



## Research article

# Can the application of certain music information retrieval methods contribute to the machine learning classification of electrocardiographic signals?



Ennio Idrobo-Ávila<sup>a,\*</sup>, Humberto Loaiza-Correa<sup>a</sup>, Rubiel Vargas-Cañas<sup>b</sup>, Flavio Muñoz-Bolaños<sup>c</sup>, Leon van Noorden<sup>d</sup>

<sup>a</sup> PSI - Percepción y Sistemas Inteligentes, Escuela de Ingeniería Eléctrica y Electrónica, Universidad del Valle, Cali, Colombia

<sup>b</sup> SIDICO – Sistemas Dinámicos de Instrumentación y Control, Departamento de Física, Universidad del Cauca, Popayán, Colombia

<sup>c</sup> CIFIEX - Ciencias Fisiológicas Experimentales, Departamento de Ciencias Fisiológicas, Universidad del Cauca, Popayán, Colombia

<sup>d</sup> IPEM - Institute for Systematic Musicology, Department of Art, Music and Theatre Sciences, Ghent University, Ghent, Belgium

## ARTICLE INFO

## Keywords:

ECG signal classification

Heart rate variability

PhysioNet physiological signals database

Music

Neural networks

## ABSTRACT

The electrocardiogram is traditionally used to diagnose a large number of heart pathologies. Research to improve the readability and classification of cardiac signals includes studies geared toward sonification of the electrocardiographic signal and others involving features related to music processing, such as Mel-frequency cepstral coefficients. In terms of music processing features, this study seeks to use music information retrieval (MIR) features as electrocardiographic signal descriptors. The study compares the discriminatory capability of the introduced features in relation to standard groups such as heart rate variability, wavelet transform, descriptive statistics, Mel coefficients and fractal analysis, evaluated using classification algorithms; the signals analyzed were extracted from public databases. The group of features extracted from wavelet transform and the MIR group showed a high level of discrimination; the best representation of the ECG signals in the study was achieved in most cases by the MIR features. Moreover, a correlation coefficient higher than 0.8 was found between a number of MIR and other feature groups, indicating a likely relationship between the electrocardiographic signals and MIR features. These results suggest the feasibility of representing the analyzed signals by music information retrieval descriptors, giving the potential to consider these electrocardiographic signals as analogues to musical signals.

## 1. Introduction

The electrocardiogram (ECG) is a medical examination that allows recording the electrical activity of the heart using contact electrodes placed in specific positions of the chest and extremities, together with a monitoring system [1]. It is one of the main physiological measures for medical diagnosis, analysis and monitoring of cardiac function [2]. Since it is related to the mechanical pumping of the heart, it provides information on its physiological state [2]. ECG represents one of the most widely used diagnostic tools in the health field, and is used in different situations from control in healthy subjects [3] to monitoring of patients in intensive care [3]. The high frequency of ECG use is due to the increase in the prevalence and incidence of cardiovascular diseases (CVD), where physicians and cardiologists have to read and analyze large numbers of

ECG records, an increasingly challenging task. It is important to note that CVD represents the leading cause of death in the world [4], with more than 17 million deaths per year in 2016, and expected to exceed 23 million in 2030 [5]. Therefore, in recent years, research has been conducted concerning ECG applications, including studies related to signal processing [6, 7], machine learning [8], ordinal statistics of patterns and symbolic dynamics [9], or using morphological and dynamic features of ECG signals [10].

In relation to ECG classification, applications and analysis, other studies looked to relate music to ECG signals. The aim of these studies is the sonification of the ECG signal [11, 12]. In this regard, some features related to music processing, such as Mel-frequency cepstral coefficients (MFCC) have been used with promising results [13]. At this point, it is worth mentioning that MFCCs are used as timbre descriptors within

\* Corresponding author.

