Automatic Quality Assessment of Capacitively-Coupled Bioimpedance Signals for Respiratory Activity Monitoring

Adnan Albaba [a, *], Ivan Castro [a], Pascal Borzée [b], Bertien Buyse [b], Dries Testelmans [b], Carolina Varon [c, d, e], Sabine Van Huffel [c, d], Tom Torfs [a]

[a] Connected Health Solutions Group at imec-Leuven, Belgium
[b] Department of Pneumology, Leuven University Centre for Sleep and Wake Disorders, UZ Leuven, Belgium
[c] Department of Electrical Engineering – ESAT, STADIUS, University of Leuven, Belgium
[d] Leuven.AI-KU Leuven Institute for AI, Belgium
[e] E-media Research Lab, Campus Groep T, KU Leuven, Belgium
[*] Corresponding author (e-mail: Adnan.Albaba@imec.be)

***Abstract* —** The objective of this work is to design an algorithm capable of classifying segments of capacitively-coupled bioimpedance (ccBioZ) for respiratory activity monitoring based on their signal quality. Such algorithm is an important building block as a pre-processing step to increase the confidence of extracted information such as respiration rate (RR). Long over-night ccBioZ recordings acquired from 12 subjects, are used for training and testing the proposed algorithm. To create a ground-truth labelling, the annotation is done manually by an experienced biomedical engineer. A total number of 52 features are extracted to capture information related to the quality of the ccBioZ segments. Five subsets of features are selected based on five different feature selection methods and tested against a full set of features to find the best trade-off between the performance of the classifier and the number of features. For classification, 19 classifiers are trained, cross-validated, and tested in three different datasets, acquired from 12 subjects: DS1 (training and validation data from 11 patients with suspected sleep apnea), DS2 (containing apneic epochs acquired from the same 11 patients), and DS3 (testing data acquired from one healthy subject). The balanced accuracy is used along with other statistical evaluation metrics.
For each test set, the best results of the quantitative evaluation came as following; DS1: Accuracy=0.91, Sensitivity=0.90, Specificity=0.95, and Balanced Accuracy=0.91. DS2: Accuracy=0.87, Sensitivity=0.88, Specificity=0.87, and Balanced Accuracy=0.88. DS3: Accuracy=0.91, Sensitivity=0.98, Specificity=0.91, and Balanced Accuracy=0.94. The results of the testing phases prove the reliability and robustness of the presented approach.

Keywords: electrical bioimpedance, BioZ, respiratory activity, SQI, signal quality, capacitive bioimpedance, quality estimation, classification, feature selection

I. Introduction

Vital sign monitoring helps the evaluation of important information related to the body's functions. Along with blood pressure, body temperature and heart rate, breathing rate is one of the primary vital signs, and it can be used for predicting patient deterioration conditions [1]. Several methods like air flow, tissue composition, and chest displacement [2] are used for extracting respiration-related information. However, impedance pneumography, which obtains the respiratory activity and extracts the respiration rate (RR) by obtaining the bioimpedance of the chest, can be more suitable for unobtrusive and long-term monitoring applications [3].
Bioimpedance deals with the ability of a living tissue to impede electrical current flow [4]. Bioimpedance applications include electrical impedance tomography [5], body composition analysis [6], cardiac output monitoring [7], and respiration activity monitoring [8].

Capacitively coupled acquisition of biopotential measurements has made promising progress in the last decade, making this method a possible solution towards unobtrusive monitoring of electrophysiological signals [9]. In this context, capacitively-coupled bioimpedance (ccBioZ), has been reported for unobtrusive monitoring of vital signs [10–14]. The BioZ readout can be de-modulated to two output components, which provide information about the resistive part (i.e. I channel) and the reactive part (i.e. Q channel) of the impedance [4]. I and Q signals are in quadrature, i.e. I (0^o) and Q (90^o) [4]. Conceptually, the information in the resistive part of the measured body impedance (i.e. given by the I channel) is sufficient for extraction of respiratory activity in impedance pneumography measurements. However, ccBioZ experiments [15] have shown that the Q channel may also contain respiratory information, because of possible phase-shift occurrences. Moreover, the lack of clear understanding of the origin of the respiratory impedance signal [16] makes it difficult to determine which channel (i.e. I or Q) contains most of the respiratory activity information. Figure 1 depicts three different ccBioZ readout cases, in which the respiratory activity happens to be reflected by either one or both I/Q channels.
In addition, motion artifacts (MA), due to free-moving conditions, affect the performance of any later processing or data mining steps of the input signals (e.g. sleep apnea detection and RR estimation). It is worth mentioning that only few researchers reported ccBioZ measurements [10–14,17], from which even fewer measurements were done in real-life scenarios [17]. Moreover, ccBioZ signals can be contaminated with baseline wandering (since there is no galvanic body biasing) or signal loss/saturation. Figure 2 depicts the effects of different sources of noise on the acquired ccBioZ signals.
The signal noise problem can be addressed using three different approaches [18]: 1) Apply signal processing techniques to denoise the signal. 2) Detect noise episodes in the signals and identify the contaminated segments as low-quality segments. 3) Combine both methods (i.e. noise detection and denoising). This work focuses on the second approach, which, besides off-line applications for removing low-quality segments from the signals, can also be used in the future for real-time processing algorithms, such as electrode selection, and low-level postprocessing steps, such as I/Q channel selection, data quality verification and data mining.

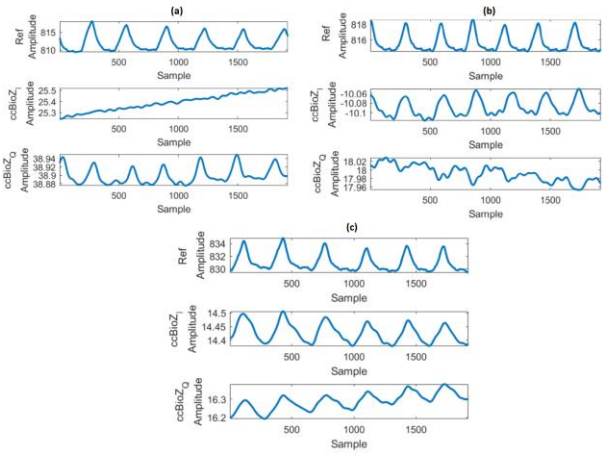


Figure 1: An example of ccBioZ readout signals aligned with the corresponding reference respiration signals. Respiration signal reflected in (a) Q channel, (b) I channel, and (c) both I and Q channels.

