# Electrocardiogram Quality Assessment with Autoencoder

Jan Garus[1,2(✉)] , Mateusz Pabian[1,2] , Marcin Wisniewski[1] ,
and Bartlomiej Sniezynski[2]

[1] Healthcare Department, Comarch SA, al. Jana Pawła II 39a,
31-864 Krakow, Poland
jan.garus@comarch.pl
[2] AGH University of Science and Technology, al. Mickiewicza 30,
30-059 Krakow, Poland
https://www.comarch.pl/healthcare/
https://www.agh.edu.pl

**Abstract.** ECG recordings from wearable devices are affected with a relatively high amount of noise due to body motion and long time of the examination, which leads to many false alarms on ill-state detection and forces medical staff to spend more time on describing each recording. ECG quality assessment is hard due to impulse character of the signal and its high variability. In this paper we describe an anomaly detection algorithm based on the Autoencoder trained on good quality examples only. Once trained, this neural network reconstructs clean ECG signals more accurately than noisy examples, which allows to distinguish both classes. Presented method achieves a normalized F1 score of 93.34% on the test set extracted from public dataset of 2011 PhysioNet/Computing in Cardiology Challenge, outperforming the solution based on the best competition participants. In contrary to many state-of-the-art methods it can be applied even on short, single-channel ECG signals.

**Keywords:** ECG · Quality assessment · Autoencoder

## 1 Introduction

Electrocardiogram (ECG) signal quality (SQ) assessment has been a topic of 2011 PhysioNet/Computing in Cardiology (CinC) Challenge [1,2]. Despite a few years that have passed, it is still an issue, which emerges in the CinC Challenge 2017 [3], where a "noisy" class was among 4 output possibilities, causing the participants a lot of problems.

ECG quality classification is an important issue in a case of remote heart monitoring, especially when system is dedicated to simply record the data, without online analysis. In such a case, a patient does not have an immediate feedback about correctness of the lead placement etc. After a 24-h examination it may turn out that most of the signal is useless.

Automatic rhythm analysis and beat-to-beat classification of the ECG signal is also substantially affected with noisy recordings. Those signals often trigger false alarms, destroying even the best positive predictivity achieved on signals from standard databases (which are rather clean), like the MIT-BIH [4]. For example, a high-frequency noise may be incorrectly interpreted as heart beats, leading to false tachycardia alarm. Similarly, extremely low signal amplitude or sensor disconnection may be classified as an asystole.

Simple methods, like detection of lead disconnection are insufficient regarding a wide range of possible signal disruption.

The main contribution of this paper is to apply the anomaly detection approach to the ECG signal quality assessment. In this setting only typical ECG shapes are learned by the model, while noisy examples are treated literally as anomalies, in contrary to standard classifiers which requires a large, well annotated database containing all possible types of noise. Experimental results show that the proposed approach was more accurate than a classical one.

The paper is organized as follows. State-of-the-art methods are presented in Sect. 2. Section 3 contains brief introduction into ECG nomenclature and techniques used in our solution. Details of our one-class classifier approach are given in Sect. 4. The evaluation is presented in Sect. 5, including the comparison to challenge entries in Subsect. 5.5.

## 2   Related Work

As we have already mentioned, ECG signal quality assessment was a subject of the 2011 PhysioNet/CinC Challenge [2].

Among the challenge competitors, an excellent paper [5] from Gari Clifford *et al.*is worth an attention. A wide range of features (named *quality indices*) has been investigated, including the time and frequency domains, a statical distribution and strictly ECG-related aspects, *i.e.*differences between QRS-detection results gathered on separated channels and from two different algorithms. Proposed quality indices were used in many works, including an extension for arrhythmia [6] and more recent papers, like [7] which combines them with additional entropy-based and Lempel-Ziv compression-based features.

Two-step algorithm [8] utilizes covariance, mean variance, peak-to-peak value and maximum value in the first phase, then a covariance and time-delayed covariance supplies the SVM classifier. The same group of authors almost won the challenge with a procedure of 5 simple if-then rules [9], which exams an amplitude change over the time, the isoline drift and a standard deviation of the normalized signal. Speaking on the winners, most of the signal properties used in [10] also concern the amplitude: QRS vs. noise amplitude, peak-to-peak value, spikes, but also volume of flat signal fragments and number of crossing points in between channels.

Another rule-based algorithm was presented by [11], which utilizes 4 Signal Quality Indicators: straight line and huge impulse detection, the Gaussian noise detection and the detection of abnormalities *RR intervals*.

An approach of the DSP transformations followed by thresholding has been also proposed in many papers. In [12] low and high frequencies are separated and checked against arbitrary thresholds, as well as signal magnitude and a number of zero-passes. Follow-up model extends it with the modified Empirical Mode Decomposition [13,14]. Wavelet decomposition can be used in the same manner [15]. Thresholding of the signal saturation, energy and baseline change over the time was proposed in [16].