E-mail address: [ennio.idrobo@correounivalle.edu.co](mailto:ennio.idrobo@correounivalle.edu.co) (E. Idrobo-Ávila).

music information retrieval (MIR) [14, 15], a field usually referred to as analysis of audio content, whose objective is the extraction of information from audio signals - digital recordings of music and audio in general [16]. MIR provides information about recognition of musical instruments, classification of musical phrases or melodies, and rhythm and high level-based music retrieval [17]. MIR analysis is also focused on recognition of emotions in music [18] and research continues concerning ECG and emotion in relation to music. Applications have been developed in which heart rate emotion data was used for generating music [19]. More recent research continues to propose methods for recognizing emotions elicited by music through ECG [20].

Even though prior research has explored a limited number of digital music processing features as ECG signal descriptors, until now the methods implemented only allowed ECG signal characterization via MFCC features using public databases in order to classify ECG as normal or abnormal. Therefore, studies in which more elements of musical analysis are incorporated must still be considered to see whether MIR features can allow better discrimination between ECG signals [13]. This study hypothesizes that ECG signals can indeed be represented through MIR features, which could serve to discriminate between different cardiopathies. In this paper, an approach that uses MIR as ECG features is presented to classify between inferior and anterior myocardial infarction, T-wave alternation, and normal ECG signals. MIR features of different natures are considered and compared with those features commonly used in cardiac signal analysis as features related to descriptive statistics [21], wavelet coefficients [22], heart rate variability [23], and fractal analysis [24]. In order to compare the descriptive capacity of each set of features, different machine learning algorithms were used, and the performance obtained with each set was compared. A correlation analysis was performed to associate the MIR features with those already known in the context of the ECG signals.

The remainder of this document is organized as follows: Section 2 presents a description of the databases used, along with the stages of the experimental procedure; Section 3 reports the results, in which the best outcomes can be observed; Section 4 covers discussion of the results; and finally, Section 5 shows the conclusions of this study and some ideas for future research.

## 2. Materials and methods

### 2.1. Database description

This study makes use of ECG signals from two publicly available databases - the MIT-BIH and the C-database, both belonging to the PhysioNet databases [25]. The MIT-BIH database [26] was included since it was considered in previous research to study MFCC coefficients as descriptors of ECG signals [13] and thus enabled comparisons to be made. The C-database was included as a challenge to the MIR features since this

database contains signals with more subtle differences between its classes, which poses difficulties for the classification stage. In both databases, Lead II of the ECG signals was selected since this lead registers the inferior electrical activity of the heart [27].

The MIT-BIH database is composed of signals belonging to three classes, arrhythmia (ARR), congestive heart failure (CHF), and normal sinus rhythm (NSR). These classes comprised 30 instances, each with a 512-second duration and sampling frequency of 128 Hz. The C-database involved the PTB Diagnostic ECG Database (PTB) [25, 28, 29] where signs of inferior and anterior myocardial infarction and normal signals were used. Additionally, the T-Wave Alternans Challenge Database (TWA) [25, 30, 31] was included where T-wave alternating signals were used. Thus, C-database was constructed with four classes: normal signals (N), signals with alternating T-wave (T-wave), signals with inferior myocardial infarction (IMI) and signals with anterior myocardial infarction (AMI). The classes N, T-wave, IMI and AMI consisted of 48, 41, 38, and 39 instances respectively, 166 instances in total. All signals are of 60 s duration and a sampling frequency of 1000 Hz, except for T-wave class signals, which were acquired with a sampling frequency of 500 Hz.

### 2.2. Experimental procedure

Once ECG derivations and anomalies were chosen and classified, the general experimental procedure was proposed. This procedure consisted of four stages: preprocessing, dataset augmentation, feature extraction and classification (Figure 1). The feature extraction stage was divided into two substages: extraction and selection, the latter to find out how each feature contributes to description of the database. For this, a new classification according to this ranking was performed, to observe the changes in the classification metrics with the number of best features; and a correlation between MIR features and other ECG features was also performed, to find any relationship(s) between them.

#### 2.2.1. Preprocessing

The preprocessing stage allows both conditioning the signals and eliminating unwanted elements such as certain frequential components. In this case, the preprocessing stage consisted of two sub-stages: resampling and baseline wander correction.