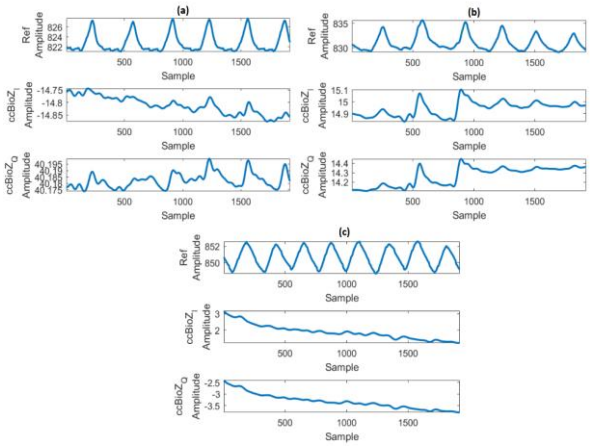


Figure 2: The effects of different types of noise on the ccBioZ signal. A ccBioZ segment contaminated with (a) Motion artifacts, (b) baseline changes, and (c) signal loss caused by low electrode coupling.

To our knowledge, no previous work was done on signal quality assessment for ccBioZ. This is mainly because this technology (capacitively-coupled bioimpedance acquisition) is considered relatively new and neither the hardware nor the ccBioZ signals are available as open-source access data. However, algorithms for evaluating the quality of sinusoidal-like (pulsatile) waveforms, e.g. photoplethysmography (PPG) and electrocardiography derived respiration (EDR), with high periodicity and dominant frequency are reported in the literature. Nemati S. et al. [19] used Hjorth descriptors as indicators of respiratory signals' quality to modify the noise covariance matrix of a Kalman filter and improve RR estimation. Sukor et al. [20] evaluated the quality of PPG signals using waveform morphology analysis approach. Li Q. et al. [21] used direct matching, linear resampling, dynamic time warping, and clipping detection along with a multi-layer perceptron neural network for assessing the utility of PPG signals. Orphanidou C. et al. [22] used threshold-based peak detection and adaptive template matching scheme to assess the quality of ambulatory PPG signals. Birrenkott D. A. et al. [23] used Fast Fourier Transform (FFT), autoregression, autocorrelation, and the Hjorth parameters as respiratory quality indices. Jorge J. et al. [24] used the Hjorth descriptors along with pre-tuned thresholds to assess the accuracy of RR estimates acquired from premature infants using chest impedance pneumogram. Naeini E. K. et al. [25] used conventional neural networks (CNN) to classify the quality of PPG signals in real-time. Liu S. H. et al. exploited fuzzy neural network [26] and deep CNNs [27] to improve the stroke volume measurements by assessing the quality of PPG signals.

The rest of this paper is structured as follows: in Section II, we present the research pipeline, which led to the proposed approach and summarize its main steps. The results are then presented and discussed in Section III and Section IV respectively. Conclusions are drawn in Section V. A table of acronyms can be found in Section VII.

II. Methods

A. Datasets and Annotations

Three different datasets were used in this work:

1) DS1: consists of 7000 segments of 30-second long ccBioZ from overnight recordings acquired from 11 patients with suspected sleep apnea who underwent simultaneous polysomnography at the sleep laboratory of the University Hospitals Leuven. Sleep apnea events were not included in this dataset.
2) DS2: consists of 6758 segments of 30-second long ccBioZ from overnight recordings which only contain sleep apnea events. These recordings were acquired from the same 11 patients, but not included in DS1. The annotation of the apneic events was done by clinicians based on the corresponding polysomnography recordings.
3) DS3: 13394 segments of 30-second long ccBioZ from 10 overnight recordings acquired from a healthy volunteer.

All recordings were acquired using the system presented in [10]. A reference respiration signal was simultaneously acquired to be used for annotation, see Table 1.

Table 1: ccBioZ datasets: acquisition setups and annotations.

Datasets	Fs (Hz)	Reference signals	Number of segments (considering I and Q channels as separate inputs)	Labels
DS1	64	Thoracic belt (standard polysomnography device)	7000	964 Good/ 2543 Acceptable/ 3439 Bad
DS2	64	Thoracic belt (standard polysomnography device)	6758	250 Good/ 754 Acceptable/ 5754 Bad
DS3	32	NeXus-10 MKII® from Mind Media, Herten, The Netherlands	9718	383 Good/ 1219 Acceptable/ 8116 Bad
	64		3676	596 Good/ 552 Acceptable/ 2528 Bad

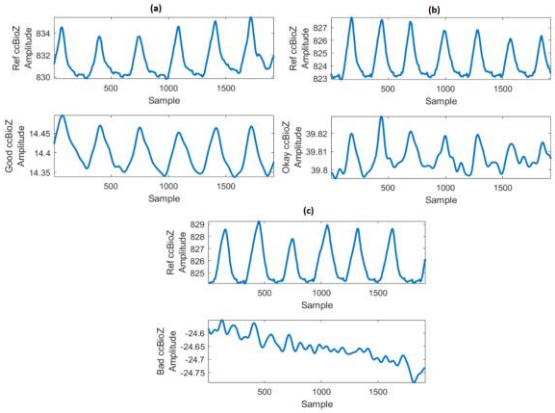


Figure 3: Example of ccBioZ segments, labelled as (a) Good, (b) Acceptable, and (c) Bad quality segments compared to their corresponding reference signals.

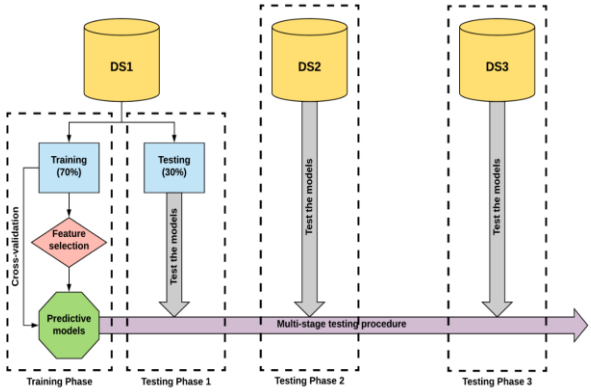


Figure 4: High-level pipeline of the presented work.

