

# A review of methods for the signal quality assessment to improve reliability of heart rate and blood pressures derived parameters

Nicolò Gambarotta<sup>1</sup> · Federico Aletti<sup>1</sup> · Giuseppe Baselli<sup>1</sup> · Manuela Ferrario<sup>1</sup>

Received: 16 December 2014 / Accepted: 29 January 2016 / Published online: 23 February 2016  
© International Federation for Medical and Biological Engineering 2016

**Abstract** The assessment of signal quality has been a research topic since the late 1970s, as it is mainly related to the problem of false alarms in bedside monitors in the intensive care unit (ICU), the incidence of which can be as high as 90 %, leading to alarm fatigue and a drop in the overall level of nurses and clinicians attention. The development of efficient algorithms for the quality control of long diagnostic electrocardiographic (ECG) recordings, both single- and multi-lead, and of the arterial blood pressure (ABP) signal is therefore essential for the enhancement of care quality. The ECG signal is often corrupted by noise, which can be within the frequency band of interest and can manifest similar morphologies as the ECG itself. Similarly to ECG, also the ABP signal is often corrupted by non-Gaussian, nonlinear and non-stationary noise and artifacts, especially in ICU recordings. Moreover, the reliability of several important parameters derived from ABP such as systolic blood pressure or pulse pressure is strongly affected by the quality of the ABP waveform. In this work, several up-to-date algorithms for the quality scoring of a single- or multi-lead ECG recording, based on time-domain approaches, frequency-domain approaches or a combination of the two will be reviewed, as well as methods for the quality assessment of ABP. Additionally, algorithms exploiting the relationship between ECG and pulsatile signals, such as ABP and photoplethysmographic recordings, for the reduction in the false alarm rate will be presented. Finally, some considerations will be drawn taking into

account the large heterogeneity of clinical settings, applications and goals that the reviewed algorithms have to deal with.

**Keywords** False alarm detection · Signal quality · ECG · Arterial blood pressure · Heart rate

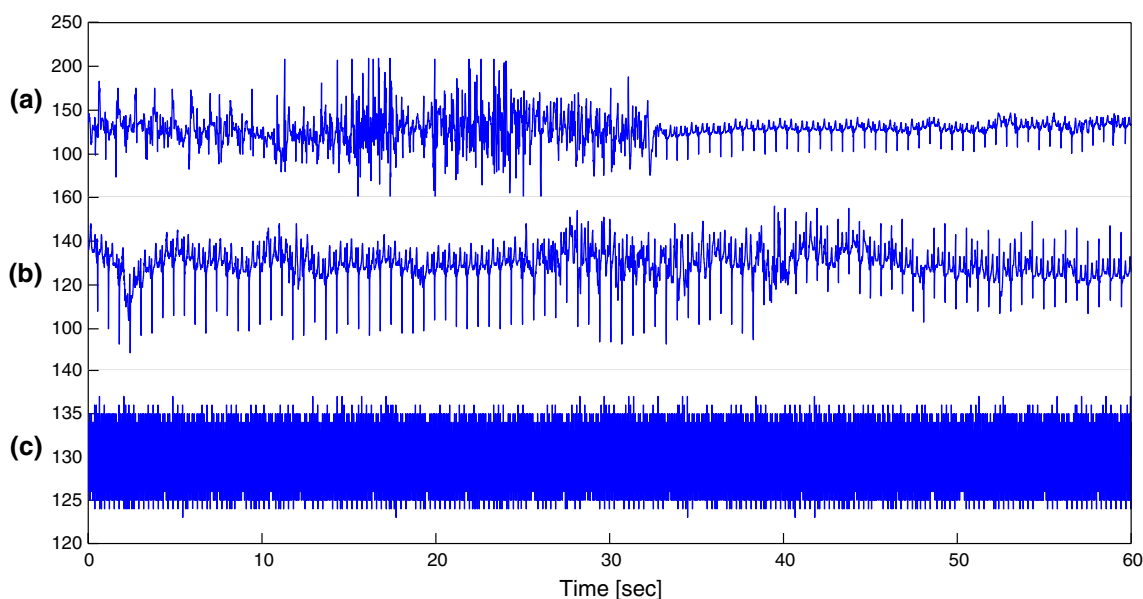
## 1 Introduction

The assessment of signal quality has been a research topic [14, 16, 17] related to the problem of false alarms in bedside monitors in the intensive care unit (ICU) for many years. Currently, bedside monitors generate a great number of potential alarms, mainly based on parameters derived from electrocardiogram (ECG), arterial blood pressure (ABP), respiratory and the photoplethysmographic (PPG) signal. The incidence of false alarms in the ICU can be as high as 90 %, as reported by several studies [1, 27, 41].

The overload of alarms which the ICU staff is exposed to can lead to alarm fatigue and to a possible desensitization to potentially life-threatening alerts, due to a drop in the overall level of attention [7]. The improvement in the quality of care also relies, for example, on the suppression of arrhythmia-dependent false alarms. A robust beat detection is therefore of paramount importance, and the reliable detection of heart rate is extremely dependent on the quality of the recording. The estimate of the heart rate (HR) can be performed not only on the ECG recording, but also on pulsatile signals such as the ABP or PPG. Reliable information about the heart rhythm can be then obtained from more than one signal. Therefore, the availability of efficient algorithms for the signal quality assessment of different signal channels represents an important progress.

✉ Manuela Ferrario  
manuela.ferrario@polimi.it

<sup>1</sup> Department of Electronics, Information and Bioengineering, Politecnico di Milano, P.zza Leonardo da Vinci 32, 20133 Milan, Italy



**Fig. 1** Three-lead ECG recording: *a* lead affected by noise and artifacts, with some recognizable beats, *b* good quality lead, *c* lead completely corrupted by noise. Notice the importance of both the lead selection and the beat-to-beat quality assessment to extract reliable RR series

Furthermore, current commercial monitors compute various parameters from the above-mentioned waveforms. The quality assessment of the recorded pulsatile signals represents the first step toward the improvement in the performance of derived indices, given that noise or poor quality signals greatly affect the parameter estimation and thus the reliability of thresholds proposed for these parameters [11].

In addition, the occurrence of irrelevant or false alarms [38, 39] could be potentially reduced by using good signal quality intervals to adapt thresholds to the specific patient needs.

As a further evidence of the importance and timeliness of this still unresolved topic of investigation, the 2015 PhysioNet/CinC Challenge focused on the development of algorithms to reduce the incidence of false alarms in the ICU [43].

In this work, we present a review of state-of-the-art algorithms for the quality control of long clinical ECG recordings, both single- and multi-lead, and of the ABP signal.

