Early Detection of Sepsis Induced Deterioration Using Machine Learning

Dal Canton, Francesco; Wiering, Marco; Quinten, Vincent

Published in:
BNAIC 2018

DOI:
10.1007/978-3-030-31978-6

Publication date:
2019

Early Detection of Sepsis Induced Deterioration Using Machine Learning

Francesco Dal Canton[1], Vincent M. Quinten[1,2], and Marco A. Wiering[1(✉)]

[1] University of Groningen, 9700 AB Groningen, The Netherlands
m.a.wiering@rug.nl
[2] University Medical Center Groningen, 9713 GZ Groningen, The Netherlands

Abstract. Sepsis is an excessive bodily reaction to an infection in the bloodstream, which causes one in five patients to deteriorate within two days after admission to the hospital. Until now, no clear tool for early detection of sepsis induced deterioration has been found. This research uses electrocardiograph (ECG), respiratory rate, and blood oxygen saturation continuous bio-signals collected from 132 patients from the University Medical Center of Groningen during the first 48 h after hospital admission. This data is examined under a range of feature extraction strategies and Machine Learning techniques as an exploratory framework to find the most promising methods for early detection of sepsis induced deterioration. The analysis includes the use of Gradient Boosting Machines, Random Forests, Linear Support Vector Machines, Multi-Layer Perceptrons, Naive Bayes Classifiers, and k-Nearest Neighbors classifiers. The most promising results were obtained using Linear Support Vector Machines trained on features extracted from single heart beats using the wavelet transform and autoregressive modelling, where the classification occurred as a majority vote of the heart beats over multiple long ECG segments.

Keywords: Sepsis · Machine Learning · Bio-signals · Health care

1 Introduction

Sepsis is a life-threatening organ dysfunction caused by an uncontrolled reaction to infection by the organism [1] that can lead to organ failure, septic shock, and death [2]. Common symptoms of sepsis include higher heart rate and respiratory rate, and abnormal changes in bodily temperature [3]. Sepsis is one of the most common causes for mortality among chronically ill patients, and it is estimated that sepsis affects at least 240 people out of 100,000 in the United States, while severe sepsis affects between 51 and 95 out of 100,000 [4]. Most patients affected by sepsis are admitted to the hospital through the Emergency Department (ED), and it was shown that approximately 20% of patients admitted to the ED with infection or sepsis deteriorate [5].

Early detection of sepsis induced deterioration is extremely valuable since it allows for fast and effective treatment. In [6] it was shown that each hour of delay

M. Atzmueller and W. Duivesteijn (Eds.): BNAIC 2018, CCIS 1021, pp. 1–15, 2019.
https://doi.org/10.1007/978-3-030-31978-6_1

in the application of appropriate treatment is correlated with a mean increase in mortality of 7.6%. Nevertheless, despite the intensive research in the field, it is still not clear how the onset, progress, and response to treatment of sepsis can be accurately monitored [7].

The traditional approach for tracking sepsis onset and development is to use discrete values describing vital signs and non-specific symptoms [3]. More recently, measures obtained from Heart Rate Variability (HRV) have been gathering research interest. Although at present the most successful studies in this area concerned sepsis development in neonates [8], some studies have been carried out to explore the predictive potential of HRV measures in adults [9,10]. In 2017 the SepsiVit study was started at the University Medical Center of Groningen (UMCG), which involves a long term data collection program, and aims at determining whether HRV measures can provide a reliable source of information for predicting deterioration in patients with suspected sepsis in the ED [11].

The current study focuses on the potential of Machine Learning based algorithms paired with the use of raw Electrocardiograph (ECG), Plethysmograph, and Respiratory Rate bio-signals collected during the SepsiVit study at the UMCG as sources of information for early detection of patient deterioration due to sepsis. Seven different Machine Learning classifiers are tested and their classification accuracies are compared across three different feature extraction methods. The first two methods involve Histograms of Derivatives (HOD) of the bio-signals, while the third one uses morphological features of heart beats extracted using the wavelet transform and autoregressive modelling as applied in [12]. The third feature extraction method was also tested in a majority vote fashion across 5 min long signal windows and 1 h long signal windows.

This paper is organized as follows. Section 2 describes the dataset in more detail. Section 3 illustrates the three feature extraction methods used to process the dataset. Section 4 lists and explains the machine learning models and how they were applied. Section 5 describes the experimental setup and the obtained results, while Sect. 6 concludes the paper.

2 Dataset

The dataset used in this research was collected at the ED of the UMCG according to the protocol of the SepsiVit study. All patients included in the study (i) are more than 18 years old, (ii) present a suspected infection or sepsis, (iii) show two or more systemic inflammatory response syndrome criteria as defined by the International Sepsis Definitions Conference [13], and (iv) provided written informed consent. Patients are not included in the study in case of (i) known pregnancy, (ii) when the patient is not admitted to the hospital from the ED or is transfered to another hospital or care facility, and (iii) in case of previous cardiac transplantation [11]. While the aim of the SepsiVit study is to collect data from 171 patients, the collected and labeled data at the time of the current study includes 132 patients (84 males; average age 61.5 years; median age 63.5 years; average missing data 53%).