To annotate the ccBioZ segments in the three datasets, an experienced biomedical engineer labelled each segment in terms of signal quality to three different labels:

1) Good quality: The ccBioZ signal is in-phase with the reference respiration signal and contains prominent peaks which are similar in position, number, and relative amplitude to those in the reference signal.
2) Acceptable quality: The ccBioZ waveform follows the pattern of the corresponding reference signal, but with added noise oscillations. However, respiration cycles should be identifiable, and a maximum of one missed cycle is allowed.
3) Bad quality: The ccBioZ waveform does not follow the pattern of the corresponding reference signal, or more than one respiration cycle may be missed by an automatic detector.

Figure 3 depicts examples of ccBioZ segments with Good, Acceptable, and Bad qualities along with their corresponding reference signals.

It is worth mentioning that the data used in this work was acquired with respect to the guidelines and regulations for experiments in human subjects. All subjects, who volunteered in the experiments within this work, signed an informed consent. To enable a low-threshold (LT) binary classification evaluation scheme, all three-level labelling sets (good/acceptable/bad quality) were transformed to a binary labelling set (good/bad quality), which was obtained by considering segments with either good or acceptable quality labels to be labelled as good quality segments (detectable respiration cycles), while those of bad quality labels remain the same (undetectable respiration cycles).

B. Feature Space and Extraction:

Before extracting the features, the ccBioZ recordings were pre-processed as follows:

1) Segmentation: each ccBioZ recording was segmented into 30-seconds segments.
2) Amplitude normalization: the amplitude of the ccBioZ signal was normalized in the minimum-maximum range within each segment, according to Eq.(1), where $y_{norm}(t)$ corresponds to a point in the normalized ccBioZ segment, and $y(t)$ is a point in the original ccBioZ segment, y.

$$y_{norm(t)} = \frac{y(t) - \min(y)}{\max(y) - \min(y)}, t \in \{0, .., 30\}\, s \quad (1)$$

3) Filtering: a 4th order Butterworth bandpass filter was used with an upper cut-off frequency at 1.2 Hz and lower cut-off frequency at 0.1 Hz.

A total of 52 features were extracted to characterize the signal quality of the ccBioZ segments, by capturing statistical measures as well as nonlinear dynamics from the ccBioZ segments in both time and frequency domain. Methods like zero-crossing rate, maximum fractal length, kurtosis, skewness, and other statistical measures (e.g. mean, median, and mode), were utilized for analysing the time-domain information of the ccBioZ signal. Moreover, power spectral density (PSD) analysis, Hjorth parameters, and FFT analysis, were utilized for analysing the frequency-domain information of the ccBioZ signal. Peak-detection and frequency-based RR estimation as well as morphological features of the detected respiration cycles, were also utilized in this work. For simplicity, a brief explanation is given for the most relevant (i.e. best performing) features. Features 2, 6 and 7 are extracted as the ratio of the PSD of the ccBioZ segment in a narrow band of interest and a broader one which includes the characteristic frequencies of MA, for different narrow and broad bands respectively ([0.1-0.7]Hz vs [0-2]Hz; [0.15-0.5]Hz vs [0.15-2]Hz; [0.05-0.6]Hz vs [0.05-2]Hz). The slope sign change and the maximum fractal length of the segment are included as Features 11 and 16, respectively. Feature 18 is the peak-based RR of the segment, and Feature 22 is the standard deviation of the RR-intervals. Feature 24 represents the mean value of the relative widths of the detected respiration cycles. Features 28, 34, and 38 represent the mean, median, and mode values of the skewness of the detected respiration cycles in the segment. The FFT-based energy distribution within the respiratory frequency range is calculated in Feature 46. In addition to the mentioned best performing features, Table 5, in section VII, presents a brief definition of the features.

C. Data Separation:

As illustrated by Figure 4, three datasets (DS1, DS2, and DS3) were used in the testing process. In the case of DS1, the set of extracted features was randomly split into a training (70% of the observations) and a testing (30% of the observations) subsets. The training subset was then used to perform the feature selection process using five different approaches (see Subsection II-E), which were validated using the testing subset. After that, the selected feature subsets, were used to train and cross-validate (CV) the classifiers. All trained classifiers were then tested using the test subset, which was not involved in the feature selection/training phase. Next, the best performing models were further tested using DS2, to evaluate performance in case of segments that include apneic epochs, and DS3, to further validate the results using data acquired from a healthy volunteer. Such approach provides an additional validation and allows to verify the robustness of the models.

It is worth mentioning that a separate classifier could be trained for each ccBioZ channel (i.e. I or Q channels). However, it was found that unified classifiers for both I and Q channels produced similar results to those specific for a single channel. Therefore, only unified models for both I and Q ccBioZ channels were considered.

D. Performance Evaluation:

The performance of the classification models was quantitatively evaluated using the common statistical metrics for binary classification problems, namely: accuracy (Acc), sensitivity (Sen), specificity (Spe), precision (Prec), area under the curve (AUC), and F1-score (F1), all of which are mathematically defined in [28]. Since most of our datasets are unbalanced, i.e. number of bad and good labelled segments are not equal, the balanced accuracy (BAcc) metric [29], see Eq.(2), was employed for choosing the best performing models. This metric reflects both sensitivity and specificity on a single value, which provides a broader representation of the classifier's performance.

$$BAcc = \left(\frac{Sen+Spe}{2}\right) \qquad (2)$$

E. Feature Selection:

In this work, five different approaches for selecting the most relevant set of features were compared:

1) Neighbourhood Component Analysis (NCA):
   NCA is a non-parametric method for reducing the dimensionality of data, based on nearest-neighbour methods. This is achieved by maximizing the expected leave-one-out classification accuracy with a regularization term [30]. The MATLAB function *fscnca* was used for implementing the NCA-based feature selection approach. This function tries to learn the weights of the features to minimize the regression loss over the data that is used for selecting the features. The absolute weight value of a feature is a measure of how important this feature is, i.e. the closer to zero the less important. To ensure robustness, 100 iterations of the process were performed using the same data input. The resulting weights were then averaged for each feature. The features with absolute averaged weight higher than 0.5 were selected.
2) Correlation-based Feature Selection (CFS):
   The best set of features should contain entries which are highly correlated with the targets and uncorrelated with each other. In order to find the best subset of features, CFS calculates the merit (S_n) of a series of subsets consisting of n features using Eq.(3), where n is the number of all features, $\overline{r_{ff}}$ is the mean value of all feature-to-feature correlations, and $\overline{r_{cf}}$ is the mean value of all feature-to-class correlations within the feature subset [31].

$$S_n = \frac{n \times \overline{r_{cf}}}{\sqrt{n+n(n-1)\overline{r_{ff}}}} \qquad (3)$$