Simple signal autocorrelation within 5-s windows is utilized in [17].

Many methods utilize the cross-channel information to a much greater extend. A correlation between channels as well as the standard deviation of the signal autocorrelation is used in [18]. Similarly, the channel covariance matrix is used as random forest input in [19]. Authors of [20] compare each input channel to the reconstruction based on all others signal leads. In out method we also reconstruct the signal, but we use each separate channel do generate its own reconstruction. Therefore our approach can be applied for both multiple- and single-channel signals.

Some authors proposed yet another comparison-based algorithms, where the extracted QRS is compared against templates obtained with clustering [21] or statistical methods [22]. Another work [23] presents a method based on the cepstral analysis.

Neural networks are used in many papers, including Convolutional NN in [24] and Self Organizing Maps in [25]. Similarly to our work, an Autoencoder network was used in [26], however authors focused on ECG correction task (as a step of human identification procedure) rather than explicit assessment of the signal quality. In this paper we concentrate on the classification aspect. Furthermore, more complex AE architectures are evaluated in our work.

## 3    Materials and Methods

In this section we present a brief introduction into ECG signal, followed by key concepts and algorithms which form the foundations of our solution.

### 3.1    ECG Signal

Electrocardiogram (ECG) signal reflects how the heart is functioning. Data is acquired from a number of physical electrodes (varying from 2 to 10), then those physical readings are subtracted in pairs, forming signal channels used in further analysis, often referred as *leads*. In stationary examination, 12-lead record is a golden standard, but in case of mobile devices, fewer electrodes are usually available, resulting in a smaller number of data channels, sometimes even a single lead.

Healthy condition is called *the normal sinus rhythm*, which idealized waveform is illustrated in the Fig. 1. Consecutive fragments (waves) of the signal are named with letters P, Q, R, S and T.
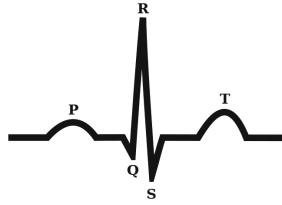
**Fig. 1.** ECG signal of the normal sinus rhythm.

Waves Q, R and S form the *QRS complex*, the most characteristic part of the heartbeat. Its presence, even disturbed by low signal quality or some medical condition is crucial in the ECG analysis. Waves Q and S may have too low amplitude to be visible, but the R should be easily recognized. In some heart conditions, the peak of the R wave is forked. The time distance between two successive R waves is commonly referred as the *RR interval*. It changes naturally in certain range as heart speeds up and slows down due to breathing and physical effort, but too big fluctuations may indicate both arrhythmia as well as problems with data acquisition. Waves P and T may be easily recognizable or barely visible or absent, depending on the signal quality and medical condition. Depending on the actual lead being analyzed, waves may be positive or negative in regard to the baseline. This is called the *polarization.*

Apart from medical problems, ideal ECG waveform may be disrupted by a number of causes, which includes: muscle movement, electrode shifts, skin contact changes and external electrical interference. Those disruptions or noises in the ECG signal are referred as *artifacts.*

Baseline of the raw ECG signal is not as flat as presented in the Fig. 1. In reality it is affected with a low frequency component, which arises from breathing, patient movement and electrical charge of electrodes. This so-called *baseline wander* not necessarily means bad signal quality and usually can be easily filtered out.

### 3.2   Anomaly Detection

Problem of the ECG signal quality assessment may be treated as a binary classification, with classes: *good quality* (GQ) and *bad quality* (BQ). An obvious solution is to build a dichotomous classifier and train it with a sufficient amount of examples of both classes.

However, it is hard to provide a balanced training set for this problem, which is required by a vast majority of classifiers to avoid classification bias [27]. There are too few BQ examples available. Moreover, due to irregularity and a variety of the artifacts' origin in the ECG signal, even smart duplication techniques, like *SMOTE* [28], are incapable of generating a sufficient training set which would cover the entire space of the BQ class.

Instead of working on some modifications of this dichotomous classifier, problem can be simplified by focusing on a GQ only. GQ is also a broad class, but

can be described in a significantly more precise way than BQ, as a concatenation of all heart beat morphologies and rhythms being considered by physicians.

Such an approach is named an *anomaly detection* or a *one-class classification* [29]. As the first name emphasizes, this technique is primarily dedicated to detection of some extremely rare states. The latter explains what knowledge it achieves: it learns how to describe a single class and nothing more.

Using this method for SQ is a different way of thinking: rather than learning how to distinguish GQ and BQ, we are telling our system only how GQ examples look like. If it encounters some significantly different example, it would be BQ.