## 2 Methods for the assessment of ECG quality

The ECG signal is often corrupted by noise, which can be within the frequency band of interest and can manifest similar morphologies as the ECG itself. The sources of noise and artifacts that typically affect the electrocardiogram acquisition can be grouped in device malfunctions, baseline wander, electrode–skin contact defects or detachment, power-line interference, motion artifacts [8, 10].

Figure 1 shows an example of a three-lead ECG recording where noise and artifacts corrupt the ECG either partially (Fig. 1a, b) or completely (Fig. 1c). Thus, it is very important not only to choose the best channel to estimate the heart rate, but also to evaluate the reliability of the detection of each single beat. Several algorithms have been proposed for the quality scoring of a single- or multi-lead ECG recording, based on time-domain approaches, frequency-domain approaches or a combination of them.

### 2.1 Time-domain algorithms

Time-domain methods are mainly based on the knowledge of the morphological features of a physiological ECG signal, and they consist in the application of amplitude thresholds or in the estimation of statistical indices.

A first method to evaluate the noise content in an ECG recording was proposed by Moody and Mark [34]. In their work, they addressed the problem of how to discard highly corrupted ECG leads before proceeding with beat classification [34]. They estimated the noise level on each lead as the mean absolute error of a linear interpolation between the endpoints of the segment beginning halfway between two consecutive R peaks. They proposed a linear discriminant function (LDF) based on the ratio between the peak-to-peak amplitude and the estimated noise level from each lead. The LDF value represents the relative quality of each lead in comparison with the other ones, and it becomes large when one channel is highly corrupted by artifacts.

Some years later, in 1989, Moody and Mark designed a new algorithm to identify noise by adopting the Karhunen–Loève transform (KLT) to analyze the ECG signal [35], using the residual error of a truncated KLT as an estimate of the noise content of the signal. Wang [42] proposed a method to detect both physiological (i.e., axis shift, QRS morphology and amplitude variation, atrial fibrillation/flutter) and non-physiological noise (i.e., line noise, baseline wander, electrode motion artifact, muscle artifacts) on a single-lead ECG. The algorithm is based on the calculation of the area under the ECG curve, assuming that differences between values of adjacent QRS complexes are small for a noise-free lead and large and highly variable for corrupted signals. A mismatch function was proposed in order to assess the differences of the normalized area. From these values, a cumulative histogram of difference is estimated. This curve rises faster for high-quality signals; therefore, the signal quality score is provided by the displacement of the curves. Running a single-lead arrhythmia detection algorithm on 44 records from the MIT-BIH database, this method obtained an average 2.56 % of false positive in the premature ventricular contraction (PVC) detections on the ‘bad quality’ leads versus a 0.47 % false positive rate on the leads marked as ‘high quality’ by this algorithm, hence demonstrating the effectiveness of the proposed method in improving arrhythmia detection performance.

More recently, Hayn et al. [19, 20] developed two different algorithms for classifying multi-lead ECGs. Both methods consist in several rules that can be grouped in three classes: (a) basic ECG quality properties such as amplitude, spike features and constant portions; (b) number of crossing points between different leads; and (c) a comparison between QRS and non-QRS amplitude. In the first work [19], the ECG quality is assessed by means of seven criteria.

Criterion A1—more than 40 % of the signal shows amplitude larger than  $\pm 2$  mV

Criterion A2—portion of samples close to spikes is higher than 40 % of the signal

Criterion A3—more than 80 % of samples have amplitude equal to the preceding ones.

Criterion A4—more than 68.5 % of the signal is marked as potentially bad according to criteria A1–A3.

Criterion B—a lead crosses any of the others more than 48 times.

Criterion C—a measure of QRS quality detection based on signal-to-noise ratio (i.e., amplitude of the lowest QRS complex detected divided by the highest amplitude of non-QRS signal portions), maximum QRS amplitude and regularity of the rhythm. The criterion is met if this measure is  $< 0.2$

Criterion D—number of lead crossing points of the second worst channel (sorted by the number of lead crossing points) is higher than 23 or the quality measure of the second worst channel (sorted by the quality measure) is lower than 0.065.

The ECG is marked as unacceptable if criterion B or C is fulfilled or if criterion D and any of criteria A are fulfilled. The algorithm correctly classified 91.6 % of the analyzed recordings [19].

The method was simplified [20], using only criteria A1–A4 and B in order to avoid the computational complexity of the QRS detector. It is based on the same approach as the original method: Criteria A4 and D are removed, some thresholds are modified and criterion C is calculated using a QRS amplitude range instead of a threshold on the maximum amplitude. The simplified algorithm successfully classified 87.3 % of the ECGs. The testing of various combinations of all the criteria above obtained sensitivity ranging from 72.9 to 84 %, specificity from 96.1 to 98.2 % and accuracy from 92.5 to 93.4 % [20].

The algorithm developed by Kuzilek et al. [26] for the 2011 CinC Challenge is based on simple rules as well, which consist in applying thresholds to the statistical, such as variance and maximum signal values and amplitude range of the signal. The thresholds are estimated using basic features and a priori knowledge of ECG signal, and they are tuned to improve the sensitivity. A nonlinear support vector machine (SVM) classifier was proposed to improve specificity to reject portions of noisy ECG not identified in the previous step. The SVM works in the feature space composed by the covariance and time-lag covariance matrices, and returns a quality score; scores derived from the two steps are then combined via weighted sum, and the final score is compared to a threshold for the final decision of acceptability. The algorithm can be further tuned by adjusting weights and threshold values, and it properly marked 83.3 % of the non-acceptable records in the test set of the Challenge.

In the same context of the 2011 Computing in Cardiology Challenge, Moody [33] proposed a time-domain algorithm for ECG quality control, based on three simple criteria to detect flat portions of the signal due to disconnection of the electrodes and segments affected by low amplitude, baseline drift, noise spikes and frequent and large changes. The final decision about the quality of the signal is based on a combination of simple criteria. A 32-ms interval is marked as quiet if the range of values of the signal is lower than 0.1 mV. A record is considered of bad quality if: (1) at least two signals are classified as constant for more than 200 ms; (2) at least one signal in the record has a percentage of quiet intervals not in the range 64–96 %; and (3) at

least one signal has an overall range of  $<0.2$  millivolts, or more than 15 millivolts. The algorithm successfully classified 89.6 % of the records in the Challenge test set.

A completely different approach was proposed instead by Kalkstein et al. [25]. Their method is not based on prior medical knowledge, but it is mainly based on general features of the data, including correlation between different leads, signal energies and direction of each lead. The set of features is then used to train a classifier composed by a combination of ensemble decision trees and K-Nearest Neighbors. The algorithm obtained an accuracy of 92.95 % in classifying the ECGs based on their usability.

Also in 2011, Chudacek et al. [9] designed a binary algorithm aiming to discern between acceptable and non-acceptable recordings. The method is conceived to be simple and have low computational burden and relies on the sequential application of simple rules to detect missing leads, poor electrode contact, high amplitude artifacts, isoline drift and noisy leads. This simple algorithm correctly classified 87.2 % of the analyzed ECGs.

