Patient-individual Morphological Anomaly Detection in Multi-lead Electrocardiography Data Streams

Alexander Acker*, Florian Schmidt*, Anton Gulenko*, Reinhard Kietzmann† and Odej Kao*
*Technische Universität Berlin (TU Berlin, Complex and Distributed IT Systems (CIT)), 10587 Berlin, Germany
†AthenaDiaX GmbH, 16816 Neuruppin, Germany

***Abstract*—Cardiac diseases like myocardial infarction, which possibly result in cardiac death, are still a relevant topic. To achieve recognitions in early stages, long term ECG monitoring devices are used. Such devices produce large amounts of data, either directly streamed or stored in databases. Manually analysing this data by experts is inefficient. Thus, automated preprocessing methods are needed to minimize the temporal effort dedicated to the inspection. The proposed method helps to identify morphological anomalies within the ECG data stream. It determines a set of meaningful time series features based on a Kolmogorov-Smirnov test (KST) and after that, applies the BICO online clustering algorithm. Thereby, the system learns the patient-individual PQRST-complex segment morphologies and after that, uses the learned models for detecting anomalies within the ECG data stream. For evaluation, real world patient data was used, which was previously tagged by electrophysiologists. As a result, the KST selected set of features was revealed to be especially suitable for analysing ECG data streams, resulting in average sensitivity rates of 98.82% and average specificity rates of 98.13%.**

***Keywords*-ECG data streams; anomaly detection; clustering; machine learning; time series;**

I. Introduction

One of the most frequent cardiac diseases is the myocardial infarction. It is responsible for most cardiac deaths [1] and may proceed unnoticed by the affected persons in early stadia [2]. An automated early detection of indicating symptoms can help physicians to apply appropriate treatments before the condition becomes life threatening. To analyse this and other cardiac diseases, an essential part of today's technical equipment in medical environments are ECG recorders. Thereby, the electrophysiological impulses from the myocardium are measured using multiple electrodes (leads) attached to the skin surface. The resulting ECG signal consists of cyclic PQRST-complexes, which corresponds to the cardiac cycle [3]. The sampling of this bioelectric signal at different body locations produces a digital time series sequence, which is used by cardiologists or electrophysiologists to propose diagnoses and select treatments. Depending on the concrete cardiac disease, pathological ECG signal sections can be either arrhythmically or morphologically abnormal or both [3]. Taking the previous example of myocardial infarction, it unveils itself as a temporary morphological deviation within the ST-segments [3]. This temporariness makes it hard to be detected in sporadic applied ECG recordings. A more frequent monitoring raises the probability of detecting such irregularities.

The consequence of the intention to be able to detect such diseases in an early stadium is an ongoing trend towards none invasive compact sensors, which are placed on the body surface [4]. It allows the continuous monitoring of a broad coverage of patients. The measured ECG data are either directly transferred to centric databases (telemetry) or stored on internal memory devices for later transmission or analysis (long term recordings) [4].

The drawback of the concept of extensive sensor supply is the arise of a vast amount of medical data, which has to be processed. ECG signal data which are recorded, but not analysed due to a lack of time, are useless in practical applications. Therefore, automated ECG analysis systems are developed to save valuable expert working time. Often, these systems do not aim to replace the human expert, but give hints to segments in the ECG recordings, which could be of interest and need further revision. Regarding telemetry and big data applications, these analysis systems are required to work with streamed data. The challenge of identifying segments of interest can be regarded as an anomaly detection problem.

Several constraints must be considered developing such systems. First, a vast amount of sensor devices and recording methods are available resulting in differently measured signals. Second, patients may have different base ECG morphologies as they may suffer from different cardiac diseases. For such patients, these chronic influences within the ECG signal should not be considered as anomalies.

Based on the described challenges, we developed a system for a patient-individual detection of morphological anomalies in continuous ECG data streams. Other than most related systems we propose a methodology for a fine-grained localization of an occurring anomaly within the PQRST-complex. Our key contributions are:

- Design and implementation of a system concept for analysing continuous multi-lead ECG data streams,
- developing an approach for automatically selecting an optimal time series feature set for ECG data,
- proposition of an methodology for morphological anomaly detection and
- evaluating the presented method to real world patient

978-1-5386-2715-0/17/$31.00 ©2017 IEEE

data.