### 3.3    Autoencoders

An *autoencoder* (AE) or *Encoder-Decoder* [30–32] is a neural network, which can be used as an anomaly detection algorithm.

It consists of following layers: input, *encoder* layer(s), *latent* representation (aka *hidden* or the *code*), *decoder* layer(s), reconstruction (output). The encoder and decoder subnetworks are mirrors of each other, working in the opposite way. In order to speed up learning procedure their weights can be shared (also referred as tied). The entire network is trained to reproduce values from the input to the output, by penalizing differences between original and reconstructed signal in a cost function.

In order to prevent the AE from "cheating" by simply passing the values through entire network, additional restrictions are involved [33]. The simplest method is to form the network in a hourglass shape with a hidden representation much smaller than input and reconstruction layers. Another popular approach, the *Denoising AE* [34] corrupts the input data with a noise (in training only) while the cost function still compares the reconstruction against the original data.

Alternatively, the latent representation may be forced to be as simple as possible by minimizing the number of non-zero elements. It can be achieved by adding the sparsity measure [35] to the cost function which penalizes every non-zero element in the latent vector. Makhzani and Frey proposed the *k-Sparse Autoencoder* [36], which introduces an additional, non-linear layer, right after the latent vector. This layer passes only $k$ greatest values from the hidden representation and replaces rest of them with zeros.

Strictly speaking, autoencoders belong to the *unsupervised learning* algorithms, as no class labels are required for training. However, if examples from one class only (GQ, in our case) would be used for training, we expect that AE will learn to reproduce other examples of this class in a significantly better extent than representatives of unseen class (BQ). This approach—applied to gamma ray readings—has been used by Sharma *et al.*in [29]. Some variants of this technique were introduced by An and Cho in [37].

Our idea is to detect BQ by comparing signal reconstruction error against threshold determined using a distinct validation set.

---

**Algorithm 1:** Algorithm overview

---

split signal into 10s-fragments
signal preprocessing
**for** *channel x* **in** *signal* **do**
    **for** *fragment z* **in** *x* **do**
        $\tilde{z} \leftarrow$ autoencoder$(z)$
        $d \leftarrow$ reconstruction_error$(z, \tilde{z})$
        **if** $d$ ¡ *threshold* **then**  mark $z$ as Good Quality
        **else**  mark $z$ as Bad Quality

---

## 4   Proposed Solution

General steps of the algorithm are presented in Algorithm 1, followed by sub-sections explaining details of this solution.

After initial split into 10 s-fragments (if database is not already in this form), signal preprocessing is applied. This step includes final extraction of 0.8 s, 1-channel fragments of the signal, positioned around supposed QRS complexes, but also few normalization operations, described in Subsect. 4.1.

Next, each 0.8 s fragment is processed with the autoencoder (Subsect. 4.2). Obtained output is then compared against the input signal (Subsect. 4.3). Finally, the quality of 0.8 s fragment is assigned using the BQ threshold calculated in the supervised phase of the training, as described in Subsect. 5.3.

### 4.1   Data Preprocessing

Our preliminary results showed, that it may be too hard for an autoencoder to reconstruct 10 s of ECG input. Moreover, according to our business requirements, higher time resolution of the results was needed. To address those circumstances, we decided to analyse a single-lead signal containing approximately a single heartbeat, which takes on average about 0.8 s .

As described in Algorithm 2, signal is resampled to 125 Hz and its amplitude is restricted to an interval of −6 mV to 6 mV, then the mean value is subtracted. The input 10 s-signal (one data channel at the time) is divided into 2.5 s-sections, from which 0.8 s-fragments (positioned around the largest peak) are finally extracted.

Few optional operations were also introduced: baseline cancellation and two types of normalization. For baseline cancellation we used a cascade of 600 ms and 200 ms median filters (similar to [38], but filters are in reversed order). Normalization of the QRS polarization ensures that the peak at the 20th sample is always positive (it may be different, depending on the actual ECG channel). Normalization to the interval of $[0, 1]$ assigns 0 to the minimum sample in the fragment, 1 to its maximum sample and accordingly modifies the rest of the samples to fit in this interval, keeping the signal shape.