In 2012, Jekova et al. [21] suggested a more complex approach to the identification of the most common sources of noise in multi-lead ECG recordings. Their method is based on signal slopes and amplitudes in different frequency bands to detect six different types of noise and artifacts: flat signal, low amplitude leads, peak artifacts, baseline wander, high-frequency/electromyography noise, power-line interferences. The global quality score thus ranges from 0 (absence of noise) to 6 and is optimized by adjusting the thresholds in order to obtain the best performances. The final algorithm was presented in two different versions: the version with high sensitivity ( $Se = 98.7\%$ ,  $Sp = 80.9\%$ ), which represents a reliable alarm for technical recording problems, and a version with high specificity ( $Se = 81.8\%$ ,  $Sp = 97.8\%$ ), which validates noisy ECGs with acceptable quality during short recordings, being less sensitive to incidental artifacts.

Johannesen and Galeotti [23] identified a similar set of common ECG disturbances and designed a two-step algorithm capable of quantifying noise. The first step consists in the detection of lead connection problems, which generate macroscopic errors, such as signal absence and saturation. The second step consists in scoring the ECG quality by quantifying and then removing baseline wander, powerline and muscular noises. For each noise source, a noise measurement was defined for every beat as the highest value across leads. Finally, the global noise measurement was obtained through summation across all beats for powerline noise and baseline wander and across sinus beats only for muscular noise; the final score was then converted to a continuous scale ranging from 0 to 10. The proposed algorithm obtained a classification accuracy of 90.0 % in

rejecting ECGs with macroscopic errors, giving a score for the overall quality.

More recently, Naseri and Homaeinezhad [36] designed a three-step algorithm whose performance relies on the assessment of energy-based and concavity-based quality of the signal, and on the correlation between different leads. First, baseline wanders and high-frequency disturbances are removed. Second, a single energy-based criterion is used to identify admissible full range, artifacts and no-signal sections of the signal. The method consists in adopting energy lower and upper bounds calculated from a set of reference ECG signals. Regarding the concavity, the rationale is based on the hypothesis that a good quality ECG is composed by QRS complexes with high concavity and isoline, whose concavity value tends to zero. In a bad quality signal with high-energy disturbances, isolines are supposed to be missing, thereby causing the energy of the normalized concavity curve to be out of the admissible range. Although the energy-concavity index (ECI) analysis performed well for high-energy artifacts, it is inherently inadequate for low-energy ones. To solve this issue, a third step based on the correlation between leads is thus proposed. A target lead is reconstructed using the other ECG leads by an artificial neural network. The reconstructed lead is then compared to the target lead by computing two indices: a fitness score based on the residuals obtained by subtracting the target signal from the reconstructed one and a score assessed from the correlation function between the two signals. The computed indices are then compared to fixed thresholds. The correlation-based approach was able to detect all kinds of noise overlooked by the ECI analysis. The algorithm obtained a final classification accuracy of 93.6 %.

## 2.2 Frequency-domain algorithms

Frequency-domain methods are based on the analysis of the spectral features of the ECG signal in order to identify its noise content. Zaunseder et al. [44] proposed an algorithm based on an ensemble of decision trees, which use simple spectral features computed in the 0.5–40 Hz band of the signal by separating its content into low-frequency noise band (0–0.5 Hz) and high-frequency noise band (45–250 Hz). The complete feature space is obtained by computing various parameters such as the power in each frequency band and the signal-to-noise power ratios for each channel, both on the complete signal and on 2.5 s segments. The number of features is then reduced to only three values (the mean value and the extremes, i.e., the minimum and the maximum values) instead of one value per channel. The algorithm correctly classified 90.4 % of the records, and a similar result was also obtained using only a single decision tree, which permits an easy interpretation.

Johannesen improved the previously proposed ECG quality metric by including a set of features such as lead fail (constant derivative of zero for all the samples in a lead), global high and low-frequency noise, QRS detection noise and high and low-frequency noise content of sinusual beats [22]. His algorithm correctly classified 79 % of the ECGs in the 2011 CinC test set, which was a worse performance in comparison with the training set; this could be due either to the thresholds assessment overfitting on the training data or to the computational time.

### 2.3 Algorithms based on time and frequency analysis

A number of algorithms proposed for the ECG quality assessment use a combination of frequency- and time-domain features, ranging from signal amplitude to statistical indices. Allen and Murray [3] proposed a method based on seven simple measures, six of which take into account the frequency content of the ECG in different bandwidths ranging from 0.05 to 100 Hz, while the seventh criterion is the number of times the signal exceeds the  $\pm 4$  mV range (out-of-range events, ORE) in order to account for gross electrode movement. They found the frequency content in the range 0.05–0.25 Hz together with the rate of OREs to be the most quality dependent among the proposed indices and useful for a gross quality analysis.

Clifford et al. [12, 13] developed two algorithms capable to classify both 12-lead and single-lead records. Both works proposed a set of six features for training different classifiers in order to distinguish acceptable ECGs from unacceptable ones. In the first algorithm, the features consist in the following parameters:

iSQI: the percentage of R peaks detected on a single lead which were detected also on all leads.

bSQI: the percentage of R peaks detected by the algorithm named *wqrs* that were also detected by a second algorithm named *eplimited* (two open source algorithms for QRS detection [18, 45]).

kSQI: the kurtosis of the distribution.

sSQI: the skewness of the distribution.

pSQI: the percentage of the signal  $x$  which appears to be a flat line ( $dx/dt < \mu$  where  $\mu = 1$  mV).

fSQI: the ratio of spectral power in two different bands  $P(5-20 \text{ Hz})/P(0-F_n \text{ Hz})$ , where  $F_n$  is the Nyquist Frequency.

These parameters were used with two classification approaches: a single classifier trained on all 12 leads combined and 12 separate classifiers trained on the individual leads. Several machine learning algorithms were considered to classify the data as acceptable (1) or unacceptable (−1): support vector machine (SVM), naive Bayes (NB),

multilayer perceptron (MLP) and linear discriminant analysis (LDA). In the 12-lead classifier, the input data consisted of 72 features, whereas the single-lead classifiers were trained on the six features extracted for each lead individually. For the 12-leads classification, the three best performing classifiers are combined to perform a majority vote, while in the single-lead approach the outputs of all the classifiers are combined and the resulting score compared to a threshold. The proposed algorithm obtained an accuracy of above 95 % on the test set.