For each patient, high sample rate vital signs are recorded with a bedside patient monitor (Philips IntelliVue MP70 System with MultiMeasurement Module using custom software based on the Philips IntelliVue Data Export Interface Protocol). The data includes time series data of ECG (500 Hz), Plethysmograph (125 Hz), and Respiratory Rate (62.5 Hz) bio-signals recorded for up to 48 h since admission to the ED. No imputation strategy is used to recover missing data due to the complexity and unpredictability of the bio-signals involved. The electrodes for recording the ECG signals are placed according to the EASI configuration [14], and in particular the data from Lead II is used for this analysis. After the data is collected, the outcomes for the patient's condition are recorded. Specifically, five outcomes are monitored: whether the patient (i) had to be transferred to the Intensive Care Unit (ICU), (ii) died in the hospital, (iii) developed kidney failure, (iv) developed liver failure, or (v) developed respiratory failure. Since the goal of this analysis is to provide a tool for early sepsis deterioration, each patient was labeled as 'deteriorating' if they registered positive to any of these five outcomes, and 'healthy' otherwise. The proportions of the two groups are specific to each feature extraction method depending on the amount of usable data, and are mentioned in the respective subsections of the paper.

3 Feature Extraction Methods

The detection of early signs of sepsis induced deterioration using bio-signals requires a procedure of feature extraction from the raw data, so that each extracted feature vector represents a segment of the original data. With this in mind, a good feature extraction procedure should yield feature vectors that are most similar among the same class and most different across different classes.

The three feature extraction methods described in this section are compared with the ones currently being developed as a part of the SepsiVit study, which were obtained exclusively from the ECG signal, after the removal of technical and physiological artifacts [15]. They include HRV measures as described in [16], and geometrical features of the R-R intervals [17].

3.1 Histograms of Derivatives

The first approach involves the extraction of the distribution of the first and second order derivatives of the available signals, or Histograms of Derivatives (HOD). This method is conceptually close to the Histogram of Oriented Gradients strategy used in image processing [18]: the objective is to obtain the frequency distribution of change in signal intensity across a signal segment. The derivative of a function at a specific input value is defined as the slope of the tangent line to the graph of the function at that point. In the case of the digital signals used in this study, an approximation of the derivative function is computed as:

$$\frac{dx}{dt} = \frac{x_{t+h} - x_t}{h} \tag{1}$$

where h is the unit interval between consecutive samples. For each of the three signals used in this study, h is set to 1 since the time between consecutive samples in each signal is constant.

The first step of this procedure is, for each patient's bio-signals (i.e. ECG, Plethysmograph, and Respiratory Rate), to extract all simultaneous 5-minute long signal segments that don't contain any missing data. The result is a collection of 5-minute long data triplets containing the three bio-signals. The length of 5 min for each signal window was chosen experimentally as it produced improved classification accuracies compared to a length of 30 min. This choice was also guided by the convenience of requiring only 5 min of recorded signal before attempting detection of sepsis induced deterioration, which would speed up the potential application of treatment.

At this stage, the first and second derivatives of each signal segment are computed. Given each signal in each data triplet, Eq. 1 was applied across the whole signal segment. The result is 6 signals, two for each type of bio-signal, of which one is the first order derivative, and the other is the second order derivative, computed by applying Eq. 1 on the computed first derivative. A plot representing an example of first and second order derivatives computed in such fashion is shown in Fig. 1.

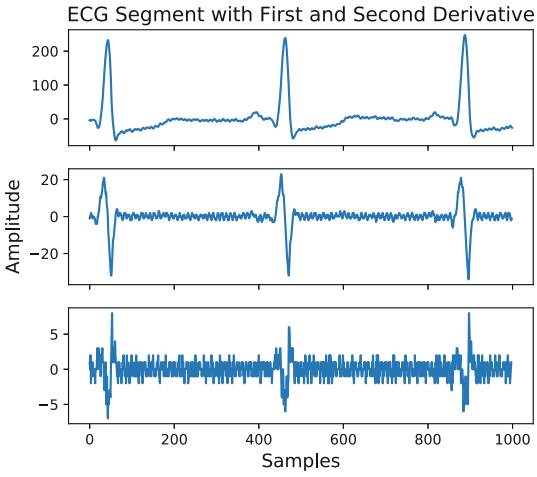

Fig. 1. Plot showing first and second order derivatives of an ECG signal segment taken from the SepsiVit dataset.