---

**Algorithm 2:** Signal preprocessing of 10s-fragment

---

**for** *channel x* **in** *signal* **do**

    $x \leftarrow$ resample $x$ to $125Hz$

    **if** *option_cancel_baseline* **then** $x \leftarrow$ cancel baseline in $x$

    $x \leftarrow x - mean(x)$

    **for** *sample $x_i$* **in** *x* **do**

        $x_i \leftarrow max(x_i, -6mV)$

        $x_i \leftarrow min(x_i, 6mV)$

    **for** *i* **in** 0 .. 3 **do**

        $y \leftarrow x_{i*312 \,..\, (i+1)*312}$ // extract 2.5$s$ fragment

        $j \leftarrow argmax(abs(y))$

        $z \leftarrow y_{j-20 \,..\, j+80}$ // extract 0.8$s$ from each 2.5$s$ fragment

        $z \leftarrow z - mean(z)$

        **if** *option_normalize_polarization* **then**

            **if** $z_{20} < mean(z)$ **then** $z \leftarrow -z + max(z)$

        **if** *option_normalize_to_inverval_from_0_to_1* **then** $z \leftarrow \frac{z-min(z)}{max(z)-min(z)}$

        **save** $z$

---

### 4.2 Investigated AE Variants

We have implemented our anomaly detection model using following variants of the Autoencoder network: simple Autoencoder (no constraints) (AE), Denoising Autoencoder (DAE) [34], Sparsity Autoencoder (SpAE) [35] and k-Sparse Autoencder (kSpAE) [36].

Each model listed above has been manually optimized in terms of the size of the latent representation and its variant-specific parameters in order to maximize the accuracy of the signal quality assessment. The best configurations are listed in Table 1.

### 4.3 Reconstruction Error

We have defined and implemented following measures of dissimilarity between input signal $x$ and it's reconstruction $\tilde{x}$:

$$
\begin{array}{ll}
E_{abs} = |x - \tilde{x}| & E_{abs\_pf3} = E_{abs} + \sqrt{|E_{abs} - E_{med3}|} \\
E_{pow} = {E_{abs}}^2 & E_{sqrt} = \sqrt{E_{abs}} \\
E_{med3} = med_3(E_{abs}) & E_{sqrt3} = \sqrt[3]{E_{abs}}
\end{array} \tag{1}
$$

where $med_k$ denotes a median filter of the length $k$.

Based on those element-wise errors, *signal reconstruction errors* are defined as the average value of corresponding measure:

$$
d_* = \sum_{i=1}^{n} \frac{E_{*i}}{n} \tag{2}
$$

where $n$ is the length of samples in signal $x$ and $E_{*i}$ is the value of measure $E_*$ for $i$-th sample.

## 5    Evaluation

Details on the database preparation, training and numerical results are provided in this section. Comparison to the algorithm based on the state of the art solution is also given in the last Subsect. (5.5).

### 5.1    Databases

In the unsupervised part of the model (*i.e.* training the autoencoder) we were using the ECG examples obtained from the following databases: MIT-BIH [4], incartdb [1], ltafdb [39], svdb [40] and AHA DB [41]. Those databases are rather noise-free, so good signal quality was assumed for the entire set. We used channels 1 and 2 from those databases, except the *incartdb* where we have chosen leads 2 and 12. Sampling frequencies were normalized to $fs = 125$ Hz .

In the supervised stage, which was the threshold calculation, we used labelled signals extracted from the PhysioNet/CinC Challenge 2011 training set (1000 records) and *Event 2* open test set B (500 records). The concatenated set was randomly split into training (for threshold selection) and test (final algorithm effectiveness) sets in a ratio of 50:50. Original data was 12-lead, 10-s, $fs = 500$ Hz.

### 5.2    Reference Classification

Due to the signal manipulation (separating channels and sampling 0.8 s fragments), existing reference classification of the labelled database was outdated. We have manually classified those examples into 3 classes with following criterion: *GQ (good quality)*—visible QRS complex but also clear P and T waves (if present), *AQ (acceptable quality)*—visible heartbeats, but less clear shape of the QRS complexes, *BQ (bad quality)*—heartbeats unrecognizable or presence of high-amplitude distortions. Visualizations of the resulting sets are presented in Fig. 2.

During threshold calculation and final model evaluation AQ examples were incorporated into GQ, because there was no business need to distinguish them and their shapes were quite similar.

### 5.3    Training

Autoencoders were trainereconstructed as an g set (Subsect. 5.1), which was split into training, validation and testing subsets in a ratio of 70:15:15.