For this propose, we identified a segment-specific representative and meaningful set of time series features. Based on these, the system performs a semi-supervised learning phase, after which it is able to identify abnormal signal segments.

Outline. The remainder of the paper is structured as follows. In section II we provide background and related work on systems used for ECG data analytic. Section III presents our idea of detecting morphological anomalies in ECG data streams based on identified meaningful time series feature sets. Section IV summarizes the evaluation of our system. Section V concludes this paper.

II. Related Work

The idea of automated ECG signal analysis is a widely respected field of research. Starting from the automated detection of R peaks, determining RR-intervals and based on that, performing arrhythmical anomaly analysis were the first steps towards reducing the workload of cardiologist and electrophysiologist experts [5]. For the more challanging morphological anomaly detection, batch data analysis methods exist based on different algorithmic approaches [6], [7]. To meet the requirements of modern monitoring systems, the recent research focus moved towards online stream data analysis, enabling the processing of continuous and/or large data quantities.

One proposed approach to detect morphologically abnormal ECG segments is to specifically model fixed signal patterns corresponding to a certain anomaly type. After that, these patterns shoulde be identified within continuous ECG data streams. Wagner [8] defined ventricular extrasystoles and ventricular tachycardias as the phenomenons of interest for his system. Ying's system [9] looks for the occurrence of atrial fibrillations. Although, the proposed systems were able to detect the specified ECG sections, they lack flexibility due to their disability to detect other morphological deviations. Furthermore, the system is provided with a fixed model for abnormal morphologies, making it impossible to self-adjust to patient-individual characteristics.

A more flexible approach was developed by Kiranyaz et al. [10]. It is based on a convolutional neural network and consists of a short offline supervised learning phase, where expert's annotated ECG data are used for learning. After that, the system is able to detect and annotate an incoming ECG data stream accordingly. Although reducing the workload of an attending expert, the system still requires a fully supervised initial learning phase, which cannot be applied on streamed data. Furthermore, its main advantage of being able to perform a specific medical classification of PQRST-complexes, is also a major disadvantage. The set of learning data must contain all possible classes for the later classification, which cannot be assumed in a real world environment.

The approach of Ngo and Veeravalli [11] is a further step towards online ECG data stream analysis. Their system implements a continuous unsupervised learning method and therefore is able to adapt to concept drifts in the ECG signal. They use the heuristic that normal signal segments occur more often than anomalies. Thus, a clustering on small data set of 30 seconds is applied, identifying the dominant morphology type. During the anomaly detection, the system checks for significant deviations from the current dominating morphology and marks those sections as anomalies. A drawback of their system is the fact that it is able to learn only one normal morphology type. Especially for patients chronically suffering from e.g. intermitted bundle branch block this approach will produce many false positive detections. Furthermore, they aim to apply their system on only one ECG lead, abandoning possible advantages of multi-lead recording systems.

All of the currently published online data analysing system proposals provide an anomaly detection on whole PQRST-complex sections. None of them went into finer granularity and cannot pin occurring anomalies to specific PQRST-segments. Furthermore, most of the analysed systems apply their analysis on only one ECG signal lead.

III. Proposed Methodology

The proposed system takes an L-lead ECG signal as input, where L is a none-empty finite set of size n containing all lead types of a certain recoding. For the analysis, the signals are defined as a continuous n-dimensional time series S:

$$S = \begin{bmatrix} s_t^{L[i]} & s_{t+1}^{L[i]} & s_{t+2}^{L[i]} & \dots \\ s_t^{L[i+1]} & s_{t+1}^{L[i+1]} & s_{t+2}^{L[i+1]} & \dots \\ s_t^{L[i+2]} & s_{t+1}^{L[i+2]} & s_{t+2}^{L[i+2]} & \dots \\ \vdots & \vdots & \vdots & \\ s_t^{L[n]} & s_{t+1}^{L[n]} & s_{t+2}^{L[n]} & \dots \end{bmatrix}, \qquad (1)$$

where each $s_t^{L[i]}$ describes signal value samples at time t from the lead type $L[i]$ with $1 \leq i \leq n$. Based on this, the system performs the analysing steps as depicted in fig. 1. First, several preprocessing steps are applied, which are explained in more detail in section III-A. This results in time series segments, that are used for the feature calculation module. It generates a set of representative features and passes the results to either the feature engineering or anomaly detection module. Former is used to select a meaningful subset of time series features. This is not part of the stream data processing and is executed only once based on a small set of ECG data and hereinafter used for the anomaly detection. The approach is described in section III-B in more detail. After that, the feature calculation module generates the previously identified meaningful features, which are used as input for the anomaly detection module. There, the