   The greedy forward selection method was used. Starting from a null set, each unselected feature was iteratively added to the feature subset. The feature that accounted for the highest S_n was kept. The procedure was then repeated with the remaining features. The values of S_n in all steps were then ranked, and the subset with higher merit was selected as the best feature subset.
3) Recursive Feature Elimination with Support Vector Machine (SVM-RFE):
   SVM-RFE elects the best feature subset by comparing the performance of a linear SVM (see Subsection II-F for more information on SVM) which was trained using different subsets of features. Unlike in the CFS method, backward search strategy was implemented in the SVM-RFE method. Starting from the full set of features, the algorithm randomly and sequentially eliminated one feature and resulted in a candidate subset of n features at each step. The SVM was then trained and validated using a 5-fold CV scheme, on all possible subsets of m features ($m \in \{1, \dots n\}$), where n is the number of all features. After that, the criterion function α was calculated on each fold and then averaged, see Eq.(4). For each candidate subset of m features, the best performing was the one that minimized α, which measures the loss in the balanced accuracy, while the least performing feature was eliminated on the next iteration.

$$\alpha = \left(1 - \frac{Sen+Spe}{2}\right) \qquad (4)$$

   Once all the best performing feature subsets were found, the optimal subset of features could be chosen as a trade-off between the corresponding balanced accuracy and the number of features, by manually choosing the optimal point on a *knee-plot* that depicts the balanced accuracies (μ), see Eq.(5), versus the number of features in the corresponding feature subsets.

$$\mu = 100 - (\alpha \times 100) \qquad (5)$$

The MATLAB function *sequentialfs* was used to implement the SVM-RFE method.

4) Hybrid Pipeline (Hybrid):
Three approaches were used in series to select the best feature subset: a) Wilcoxon signed-rank test (WSRt), b) CFS, and c) SVM-RFE.
The WSRt method is a non-parametric method that statistically verifies whether two samples are taken from different populations. It tries to disprove the null hypothesis (H_0) that the differences between pairs of observations follow a symmetric distribution around zero, at a given significance level defined by the *p-value* (i.e. the probability of the occurrence of H_0). Eq.(6) holds if H_0 was proved true.

$$H_0 : P(X > 0) = P(0 > X) \qquad (6)$$

A random sample of observations, labelled of bad quality segments, was chosen from the training dataset and tested against a sample of good quality labelled segments for each feature using the MATLAB function *signrank*. A *p-value*=0.01 was used as a threshold for rejecting H_0. If H_0 holds (i.e. good and bad segments are statistically similar), the feature under evaluation is excluded.
It is worth mentioning that both WSRt and CFS are filter methods (i.e. features that are not relevant according to some metric are filtered out of the set), while the SVM-RFE is a wrapper method (i.e. a classifier is trained with consecutive subsets of features and the subset leading to the highest performance outcome is chosen).

5) Individual Feature Evaluation (Ind_Ev):
Here, two different approaches were used to indicate the importance of each feature by evaluating its performance individually, i.e. how good a feature is at differentiating between a good and a bad-quality segment using a simple threshold-based classifier:
   a) Receiver Operating Characteristics (ROC) curve: this was done by obtaining the ROC curve from each feature for the training dataset and finding the most cost-effective point in the ROC curve, using the MATLAB function *ROCOut* [32]. After that, the testing dataset was used to evaluate the performance of the tuned binary classifier to prevent overfitting.
   b) Decision Trees: by training a one-level decision tree (see Subsection II-F for more information on decision trees) using the training dataset. Next, the performance of the tuned decision tree was evaluated using the testing dataset to prevent overfitting.

F. Classification:

To find the best performing classifiers, our approach was to train 19 different classification models, each with differently tuned hyperparameter settings. Four supervised machine learning methods were exploited in this process:

- Decision Trees: a predictive model that uses a tree-like graphical representation to predict responses to certain observations [33]. The MATLAB function *fitctree* was used to train Fine, Medium, and Coarse decision trees models, with Gini's diversity index as that split criterion, and a maximum number of splits equal 100, 20, and 4, respectively.
- K-Nearest-Neighbors (KNN): a non-parametric approach that uses *K* training examples to predict the class of certain observations [34]. The MATLAB function *fitcknn* was used to train the KNN models. Refer to Table 2 for a detailed list of all KNN classifiers and their corresponding hyperparameters' settings.
- Support Vector Machine (SVM): SVM uses a hyperplane as the decision boundary to separate different classes of observations [35]. The MATLAB function *fitcsvm* was used to train the SVM models. Refer to Table 2 for a detailed list of all SVM classifiers and their corresponding hyperparameters' settings.
- Ensemble learning methods: several weak learning methods are combined to one ensemble predictor, yielding a higher performance predictor, better than what could have been obtained using any of the forming learning algorithms alone [36]. The MATLAB function *fitcensemble* was used to train Boosted, Bagged, and RUSBoosted trees as well as Subspace KNN models, all of which using a number of learning cycles equal to 30.

By altering the hyperparameters of each learning algorithm, we ended up with 19 different trained models (covering more common possibilities for hyper parameter tuning, per algorithm). A 5-Fold CV scheme was employed for training/validation using the training dataset.

Table 2: Models and hyperparameters.

Model	Hyperparameter	Model	Hyperparameter
Linear SVM	BoxConstraint: 1 / KernalScale: Auto KernalFunction: Linear PolynomialOrder: N.A.	Fine KNN	Distance: Euclidean DistanceWeight: Equal Exponent: N.A. / NumNeighbors: 1
Quadratic SVM	BoxConstraint: 1 / KernalScale: Auto KernalFunction: Polynomial PolynomialOrder: 2	Medium KNN	Distance: Euclidean DistanceWeight: Equal Exponent: N.A. / NumNeighbors: 10
Cubic SVM	BoxConstraint: 1 / KernalScale: Auto KernalFunction: Polynomial PolynomialOrder: 3	Coarse KNN	Distance: Euclidean DistanceWeight: Equal Exponent: N.A. / NumNeighbors: 100
Fine Gaussian SVM	BoxConstraint: 1 / KernalScale: 2 KernalFunction: Gaussian PolynomialOrder: N.A.	Cosine KNN	Distance: Cosine DistanceWeight: Equal Exponent: N.A. / NumNeighbors: 10
Medium Gaussian SVM	BoxConstraint: 1 / KernalScale: 7.9 KernalFunction: Gaussian PolynomialOrder: N.A.	Cubic KNN	Distance: Minkowski DistanceWeight: Equal Exponent: 3 / NumNeighbors: 10
Coarse Gaussian SVM	BoxConstraint: 1 / KernalScale: 31 KernalFunction: Gaussian PolynomialOrder: NA	Weighted KNN	Distance: Euclidean DistanceWeight: SquaredInverse Exponent: NA / NumNeighbors: 10

III. Results

A. *Feature selection*

As mentioned previously in Subsection II-E, five different approaches were exploited for selecting the best subset of features. The number of features selected by the Hybrid, NCA, Ind_Ev, CFS, and SVM-RFE methods are 8, 13, 13, 14, and 27, respectively. Figure 5 depicts the results for the NCA-based feature selection. Moreover, Figure 6 depicts the knee-plot which is generated in the Hybrid test for feature selection. Figure 7 summarizes the results for the feature selection phase.