For each combination of the autoencoder variant (Sect. 4.2) and the measure of reconstruction error (Sect. 4.3) we calculated the BQ threshold using the labelled training set. It is set at the crossing point of the histograms made exclusively for GQ (including AQ) and BQ examples (using a half of the labelled set, leaving the other half for testing). This step is illustrated in Fig. 3. To avoid classification bias, number of signals from a minority class (BQ) was upscaled.
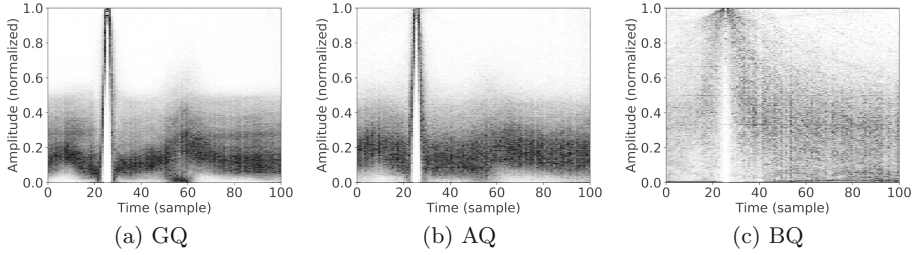
**Fig. 2.** Visual summaries of the each class' signals in the quality-labelled set. Baseline cancellation, 0–1 normalization and the normalization of R-peak polarization were applied to all signals in each class (see Sect. 4.1). Then, for each time index (horizontal axis), histogram of values is calculated and drawn along the vertical axis. Based on those visualizations, signals from the GQ and AQ classes are quite similar within those subsets, in opposite to much more diverse BQ subset.
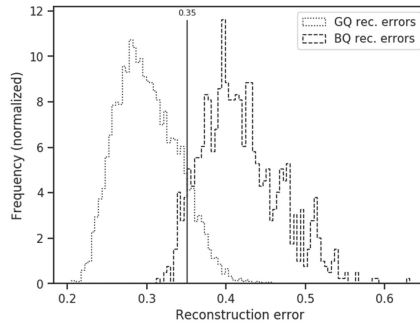


**Fig. 3.** Histogram of GQ vs BQ reconstruction error. Number of examples in BQ class is upscaled to fit the other class. Obtained for SpAE model with $d_{sqrt3}$ as reconstruction error measure. BQ threshold is calculated at the crossing point of both histograms, *i.e.* at $d_{sqrt3} = 0.35$. This setting gives classification accuracy $F1_{norm} = 91.62\%$.

## 5.4   Results

Experimental results are summarized in this subsection. Examples of GQ and BQ signals and their reconstructions are illustrated in Fig. 4.

Table 1 presents selected results, namely the mean reconstruction errors (average of the *Mean Square Error* (MSE) along the set, *i.e.* $\overline{d_{pow}}$) on the autoencoder testing set (unlabelled) and classification results on the labelled test set. For the sake of readability results of only two measures are listed: $d_{abs}$ and the one that gives the best classification results (for given architecture).

Due to imbalance in a number of GQ and BQ examples (the latter group is less than 15% of all available data), we used *normalized F1-score*, proposed in [42], to evaluate the classification. This accuracy measure gives a better insight in the anomaly detection problem, were anomalies are rare by definition, which
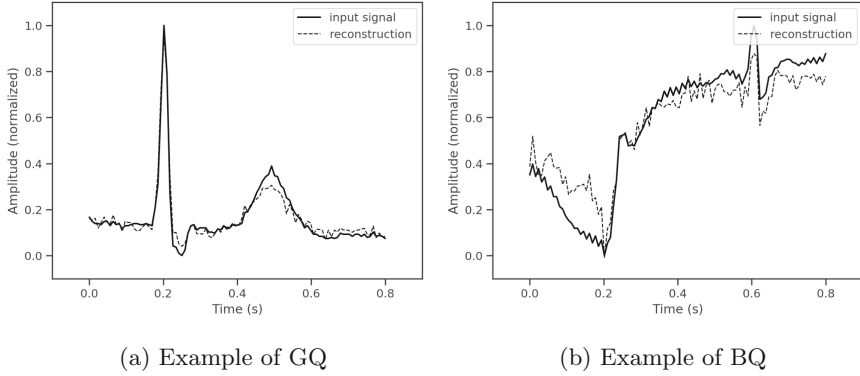
(a) Example of GQ                    (b) Example of BQ

**Fig. 4.** Examples of the GQ and BQ signals and their reconstructions. GQ: The QRS complex, as well as the T wave are quite accurately reproduced which effect in a small reconstruction error. BQ: Signal drop before 0.2 s is reconstructed as an ill-formed QRS complex which is significantly different than input signal, leading to high reconstruction error. Interestingly, the AE correctly reproduces the signal around 0.6 s which may be a true QRS complex.

heavily affects standard F1-score. Following the suggestions in [42], we present both F1 and normalized F1 scores.

The best classification result $F1_{norm} = 93.34\%$ was obtained by the simplest model, which was the plain AE without any constraints, with the latent layer wider than the input (150 neurons vs. 100). Such a model was expected to copy values from the input to output without generalizing them, making no difference in reconstruction error between examples of both classes. Indeed, its mean reconstruction error is very low (0.0025), but surprisingly the error distribution allows to distinguish BQ examples from GQ with a relatively high effectiveness. The best results were obtained using the $d_{abs}$ as reconstruction error measure.