In order to obtain the frequency distribution of each derivative, a 20-bin frequency histogram is computed for each of the 6 derivative signals. In order to exclude outliers, the extrema of each histogram are computed as follows. For each of the 6 derivative signals, the minimum and maximum values are collected across the whole dataset, for a total of 12 values. A 95% interval is then calculated for each of the 12 resulting lists of values. The lowest value in the 95% interval was chosen for the minimum of each histogram, while the maximum value in the 95% interval was chosen for the maximum of each histogram. The values found with this method are reported in Table 1.

The result was six 20-bin histograms, three for the first derivative of ECG, Plethysmograph, and Respiratory Rate, and three for their second derivatives, for each 5-minute long data segment. Each of these histograms was then centered (by subtracting the mean) and scaled (by dividing by the standard deviation).

These six histograms were then concatenated so that the first three vectors were the histograms of the first derivative of ECG, Plethysmograph, and Respiratory Rate histograms, while the last three were the histograms of the second derivatives in the same order.

The last step of the feature extraction process involved, for the ECG signal contained in each of the data triplets, extracting the mean and the standard deviation of the Heart Rate, $\mu(HR)$ and $\sigma(HR)$. These two values were appended to each concatenated frequency histogram vector to produce a 122-dimensional feature vector. Only patients that had at least one uninterrupted 5-minute long window containing all three bio-signals were included in this procedure. This feature extraction method yielded 14,389 feature vectors from 89 different patients. Out of the total number of data triplets, 50.8% came from patients marked as 'deteriorating'.

3.2 Δ of Histograms of Derivatives

The second feature extraction approach is largely based on the one described in Subsect. 3.1. The objective of this method is to obtain a measure of the change between the HODs of consecutive 5-minute long data triplets. Initially all pairs of consecutive 5-minute long data triplets are collected, so that in each pair the second triplet directly follows the first one in the time domain. The two 122-dimensional feature vectors for both data triplets are then extracted according to the procedure described in Subsect. 3.1. The final feature vector is then computed as the element-wise difference between the two vectors as:

$$fv_{\Delta} = fv_t - fv_{t-1} \tag{2}$$

where fv_{t-1} and fv_t are the feature vectors extracted from the first and second data triplets respectively. Only patients that had at least one uninterrupted 10-minute long window containing all three bio-signals were included in this procedure. This feature extraction procedure yielded 13,110 feature vectors from 88 different patients. Out of the total number of data triplets, 50.5% came from patients marked as 'deteriorating'.

3.3 Wavelet Transform and Autoregressive Modelling

The last feature extraction procedure involves using the wavelet transform and autoregressive modelling on exclusively the ECG signal. This approach relies on extracting morphological features from individual heart beats, replicating the approach found in [12]. This procedure required a preprocessing step of noise removal from the ECG signal and extraction of all available heart beats (done with the Python package Biosppy 0.5.1), where the R-peaks were detected using Hamilton's approach [19]. Each heart beat is extracted in the form of an array of 300 samples, where the R-peak occurs at the 100^{th} sample. An example of a series of extracted heart beats is shown in Fig. 2.

Table 1. Extrema of each of the 6 frequency histograms, computed for the SepsiVit dataset by considering the 95% interval for each minimum and maximum value in each derivative signal.

	1^{st} derivative		2^{nd} derivative	
	Min	Max	Min	Max
ECG	−348	343	−307	307
Pleth.	−756	768	−511	518
Resp.	−681	722	−523	676

Due to memory limitations of the computer used when running the Machine Learning algorithms, a sample of 10,000 heart beats was selected for each patient to be used in the study. The sample of heart beats for each patient was selected by (1) extracting all heart beats for that patient, and (2) keeping 10,000 evenly spaced heart beats across all heart beats of the patient ordered in the time domain. This was done to ensure that, for each patient, heart beats from all stages of their stay in the hospital were available.

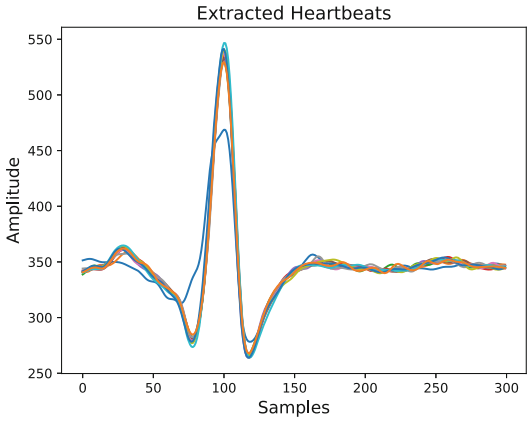

Fig. 2. Plot showing exemplar heart beats extracted from an ECG segment taken from the SepsiVit dataset, after noise removal has been applied. The different colors represent the different heart beats. (Color figure online)

A time-frequency decomposition of each heart beat was then produced using the wavelet transform as done in [12], which has been shown to be a good tool for QRS complex detection [20].