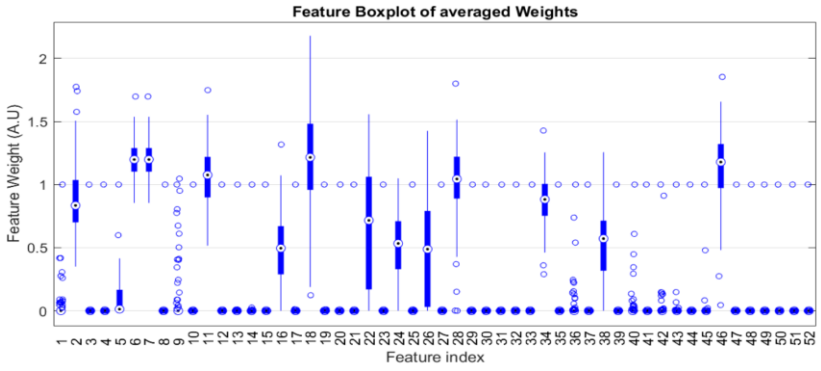


Figure 5: Boxplot of averaged NCA feature weights.

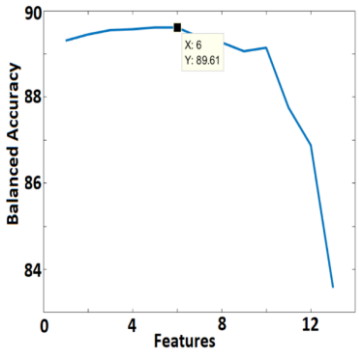


Figure 6: Knee-plot for the Hybrid test.

Hybrid		■	■								■								■						
NCA		■				■	■				■					■		■			■		■		■
Ind_Ev	■	■			■	■	■				■							■					■	■	
CFS		■	■	■							■								■		■				
SVM-RFE	■	■	■			■	■		■		■		■			■		■	■		■	■		■	■
T/F	T	F	F	T	T	F	F	T	T	T	T	T	T	T	T	T	T	T	F	T	T	T	T	T	T	T
Feature #	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26
Hybrid	■															■			■	■						
NCA		■						■				■								■						
Ind_Ev										■				■	■					■						
CFS		■	■			■		■									■		■	■	■					
SVM-RFE		■	■												■	■	■	■	■	■					■	
T/F	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	F	F	F	F	T	T	T	T	T	T
Feature #	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45	46	47	48	49	50	51	52

Figure 7: Results for the five different feature selection methods. A blue box denotes that the corresponding feature was selected by the corresponding feature selection method. Letters T and F denote if the corresponding feature performs a time or a frequency domain analysis, respectively. Red borders indicate that the corresponding feature was selected by all feature selection methods.

B. *Classification:*

As previously mentioned in Subsection II-F, 19 different classifiers were utilized in this work. The reliability of the performance of each model was assessed during the training phase with 5-Fold CV. To statistically evaluate the performance of the models, three phases of testing were applied: 1) using the test dataset which was partitioned from DS1, 2) using DS2 to study the effect of the apneic episodes on the models' performance, and 3) using DS3 for extended validation with data from a new subject (see Figure 4). Table 3 summarizes the results of the testing phases using the LT labelling scheme. For simplicity, only the results of the best performing models (in terms of BAcc) were included.

To proceed with the testing phases 2 and 3, a classifier (Fine Gaussian SVM 'NCA feature subset') was selected. This model provided good trade-off between the dimensionality of the feature subset and the performance of the classification model. Figure 8 depicts the confusion matrices for the test phases 1, 2, and 3 using the selected model with the LT labelling scheme.

In addition, the proposed solution was tested using a high-threshold (HT) binary labelling scheme (i.e. good segments are labelled as good, while both acceptable and bad segments are labelled as bad), by implementing the same pipeline, illustrated by Figure 4. The best results (in terms of BAcc) in each testing phase were produced by a Medium Gaussian SVM classifier and came as following: DS1: Accuracy=0.84, Sensitivity=0.87, Specificity=0.84, and Balanced Accuracy=0.85. DS2: Accuracy=0.87, Sensitivity=0.88, Specificity=0.87, and Balanced Accuracy=0.88. DS3: Accuracy=0.91, Sensitivity=0.98, Specificity=0.91, and Balanced Accuracy=0.94.



Figure 8: Confusion matrices for (a) test phase 1, (b) test phase 2, and (c) test phase 3 using the Fine Gaussian SVM 'NCA feature subset' with LT labelling scheme.

Table 3: Summarized results of the best performing classifiers for testing phases 1, 2, and 3.

Dataset	Feature set	Model	CV BAcc [mean, std]	BAcc	Sen	Spe	Prec	F1	Acc
DS1	All features	Medium Gaussian SVM	[0.91, 0.01]	0.91	0.88	0.95	0.94	0.91	0.91
	NCA	**Fine Gaussian SVM**	**[0.90, 0.01]**	**0.91**	**0.89**	**0.93**	**0.93**	**0.91**	**0.91**
	CFS	Medium Gaussian SVM	[0.89, 0.01]	0.90	0.87	0.93	0.92	0.90	0.90
	SVM	Fine Gaussian SVM	[0.90, 0.01]	0.91	0.90	0.93	0.93	0.91	0.91
	Hybrid	Medium Gaussian SVM	[0.89, 0.01]	0.89	0.87	0.92	0.92	0.89	0.89
	Ind_Ev	Linear SVM	[0.90, 0.01]	0.90	0.89	0.91	0.91	0.90	0.90
DS2	NCA	Fine Gaussian SVM	N.A.	0.82	0.84	0.80	0.42	0.56	0.80
DS3	NCA	Fine Gaussian SVM	N.A.	0.94	0.98	0.87	0.65	0.78	0.89

IV. Discussion

The proposed methods for selecting the best features showed significant reduction in the dimensionality of the feature set, see Figure 7. In the SVM-RFE algorithm, the manual process of choosing the optimal point on the *knee-plot* has automated alternatives, however, this manual approach made it more flexible to trade between the number of selected features and balanced accuracy of the classifier. It is worth mentioning that the SVM-RFE algorithm may result in biased selection, e.g. a feature that is randomly correlated with the class targets would be selected by the SVM-RFE algorithm. Such features would enhance the performance of specific datasets, while failing with others. This explains the higher dimensionality of the SVM-RFE feature set, compared to the other feature sets. On the other hand, in the Hybrid pipeline, both the WRSt and CFS steps minimized this bias, as the number of selected features within the WRSt, CFS, and SVM steps were 47, 13, and 8, respectively.