In the second algorithm proposed by Clifford et al. [13], two criteria were added in place of fSQI: the relative power in the QRS complex band  $P(5-20 \text{ Hz})/P(0-40 \text{ Hz})$  and the relative power in the baseline band  $(1-P(0-1 \text{ Hz}))/P(0-40 \text{ Hz})$ . The comparison of the classifiers was in this case limited to SVM and MLP, and the best performance for the single-lead approach was achieved with 96.5 % of accuracy and by considering only four criteria to train the SVM classifier (kSQI, bSQI and the relative powers in the QRS and in the baseline bands). A similar performance was obtained by using a 12-lead approach with five criteria: kSQI, sSQI, bSQI, pSQI and the relative power in the baseline band.

A similar set of criteria was then used by Behar et al. [5] with slight modifications. In particular, they proposed rSQI as the ratio of the number of beats detected by the algorithms *eplimited* and *wqr*, and pcaSQI as the ratio of the sum of the eigenvalues associated with the five principal components over the sum of all eigenvalues obtained by principal component analysis (PCA) applied to the time-aligned ECG cycles detected by the *eplimited* algorithm. These features were used to train a SVM classifier, obtaining values of accuracy of 97 and 93 % on the test dataset and on the arrhythmia set, respectively.

The authors evaluated the possibility to integrate their algorithm with information derived from other pulsatile signals in order to reduce false alarms in the ICU, so they extended their previous work [6] by studying the association between the quality assessed by the classifier and the alarm status (true vs. false alarm). An alarm was considered false if the quality of the best lead of a record was inferior to a threshold. The results were good for ectopic beats, tachycardia rhythms and atrial fibrillation, as well as for sinus rhythm, but the false alarm suppression on arrhythmic rhythms was not accurate enough.

Further, Liu et al. [31] adopted a similar approach and developed an algorithm to evaluate the acceptability of both single- and 12-lead ECG recordings. They used four binary flags to highlight the presence of four different quality problems:

Straight-line signal: the signal is divided into 2-min windows and a threshold is applied to the standard deviation of the ECG derivative.

**Huge impulse:** the signal is divided into 1-min segments and a threshold is applied in order to detect impulses.

**Gaussian noise:** the signal is divided into 1-min segments and a threshold is applied to the sample entropy values and to frequency ratio SF/NF where SF is the signal-frequency component (0.05 through 40 Hz) and NF is the noise-frequency component (more than 40 Hz).

**R-peak detection errors:** abnormalities in the RR intervals are detected using the impulse rejection filter, a previously developed algorithm [32].

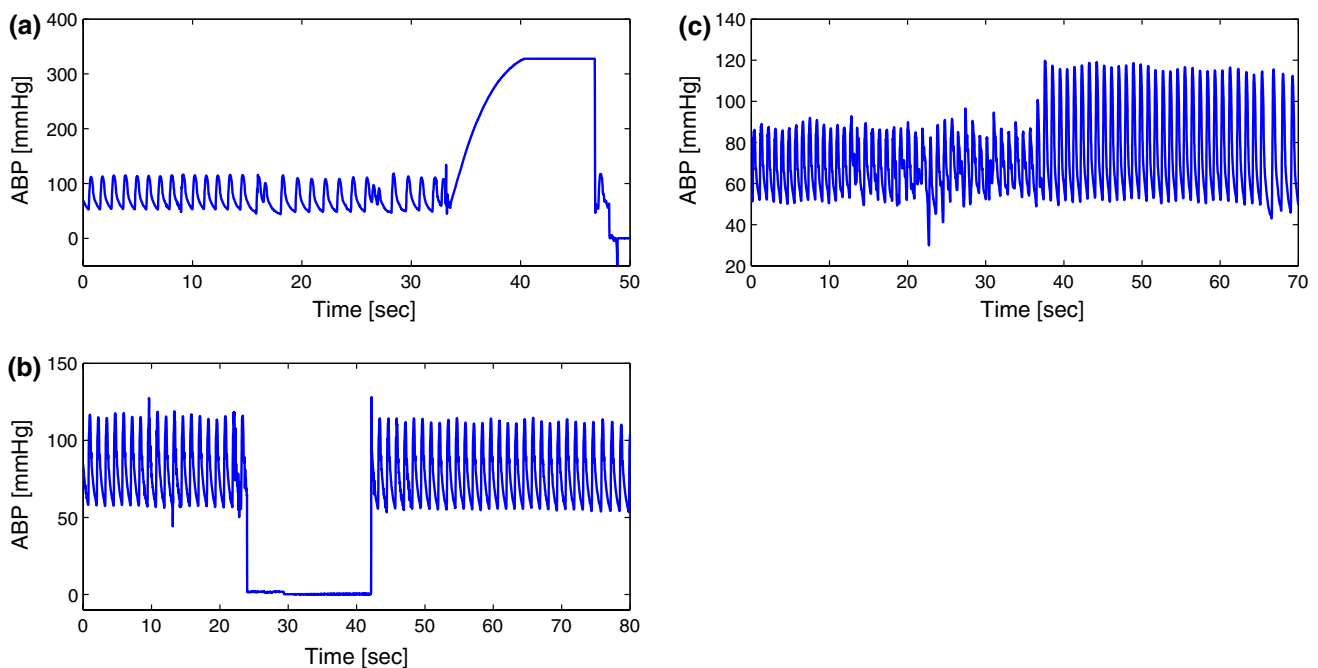
The authors then combined the flags and the number of noisy segments to obtain a single-lead signal quality index (SSQI) ranging from 0 to 1 (bad to good quality). The SSQIs are then summed up across leads to generate a 12-lead signal quality index. The algorithm scored 90.67 % of sensitivity and 89.78 % of specificity in classifying records as acceptable or unacceptable.

Di Marco et al. [15] identified seven causes of poor quality and quantified them through the analysis of QRS amplitudes and joint-time frequency (JTF) representation of the temporal–spectral distribution of the ECG energy. Baseline drift, constant amplitude, QRS artifacts, spurious spikes, white noise and signal saturation are identified using time and frequency marginal energy computed in four different frequency bandwidths. QRS amplitude is defined here as the median value of the peak-to-nadir amplitude difference of the QRS complexes detected. The

classification problem is then addressed following two different approaches. The first one is a classification based on a cascade of single-condition decision rules. The five indices derived from JTF analysis and QRS amplitude are compared to fixed thresholds to generate six conditions per lead. This classification scored 90.4 % of accuracy. The second approach is a supervised learning classifier applied to a subset of the original 60 features, obtained by means of a genetic algorithm. This approach obtained 91.4 % of accuracy.

### 3 Methods for assessment of ABP quality

Similarly to ECG, the ABP signal is often corrupted by non-Gaussian, nonlinear and non-stationary noise and by artifacts, especially in records collected in the ICU. Moreover, several important parameters are derived from ABP, such as systolic blood pressure or pulse pressure, which makes it crucial to record a high-quality continuous ABP waveform. In fact, artifacts and noise strongly affect the reliability of these derived variables, and the filtering process must take into account the possible different artifact types. Movements or mechanical disturbance on the transducer/line can produce the following typical artifacts: saturation to a maximum (Fig. 2a), saturation to a minimum (Fig. 2b), reduced pulse pressure (Fig. 2c), square wave, high-frequency artifacts, impulse artifacts.



