshold is better than the universal threshold to discriminate noise.

For threshold selection the optimum point is the point where the minimum distance line (dashed line in figure 4) intersects with the ROC curve [17], [18]. From figure 4 "EMD" base signal has the best optimum point as it lies closer to the point $(0, 1)$, but the AUC of "bior1.5" is greater than "EMD". This is a limitation of ROC curves [18], which makes picking between "bior1.5" and "EMD" not as obvious. In terms of computational speed "bior1.5" is the preferred base signal.

ISBN 978-90-827970-1-5 © EURASIP 2018

One important result is the similarity of the ROC curves with the fixed threshold. There is a best base signal in terms of average AUC, however for a different criteria including computational complexity and simplicity in implementation there are other viable choices.

Balancing and various restriction imposing on the subsets do not change the average- ROC curves and AUC significantly. Nevertheless to do a proper balancing of the classes is difficult in a sample-level evaluation. This is because an ECG segment may contains different noise classes whose durations vary. Restricting the subsets to mutally exclusive contain data from one person within a day or for all 3 days have been tested and the results do not change, i.e. the variation between AUC, between the subsets are some what consistent.

VI. Further Works

The main limitation of the proposed noise detection is that it cannot detect noise with shorter duration than the threshold d_{th}. To tackle this deficiency, new methods need to be developed. Until now the focus has been noise detection. However, one is more interested in which (usable) region is not subjected by noise, or a reconstructed region, such that it is possible to extract features. To reconstruct a region, a leader-follower clustering algorithm [19] could be used to extract the dominant QRS complex. However, a signal duration of 10s is a limitation. If there exists noise already covering most of the duration of the signal it is not always possible to create a dominant cluster. Even if there exist a dominant cluster, the QRS-morphology might not be the desired one.

VII. Conclusion

The fixed threshold (with additive offset) model gives an overall much better AUC, suggesting it can discriminate noise better than the universal threshold (with multiplicative offset). The average ROC curves between the base signals are quite similar. However, the best base signal is "bior1.5" with an averaged AUC of 0.862. The "EMD" base signal has a better optimal point. Though, overall with computational speed taken into consideration "bior1.5" is the preferred choice. Though, other base signals are also viable when other criteria is used such as simplicity in implementation and computational complexity. The "bior1.5" base signal with the fixed offset will be used on the whole data set for further works.

References

[1] L. Johannesen and L. Galeotti, "Automatic ECG quality scoring methodology: mimicking human annotators," *Physiological Measurement*, vol. 33, no. 9, p. 1479, 2012. [Online]. Available: http://stacks.iop.org/0967-3334/33/i=9/a=1479

[3] D. P. Tobón V., T. H. Falk, and M. Maier, "Ms-qi: A modulation spectrum-based ECG quality index for telehealth applications," *IEEE Transactions on Biomedical Engineering*, vol. 63, no. 8, pp. 1613–1622, Aug 2016.

[2] J. Lee, D. D. McManus *et al.*, "Automatic motion and noise artifact detection in holter ECG data using empirical mode decomposition and statistical approaches," *IEEE Transactions on Biomedical Engineering*, vol. 59, no. 6, pp. 1499–1506, June 2012.

[4] I. Silva, G. B. Moody, and L. Celi, "Improving the quality of ECGs collected using mobile phones: The physionet/computing in cardiology challenge 2011," in *2011 Computing in Cardiology*, Sept 2011, pp. 273–276.

[5] P. Kligfield, L. S. Gettes *et al.*, "Recommendations for the standardization and interpretation of the electrocardiogram," *Circulation*, vol. 115, no. 10, pp. 1306–1324, 2007. [Online]. Available: http://circ.ahajournals.org/content/115/10/1306

[6] A. Cohen and J. Kovacevic, "Wavelets: the mathematical background," *Proceedings of the IEEE*, vol. 84, no. 4, pp. 514–522, Apr 1996.

[7] N. E. Huang, Z. Shen *et al.*, "The empirical mode decomposition and the hilbert spectrum for nonlinear and non-stationary time series analysis," *Proceedings of the Royal Society of London A: Mathematical, Physical and Engineering Sciences*, vol. 454, no. 1971, pp. 903–995, 1998. [Online]. Available: http://rspa.royalsocietypublishing.org/content/454/1971/903

[8] U. Satija, B. Ramkumar, and M. S. Manikandan, "Automated ECG noise detection and classification system for unsupervised healthcare monitoring," *IEEE Journal of Biomedical and Health Informatics*, 2017.

[9] M. E. Torres, M. A. Colominas *et al.*, "A complete ensemble empirical mode decomposition with adaptive noise," in *Acoustics, speech and signal processing (ICASSP), 2011 IEEE international conference on*. IEEE, 2011, pp. 4144–4147.

[10] J. A. Hanley and B. J. McNeil, "The meaning and use of the area under a receiver operating characteristic (roc) curve." *Radiology*, vol. 143, no. 1, pp. 29–36, 1982.

[11] U. Satija, B. Ramkumar, and M. S. Manikandan, "Low-complexity detection and classification of ECG noises for automated ECG analysis system," in *2016 International Conference on Signal Processing and Communications (SPCOM)*, June 2016, pp. 1–5.

[12] C. Cai and P. d. B. Harrington, "Different discrete wavelet transforms applied to denoising analytical data," *Journal of Chemical Information and Computer Sciences*, vol. 38, no. 6, pp. 1161–1170, 1998. [Online]. Available: http://dx.doi.org/10.1021/ci980210j

[13] H. A. Kestler, M. Haschka *et al.*, "De-noising of high-resolution ECG signals by combining the discrete wavelet transform with the wiener filter," in *Computers in Cardiology 1998. Vol. 25 (Cat. No.98CH36292)*, Sep 1998, pp. 233–236.

[14] M. K. Mihcak, I. Kozintsev, and K. Ramchandran, "Spatially adaptive statistical modeling of wavelet image coefficients and its application to denoising," in *1999 IEEE International Conference on Acoustics, Speech, and Signal Processing. Proceedings. ICASSP99 (Cat. No.99CH36258)*, vol. 6, Mar 1999, pp. 3253–3256 vol.6.

[15] M. K. Mihcak, I. Kozintsev *et al.*, "Low-complexity image denoising based on statistical modeling of wavelet coefficients," *IEEE Signal Processing Letters*, vol. 6, no. 12, pp. 300–303, Dec 1999.

[16] J. Fridrich, "Digital image forensics," *IEEE Signal Processing Magazine*, vol. 26, no. 2, pp. 26–37, March 2009.

[17] J. V. Carter, J. Pan *et al.*, "Roc-ing along: Evaluation and interpretation of receiver operating characteristic curves," *Surgery*, vol. 159, no. 6, pp. 1638 – 1645, 2016. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0039606016000660

[18] J. M. Lobo, A. Jiménez-Valverde, and R. Real, "Auc: a misleading measure of the performance of predictive distribution models," *Global ecology and Biogeography*, vol. 17, no. 2, pp. 145–151, 2008.

[19] L. Sörnmo and P. Laguna, *Bioelectrical signal processing in cardiac and neurological applications*. Academic Press, 2005, vol. 8.

ISBN 978-90-827970-1-5 © EURASIP 2018