Both Denoising AE (with Gausssian and Salt-and-Pepper noise) produced slightly higher reconstruction errors (0.0044 and 0.0104 respectively) and gave the classification results worse by almost 3%. The Sparsity AE's mean reconstruction error is also greater (0.0168), but model was better in classification task, achieving $F1_{norm} = 92.02\%$. Finally, the k-Sparse AE with only 5 non-zero values in 125-wide latent layer gave a significantly worse reconstruction (mean error is 0.4135) but almost the highest classification score of $F1_{norm} = 93.21\%$. In this model the most effective reconstruction error measure was the $d_{abs\_pf3}$, which incorporates the penalty for local fluctuations of $E_{abs}$.

To summarize, presented results prove the correctness of proposed approach. The best models utilize the unconstrained AE and the k-Sparsity AE, both reaching over 93% accuracy expressed in $F1_{norm}$ and $F1 \approx 73\%$.

**Table 1.** Comparison of selected reconstruction errors and classification results on the *test set* for different variants of AE model. Following abbreviations were used: *AE* – autoencoder, *lat. w.* – latent width, *BC* – baseline cancellation, *NP* – normalize R-peak polarization, *rec. err* – reconstruction error ($\overline{d_{pow}}$) on unlabelled evaluation set, *F1-norm* – normalized F1-score [42], *simAE* – simple AE (without any constraints), *DAE-Ga* – Denoising AE with Gaussian noise of intensity $\alpha$, *DAE-SP* – Denoising AE with Salt and Pepper noise of intensity $\alpha$, *SpAE* – Sparsity AE with sparsity rate $\rho$ and penalty weight $w$, *kSpAE* – k-Sparse AE with $k$ non-zero elements in latent vector and regularization L2 of the latent layer parametrized by $\lambda$.

| Model | | | | | AE rec. | Classification | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| AE | lat. w. | BC | NP | params | err | $d_{abs}$ | | best $d_*$ | | |
| | | | | | | $F1$ | $F1_{norm}$ | $d_*$ | $F1$ | $F1_{norm}$ |
| simAE | 150 | Yes | No | | 0.0025 | 0.7347 | 0.9334 | $d_{abs}$ | 0.7347 | 0.9334 |
| DAE-Ga | 100 | Yes | No | $\alpha = 0.1$ | 0.0044 | 0.6662 | 0.9059 | $d_{abs}$ | 0.6662 | 0.9059 |
| DAE-SP | 5 | No | No | $\alpha = 0.1$ | 0.0104 | 0.6596 | 0.9076 | $d_{abs}$ | 0.6596 | 0.9076 |
| SpAE | 150 | No | Yes | $\rho = 0.005$ $w = 0.25$ | 0.0168 | 0.6809 | 0.9202 | $d_{abs}$ | 0.6809 | 0.9202 |
| kSpAE | 125 | Yes | Yes | $k = 5$ $\lambda = 0.0005$ | 0.4135 | 0.6300 | 0.9049 | $d_{abs-pf3}$ | 0.7261 | 0.9321 |

## 5.5   Comparison to State of the Art

In contrary to 2011 PhysioNet/CinC Challenge, our solution operates on much shorter, and single-channel data (see Sect. 5.1). For this reason, a direct comparison of reported results is not possible, but we evaluated the challenge-winning algorithm [5] by building a similar Random Forest (RF) classifier using those features, which can be applied on our data, namely: frequency-domain (power spectral density in few bands and the relative power in QRS complex and in the baseline), time-domain (number of quantile crossings, percentage of flat signal, peak-to-peak amplitude) and statistical (skewness and kurtosis of the samples' distribution).

Labelled dataset was split identically as for the AE-based model (Subsect. 5.3) into validation (here: training) and test sets. After feature calculation, artificial BQ feature vectors were generated using the SMOTE algorithm [28], in order to balance the number of examples of both classes.

Finally, the RF classifier achieved $F1 = 0.3791$ and $F1_{norm} = 0.7231$, which is a significantly worse performance than results of proposed AE solution.

## 6   Conclusions and Further Work

We proposed a reliable ECG signal quality assessment algorithm, based on the anomaly detection approach, utilizing the autoencoder neural network.

Presented testing results confirm a high effectiveness of this approach. Its internal structure is rather simple, as no features nor rules has to be designed, so it can be easily applied to other kinds of signals. Thanks to unsupervised

autoencoder training, only a small amount of data has to be annotated. This property is a significant advantage over classic classifier-based approach.

Our further efforts will concern on the relation between autoencoder reconstruction error during the training on the unlabelled data and the final classification accuracy. Other autoencoder architectures are also in our scope of interest.