**2.2.1.1. Resampling.** The objective of the resampling stage is to put all signals of the dataset in as similar conditions as possible, i.e. similar frequential content. All signals in the C-database were resampled to the minimum sample frequency of the original dataset; namely, 500 Hz. The resampling process was not applied to the MIT-BIH database since all the signals had the same sampling rate.

**2.2.1.2. Baseline wander correction.** After resampling the signals, a baseline wander correction process was applied. This procedure was

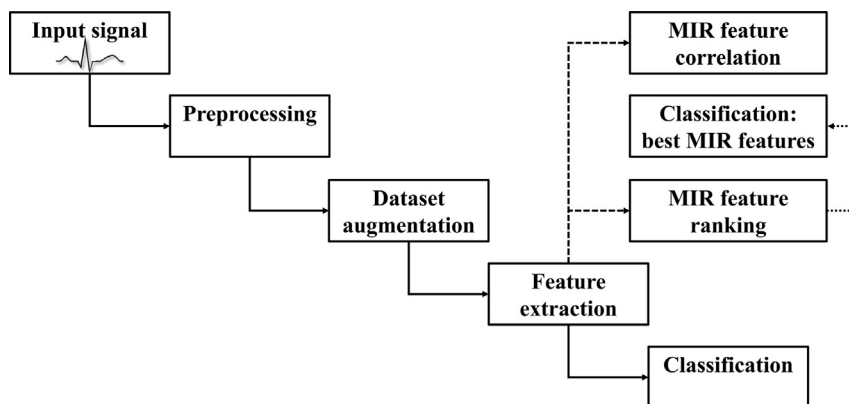
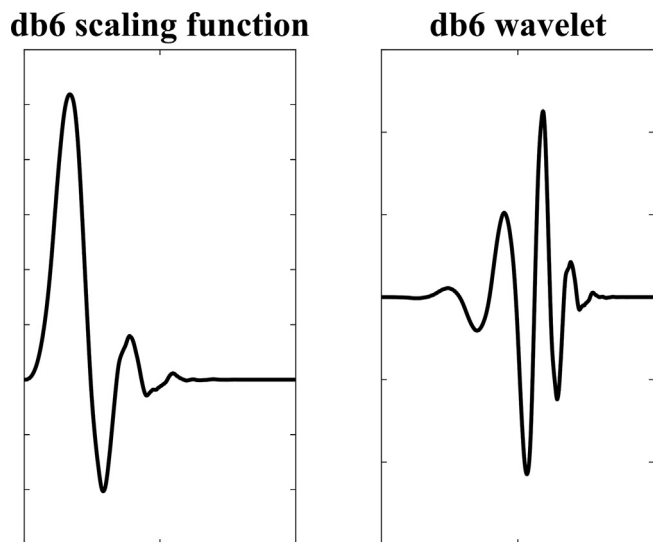


Figure 1. Block diagram of the experimental procedure.

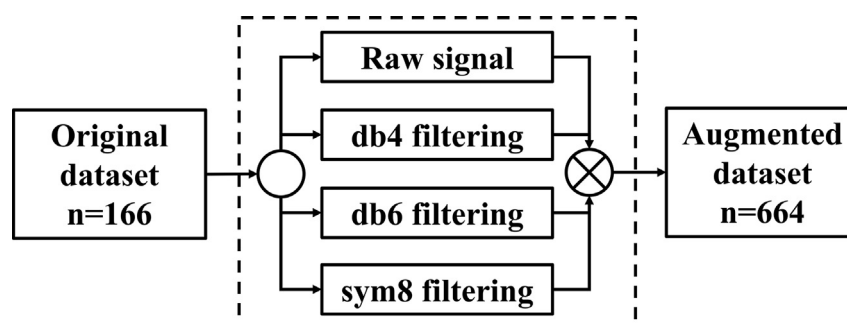


**Figure 2.** Mother wavelet Daubechies 6. For more information on mother wavelets see [33].

performed to remove noise produced by the breathing or movement of subjects [32], which produces problems as it overlaps with the ST segment [32]. Baseline wander correction was applied to both databases using discrete wavelet transform (DWT) with the same filter bank: Daubechies 6 (Figure 2) [33]. This mother wavelet was selected because it is one of the most similar to a normal ECG signal. Wavelet decomposition was performed on ten levels of the input signal, and a frequency band between 0 and 0.49 Hz was removed, corresponding to baseline drift [34]. Following this decomposition, signal reconstruction was performed from approximation coefficients of the last level and all the detail coefficients of all decomposition levels were equaled to zero. As a result of this process, a low-frequency signal represented the baseline of the input signal. Finally, it was necessary to subtract the signal obtained from wavelet analysis of the input signal, in order to correct the baseline.