The wavelet transform is an operation that represents a signal with a series of coefficients which describe the energy distribution of the signal across both time and frequency. The continuous wavelet transform (CWT) of a continuous signal is defined as [21]:

$$CWT_x(b,a) = \frac{1}{\sqrt{|a|}} \int_{-\infty}^{\infty} x(t) g\left(\frac{t-b}{a}\right) dt \tag{3}$$

where the wavelet $g(t)$ satisfies the conditions reported in [22]. a and b ($a, b \in \Re, a \neq 0$) are the dilation and translation parameters. The chosen wavelet, which in the case of this study is the Daubechies wavelet of order 8, as done by Qibin and Liqing [12], is compressed or expanded depending on the value of a, in such a

way that coefficients can be extracted to describe the morphology of the signal at different frequency ranges. The high computational complexity of this approach can be reduced by discretising one or both parameters of the function. The case where a is discretised is defined as the dyadic wavelet transform D_yWT. a is discretised along the dyadic sequence 2^i ($i \in \mathbb{N}$) [20]. D_yWT is then defined as:

$$D_yWT_x(b, 2^i) = \frac{1}{\sqrt{2^i}} \int_{-\infty}^{\infty} x(t) g\left(\frac{t-b}{2^i}\right) dt \tag{4}$$

The dyadic wavelet transform was consequently applied to all heart beat signals (done with the Python package pywt 1.0.6 [23]). A required parameter for the operation was the decomposition level, which influences the frequency ranges extracted from the signal. The chosen decomposition level was 4 as done in [12]. The wavelet transform decomposition yielded four detail coefficients d_1, d_2, d_3, d_4 and the vector of approximation coefficients a_4. The detail coefficients represent the high frequency parts of the ECG signal, while the vector of approximation coefficients a_4 represent the lower frequency changes in each heart beat, corresponding with the main features of the QRS complexes. For each heart beat, the vector a_4 contained 32 points.

The second step was the extraction of the coefficients of an autoregressive model trained on each heart beat. An autoregressive model of order p of a signal $x[n]$ is defined as the linear combination of the p previous samples in the signal, and can be expressed as:

$$x[n] = \sum_{i=1}^{p} a[i]x[n-i] + e[n] \tag{5}$$

where $a[i]$ is the i^{th} coefficient and $e[n]$ is white noise with mean zero [12]. The number of coefficients p was chosen to be 14 using the Akaike Information criterion [24], so that the 14 coefficients a_{ar} of the autoregressive model were extracted from each heart beat (done with the Python package statsmodels 0.9). The two obtained vectors $a_4 = \{w_1, \ldots, w_{32}\}$ and $a_{ar} = \{a_1, \ldots, a_{14}\}$ were then concatenated to form the feature vector for that heart beat. Only patients whose ECG signal contained at least one heart beat detectable using Hamilton's approach [19] were included in this procedure. This feature extraction procedure yielded 1,155,997 feature vectors from 123 different patients. Out of the total number of data triplets, 44.9% came from patients marked as 'deteriorating'.

Due to the large number of feature vectors obtained with this method, Principal Component Analysis (PCA), a common feature reduction procedure, was used to compress the dimensionality of the feature vectors from 46 to 10 dimensions [25]. PCA involves projecting a set of vectors across the dimension with the maximal variance, in order to reduce the number of dimensions while preserving the maximal amount of information regarding the distribution of the vectors. For each test, PCA was applied by fitting it on the training split of the data, and then applying it to both the training and the testing splits of the data.

4 Machine Learning Methods

All algorithms described in this section were implemented in Python using the package scikit-learn 0.19.1 [26]. The dataset was split into training and testing/validation sets using 90% and 10% of the data respectively. The strategy used for splitting the dataset was group 10-fold cross-validation, so that 10 iterations of testing were performed for each algorithm. An important property of the group k-folds strategy for dataset splitting is that no data from the same patient occurred in different folds, so as to eliminate overfitting over single patients. The results as reported in Sect. 5 consist of the mean classification accuracy for the tuned models across the 10 training iterations, along with its standard deviation. The accuracy was computed as the number of correct classifications over all classification attempts. For the Linear Support Vector Machine, weighted k-Nearest Neighbors, and Multi-Layer Perceptron, the data must be scaled. A MinMax scaler, which scales each feature to an interval $[0, 1]$, was chosen experimentally as it yielded better results compared to a standard scaler. For each training fold the scaler was fitted on the training split of the dataset, and consequently applied to both the training and the testing split. Class scaling was applied to the two classes in the training phase for all classifiers except for the Multi-Layer Perceptron and the Weighted k-Nearest Neighbors, in order to normalise the impact

Table 2. Parameters used for each of the classifiers. The feature extraction methods are, in order: Histograms of Derivatives (HOD, see Subsect. 3.1), Difference of Histograms of Derivatives (HOD_Δ, see Subsect. 3.2), wavelet transform and autoregressive modelling (HB, see Subsect. 3.3), and using the HRV measures extracted as part of the SepsiVit study (SV). The classifiers are, in order: Linear Support Vector Machine (SVM), Random Forest (RF), Gradient Boosting Machine (GBM), Weighted k-Nearest Neighbors (WkNN), Multi-Layer Perceptron (MLP), and Linear Regression (LR).