**Fig. 2** Some of the most common ABP artifacts: **a** saturation to a maximum, mostly due to the flushing of the *arterial line*, **b** square wave, usually caused by a fast flush test of the sensor, **c** reduced pulse pressure, possibly due to a clot in the *arterial line*

To take into account all these types of artifacts, Li and Clifford [28] proposed an ABP signal quality index based upon the combination of two previously proposed quality measures: wSQI [29] and jSQI [30]. The former was designed to reject low-quality ABP segments and the latter to set physiological constraints to the features extracted from the pressure signal.

The algorithm named wSQI is based on pulse detection followed by the beat-by-beat classification of the ABP waveform [29], performed through a fuzzy approach. Various morphological features of the signal, mostly amplitudes and slopes, are computed to form a set of features, updated with following beats. The algorithm also exploits the relationship between the ABP and the ECG waveforms, and it provides a global index to assess the reliability of heart beat detection from the available signals (ECG and ABP).

The open source algorithm named jSQI is based on the beat-to-beat evaluation of the features extracted from the ABP waveform by means of abnormality criteria regarding noise level, physiologic ranges and beat-to-beat variations [30]. Li et al. [30] evaluated the quality of the ABP waveform by combining the previously described algorithms. The value of wSQI is maintained if also jSQI indicates a good quality signal, otherwise it is scaled by a coefficient.

The signal abnormality index (SAI) developed by Sun et al. [40] has also been used by Asgari et al. [4] as a reference method both to evaluate their algorithm and to improve the performance through the introduction of further abnormality conditions. The proposed method is based on the singular value decomposition (SVD), used to decompose the ABP pulses into signal and noise subspace. Each ABP pulse is then validated by comparing the ratio of the energy in the signal subspace over the energy in the noise subspace with a threshold. The algorithm obtained 94.69 % of accuracy, definitely higher than the 88.69 % obtained by the jSQI algorithm alone. The performance on the single beat validation is then maximized by integrating the SVD algorithm with the abnormality conditions and this algorithm obtained the true positive rate of 99.05 % and the false positive rate of 3.92 %.

#### 4 Reduction in false alarm: quality of heart rhythm estimation

As previously discussed, one of the most important tasks of the signal quality assessment is the reduction in false alarms in the ICU, generated either by the ECG or the ABP signal. In addition to the above-mentioned quality-based algorithms for signal classification and validation, some authors have exploited the availability to use pulsatile signals such as ABP and PPG and their correlation with the ECG in order to reduce false arrhythmias alarms. Pulsatile

signals provide indeed redundant and dependent information about heart rate, and they can therefore be used for cross-validation in order to avoid false alarms of life-threatening events related to heart rhythm.

In 2004, Zong et al. [46] proposed an algorithm to reduce the ABP alarms by using the relationship between the ABP and the ECG signal. After assessing the quality of the ABP signal as previously described, two ECG-based parameters are determined: ECG–ABP delay time and ECG rhythm. These two indices are then used to estimate the reliability of the ECG and to modify the signal score computed in the previous stage. The algorithm scored a sensitivity of 99.8 % and a positive predictive value of 99.3 % in reducing false alarms generated by ABP artifacts.

Ali et al. [2] proposed a general method to compare the information contained in two correlated signals specifically in ABP and ECG waveforms. The algorithm is based on the assumption that the two signals are highly correlated and that physiological disturbances would affect them both. The representation of ABP and ECG on the same plot produces a morphogram, which normally has a recurring shape. Artifacts are then detected as perturbations in the plot contour. The algorithm was tested on a set of manually annotated ABP true alarms, and a sensitivity of 90 % and a specificity of 99 % were obtained.

The work of Li et al. [29] followed the same rationale and aimed at obtaining a reliable estimation of the heart rate. The proposed algorithm performs first a quality assessment of ABP and ECG and successively implements the Kalman filters in order to estimate one HR series from ECG and a second HR series from ABP separately. The HR estimates are merged using all available sources of information, weighted by the inverse of the Kalman innovation of the signal, which is a measure of the novelty in a signal. The hypothesis is that an artifact is likely to generate a novel value, unless more than one information source indicates a change in the heart rhythm itself. The algorithm was tested on bradycardia and tachycardia episodes, obtaining a correct tracking performance, respectively, of 99 and 99.9 % and an artifact tracking percentage, i.e., percentage of false alarm, of 17 and 35 %.

In 2012, Li and Clifford [28] proposed a new method, adding the information provided by the PPG signal in addition to ECG and ABP signals. They compared three different approaches for false alarm suppression. The first one is based on PPG signal only; the second approach is based on a Kalman filter with the HR estimated on the three available signals and their respective signal quality scores. Finally, the third approach is based on a relevance vector machine (RVM) with a subset of features selected using a genetic algorithm. The algorithm was tested for asystole, extreme bradycardia (EB), extreme tachycardia (ET) and ventricular tachycardia (VT). The best performances in

terms of false alarm suppression rate were obtained using the PPG method for the asystole (83.1 %), the RVM for VT (30 %) and the Kalman filter for EB and ET (95.2 % and 46.0 %, respectively).

Aboukhalil et al. [1] developed an algorithm to reduce false arrhythmia alarms which integrates the ECG analysis by using the ABP and the signal abnormality index previously introduced by Sun et al. [40]. If the quality is high enough, each alarm generated on the ECG is processed and confirmed or suppressed by comparing the information contained in the pressure waveform. The algorithm obtained a false alarm suppression rate of 93.5 % for asystole, 81.0 % for EB, 63.7 % for ET, 58.2 % for ventricular fibrillation and 33.0 % for VT. The true alarms suppression rate was 0 % except for VT (9.4 %).