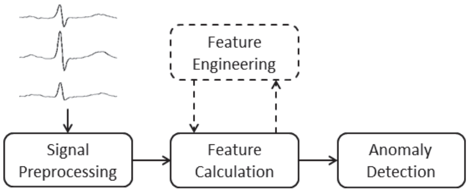


Figure 1. Main processing steps of the system.

classification is performed, producing a binary target tagging for each ECG signal segment (see section III-C).

A. Signal Preprocessing

At first, a modified R-peak detection algorithm proposed by Engelse [12] is used to identify each PQRST-complex. As an addition, we previously apply a digital first order 0.5 Hz high pass filter and a third order 35 Hz low pass filter in order to reduce baseline wanderings and noise. After that, as proposed by Engelse [12], the approximative second derivative of each ECG lead signal is calculated based on eq. (2), where h denotes which prior and subsequent time series value should be used to calculate the current derivative value. We set h to 1.

$$f''(x) \approx \frac{f(x+h) - 2f(x) + f(x-h)}{h^2} \tag{2}$$

Next, we adjusted the proposed simple threshold method, resulting in a lookahead adaptive threshold method, which is applied on the signal's second derivatives to mark the R-peaks. After that, a checking for synchronicity of the marker locations between each signal lead was added. Only marks occurring in at least two lead signals with a time jitter lower than 8 ms are accepted as valid R-peaks. Finally, these markers are used to extract the surrounding PQRST signal area and split each one into a P-, QRS- and T-segment.

B. Feature Engineering

The preliminary task of the proposed methodology is to determine a set of representative and meaningful features respectively for each ECG signal segment, which should be used for the anomaly detection. Representative means that the feature set is calculated based on the sampled signal values and the size of the feature set is smaller than the number of samples in the signal segment. For a given set of features $X_1, \ldots, X_i, \ldots, X_n$, meaningfulness is defined by the terms proposed by Christ et al. [13] and Radivojac et al. [14]. Based on a binary target set Y it is stated that the meaningfulness for a classification process based on a metric X_i can be measured by the difference between the conditional density functions $f_{X_i|Y=y_0}$ and $f_{X_i|Y=y_1}$, which respectively represent the statistical distribution of X_i under a given target Y. Considering the anomaly detection problem of the ECG signal, y_1 represents "abnormal" and y_0 "individually normal" segments. This consideration is implemented into a statistical test based on the following hypothesis:

$$\begin{aligned} H_0^i &= \{f_{X_i|Y=y_1} = f_{X_i|Y=y_2}\}, \\ H_1^i &= \{f_{X_i|Y=y_1} \neq f_{X_i|Y=y_2}\}. \end{aligned} \tag{3}$$

Therefore, the continuous feature binary target Kolmogorov-Smirnov test (KST) was applied to compare the two empirical distribution functions. This results in a p-value, which measures the probability that a feature X_i is meaningful for predicting Y. Based on the proposed hypothesis in eq. (3), small p-values reveal meaningful features for the classification process. This test was applied on a pool of 85 features, out of which the ones with the lowest p-value were selected. The time series feature calculations were adapted from the work of Christ et al. [13]. The results of the selection are listed in table II.

The feature calculation is based on two time series. The first consists of the sample value set S representing the current ECG segment. For the second, a sliding window containing the previous w segments is used to calculate a sample-wise arithmetic average signal segment $\overline{S}^w$. Based on the described feature selection approach and time series input, the following features were selected for the subsequent anomaly detection task, whereby respectively annotated entries are adopted from [13]:

1. Arithmetic mean: $mean(S)$ [13],
2. Standard deviation: $std(S)$ [13],
3. Skewness: $skewness(S)$ [13],
4. Minimal sample value: $minimum(S)$ [13],
5. Below-above mean ratio: $\frac{dataPointsAboveMean(S)}{dataPointsBelowMean(S)}$ [13],
6. First relative location of the maximal value: $first\_index\_max(S)$ [13],
7. Slope sum of the signal sample values in S, defined as the sum of all differences of adjacent values $\sum_{i=0}^{n-2} s_{i+1} - s_i$.
8. Absolute slope sum of the signal sample values in S. This calculation is based on the same principle as 7., but differs at two aspects. It calculates the absolute slope value between two sample values and the distance between the values can be defined by a parameter d through $\sum_{i \in I} \sqrt{(s_{i+d} - s_i)^2}$, where $I = [0, d, 2*d, 3*d, \ldots, m]$ is the set of relevant indices and the last index m is calculated by $m = n - (n \; mod \; d) - d$. n denotes the size of S and mod is the modulo operation.
9. The sample-wise average absolute deviation between S and $\overline{S}^w$. Therefore, the element-wise absolute difference between each sample value in S and $\overline{S}^w$ is calculated. After that, the arithmetic average over all absolute differences is computed $\frac{1}{n} * \sum_{i=0}^{n-1} \sqrt{(s_i - \overline{s}_i^w)^2}$.

Based on these feature selection results, the system was configured to perform its stream based anomaly detection. Therefore, the feature generation module is responsible for the feature calculation and normalization. A min-max scaling was used to normalize the feature values.

Table I
FEATURE SELECTION RESPECTIVELY FOR EACH ECG SIGNAL SEGMENT BASED ON p-VALUES CALCULATED BY A KST.

Seg.	Features	p-values
P	first relative max. location	0.079
	below-above mean ratio	0.108
	skewness	0.201
	arithmetic mean	0.228
	slope sum	0.245
QRS	sample-wise average absolute deviation ($w = 16$)	0.192
	minimal value	0.267
	below-above mean ratio	0.415
	arithmetic mean	0.442
T	sample-wise average absolute deviation ($w = 22$)	0.098
	standard deviation	0.388
	absolute slope sum ($d = 5$)	0.403

C. Anomaly Detection

The incoming input data can be formally described by vectors of normalized feature values arriving at a certain time t: $V_{l,j}(t) = (x_1, x_2, \ldots, x_n)$, where l is the ECG signal lead, j is the PQRST-complex segment and n is the total number of features. Each of these vectors are used as an input for the anomaly detection process, which includes two operation phases. During the first phase the system creates a representative model for all ECG segment morphologies, that are not considered as anomalies. This is done respectively for each ECG lead and segment type. Thus, respective models $M_{l,j}$ are created, representing all individually normal morphologies. Therefore, the system can be considered to learn semi-supervised as it uses a seed of expert knowledge to determine the normal signal segment morphologies but after that, is able to detect anomalies on its own.

For the creation of these models, a clustering algorithm for continuous data streams was applied. For our method we used the stream clustering algorithm BICO proposed by Fichtenberger et al. [15]. It adapts the concept of BIRCH's CF (clustering feature) trees [16] and provides it with a theoretical concept of coresets. Like the well known k-means clustering, BICO computes centroids with an additional radius, resulting in spherical clusters. Based on two dimensional data points, the concept of this approach is depicted in fig. 2. The black dots were used to create the sphere clusters during the learning phase. Based on this, incoming data samples are checked to be assigned to one of the clusters. Points outside of any sphere are considered as anomalies.

We used the BICO algorithm implementation of Bifet et al. [17], which is able to operate in two processing modes for incoming feature vectors $V_{l,j}(t)$. A vector can either be used to built the tree or the algorithm can check whether it lies within the radius of a cluster. Former mode is used during the learning phase. Thereby, the clusters for all individually normal ECG segment morphologies are created. Note that this can include several morphology types. After the initial learning phase, the system switches toward the anomaly detection operation mode. Each incoming feature vector $V_{l,j}(t)$ is compared to the corresponding model $M_{l,j}$. If a cluster exists, which can absorb the vector, the corresponding ECG signal segment is considered not to be an anomaly. Otherwise, it is marked as abnormal. After that, the segments of the same type belonging to the same PQRST-complex are compared. If at least one of them is marked as abnormal, the whole segment set is considered to be an anomaly. The concept is depicted in fig. 3. This approach allows the detection of abnormal segment morphologies occurring only in certain ECG lead signals.