# References

1. Goldberger, A.L., et al.: Physiobank, physiotoolkit, and physionet. Circulation **101**(23), e215–e220 (2000)
2. Silva, I., Moody, G.B., Celi, L.: Improving the quality of ECGS collected using mobile phones: the PhysioNet/computing in cardiology challenge 2011. In: 2011 Computing in Cardiology, pp. 273–276, September 2011
3. Clifford, G.D., et al.: AF classification from a short single lead ECG recording: the PhysioNet/computing in cardiology challenge 2017. In: 2017 Computing in Cardiology (CinC), pp. 1–4 (2017)
4. Moody, G., Mark, R.: The impact of the MIT-BIH arrhythmia database. IEEE Eng. Med. Biol. Mag. **20**, 45–50 (2001)
5. Clifford, G.D., Behar, J., Li, Q., Rezek, I.: Signal quality indices and data fusion for determining clinical acceptability of electrocardiograms. Physiol. Meas. **33**(9), 1419 (2012)
6. Behar, J., Oster, J., Li, Q., Clifford, G.D.: ECG signal quality during arrhythmia and its application to false alarm reduction. IEEE Trans. Biomed. Eng. **60**(6), 1660–1666 (2013)
7. Liu, C., et al.: Signal quality assessment and lightweight QRS detection for wearable ECG SmartVest system. IEEE Internet Things J. **6**(2), 1363–1374 (2019)
8. Kuzilek, J., Huptych, M., Chudacek, V., Spilka, J., Lhotska, L.: Data driven approach to ECG signal quality assessment using multistep SVM classification. In: 2011 Computing in Cardiology, pp. 453–455, September 2011
9. Chudacek, V., Zach, L., Kuzilek, J., Spilka, J., Lhotska, L.: Simple scoring system for ECG quality assessment on android platform. In: 2011 Computing in Cardiology, pp. 449–451, September 2011
10. Hayn, D., Jammerbund, B., Schreier, G.: ECG quality assessment for patient empowerment in mHealth applications. In: 2011 Computing in Cardiology, pp. 353–356, September 2011
11. Liu, C., Li, P., Zhao, L., Liu, F., Wang, R.: Real-time signal quality assessment for ECGs collected using mobile phones. In: 2011 Computing in Cardiology, pp. 357–360, September 2011
12. Satija, U., Ramkumar, B., Manikandan, M.S.: A simple method for detection and classification of ECG noises for wearable ECG monitoring devices. In: 2015 2nd International Conference on Signal Processing and Integrated Networks (SPIN), pp. 164–169, February 2015

13. Satija, U., Ramkumar, B., Manikandan, M.S.: Automated ECG noise detection and classification system for unsupervised healthcare monitoring. IEEE J. Biomed. Health Inform. **22**(3), 722–732 (2018)
14. Satija, U., Ramkumar, B., Manikandan, M.S.: A new automated signal quality-aware ECG beat classification method for unsupervised ECG diagnosis environments. IEEE Sens. J. **19**(1), 277–286 (2019)
15. Hermawan, I.,et al.: Temporal feature and heuristics-based noise detection over classical machine learning for ECG signal quality assessment. In: 2019 International Workshop on Big Data and Information Security (IWBIS), pp. 1–8, October 2019
16. Redmond, S.J., Xie, Y., Chang, D., Basilakis, J., Lovell, N.H.: Electrocardiogram signal quality measures for unsupervised telehealth environments. Physiol. Meas. **33**(9), 1517 (2012)
17. Moeyersons, J., Testelmans, D., Buyse, B., Willems, R., Van Huffel, S., Varon, C.: Evaluation of a continuous ECG quality indicator based on the autocorrelation function. In: 2018 Computing in Cardiology Conference (CinC), vol. 45, pp. 1–4, September 2018
18. Martinez-Tabares, F.J., Espinosa-Oviedo, J., Castellanos-Dominguez, G.: Improvement of ecg signal quality measurement using correlation and diversity-based approaches. In: 2012 Annual International Conference of the IEEE Engineering in Medicine and Biology Society, pp. 4295–4298, August 2012
19. Morgado, E., et al.: Quality estimation of the electrocardiogram using cross-correlation among leads. BioMed. Eng. OnLine **14**, 59 (2015)
20. Naseri, H., Homaeinezhad, M.: Electrocardiogram signal quality assessment using an artificially reconstructed target lead. Comput. Meth. Biomech. Biomed. Eng. **18**(10), 1126–1141 (2015), pMID: 24460414
21. Shahriari, Y., Fidler, R., Pelter, M.M., Bai, Y., Villaroman, A., Hu, X.: Electrocardiogram signal quality assessment based on structural image similarity metric. IEEE Trans. Biomed. Eng. **65**(4), 745–753 (2018)
22. Shi, Y., et al.: Robust assessment of ECG signal quality for wearable devices. In: 2019 IEEE International Conference on Healthcare Informatics (ICHI), pp. 1–3, June 2019
23. Castiglioni, P., Meriggi, P., Faini, A., Rienzo, M.D.: Cepstral based approach for online quantification of ECG quality in freely moving subjects. In: 2011 Computing in Cardiology, pp. 625–628, September 2011
24. Zhou, X., Zhu, X., Nakamura, K., Mahito, N.: ECG quality assessment using 1D-convolutional neural network. In: 2018 14th IEEE International Conference on Signal Processing (ICSP), pp. 780–784, August 2018
25. Ghosal, P., Sarkar, D., Kundu, S., Roy, S., Sinha, A., Ganguli, S.: ECG beat quality assessment using self organizing map. In: 2017 4th International Conference on Opto-Electronics and Applied Optics (Optronix), pp. 1–5, November 2017
26. Karpinski, M., Khoma, V., Dudvkevych, V., Khoma, Y., Sabodashko, D.: Autoencoder neural networks for outlier correction in ECG- based biometric identification. In: 2018 IEEE 4th International Symposium on Wireless Systems within the International Conferences on Intelligent Data Acquisition and Advanced Computing Systems (IDAACS-SWS), pp. 210–215 (2018)
27. Bellinger, C., Drummond, C., Japkowicz, N.: Beyond the boundaries of SMOTE. In: Frasconi, P., Landwehr, N., Manco, G., Vreeken, J. (eds.) ECML PKDD 2016. LNCS (LNAI), vol. 9851, pp. 248–263. Springer, Cham (2016). https://doi.org/10.1007/978-3-319-46128-1_16
28. Chawla, N., Bowyer, K., Hall, L., Kegelmeyer, W.: SMOTE: synthetic minority over-sampling technique. J. Artif. Intell. Res. (JAIR) **16**, 321–357 (2002)