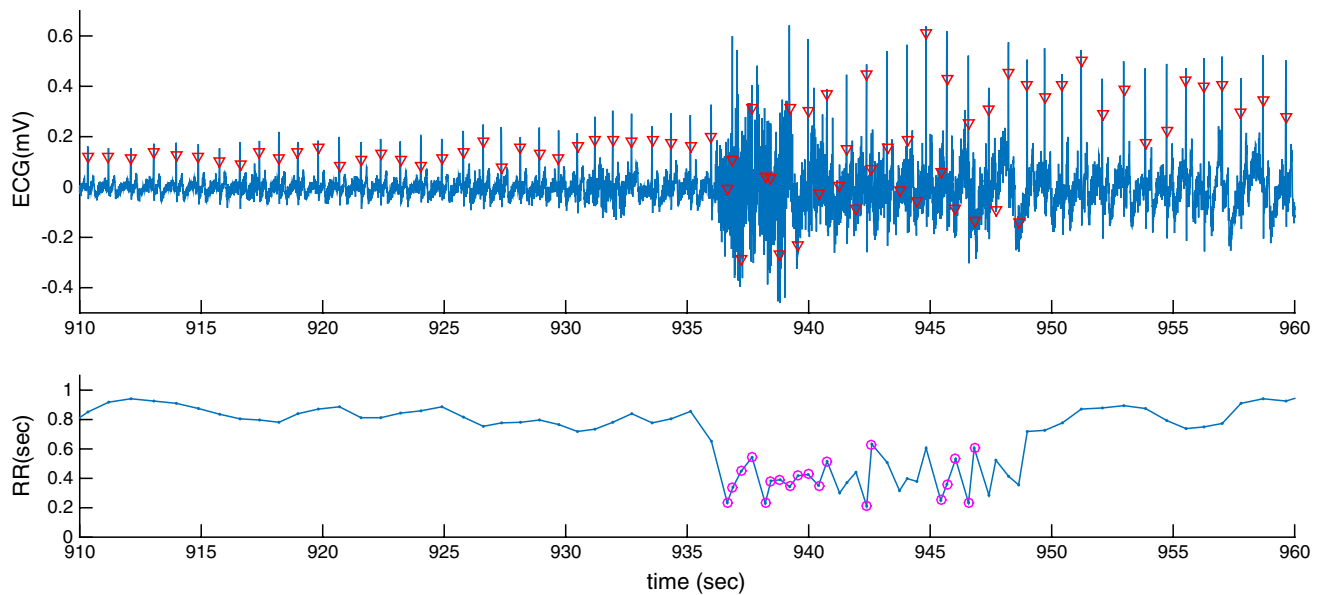
Recently, Johnson et al. [24] and Pimentel et al. [37] proposed a multimodal heart beat detection. The idea behind these approaches is to compare the information conveyed by two correlated signals, specifically ABP and ECG waveforms. In the first approach, the quality of the ECG tracing (ECG SQI) is based upon the agreement between two distinct peak detectors. The quality of the ABP waveform (ABP SQI) is evaluated at various times during the cardiac cycle and it is considered good if the ABP values are physiologically plausible. Detections from these two signals were merged by selecting the time instant of R peak occurrence or of heart contraction detected from the signal with a higher quality.

In the second approach [37], features based on the wavelet transform, signal gradient and signal quality indices are

**Table 1** Signal quality algorithms and their clinical context of application

Paper	Signals	Clinical context
<i>Entire recording or segment quality assessment</i>		
Moody and Mark [34]	2-lead ECG	Arrhythmia analysis
Wang [42]	Single-lead ECG	Arrhythmia detection/best lead selection
Hayn et al. [19]	Multi-lead ECG	Telehealth/mHealth
Hayn et al. [20]	Multi-lead ECG	Telehealth/mHealth
Kuzilek et al. [26]	12-lead ECG	Telehealth/mHealth
Moody [33]	Multi-lead ECG	Telehealth/mHealth
Kalkstein et al. [25]	Multi-lead ECG	Telehealth/mHealth
Chudacek et al. [9]	Multi-lead ECG	Telehealth/mHealth
Jekova et al. [21]	Multi-lead ECG	Diagnostic ECG (artefacts detection)
Johannesen and Galeotti [23]	Multi-lead ECG	Diagnostic ECG (artefacts detection and noise quantification)
Naseri and Homaeinezhad [36]	Multi-lead ECG	General purpose (continuous quality scoring)
Zaunseder et al. [44]	Multi-lead ECG	General purpose
Johannesen [22]	Multi-lead ECG	Telehealth/mHealth
Allen and Murray [3]	Single-lead ECG	False arrhythmia alarm reduction (ICU)
Clifford et al. [10]	Single-lead/12-lead ECG	Ambulatory recording
Clifford et al. [13]	Single-lead/12-lead ECG	Ambulatory recording
Behar et al. [5]	ECG	Telehealth/mHealth
Behar et al. [6]	Single-lead ECG	False arrhythmia alarm reduction (ICU)
Liu et al. [31]	Single-lead/12-lead ECG	Telehealth/mHealth
Di Marco et al. [15]	12-lead ECG	Ambulatory recording
Ali et al. [2]	ABP, ECG	False pressure alarm reduction (ICU)
Li et al. [30]	ABP (beat-to-beat), multi-lead ECG	Robust heart rate estimation
Li and Clifford [28]	ABP, PPG, ECG	False arrhythmia alarm reduction (ICU)
Aboukhalil et al. [1]	ABP, ECG	False arrhythmia alarm reduction (ICU)
<i>Beat-to-beat quality assessment</i>		
Moody and Mark [33]	Single-lead ECG	Arrhythmia analysis
Li et al. [30]	ABP	Robustness of ABP derived parameters
Zong et al. [46]	ABP, ECG	False pressure alarm reduction (ICU)
Sun et al. [40]	ABP	Robustness of ABP derived parameters
Asgari et al. [4]	ABP	Robustness of ABP derived parameters
Pimentel et al. [37]	ABP, ECG	Robust Heart Rate estimation
Johnson et al. [24]	ABP, ECG	Robust Heart Rate estimation





**Fig. 3** A beat-to-beat signal quality index applied to a noisy ECG. In the *upper panel*, the ECG is displayed with the identified R peaks. The *lower panel* shows the RRI series; the beats coming from noisy ECG and thus misplaced are highlighted by circles

extracted from the ECG and ABP waveforms and used in a hidden semi-Markov model approach.

## 5 Discussion and conclusion

The present review intends to draw attention to the large heterogeneity of applications in diverse clinical settings that may take advantage of a robust signal quality assessment. Table 1 summarizes the methods reviewed in this work and their applications, such as best lead selection, arrhythmia detection, reduction in false arrhythmia alarms, reduction in false pressure alarms and improvement in heart rate estimation. Given the multiplicity of possible applications and goals, the scores used to evaluate the algorithms are heterogeneous and the performances are thus difficult to compare, being dependent on the specific task the algorithms are developed for. Moreover, some algorithms are quite flexible and the parameters can be set and adapted according to the specific application, as in the case of the method proposed by Jekova et al. [21].

Most of the algorithms are tested on freely available databases, which are annotated and represent the gold standard, such as the ECG dataset with arrhythmias annotations. One of the most important examples of these open access repositories is the PhysioNet Web site, which is a public service of the PhysioNet Resource funded by the National Institute of Biomedical Imaging and Bioengineering (NIBIB) and the National Institute of General Medical Sciences (NIGMS) at the National Institutes of Health

(NIH) and a precious source of multi-parameter cardiopulmonary, neural and other biomedical signals [43].

The possibility to test different methods on the same data would on the one side permit a straightforward comparison among the algorithms; on the other side, this could represent a limitation, as the algorithm improvements could be in part database driven.