Figure 7 shows that only three features were selected by all feature selection methods, which are feature 2, 11, and 46. The average number of selections for a frequency domain feature was 3, compared to 1.39 for a time domain feature, which can be attributed to the highest number of time domain features (43) compared to frequency domain features (9).
Individually, features 6 and 46, both of which perform a frequency domain analysis, scored the highest performance in the Ind_Ev feature selection. However, all the feature subsets included both time and frequency domain features.

Amongst all trained classification models, the SVM-based models produced the best results. In particular, the Fine Gaussian SVM model, along with the NCA feature subset, provided the best combination between feature subset dimensionality and the performance metrics of the classifier, see Table 3, thanks to its robust implementation, see Subsection II-E.

Higher BAcc was reported in testing phase 1 for LT labelling scheme than HT labelling scheme in DS1, see Subsection III-B. This was due to the highly unbalanced HT labelling scheme compared to the LT labelling scheme, as well as the additional challenge of differentiating 'Acceptable' signals against 'Good' signals. On the other hand, the classifiers performed better in testing phase 2 for HT labelling scheme, due to the presence of apneic epochs in DS2. In testing phase 2 results, a relative drop in the performance of the classifier was notable, as no focus was put in the cases with apneic events within the training phase. However, such results showed potential in extending the application of the proposed algorithm to cover detection of apneic events. The high performance in testing phase 3, using both HT and LT labelling schemes, proved the reliability and robustness of the proposed algorithm, as the data in DS3 was acquired from a different subject from those in the training dataset. Low precision values were reported in test phases 2 and 3. This is due to the highly unbalanced DS2 and DS3. In addition, some of the misclassified samples are those of "Acceptable" signal quality, which could be pictured as a gray-area, see Figure 9.

The use of a binary-class scheme enables the user to manually tune the level of strictness of the classification step. Such design decision would depend on the end use of the whole application. For RR estimation, we used a LT labelling approach, as respiration cycles are detectable within the good and acceptable quality bands. A high-threshold labelling scheme would be necessary for other applications such as detection of apneic events, with the additional challenge of the lower performance when aiming to classify good quality segments with an apneic event.
As mentioned in Subsection II-B, minimal signal denoising was applied in the pre-processing step. Moreover, no normalizing was done to the extracted features over long ranges before feeding them into the classifiers, enabling them to be possibly deployed in real-time applications.

The engineering approach used for training the classifiers (i.e. training 19 different classifiers), resulted in a wide range of comparable working classifiers. However, such approach puts less focus on individualizing the classifier (i.e. fine tuning the hyperparameters).

The proposed solution dealt with the quality assessment of ccBioZ signals, acquired using non-contact electrodes in a realistic (sleep) environment under free-moving conditions. To our knowledge, no previous work on this topic was published before. Therefore, a fair comparison with the previous related work (see Section I) is difficult to perform. However, the performance of the proposed solution (BAcc up to 0.94) falls within the generally accepted range in the literature for pulsatile waveforms, e.g. PPG and EDR, with high periodicity and dominant frequency: Li Q. et al. [21] reported BAcc of

0.91, Orphanidou C. et al. [22] reported BAcc of 0.93, Liu S. H. et al. in [26] and [27] reported BAcc of 0.88 and 0.91, respectively. Moreover, the reliability and robustness of the proposed solution was tested on a large amount of data with varying characteristics (i.e. three different testing datasets), compared to what was reported in the literature.
The focus of this work was put on identifying ccBioZ segments using binary classifiers. Multi-class classification approaches can be investigated in future work, especially because all three datasets were labelled in a multi-level quality scheme. The labelling as well as the training steps were done to fulfil the goal of classifying the quality of the ccBioZ signals. Therefore, medical applications, like detection of apneic events, were not considered. Another limitation in the presented work is that the annotation (i.e. labelling) step was done by only one experienced biomedical engineer. This research however is ongoing and will be extended in the future.

V. Conclusion

An automated machine-learning-based approach for assessing the quality of ccBioZ signals was demonstrated, tested, and statistically evaluated, with resulting BAcc up to 0.94. A full-package solution for the ccBioZ signal quality classification problem was offered by the proposed pipeline. Moreover, five alternative methods for selecting the best subset of features were exploited, enabling a good trade-off between the dimensionality of the feature subset and the performance of the classifier. A multi-stage testing pipeline, composed of three testing phases, was applied for statistically evaluating the classifiers. Three large datasets of ccBioZ acquired with varying characteristics were utilized.
Extending the presented work to cover different sources of sinusoid-like medical signals (e.g. radar-based respiration signals, PPG signals, and EDR signals) is possible in a future work.

VI. Acknowledgements

The authors would like to thank Neide Simões-Capela for her guiding inputs and useful advice. This work was supported by imec funds 2019-2020. S. Van Huffel and C. Varon were supported by the VLAIO grant 150466: OSA+ and the funding from the Flemish government (AI Research Program).

VII. Appendix

(a)

Classes	Good (1,2)	Bad (3)
1	283 (97.59%)	7 (2.41%)
2	651 (85.43%)	111 (14.67%)
3	74 (7.06%)	974 (92.94%)

(b)

Classes	Good (1,2)	Bad (3)
1	235 (94.00%)	15 (6.00%)
2	604 (80.11%)	150 (19.89%)
3	1163 (20.21%)	4591 (79.79%)

(c)

Classes	Good (1,2)	Bad (3)
1	976 (99.69%)	3 (0.31%)
2	1714 (96.78%)	57 (3.22%)
3	1424 (13.38%)	9220 (86.62%)

Figure 9: Detailed confusion matrices for (a) test phase 1, (b) test phase 2, and (c) test phase 3. Classes 1, 2, and 3 represent good, acceptable, and bad, respectively. Both red and orange cells are the misclassified labels, with red being the most unwanted error band.