29. Sharma, S., Bellinger, C., Japkowicz, N., Berg, R., Ungar, K.: Anomaly detection in gamma ray spectra: a machine learning perspective. In: 2012 IEEE Symposium on Computational Intelligence for Security and Defence Applications, pp. 1–8, July 2012

30. Lecun, Y.: PhD thesis: Modeles connexionnistes de l'apprentissage (connectionist learning models). Universite P. et M. Curie (Paris 6) (6 1987)

31. Bourlard, H., Kamp, Y.: Auto-association by multilayer perceptrons and singular value decomposition. Biol. Cybern. **59**(4), 291–294 (1988)

32. Hinton, G.E., Zemel, R.S.: Autoencoders, minimum description length and Helmholtz free energy. In: Cowan, J.D., Tesauro, G., Alspector, J. (eds.) Advances in Neural Information Processing Systems, vol. 6, pp. 3–10. Morgan-Kaufmann (1994)

33. Goodfellow, I., Bengio, Y., Courville, A.: Deep Learning. MIT Press (2016). http://www.deeplearningbook.org

34. Vincent, P., Larochelle, H., Lajoie, I., Bengio, Y., Manzagol, P.A.: Stacked denoising autoencoders: learning useful representations in a deep network with a local denoising criterion. J. Mach. Learn. Res. **11**, 3371–3408 (2010)

35. Zhou, L., Yan, Y., Qin, X., Yuan, C., Que, D., Wang, L.: Deep learning-based classification of massive electrocardiography data. In: 2016 IEEE Advanced Information Management, Communicates, Electronic and Automation Control Conference (IMCEC), pp. 780–785 (2016)

36. Makhzani, A., Frey, B.J.: k-Sparse autoencoders. CoRR abs/1312.5663 (2013). http://arxiv.org/abs/1312.5663

37. An, J., Cho, S.: Variational autoencoder based anomaly detection using reconstruction probability. In: Special Lecture on IE (2015)

38. Chazal, P., Heneghan, C., Sheridan, E., Reilly, R., Nolan, P., O'Malley, M.: Automated processing of the single-lead electrocardiogram for the detection of obstructive sleep apnoea. IEEE Trans. Bio-med. Eng. **50**, 686–96 (2003)

39. Sahakian, A.V., Petrutiu, S., Swiryn, S.: Abrupt changes in fibrillatory wave characteristics at the termination of paroxysmal atrial fibrillation in humans. EP Europace **9**(7), 466–470 (2007)

40. Greenwald, S., Patil, R., Mark, R.: Improved detection and classification of arrhythmias in noise-corrupted electrocardiograms using contextual information (1990)

41. American Heart Association ECG Database. www.ecri.org/american-heart-association-ecg-database-usb/. Accessed 15 January 2019

42. Jeni, L., Cohn, J., De la Torre, F.: Facing imbalanced data - recommendations for the use of performance metrics. In: 2013 Humaine Association Conference on Affective Computing and Intelligent Interaction, vol. 2013, September 2013