Aside from the setting-specific goals, the clinical context in which the ECG is recorded is also a major factor to be considered when selecting a quality assessment algorithm, as the recording system and the environment highly affect the level of noise and artifacts that may alter the signal. A simple visual inspection can reveal the great difference in terms of noise between an ambulatory ECG and a mobile phone-collected one, for instance.

Finally, a thorough assessment of the quality of a cardiovascular signal is of even greater importance when the final outcome of signal processing consists of derived parameters, such as time series of heart rate, systolic, diastolic, mean, pulse pressure values and their power spectra or other dynamic indices.

Figure 3 shows that the analysis of the series of RR intervals (RRI) alone would lead to wrong conclusions based on the heart rate. From the inspection of the ECG signal in the upper panel, it is obvious that the aforementioned alterations in RRI are due to the noise in the ECG, which leads to a misplacement of the R peaks rather than to an actual episode of tachycardia. It is therefore essential to assess the quality of the signal in order to improve the reliability of the derived parameters.

In conclusion, this review has outlined three main factors to consider in order to choose the appropriate method for the signal quality assessment: i) the clinical context the signal is acquired in; ii) the specific goal of the quality assessment; and iii) the desired reliability level of the parameters and indices derived from the original recording. This last factor is strongly related to the second one. For example, tasks such as the beat-to-beat estimation of cardiac output from ABP measurements and the reduction in ICU false alarm rate require a preventive assessment of the signal, both in terms of beat-to-beat and overall quality of the entire waveform.

The majority of the methods reviewed in this manuscript were tested on datasets including cardiovascular signals recorded by the same acquisition system or by the same device. Signals can have different noise rate or signal-to-noise ratio according to the adopted device. Current technologies permit to have pulsatile signals or ECGs recorded by wearable sensors with range of values or frequency properties different than ambulatory ECG or clinical ABP ones. New algorithms for the assessment of signal quality, which are effective also for these new monitoring systems, are necessary and represent an important field of application to extend the use of most of the currently available methods, which have been discussed in this review.

**Acknowledgments** This research was supported by the Shockomics Grant (FP7 EU Project No 602706).

**Conflict of interest** The authors declare that there are no conflicts of interest.

## References

1. Aboukhalil A, Nielsen L, Saeed M, Mark RG, Clifford GD (2008) Reducing false alarm rates for critical arrhythmias using the arterial blood pressure waveform. *J Biomed Inform* 41:442–451. doi:10.1016/j.jbi.2008.03.003.Reducing
2. Ali W, Eshelman L, Saeed M (2004) Identifying artifacts in arterial blood pressure using morphogram variability. *Comput Cardiol* 31:697–700
3. Allen J, Murray A (1996) Assessing ECG signal quality on a coronary care unit. *Physiol Meas* 17:249–258
4. Asgari S, Bergsneider M, Hu X (2010) A robust approach toward recognizing valid arterial-blood-pressure pulses. *IEEE Trans Inf Technol Biomed* 14:166–172. doi:10.1109/TITB.2009.2034845
5. Behar J, Oster J, Li Q, Clifford GD (2012) A single channel ECG quality metric. *Comput Cardiol* 39:381–384
6. Behar J, Oster J, Li Q, Clifford GD (2013) ECG signal quality during arrhythmia and its application to false alarm reduction. *IEEE Trans Biomed Eng* 60:1660–1666
7. Chambrin MC (2001) Alarms in the intensive care unit: how can the number of false alarms be reduced? *Crit Care* 5:184–188
8. Chen Y, Yang H (2012) Self-organized neural network for the quality control of 12-lead ECG signals. *Physiol Meas* 33:1399–1418. doi:10.1088/0967-3334/33/9/1399
9. Chudacek V, Zach L, Kuzilek J, Spilka J, Lhotska L (2011) Simple scoring system for ECG quality assessment on Android platform. *Comput Cardiol* 38:449–451
10. Clifford GD (2006) ECG statistics, noise, artifacts, and missing data. In: Clifford GD, Azuaje F, McSharry PE (eds) *Advanced methods and tools for ECG data analysis*. Artech House, pp 55–99
11. Clifford GD, Long WJ, Moody GB, Szolovits P (2009) Robust parameter extraction for decision support using multimodal intensive care data. *Philos Trans A Math Phys Eng Sci* 367:411–429. doi:10.1098/rsta.2008.0157
12. Clifford GD, Lopez D, Li Q, Rezek I (2011) Signal quality indices and data fusion for determining acceptability of electrocardiograms collected in noisy ambulatory environments. *Comput Cardiol* 38:285–288
13. Clifford GD, Behar J, Li Q, Rezek I (2012) Signal quality indices and data fusion for determining clinical acceptability of electrocardiograms. *Physiol Meas* 33:1419–1433. doi:10.1088/0967-3334/33/9/1419
14. Devlin PH, Mark RG, Ketchum JW (1984) Detecting electrode motion noise in ECG signals by monitoring electrode impedance. *Comput Cardiol* 11:227–230
15. Di Marco LY, Duan W, Bojarnejad M, Zheng D, King S, Murray A, Langley P (2012) Evaluation of an algorithm based on single-condition decision rules for binary classification of 12-lead ambulatory ECG recording quality. *Physiol Meas* 33:1435–1448. doi:10.1088/0967-3334/33/9/1435
16. Feldman CL, Hubelbank M, Haffajee CI, Kotilainen P (1979) A new electrode system for automated ECG monitoring. In: *Proceedings in computers in cardiology*
17. Ferguson PF Jr, Mark RG, Burns SK (1986) An algorithm to reduce false positive alarms in arrhythmia detectors using dynamic electrode impedance monitoring. In: *Proceedings of computers in cardiology*
18. Hamilton PS, Tompkins WJ (1986) Quantitative Investigation of QRS detection rules Using the MIT/BIH arrhythmia database. *IEEE Trans Biomed Eng* BME-33:1157–1165
19. Hayn D, Jammerbund B, Schreier G (2011) ECG quality assessment for patient empowerment in mHealth applications. *Comput Cardiol* 38:353–356
20. Hayn D, Jammerbund B, Schreier G (2012) QRS detection based ECG quality assessment. *Physiol Meas* 33:1449–1461. doi:10.1088/0967-3334/33/9/1449
21. Jekova I, Krasteva V, Christov I, Abächerli R (2012) Threshold-based system for noise detection in multilead ECG recordings. *Physiol Meas* 33:1463–1477. doi:10.1088/0967-3334/33/9/1463
22. Johannesen L (2011) Assessment of ECG quality on an Android platform. *Comput Cardiol* 38:433–436
23. Johannesen L, Galeotti L (2012) Automatic ECG quality scoring methodology: mimicking human annotators. *Physiol Meas* 33:1479–1489. doi:10.1088/0967-3334/33/9/1479
24. Johnson AE, Behar J, Andreotti F, Clifford GD, Oster J (2015) Multimodal heart beat detection using signal quality indices. *Physiol Meas* 36(8):1665
25. Kalkstein N, Kinar Y, Nađaman M, Neumark N, Akiva P (2011) Using machine learning to detect problems in ECG data collection. *Comput Cardiol* 38:437–440
26. Kuzilek J, Huptych M, Chudacek V, Spilka J, Lhotska L (2011) Data driven approach to ECG signal quality assessment using multistep SVM classification. *Comput Cardiol* 38:453–455
27. Lawless ST (1994) Crying wolf: false alarm in a pediatric intensive care unit. *Crit Care Med* 22:981–985
28. Li Q, Clifford GD (2012) Signal quality and data fusion for false alarm reduction in the intensive care unit. *J Electrocardiol* 45:596–603. doi:10.1016/j.jelectrocard.2012.07.015
29. Li Q, Mark RG, Clifford GD (2008) Robust heart rate estimation from multiple asynchronous noisy sources using signal quality indices and a Kalman filter. *Physiol Meas* 29:15–32
30. Li Q, Mark RG, Clifford GD (2009) Artificial arterial blood pressure artifact models and an evaluation of a robust blood