### 2.2.2. Database augmentation

**2.2.2.1. Wavelet-based shrinkage filtering.** The data augmentation stage is essential to ensure that the machine learning algorithms have enough samples from which to learn. Additional data have to replicate the conditions that were encountered during the acquisition process. The data augmentation process was applied only to the C-database, as the MIT-BIH database contained enough data to train the chosen algorithms. Data augmentation was implemented using the wavelet-based shrinkage filtering method [35], in which three mother wavelets were applied to filter the signals: Daubechies 4 (db4), Daubechies 6 (db6), and Symlets 8 (sym8). Thus, for each raw signal, three more signals were generated (Figure 3).



**Figure 3.** Database augmentation stage.

### 2.2.3. Segmentation

Before the feature extraction stage, a segmentation phase was performed. This process was carried out to analyse the signals in more detail, i.e. short periods of time. The signals were segmented using a Hamming Window. In the MIT-BIH database, the signals were divided into segments of ten seconds, without overlapping, while in the C-database, each signal was divided into nine segments of ten seconds each, with a window overlapping of 40% [36]. This interval was chosen because, in a healthy heart rate range, there are between 50 and 90 bpm (beats per minute). Thus, it is possible to have between 8 to 15 heartbeats within each window [37]. Using this amount of heartbeats, it is possible to establish a pattern behaviour in both the ECG signal and some heart rate variability (HRV) analysis [38, 39].

### 2.2.4. Feature extraction

Once the segmentation stage was carried out, the feature extraction step was performed. This process allows compact representation of data (ECG signals) through features that depict particular behaviours or patterns in the data [40]. From the MIT-BIH database (Table 1), a set of MIR features was extracted, while six groups of features were extracted from the C-database (Table 2) - MIR, Mel-frequency cepstral coefficients (MFCC), descriptive statistics, fractal analysis, HRV, and descriptive statistics from wavelet coefficients. Although MFCC could be considered part of MIR, it was omitted in this MIT-BIH analysis and was considered elsewhere mainly in order to compare the outcomes obtained from MIR features with the MFCC results obtained in [13], through study of the same MIT-BIH database.

In calculating HRV, both temporal and frequency domain features were considered. HRV was computed through detection of R-peaks; R-peaks were identified using the Pan-Tompkins algorithm [45]. Undetected peaks were marked manually. Regarding feature extraction of wavelet coefficients, three levels of decomposition, descriptive statistics for these components, and the Higuchi fractal dimension were considered. Implementation of the algorithms required in this study was carried out using Matlab [46], while MIR feature extraction was performed using MIRtoolbox [44, 47].

### 2.2.5. Feature ranking

A best features ranking was carried out using the information gain ratio [48] with the C-database. This process was performed to observe whether the classification process could be improved by considering only the best features. This process was also performed on the MIR features to determine how much each MIR feature contributes to data representation. From the ranking process, two groups were created, “Best features” (the five best-ranked features from each of the MIR, Statistics, HRV, and Wavelet groups) and “14-MIR” (the 14 best-ranked MIR features). The number 14 was selected according to a classification analysis with the best MIR features (See below: Section 3.5. Classification with best MIR features).