		HOD	HOD_Δ	HB	SV
SVM	C	11	12	15	9.5
RF	n_estimators	7,000	5,000	3,500	5,000
GBM	n_estimators	10,000	10,000	10,000	10,000
	learning_rate	0.01	0.01	0.005	0.0001
	min_samples.	10			
WkNN	n_neighbors	6	11	251	55
	p	1			
MLP	hidden_n.	31	53	7	4
	learning_rate	0.0005	0.0005	0.0005	0.001
	max_iter	3,000			
	activation	*logistic*			
LR	C	15	8	10	15
	solver	*newton-cg*			
	multi_class	*multinomial*			

of the distribution of the two classes during training. The parameter tuning for all algorithms was done by parameter grid search using cross-validation. The parameters for all algorithms are reported in Table 2.

4.1 Linear Support Vector Machine

Support Vector Machines (SVMs) are a set of supervised learning algorithms useful in classification, which is widely and successfully applied in the medical field [12,27,28]. A Linear Support Vector Machine generates a hyperplane which position and orientation is optimised to best differentiate between the two classes, and which is computed using the support vectors, which are the vectors in the training set closest to the decision hyperplane [29]. The Linear SVM model used the squared hinge loss function, which produced a classification boundary with a soft margin, yielding classification probabilities. The only tuned parameter was C, which represents the importance given to outliers during training.

4.2 Random Forest

A Random Forest is an ensemble-based algorithm which works as a combination of decision tree predictors [30]. Each tree in a Random Forest is initialised using the values of a random vector sampled independently using the same distribution. This method is more robust to overfitting compared to standard decision trees [31]. All default parameters were kept the same as the scikit-learn implementation of the algorithm [26], except for $n\_estimators$, the number of trees to be generated. As the number of trees is increased, the accuracy normally increases and eventually plateaus. In the case of the wavelet transform and autoregressive modelling feature extraction method (see Subsect. 3.3), the number of generated trees was artificially kept low to accomodate for the memory limitations of the computer used in the analysis.

4.3 Gradient Boosting Machine

The Gradient Boosting Machine algorithm is, much like the Random Forest, an ensemble-based algorithm used in classification which combines a number of weak decision tree classifiers into a strong decision tree classifier. Each decision tree is generated by combining the previous decision trees and applying a higher weight to events that are difficult to predict. The result is a gradient descent algorithm that minimizes the classification error by generating more decision trees [32]. The two parameters that were tuned for this algorithm were $n\_estimators$, the number of trees to be generated, and $learning\_rate$, which shrinks the contribution of each tree. There is a trade-off between the values of the two parameters, so they need to be adjusted to each other. For all other parameters, the defaults of the scikit-learn package were used, except for the value of $min\_samples\_leaf$, which was set to 10. This value defines the minimum number of feature vectors to be found in each leaf of the decision trees.

4.4 Weighted k-Nearest Neighbors

The Weighted k-Nearest Neighbors (WkNN) algorithm is a variation of the standard k-Nearest Neighbors classification algorithm. The latter works by, for each feature vector in the testing set, producing a majority vote across the k closest feature vectors of the training set, according to a specified distance metric. The WkNN algorithm works in a similar fashion, with the added feature that votes from each neighboring feature vector are scaled depending on their distance from the feature vector to be classified [33]. The tuned parameter was only *n_neighbors*, which is k, the number of the closest feature vectors that are taken into account for the classification. The distance metric used for this algorithm was the Minkowski distance, with the inverse scaling factor p set to 1.

4.5 Multi-Layer Perceptron

The Multi-Layer Perceptron (MLP) is a type of feedforward artificial neural network which implements the backpropagation supervised learning algorithm. The MLP implemented as a part of this study contained only one hidden layer. The amount of neurons in the hidden layer was the parameter *hidden_neurons*, tuned for each feature extraction method. The final, output layer contains a number of neurons equal to the number of classes, to which activations a Softmax function is applied in order to compute class-wise probabilities. The *learning_rate* parameter was also tuned using cross-validation [31,34]. All other parameters were kept to the defaults given by scikit-learn, except for the applied logistic activation function, and the maximum number of training iterations for the algorithm, which was set to 3,000.

4.6 Naïve Bayes Classifier

The Naïve Bayes classifier is one of the simplest probabilistic classifiers, which has the advantage of being computationally inexpensive, and has been used with success on Heart Rate Arrhythmia classification in [35]. This classifier constructs a set of probabilities, which correspond to the probability that each feature value appears among the feature vectors within a certain class. The Naïve Bayes classifier makes, however, a strong assumption of conditional independence between the features within the feature vectors [36]. This assumption rarely holds in real life scenarios, and it clearly doesn't hold for the feature vectors extracted with the procedures described in Sect. 3. For this study, the Gaussian Naïve Bayes classifier was used, which relies on the assumption that the likelihood of the features follows a Gaussian distribution. The algorithm was tested as it tends to perform well in many classification tasks, and because of its conveniently low computational complexity. This classifier requires only the prior probabilities of the two classes, computed as the proportion of each class across each complete processed dataset.