pressure and heart rate estimator. *Biomed Eng Online* 8:13. doi:10.1186/1475-925X-8-13

31. Liu C, Li P, Zhao L, Liu F, Wang R (2011) Real-time signal quality assessment for ECGs collected using mobile phones. *Comput Cardiol* 38:357–360
32. McNames J, Thong T, Aboy M (2004) Impulse rejection filter for artifact removal in spectral analysis of biomedical signals. *Conf Proc IEEE Eng Med Biol Soc* 1:145–148. doi:10.1109/IEMBS.2004.1403112
33. Moody BE (2011) Rule-based methods for ECG quality control. *Comput Cardiol* 38:361–363
34. Moody GB, Mark RG (1982) Development and evaluation of a 2-lead ecg analysis program. *Computers in cardiology* 9:39–44
35. Moody GB, Mark RG (1989) QRS morphology representation and noise estimation using the Karhunen–Loeve transform. In: *Proceedings in computers in cardiology*. IEEE, Jerusalem, pp 269–272
36. Naseri H, Homaeinezhad MR (2015) Electrocardiogram signal quality assessment using an artificially reconstructed target lead. *Comput Methods Biomech Biomed Eng* 8(10):1126–1141
37. Pimentel MA, Santos MD, Springer DB, Clifford GD (2015) Heart beat detection in multimodal physiological data using a hidden semi-Markov model and signal quality indices. *Physiol Meas* 36(8):1717
38. Schmid F, Goepfert MS, Reuter DA (2013) Patient monitoring alarms in the ICU and in the operating room. *Crit Care* 17:216. doi:10.1186/cc12525
39. Siebig S, Kuhls S, Imhoff M, Langgartner J, Reng M, Scholmerich J, Gather U, Wrede CE (2010) Collection of annotated data in a clinical validation study for alarm algorithms in intensive care—a methodologic framework. *J Crit Care* 25:128–135
40. Sun JX, Reischer AT, Mark RG (2006) A signal abnormality index for arterial blood pressure waveforms. *Comput Cardiol* 33:13–16
41. Tsien CL, Fackler JC (1997) Poor prognosis for existing monitors in the intensive care unit. *Crit Care Med* 25:614–619
42. Wang JY (2002) A new method for evaluating ECG signal quality for multi-lead arrhythmia analysis. *Comput Cardiol* 29:85–88
43. [www.physionet.org](http://www.physionet.org)
44. Zaunseder S, Huhle R, Malberg H (2011) CinC challenge—Assessing the usability of ECG by ensemble decision trees. *Comput Cardiol* 38:277–280
45. Zong W, Moody GB, Jiang D (2003) A robust open-source algorithm to detect onset and duration of QRS complexes. *Comput Cardiol* 30:737–740
46. Zong W, Moody GB, Mark RG (2004) Reduction of false arterial blood pressure alarms using signal quality assessment and relationships between the electrocardiogram and arterial blood pressure. *Med Biol Eng Comput* 42:698–706



**Nicolò Gambarotta** earned his B.Sc. and his M.Sc. from Politecnico di Milano in 2011 and 2013, respectively. During his master thesis, he studied the role of the autonomic nervous system during intestinal ischemia in rat experiments. He has been a research fellow at Politecnico di Milano since 2014 supported by the EU project Shockomics.



**Federico Aletti** earned his M.Sc. and his Ph.D. from Politecnico di Milano in 2005 and 2009, respectively. He has been a post-doctoral fellow with the BBB Lab since 2009. He has been a visiting scholar with the Physiologic Signal Processing Laboratory at Michigan State University, East Lansing, MI, USA (advisor: Prof. Ramakrishna Mukkamala), and he has been a Visiting Professor at Universidade de Nove de Julho, Sao Paulo, Brazil. He is currently a postdoc EU Marie Curie Outgoing fellow (2014–2016) at UC San Diego. The research topics of his current project are: heart damage and cardiovascular dysfunction in circulatory shock, design of technology for enteral administration of protease inhibitors.



**Giuseppe Baselli** Milano 1958, M.Sc. in Electronic Engineering at the Politecnico di Milano cum laude, in 1983. He was an assistant professor at the Università di Brescia, an associate professor (1998) and next full professor (2001) at Politecnico di Milano. He was director of the Biomedical Engineering program (2004–2009) and of the Department of Bioengineering (2010–2012) and member of the Academic Senate (2011–2012). His research interests are in the field of biomedical signal processing and in its relationships with the linear and nonlinear modeling of cardiovascular regulation, image reconstruction in oncology and molecular imaging, neuro-imaging including functional and anatomical connectivity and vascular regulation. He is currently the PI of the EU project Shockomics.



**Manuela Ferrario** received her Ph.D. degree in Bioengineering (2007) with a thesis on advanced methods and techniques for the antepartum diagnosis of fetal risk. Her postdoc fellowship, supported by Fresenius Medical Care, GmbH, was focused on the analysis of heart rate variability (HRV) and biological cardiovascular signals in hemodialysis patients. She collaborates with the Renal Research Institute, New York City, USA, and with the Hospital San Bortolo, Vicenza. She was

the unit leader of a project supported by FIRB2008 on hemodynamic monitoring during surgery in collaboration with Università di Roma Tor Vergata. She is currently an assistant professor supported by the grant FIRB2012 in collaboration with Università Vita e Salute di Milano. She is member of the scientific committee of the EU project Shockomics (2013–2017). Her research interests include mathematical modeling of physiological systems, model identification with applications to cardiovascular system, signal processing, nonlinear analysis, variability analysis of cardiovascular signals, signals classification.