Table 4: Table of acronyms

Acronym	Meaning
ccBioZ	Capacitively-Coupled Bioimpedance
RR	Respiration Rate
MA	Motion Artefact
PPG	Photoplethysmography
EDR	Electrocardiography Derived Respiration
FFT	Fast Fourier Transform
CNN	Conventional Neural Networks
Fs	Sampling Frequency
N.A.	Not Applicable
DS	Dataset
LT	Low-Threshold
PSD	Power Spectral Density
CV	Cross-Validate
Acc	Accuracy
Sen	Sensitivity
Spe	Specificity
Prec	Precision
AUC	Area Under the Curve
F1	F1-score
BAcc	Balanced Accuracy
HT	High-Threshold
NCA	Neighbourhood Component Analysis
CFS	Correlation-based Feature Selection
SVM-RFE	Recursive Feature Elimination with Support Vector Machine
Hybrid	Hybrid Pipeline
WSRt	Wilcoxon Signed-Rank test
Ind_Ev	Individual Evaluation
ROC	Receiver Operating Characteristics
KNN	K-Nearest-Neighbors
SVM	Support Vector Machine

Table 5: A brief definition of the features

Feature	Definition	Feature	Definition
1	The zero-crossing rate of the ccBioZ segments	**27**	The mean value of the kurtosis of the detected respiration cycles of the ccBioZ segments "mean(kurtosis(ccBioZ))"
2	The ratio of the PSD of the ccBioZ segment in a narrow band [0.1-0.7]Hz and a broader band [0-2]Hz	**28**	mean(skewness(ccBioZ))
3	The Hjorth parameter for the ccBioZ segments (measure of the frequency uniqueness)	**29**	The mode value of the kurtosis of the detected respiration cycles of the ccBioZ segments "mode(kurtosis(ccBioZ))"
4	Statistical measures of the normalized autoregressive parameters of a model of the ccBioZ segments	**30**	The variance of the kurtosis of the detected respiration cycles of the ccBioZ segments "var(kurtosis(ccBioZ))"
5	Statistical measures of the autocorrelation sequence of the ccBioZ segments	**31**	The std of the kurtosis of the detected respiration cycles of the ccBioZ segments "std(kurtosis(ccBioZ))"
6	The ratio of the PSD of the ccBioZ segment in a narrow band [0.15-0.5]Hz and a broader band [0.15-2]Hz	**32**	std(skewness(ccBioZ))
7	The ratio of the PSD of the ccBioZ segment in a narrow band [0.05-0.6]Hz and a broader band [0.05-2]Hz	**33**	The median value of the kurtosis of the detected respiration cycles of the ccBioZ segments "median(kurtosis(ccBioZ))"
8	The mean absolute deviation of the ccBioZ segments	**34**	median(skewness(ccBioZ))
9	The median absolute deviation of the ccBioZ segments	**35**	(mean(kurtosis(ccBioZ)) + mode(kurtosis(ccBioZ))) * std(kurtosis(ccBioZ))
10	The mean absolute value over a band of interest of the ccBioZ segments in the time domain	**36**	(mean(skewness(ccBioZ)) + mode(skewness (ccBioZ))) * std(skewness (ccBioZ))
11	The number of times in which the sign of the slope of the ccBioZ segments changes	**37**	mean(kurtosis(ccBioZ)) + mode(kurtosis(ccBioZ))
12	The root mean square of the ccBioZ segments	**38**	mean(skewness (ccBioZ)) + mode(skewness (ccBioZ))
13	The average of the exponential of the summation of the logarithm of all samples of the ccBioZ segments	**93**	mode(kurtosis(ccBioZ)) * std(kurtosis(ccBioZ))
14	The weighted (higher weight for samples in the middle of the segment) mean absolute value of the ccBioZ segments	**40**	mode(skewness (ccBioZ)) * std(skewness (ccBioZ))
15	The peak percentage rate of the ccBioZ segments	**41**	The summation of the kurtosis of the detected respiration cycles of the ccBioZ segments
16	The maximum fractal length of the ccBioZ segments	**42**	The summation of the skewness of the detected respiration cycles of the ccBioZ segments
17	The difference between RR acquired from two different methods	**43**	The ratio of the PSD of the largest harmonic to the next 5 harmonics of the periodogram of the ccBioZ segments
18	The peak-based RR of the ccBioZ segments	**44**	The ratio of the PSD of the largest harmonic to the rest of the periodogram except the next 5 largest harmonics
19	The RR based on frequency domain analysis	**45**	The ratio of the PSD of the largest harmonic to the rest of the periodogram of the ccBioZ segments
20	The RR based on a simple peak detection method	**46**	The FFT-based energy distribution within the respiratory frequency range of the ccBioZ segments
21	The mean value of RR estimates acquired from two different methods	**47**	Mean squared error of the ccBioZ segments with respect to a generated sine wave of similar number of peaks
22	The standard deviation of the RR-intervals of the ccBioZ segments	**48**	Root mean squared error of the ccBioZ segments with respect to a generated sine wave of similar number of peaks
23	Measure of agreement between different peak detectors	**49**	Summation of the squared error of the ccBioZ segments to a generated sine wave with similar number of peaks
24	The mean value of the relative widths of the detected respiration cycles of the ccBioZ segments	**50**	Root summation squared error of the ccBioZ segments to a generated sine wave with similar number of peaks
25	The mean value of the relative amplitudes of the detected respiration cycles of the ccBioZ segments	**51**	Energy index of the ccBioZ segments
26	Peak positions relative to a generated sine wave with number of peaks equals the number of detected cycles	**52**	Root energy index of the ccBioZ segments

VIII. References

[1] D.R. Hess, B.D. Medoff, Respiratory monitoring, Current Opinion in Critical Care. 5 (1999) 52.

[2] F.Q. AL-Khalidi, R. Saatchi, D. Burke, H. Elphick, S. Tan, Respiration rate monitoring methods: A review, Pediatric Pulmonology. 46 (2011) 523–529.

[3] V.-P. Seppä, J. Väisänen, O. Lahtinen, J. Hyttinen, Assessment of breathing parameters during running with a wearable bioimpedance device, in: 4th European Conference of the International Federation for Medical and Biological Engineering, Springer, 2009: pp. 1088–1091.