**Table 1.** List of MIR features extracted from ECG signals (26 features) MIT-BIH database.

Features	Description
mean pitch, standard deviation pitch, zero-crossing, low energy rate, tempo, minimum tempo, maximum tempo, mean tempo, standard deviation of tempo, pulse clarity, event density, minimum novelty, maximum novelty, mean novelty, standard deviation of novelty, key, mode, spectral spread, spectral distribution centroid, spectral roll-off, spectral skewness, spectral kurtosis, spectral flatness, spectral regularity, spectral entropy, root mean square	Although, many features were included, a greater influence of tempo-related features was expected due to the rhythmic nature of the heart. The main MIR characteristics considered here include root mean square, tempo, zero-cross, spectral flatness, and spectral spread. Root mean square is related to sound intensity [41]. Tempo is related to the tempo of a musical signal; this is estimated from the detection of periodicities in the signals [42]. Zero-crossing represents the number of changes of sign in consecutive blocks of signals [16]. Zero-crossing has high values for segments with noise and low values for tonal parts or parts with determined frequencies. Its long constant values during constant pitches indicate the presence of a fundamental frequency. Spectral flatness measures the amount of correlation structure that exists in a signal [43]; it also indicates whether the spectral distribution is smooth, associated with noise, or with peaks, associated with a tonal behavior [44]. A spectral flatness close to one represents spectrum with components in all frequency bands, similar to the spectrum of white noise. In contrast, a spectral flatness close to zero represents a spectrum with components in a limited number of frequency bands, such as those present in pure tones or the sum of sinusoidal components. Spectral spread is associated with the standard deviation of the signal spectrum around the spectral centroid; this measure is associated with perception of the timbre of an audio signal [16].

### 2.2.6. Classification

Once the features were extracted and selected, a classification process was carried out to determine whether or not the extracted features can represent particular ECG signals. Classification was implemented using six classical AI algorithms: AdaBoost, CN2 rule inducer, neural network, random forest, decision trees (Tree), and k-nearest neighbors (kNN) (Table 3). These algorithms were applied to both datasets: MIT-BIH and C-database.

**2.2.6.1. Training, test and evaluation of models.** The datasets were split into training and testing sets corresponding to 80 and 20% of the data, respectively. The datasets had a total of 4,590 and 5,976 instances related to MIT-BIH and the C-database, respectively, within which the training and test sets had respectively 3,672 and 918 instances for the MIT-BIH database, and 4,860 and 1,116 instances for the C-database. These two sets were constructed considering that data of subjects in the test set should not be in the training set. Although this approach generally produces low-performance metrics in evaluation because of inter-participant variability [49], it was selected since it is closest to real applications. Cross-validation with five folds was used for training the models. The performance of each classifier was evaluated using the area under the ROC curve (AUC), and accuracy. In the classification process, the performance of the classification algorithms was compared with each group of features, revealing the group of features which best described the ECG signals and the classifiers that achieved the best performance with the selected database.

### 2.2.7. Correlation analysis

Finally, using the Pearson correlation coefficient, an analysis was carried out of correlation between the MIR features and the other features extracted from the C-database, to establish a link between the ECG signal and the MIR features.

## 3. Results

### 3.1. Preprocessing

#### 3.1.1. Baseline correction

Applying the methodology described above, the ECG baseline was subtracted (Figure 4), resulting in enhanced and cleaner signals, with a near-constant dc level.

### 3.2. Database augmentation

Shrinkage filtering with three mother wavelets - Daubechies 4 (db4), Daubechies 6 (db6), and Symlets 8 (sym8) - augmented the database by a factor of 4 (from 166 to 664 instances), each filtering subtly modifying the original signal (Figure 5); e.g. filtering with the sym8 and db6 wavelets was noticeably smoother than with db4.