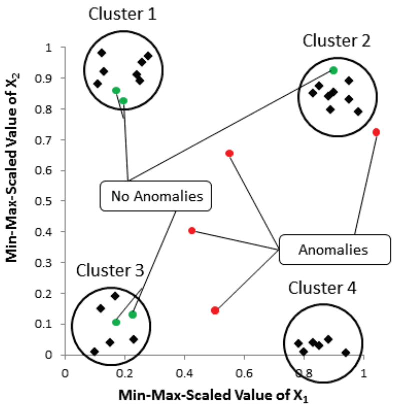


Figure 2. Clustering-based anomaly detection for a two dimensional feature vector $(X_1, X_2)^T$. Black dots represent data values, which were used to learn the model. Red dots are considered to be anomaly data. Green dots are considered not to be anomaly data.

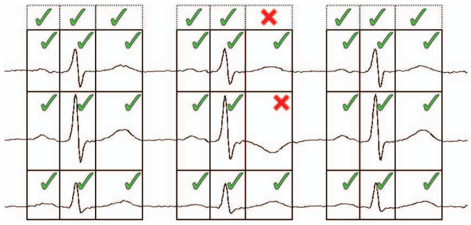


Figure 3. A snippet of a three channel ECG recording, whereupon the system performed its anomaly detection. Non-anomaly PQRST-segments are marked with a green tick, anomalies are annotated by a red cross. The image illustrates the concept of how the merging of segments from different channels is performed.

IV. EVALUATION

This section is dedicated to the evaluation of the proposed methodology from section III with respect to the requirements described in section I. A set of 50 real world patient ECG recordings were used for the evaluation. Each was previously annotated by expert electrophysiologists to determine the system performance. The recording time varies between 4 and 23 hours. The used ECG recorders measure a three channel Einthoven limb lead signal with a sample frequency of 250 Hz. During the expert annotation process, the patient-individual normal morphologies were defined respectively for each recording. After that, each abnormal PQRST-complex, considerably deviating from the defined normal morphologies, was marked. Additionally, the segments (P, QRS and T) which were affected by the abnormal deviation were tagged.

For the learning phase of the anomaly detection system, the first, not mandatory cohesive, 6 minutes of each record without anomalies were used, whereby abnormal marked segments were filtered out. The rest of each data set was used as test data during the detection phase. The aim of the evaluation procedure was to determine the impact of the feature selection on the system's detection precision. Therefore, the following three feature sets were compared in terms of their sensitivity and specificity, respectively for each ECG segment type:

- Random features,
- raw features and
- selected features.

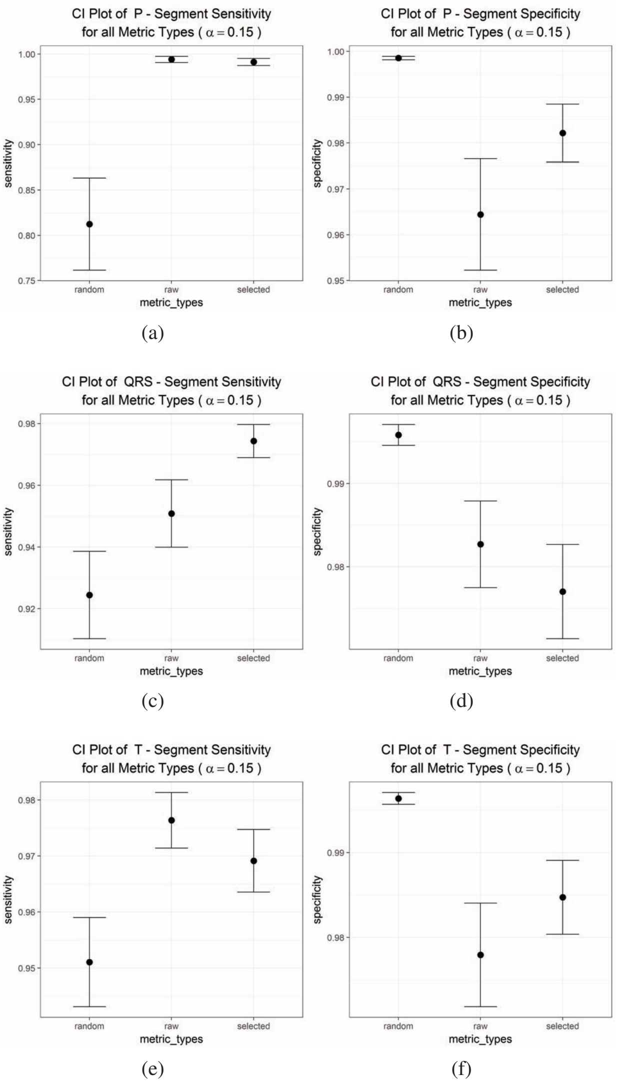
The features in the *random* set were randomly selected and are the same for all segment types. For the *raw* set, no feature generation is performed. Therefore, the anomaly detection was executed based on the raw sampled signal values. The *selected* feature set contains the features identified in section III-B. The comparison was made based on the sensitivity (se) and specificity (sp) measure defined by:

Table II
STATISTICAL VALUES USED FOR THE EVALUATION.

	Annotated Abnormal	**Annotated Normal**
Detected Abnormal	TP	FP
Detected Normal	FN	TN

$$se = \frac{TP}{TP + FN}; \quad sp = \frac{TN}{TN + FP}. \tag{4}$$

The former measures the quality of correctly detected anomalies, which indicates the risk of missing abnormal segments. The latter defines the risk of false alarms, whereby a high value poses a low false alarm probability. Both measures were calculated respectively for each of the 50 ECG measurement analyses using the different metric types. Out of these, an arithmetic mean sensitivity and specificity was calculated, which made it possible to compare the different metric configurations of the system. Those results, together with a confidence interval of $\alpha = 0.15$, are plotted in fig. 4. To determine whether the difference between the metric types was significant, an unpaired t-test was applied. A preceding Levene's test was used to determine the homogeneity of variance across every segment's anomaly and not anomaly group. It consistently revealed a differing variance. Therefore, the unpaired t-test for unequal variance with a confidence level of $\alpha = 0.15$ was applied.



(a) (b)

(c) (d)

(e) (f)

Figure 4. CI plots of the sensitivities and specificities with confidence intervals of $\alpha = 0.15$ in dependence of the used metric type for the (a and b) P-, (c and d) QRS- and (e and f) T-segments.

The t-test revealed that for all segment types the sensitivity of the raw and selected features was significantly higher than the value of the randomly picked feature set. There is no significant difference between the P-segment sensitivities of raw (0.994) and selected (0.991) feature types or respectively T-segment sensitivities of raw (0.976) and selected (0.969) feature types. For QRS-segments the sensitivity measure of the selected features set (0.974) was significantly higher than the value for the raw data values (0.951).

Regarding the specificity, the randomly picked feature set was performing significantly better than both other feature types for every segment. The difference between the QRS-segment specificities of raw data values (0.983) and selected feature types (0.977) was not significant. Respectively, no significance was revealed between the specificity results of raw data values (0.983) and selected feature types (0.977) for the T-segments. For the P-segments, the specificities of the selected feature set (0.982) performed significantly better than the raw data values (0.964). Both the sensitivity as well as the specificity measures lie at a constant high level of above 0.96.

A. Discussion

The main priority of medical data analysis systems is to strictly prevent missing health risking phenomenons. Therefore, the sensitivity of medical analysis systems usually has a higher priority than the specificity [4]. On the other hand a low specificity results in a high rate of false alarms, which lowers the acceptance of medical personal when they operate with it [4].

Based on this assumptions and the previous results, we compare the results of the selected features and the raw data values. Based on the sensitivity, the QRS-segments should be analysed based on the selected feature type. For the P-segments the specificity measure provides an argument to use the selected feature type over the raw data values. For the T-segments, no preference of features could be determined.

V. Conclusion

The proposed work introduced a novel approach for detecting abnormal signal segment morphologies in ECG data streams based on different representative time series features. By splitting each PQRST-complex into three signal segments (P-, QRS- and T-segment) it is possible to concretely localize occurring anomalies. During a preceding feature analysis, a set of representative and meaningful features for each segment type was identified based on their distribution function using the KST. These features were used during the anomaly detection to identify outlier respectively in every ECG signal lead. Therefore, the BICO clustering algorithms for data streams was applied. After an initial semi-supervised learning phase, during which the system learned all relevant morphologies respectively for every signal lead and segment type, it switches into the anomaly detection mode, whereby all morphologies, which do not fit into the learned segment models are marked as anomalies. Therefore, the system does not need to put any effort into learning specific anomaly types.