4.7 Logistic Regression

The Logistic Regression classifier is a standard linear model for classification. In this study, a multinomial logistic regression was used, which means that the probability estimates should be better calibrated per class compared to a dichotomous implementation. The classifier used the 'newton-cg' solver. The only parameter tuned using cross-validation was C, the inverse of the regularization strength α.

5 Experiments and Results

For each tuned classifier and for every testing procedure, the mean and standard deviation of the classification accuracy across the 10 folds of the cross-validation process are reported. The testing procedures were five in total. The first three involved standard classification of the feature vectors obtained with the three feature extraction methods described in Sect. 3 using cross-validation. For each of the three produced datasets, each feature vector was assigned the same label as the patient that it was extracted from. During the training phase, the classifier was trained on the training set using the correct labels. During the testing phase, each feature vector was classified as belonging to the 'deteriorating' class or to the 'healthy' class. The result of the classification was then compared with the correct label in order to compute the accuracy (i.e. the proportion of correct classifications during the testing phase).

The last two testing procedures were applied to the morphology descriptors, which are described in Subsect. 3.3). For both testing procedures, the training phase was the same as for the third testing procedure, so that the classifier could classify each heart beat as 'deteriorating' or not given its feature vector. What changed in the last two testing procedures was the testing phase. The first of the two testing procedures was done as a majority vote, where heart beats are extracted and processed for all 5-minute long ECG segments. The classification process is then applied to all heart beats in each 5-minute long ECG segment so that if 50% or more of the heart beats are classified as 'deteriorating', then the whole segment receives such classification outcome. The third testing procedure is performed in a similar fashion by taking a majority vote across 12 5-minute long ECG segments.

All testing procedures are compared to the performance of the tuned algorithms used on the HRV features extracted as part of the SepsiVit study, as mentioned in Sect. 3. All outcomes of the testing procedures are reported in Table 3.

The Histograms of Derivatives and Differences of Histograms of Derivatives methods for feature extraction did not show any promise, ranging from a mean classification accuracy of $43.1 \pm 11.9\%$ for the Multi-Layer Perceptron in the Difference of Histograms of Derivatives procedure, to $56.6 \pm 12\%$ for the Random Forests algorithm applied to the Histograms of Derivative method for feature extraction.

The best results were obtained using the Linear Support Vector Machine on the feature vectors extracted in the SepsiVit study, which had a mean accuracy

Table 3. Mean and standard deviation of the classification accuracies for all models and testing procedures. The testing procedures are, in order: Histograms of Derivatives (HOD, see Subsect. 3.1), difference of Histograms of Derivatives (HOD_Δ, see Subsect. 3.2), wavelet transform and autoregressive modelling without majority vote (HB, see Subsect. 3.3), wavelet transform and autoregressive modelling applied in a majority vote fashion over 5-minute long ECG segments (MV), wavelet transform and autoregressive modelling applied in a majority vote fashion over 12 5-minute long ECG segments (MV_2), and using the HRV measures extracted as part of the SepsiVit study (SV).The classifiers are, in order: Linear Regression (LR), Weighted k-Nearest Neighbors (WkNN), Naïve Bayes (NB), Linear Support Vector Machine (SVM), Multi-Layer Perceptron (MLP) Random Forest (RF), and Gradient Boosting Machine (GBM).

	HOD	HOD_Δ	HB	MV	MV_2	SV
LR	54.1 ± 14.3	50.5 ± 7.4	59.3 ± 9.4	60.6 ± 10.8	61.0 ± 10.6	63.0 ± 5.2
WkNN	52.8 ± 6.7	50.4 ± 6.6	55.1 ± 6.3	57.1 ± 10.9	57.8 ± 11.2	57.9 ± 5.8
NB	54.8 ± 13.3	49.7 ± 14.3	51.8 ± 10.7	54.0 ± 15.4	53.9 ± 15.9	57.9 ± 5.8
SVM	52.4 ± 13.9	50.5 ± 12.7	$\mathbf{60.9 \pm 9.1}$	$\mathbf{62.2 \pm 10.7}$	$\mathbf{62.4 \pm 10.9}$	$\mathbf{65.5 \pm 7.9}$
MLP	53.8 ± 11.1	43.1 ± 11.9	59.8 ± 12.9	57.1 ± 15.2	56.9 ± 15.9	60.3 ± 8.1
RF	$\mathbf{56.3 \pm 12}$	$\mathbf{54.8 \pm 6.7}$	55.4 ± 7.8	58.2 ± 12.2	58.5 ± 12.8	59.3 ± 6.9
GBM	54.6 ± 8.4	54.4 ± 9.0	57.6 ± 7.8	61.5 ± 13.1	61.9 ± 13.6	61.3 ± 8.5

of 65.5% and a standard deviation of 7.9%. The most promising results were obtained with the feature extraction method involving the wavelet transform and autoregressive modelling, which was only marginally improved by the majority vote testing procedures. The Linear Support Vector Machine classifier produced the best results with the data extracted in this fashion, peaking at $62.4 \pm 10.9\%$ mean classification accuracy.