**Table 2.** List of features extracted from ECG signals, C-database.

Group	Features	Total
MFCC	mfcc1, mfcc2, mfcc3, mfcc4, mfcc5, mfcc6, mfcc7, mfcc8, mfcc9, mfcc10, mfcc11, mfcc12, mfcc13, maximum mfcc, minimum mfcc, mean mfcc, mfcc variance, mfcc skewness, mfcc kurtosis	19
Descriptive statistics	maximum, minimum, mean, variance, skewness, kurtosis, median, mode, energy, entropy	10
Fractal analysis	Higuchi fractal dimension, Katz fractal dimension, Hurst exponent, HRV detrended fluctuation analysis alpha1, HRV detrended fluctuation analysis alpha2	5
HRV	mean R-R interval, root mean square of the successive differences, median of Euclidean distance, interquartile range of Euclidean distance, mean of the heart rate, probability of intervals greater than 50ms, triangular index from the interval histogram, performing triangular interpolation, correlation dimension, approximate entropy, standard deviation1 of the Poincaré plot, standard deviation2 of the Poincaré plot, ratio of standard deviation1 and standard deviation2 of the Poincaré plot, very low-frequency components, low-frequency components, high-frequency components, ratio of low and high-frequency components, power of low-frequency components, power of high-frequency components, total power, HRV detrended fluctuation analysis alpha1, HRV detrended fluctuation analysis alpha2	22
Wavelet coefficients	maximum(cfs0-7), minimum(cfs0-7), mean(cfs0-7), variance(cfs0-7), median(cfs0-7), mode(cfs0-7), energy(cfs0-7), entropy(cfs0-7), Higuchi fractal dimension (cfs0-7)	72
MIR	As described in Table 1	26
<b>Total</b>		<b>154</b>

**Table 3.** Configuration parameters of classical artificial intelligence algorithms implemented.

Classification algorithms	Configuration MIT-BIH database	Configuration C-database
AdaBoost	Base estimator: Tree, Number of estimators: 50, Learning rate: 1, Classification algorithm: SAMME.R	Base estimator: Tree, Number of estimators: 50, Learning rate: 1, Classification algorithm: SAMME.R
CN2 rule inducer	Unordered rule, Evaluation measure: entropy, Beam width: 5, Regression loss function: linear	Unordered rule, Evaluation measure: entropy, Beam width: 5, Regression loss function: linear
Neural network	Multi-layer perceptron with backpropagation. Neurons in hidden layers: 150, Activation: ReLu, Regularization: alpha = 0.002, Solver: Adam	Multi-layer perceptron with backpropagation. Neurons in hidden layers: 5, Activation: ReLu, Regularization: alpha = 3, Solver: L-BFGS-B
Random forest	Number of trees: 20	Number of trees: 10
Decision trees	Induce binary tree, Minimum number of instances in leaves: 7, Limit the maximal tree depth to: 100	Induce binary tree, Minimum number of instances in leaves: 2, Limit the maximal tree depth to: 100
K-nearest neighbors	Number of neighbors: 5, Metric: Manhattan, Weight: Distance	Number of neighbors: 3, Metric: Manhattan, Weight: Distance

### 3.3. MIR feature ranking

Ranking of all MIR features was carried out using the information gain ratio (Table 4). As expected, the tempo features best described these ECG signals.

### 3.4. Classification: evaluation of models

#### 3.4.1. Classification: MIT-BIH database

After training the algorithms with the MIT-BIH database, performance was evaluated using the AUC (Figure 6), and accuracy metrics (Figure 7) in both binary and multiclass classifications. The two classes used were normal sinus rhythm (NSR) and congestive heart failure (CHF), while arrhythmia (ARR) to make three classes. With the exception of CN2 rule inducer for 3 classes, performance across all algorithms surpassed 0.90 AUC and 0.85 accuracy. Neural network, with 0.98 AUC and 0.99 accuracy (2 classes) and 0.96 AUC and 0.99 accuracy (3 classes) outperformed all the others.