During the evaluation, the impact of choosing a meaningful set of representative time series features was shown. Based on long term ECG signal data, that were thoroughly annotated by electrophysiologists, the system was tested under different feature set configurations. The results confirmed the assumption that a automated feature selections based on the KST have an obvious positive impact on the anomaly detection quality.

References

[1] G. M. WRITING, D. Lloyd-Jones, R. Adams, T. Brown, M. Carnethon, S. Dai, G. De Simone, T. Ferguson, E. Ford, K. Furie *et al.*, "Heart disease and stroke statistics–2010 update: a report from the american heart association." *Circulation*, vol. 121, no. 7, p. e46, 2010.

[2] G. S. Marschall S. Runge and C. Patterson, Eds., *Netter's Cardiology*. Elsevier Health Sciences, 2010.

[3] J. Malmivuo and R. Plonsey, *Bioelectromagnetism - Principles and Applications of Bioelectric and Biomagnetic Fields*. Oxford University Press, 1995.

[4] R. Choudhury, *Telemetry Monitoring: Validity*. VDM, 2009.

[5] D. A. Coast, R. M. Stern, G. G. Cano, and S. A. Briller, "An approach to cardiac arrhythmia analysis using hidden markov models," *IEEE Transactions on biomedical Engineering*, vol. 37, no. 9, pp. 826–836, 1990.

[6] S. Osowski and T. H. Linh, "Fuzzy clustering neural network for classification of ecg beats," in *Neural Networks, 2000. IJCNN 2000, Proceedings of the IEEE-INNS-ENNS International Joint Conference on*, vol. 5. IEEE, 2000, pp. 26–30.

[7] M. Lagerholm, C. Peterson, G. Braccini, L. Edenbrandt, and L. Sornmo, "Clustering ecg complexes using hermite functions and self-organizing maps," *IEEE Transactions on Biomedical Engineering*, vol. 47, no. 7, pp. 838–848, 2000.

[8] G. v. Wagner, "Entwicklung von methoden zur echtzeitanalyse von ekg-signalen mit neuro-fuzzy-systemen fur anwendungsszenarien der telemedizin," Ph.D. dissertation, Karlsruhe, Univ., Diss., 2006, 2006.

[9] H. Ying, "Atrial fibrillation detection," Master's thesis, Czech Technical University in Prague, 2009.

[10] S. Kiranyaz, T. Ince, and M. Gabbouj, "Real-time patient-specific ecg classification by 1-d convolutional neural networks," *IEEE Transactions on Biomedical Engineering*, vol. 63, no. 3, pp. 664–675, 2016.

[11] D. Ngo and B. Veeravalli, "Design of a real-time morphology-based anomaly detection method from ecg streams," in *Bioinformatics and Biomedicine (BIBM), 2015 IEEE International Conference on*. IEEE, 2015, pp. 829–836.

[12] W. A. H. Engelse and C. Zeelenberg, "A single scan algorithm for qrs-detection and feature extraction," *Computers in cardiology*, vol. 6, no. 1979, pp. 37–42, 1979.

[13] M. Christ, A. W. Kempa-Liehr, and M. Feindt, "Distributed and parallel time series feature extraction for industrial big data applications," *arXiv preprint arXiv:1610.07717*, 2016.

[14] P. Radivojac, Z. Obradovic, A. K. Dunker, and S. Vucetic, "Feature selection filters based on the permutation test," in *European conference on machine learning*. Springer, 2004, pp. 334–346.

[15] H. Fichtenberger, M. Gillé, M. Schmidt, C. Schwiegelshohn, and C. Sohler, "Bico: Birch meets coresets for k-means clustering," in *European Symposium on Algorithms*. Springer, 2013, pp. 481–492.

[16] T. Zhang, R. Ramakrishnan, and M. Livny, "Birch: an efficient data clustering method for very large databases," in *ACM Sigmod Record*, vol. 25, no. 2. ACM, 1996, pp. 103–114.

[17] A. Bifet, G. Holmes, R. Kirkby, and B. Pfahringer, "Moa: Massive online analysis," *Journal of Machine Learning Research*, vol. 11, no. May, pp. 1601–1604, 2010.