Overall, the Linear Support Vector Machine was the best classifier, sometimes beaten by the Random Forest.

6 Conclusion and Future Work

The results presented in the previous section show that none of the attempted feature extraction methods are superior in their ability to encapsulate differences between the two classes and similarity among the same class compared to the HRV features extracted as part of the SepsiVit study [11]. Nonetheless, the results of this study imply that there is more useful information in the morphological descriptions of the ECG signal compared to the frequency distributions of the slopes of high frequency bio-signals.

While there was an increase in classification accuracy obtained by applying the majority vote testing strategies, the fact that the improvement was as small as 1.5% indicates that the improvement is only marginal, and given the benefits of early detection of sepsis induced deterioration [6], a classification strategy requiring less data such as the standard heart beat classification or the majority

vote across 5-minute ECG segments might be more beneficial for improving survival rates, compared to one that uses 60-minute ECG segments.

A difficulty encountered in this study was the limited size of the dataset. The low variability in the bio-signals across the data of each individual patient makes it so that the diversity in the dataset, and so the capacity of the Machine Learning algorithms to properly generalise the problem, is entirely dependent on the amount of different patients included in the study. Since reaching the target of the SepsiVit study of 171 patients (i.e. only 30% more than were available for this research) is likely not going to produce sufficient diversity in the dataset, future data collection programs are needed to further investigate the predictive potential of high frequency bio-signals for early detection of sepsis induced deterioration.

Future studies could focus on any of the following points for improvement. A more complete analysis of the feature extraction methods should be carried out: new strategies should be tested, and all strategies should be used together to produce feature vectors containing all features for each bio-signal segment. An analysis of which features contribute the most to the classification would then reveal the features that are most relevant towards the early detection of sepsis induced deterioration. Furthermore, different classifiers should be tested. Obvious candidates are Recurrent Neural Networks such as LSTMs, widely used on time series data, which nevertheless require large amounts of data for effective training, and which as such would depend on a new data collection program.

References

1. Singer, M., et al.: The third international consensus definitions for sepsis and septic shock (sepsis-3). JAMA **315**(8), 801–810 (2016). 26903338[pmid]
2. Bone, R.C., Fisher, C.J., Clemmer, T.P., Slotman, G.J., Metz, C.A., Balk, R.A.: Sepsis syndrome: a valid clinical entity. Methylprednisolone severe sepsis study group. Crit. Care Med. **17**(5), 389–393 (1989)
3. Buchan, C.A., Bravi, A., Seely, A.J.E.: Variability analysis and the diagnosis, management, and treatment of sepsis. Curr. Infect. Dis. Rep. **14**(5), 512–521 (2012)
4. Danai, P., Martin, G.S.: Epidemiology of sepsis: recent advances. Curr. Infect. Dis. Rep. **7**(5), 329–334 (2005)
5. Glickman, S.W., et al.: Disease progression in hemodynamically stable patients presenting to the emergency department with sepsis. Acad. Emerg. Med. **17**(4), 383–390 (2010)
6. Brindley, P.G., Zhu, N., Sligl, W.: Best evidence in critical care medicine early antibiotics and survival from septic shock: it's about time. Can. J. Anesth./Journal canadien d'anesthésie **53**(11), 1157–1160 (2006)
7. Dellinger, R.P., et al.: Surviving sepsis campaign: international guidelines for management of severe sepsis and septic shock 2012. Crit. Care Med. **41**(2), 580–637 (2013)
8. Moorman, J.R., et al.: Mortality reduction by heart rate characteristic monitoring in very low birth weight neonates: a randomized trial. J. Pediatrics **159**(6), 900–906.e1 (2011)
9. Ahmad, S., et al.: Continuous multi-parameter heart rate variability analysis heralds onset of sepsis in adults. PLoS ONE **4**(8), 1–10 (2009)