[4] O.G. Martinsen, S. Grimnes, Bioimpedance and bioelectricity basics, Academic press, 2011.

[5] M. Cheney, D. Isaacson, J.C. Newell, Electrical impedance tomography, SIAM Review. 41 (1999) 85–101.

[6] R.N. Baumgartner, W.C. Chumlea, A.F. Roche, Bioelectric impedance phase angle and body composition, The American Journal of Clinical Nutrition. 48 (1988) 16–23.

[7] D.P. Bernstein, Continuous noninvasive real-time monitoring of stroke volume and cardiac output by thoracic electrical bioimpedance., Critical Care Medicine. 14 (1986) 898–901.

[8] C.S. Poon, Y.C. Chung, T.T.C. Choy, J. Pang, Evaluation of two noninvasive techniques for exercise ventilatory measurements, in: Proceedings of the Annual International Conference of the IEEE Engineering in Medicine and Biology Society, IEEE, 1988: pp. 823–824.

[9] Y. Sun, X.B. Yu, Capacitive biopotential measurement for electrophysiological signal acquisition: A review, IEEE Sensors Journal. 16 (2016) 2832–2853.

[10] I.D. Castro, A. Patel, T. Torfs, R. Puers, C. van Hoof, Capacitive multi-electrode array with real-time electrode selection for unobtrusive ECG & BIOZ monitoring, in: 2019 41st Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), IEEE, 2019: pp. 5621–5624.

[11] R. Macías, M.A. García, J. Ramos, R. Bragos, M. Fernández, Ventilation and heart rate monitoring in drivers using a contactless electrical bioimpedance system, in: Journal of Physics: Conference Series, IOP Publishing, 2013: p. 012047.
[12] D.G. Abad, Development of a capacitive bioimpedance measurement system, MS Thesis. (2009).
[13] P.S. Luna-Lozano, R. Pallàs-Areny, Heart rate detection from impedance plethysmography based on concealed capacitive electrodes, in: XIX IMEKO World Congress Fundamental and Applied Metrology, 2009: pp. 1701–1706.
[14] E. Pinheiro, O. Postolache, P. Girao, Contactless impedance cardiography using embedded sensors, Measurement Science Review. 13 (2013) 157–164.
[15] I. Castro Miller, Robust capacitively-coupled electrocardiography and bioimpedance for unobtrusive health monitoring, Ph.D dissertation, College of Eng. and Sc., KUL Univ., Leuven, (2019).
[16] V.-P. Seppä, Development and clinical application of impedance pneumography technique, (2014).
[17] I.D. Castro, M. Mercuri, A. Patel, R. Puers, C. van Hoof, T. Torfs, Physiological Driver Monitoring Using Capacitively Coupled and Radar Sensors, Applied Sciences. 9 (2019) 3994.
[18] O. Such, Motion tolerance in wearable sensors-The challenge of motion artifact, in: 2007 29th Annual International Conference of the IEEE Engineering in Medicine and Biology Society, IEEE, 2007: pp. 1542–1545.
[19] S. Nemati, A. Malhotra, G.D. Clifford, Data fusion for improved respiration rate estimation, EURASIP Journal on Advances in Signal Processing. 2010 (2010) 926305.
[20] J.A. Sukor, S.J. Redmond, N.H. Lovell, Signal quality measures for pulse oximetry through waveform morphology analysis, Physiological Measurement. 32 (2011) 369.
[21] Q. Li, G.D. Clifford, Dynamic time warping and machine learning for signal quality assessment of pulsatile signals, Physiological Measurement. 33 (2012) 1491.
[22] C. Orphanidou, T. Bonnici, P. Charlton, D. Clifton, D. Vallance, L. Tarassenko, Signal-quality indices for the electrocardiogram and photoplethysmogram: Derivation and applications to wireless monitoring, IEEE Journal of Biomedical and Health Informatics. 19 (2014) 832–838.
[23] D.A. Birrenkott, M.A.F. Pimentel, P.J. Watkinson, D.A. Clifton, Robust estimation of respiratory rate via ECG-and PPG-derived respiratory quality indices, in: 2016 38th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), IEEE, 2016: pp. 676–679.
[24] J. Jorge, M. Villarroel, S. Chaichulee, G. Green, K. McCormick, L. Tarassenko, Assessment of signal processing methods for measuring the respiratory rate in the neonatal intensive care unit, IEEE Journal of Biomedical and Health Informatics. 23 (2019) 2335–2346.
[25] E.K. Naeini, I. Azimi, A.M. Rahmani, P. Liljeberg, N. Dutt, A Real-time PPG quality assessment approach for healthcare internet-of-things, Procedia Computer Science. 151 (2019) 551–558.
[26] S.-H. Liu, J.-J. Wang, W. Chen, K.-L. Pan, C.-H. Su, Classification of photoplethysmographic signal quality with fuzzy neural network for improvement of stroke volume measurement, Applied Sciences. 10 (2020) 1476.
[27] S.-H. Liu, R.-X. Li, J.-J. Wang, W. Chen, C.-H. Su, Classification of Photoplethysmographic Signal Quality with Deep Convolution Neural Networks for Accurate Measurement of Cardiac Stroke Volume, Applied Sciences. 10 (2020) 4612.
[28] T. Fawcett, An introduction to ROC analysis, Pattern Recognition Letters. 27 (2006) 861–874.
[29] D.N. Sotiropoulos, G.A. Tsihrintzis, The class imbalance problem, in: Machine Learning Paradigms, Springer, 2017: pp. 51–78.
[30] W. Yang, K. Wang, W. Zuo, Neighborhood Component Feature Selection for High-Dimensional Data., JCP. 7 (2012) 161–168.
[31] M.A. Hall, Correlation-based feature selection for machine learning, (1999).
[32] G. Cardillo, ROC curve: compute a Receiver Operating Characteristics curve, Natick (MA): Math-Works. (2008).
[33] S. Shalev-Shwartz, S. Ben-David, Understanding machine learning: From theory to algorithms, Cambridge university press, 2014.
[34] N.S. Altman, An introduction to kernel and nearest-neighbor nonparametric regression, The American Statistician. 46 (1992) 175–185.
[35] M.A. Hearst, S.T. Dumais, E. Osuna, J. Platt, B. Scholkopf, Support vector machines, IEEE Intelligent Systems and Their Applications. 13 (1998) 18–28.
[36] D. Opitz, R. Maclin, Popular ensemble methods: An empirical study, Journal of Artificial Intelligence Research. 11 (1999) 169–198.