10. Bravi, A., Green, G., Longtin, A., Seely, A.J.E.: Monitoring and identification of sepsis development through a composite measure of heart rate variability. PLoS ONE **7**(9), e45666 (2012). PONE-D-12-18432[PII]
11. Quinten, V.M., van Meurs, M., Renes, M.H., Ligtenberg, J.J.M., ter Maaten, J.C.: Protocol of the SepsiVit study: a prospective observational study to determine whether continuous heart rate variability measurement during the first 48 hours of hospitalisation provides an early warning for deterioration in patients presenting with infec. BMJ Open **7**(11), e018259 (2017)
12. Zhao, Q., Zhang, L.: ECG feature extraction and classification using wavelet transform and support vector machines. In: 2005 International Conference on Neural Networks and Brain, vol. 2, pp. 1089–1092, October 2005
13. Levy, M.M., et al.: 2001 SCCM/ESICM/ACCP/ATS/SIS international sepsis definitions conference. Crit. Care Med. **31**(4), 1250–1256 (2003)
14. Cardoso, J.F., Laheld, B.H.: Equivariant adaptive source separation. IEEE Trans. Signal Process. **44**(12), 3017–3030 (1996)
15. Peltola, M.: Role of editing of R-R intervals in the analysis of heart rate variability. Front. Physiol. **3**, 148 (2012)
16. Shaffer, F., Ginsberg, J.P.: An overview of heart rate variability metrics and norms. Front. Public Health **5**, 258 (2017). 29034226[pmid]
17. Moridani, M.K., Setarehdan, S.K., Nasrabadi, A.M., Hajinasrollah, E.: Non-linear feature extraction from HRV signal for mortality prediction of ICU cardiovascular patient. J. Med. Eng. Technol. **40**(3), 87–98 (2016). PMID: 27028609
18. Dalal, N., Triggs, B.: Histograms of oriented gradients for human detection. In: 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR 2005), vol. 1, pp. 886–893, June 2005
19. Hamilton, P.: Open source ECG analysis. Comput. Cardiol. **29**, 101–104 (2002)
20. Kadambe, S., Murray, R., Boudreaux-Bartels, G.F.: Wavelet transform-based QRS complex detector. IEEE Trans. Biomed. Eng. **46**(7), 838–848 (1999)
21. Morlet, J., Arens, G., Fourgeau, E., Glard, D.: Wave propagation and sampling theory - Part i: complex signal and scattering in multilayered media. Geophysics **47**(2), 203–221 (1982)
22. Grossmann, A.: Wavelet transforms and edge detection. In: Albeverio, S., Blanchard, P., Hazewinkel, M., Streit, L. (eds.) Stochastic Processes in Physics and Engineering, pp. 149–157. Springer, Dordrecht (1988). https://doi.org/10.1007/978-94-009-2893-0_7
23. Lee, G., et al.: Pywavelets - wavelet transforms in Python (2006). Accessed 2018
24. Akaike, H.: Information theory and an extension of the maximum likelihood principle. In: Parzen, E., Tanabe, K., Kitagawa, G. (eds.) Selected Papers of Hirotugu Akaike, pp. 199–213. Springer, New York (1998). https://doi.org/10.1007/978-1-4612-1694-0_15
25. Jolliffe, I.: Principal component analysis. In: Lovric, M. (ed.) International Encyclopedia of Statistical Science, pp. 1094–1096. Springer, Heidelberg (2011). https://doi.org/10.1007/978-3-642-04898-2
26. Pedregosa, F., et al.: Scikit-learn: machine learning in Python. J. Mach. Learn. Res. **12**, 2825–2830 (2011)
27. Li, Q., Rajagopalan, C., Clifford, G.D.: Ventricular fibrillation and tachycardia classification using a machine learning approach. IEEE Trans. Biomed. Eng. **61**(6), 1607–1613 (2014)
28. Song, M.H., Lee, J., Cho, S.P., Lee, K.J., Yoo, S.K.: Support vector machine based arrhythmia classification using reduced features. Int. J. Control Autom. Syst. **3**(4), 571–579 (2005)

29. Hearst, M.A., Dumais, S.T., Osuna, E., Platt, J., Schölkopf, B.: Support vector machines. IEEE Intell. Syst. Appl. **13**(4), 18–28 (1998)
30. Breiman, L.: Random forests. Mach. Learn. **45**(1), 5–32 (2001)
31. Hastie, T., Tibshirani, R., Friedman, J.H.: The Elements of Statistical Learning. Springer, New York (2009). https://doi.org/10.1007/978-0-387-84858-7
32. Friedman, J.H.: Greedy function approximation: a gradient boosting machine. Ann. Stat. **29**(5), 1189–1232 (2001)
33. Hechenbichler, K., Schliep, K.: Weighted k-nearest-neighbor techniques and ordinal classification (2004). Accessed 2018
34. Kriesel, D.: A brief introduction to neural networks (2007)
35. Soman, T., Bobbie, P.O.: Classification of arrhythmia using machine learning techniques. WSEAS Trans. Comput. **4**, 548–552 (2005)
36. Chan, T.F., Golub, G.H., LeVeque, R.J.: Updating formulae and a pairwise algorithm for computing sample variances. In: Caussinus, H., Ettinger, P., Tomassone, R. (eds.) COMPSTAT 1982 5th Symposium Held at Toulouse 1982, pp. 30–41. Physica-Verlag, Heidelberg (1982)