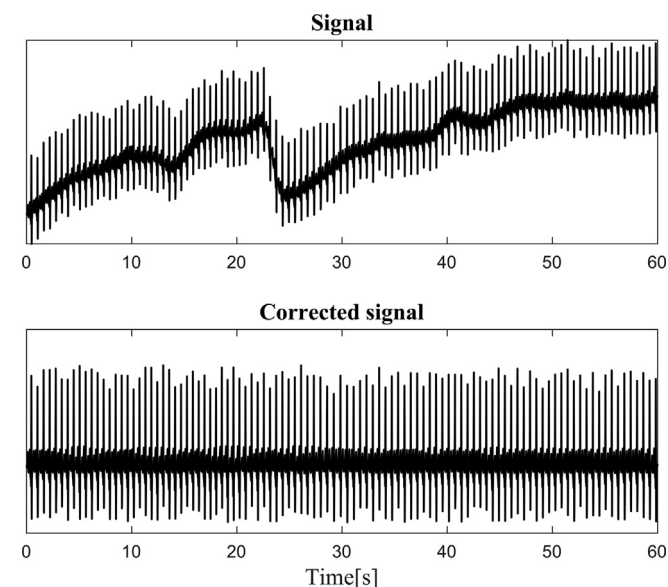


Figure 4. Example of ECG baseline correction.

#### 3.4.2. Classification: C-database

Having trained the algorithms with the C-database, performance was evaluated using AUC and accuracy. AUC again revealed a strong overall performance by neural network (Figure 8). The MIR features, in combination with neural network, outperformed all other groups. The “Best features” group was also seen to perform well. An accuracy value that exceeded 0.7 showed that MIR features in combination with neural network again performed best (Figure 9).

#### 3.5. Classification with best MIR features: C-database

Having ranked the MIR features, analysis of the classification performance was carried out using the MIR feature ranking. In this analysis, neural network was selected, as the best overall performing algorithm from previous results. Observation began with two features (Figure 10). In this performance, the values of AUC and accuracy are observed to be around 0.85 and 0.65 respectively.

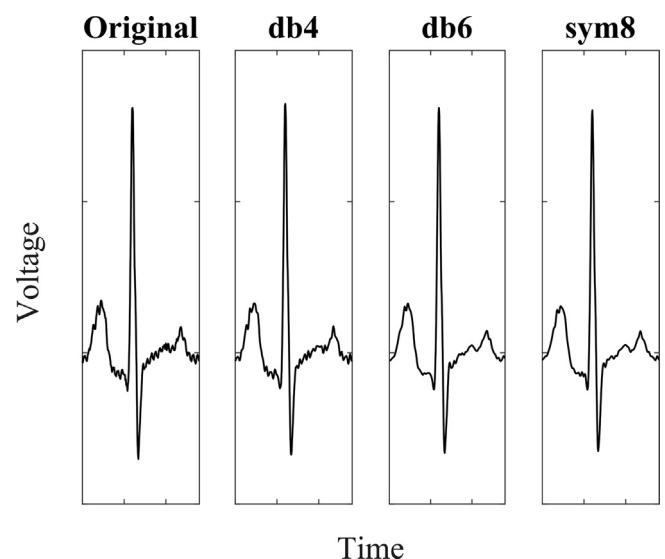


Figure 5. Example of wavelet-based shrinkage filtering.



**Table 4.** Ranking of MIR features using information gain ratio (IGR).

Ranking	MIR feature	IGR	Ranking	MIR feature	IGR
1	mean tempo	0.1181	14	standard deviation of tempo	0.0432
2	tempo	0.1137	15	spectral spread	0.0367
3	minimum tempo	0.0995	16	low energy rate	0.0324
4	pulse clarity	0.0773	17	spectral regularity	0.0296
5	root mean square	0.0731	18	spectral flatness	0.0283
6	spectral entropy	0.0708	19	standard deviation pitch	0.0240
7	spectral skewness	0.0641	20	zero-crossing	0.0233
8	spectral roll-off	0.0597	21	key	0.0165
9	spectral centroid	0.0579	22	mode	0.0060
10	spectral kurtosis	0.0569	23	standard deviation of novelty	0.0040
11	event density	0.0526	24	maximum novelty	0.0028
12	maximum tempo	0.0483	25	mean of novelty	0.0015
13	mean pitch	0.0453	26	minimum novelty	0.0009

**3.6. Correlation: MIR and ECG features**

In addition to the classification, a correlation analysis was performed using the Pearson correlation coefficient. The MIR features and their relationship to features from the other groups were considered using correlation analysis (Table 5). This approach used Landis and Koch levels of reliability [50] and relationships between features with an absolute value of correlation coefficient greater than 0.8 were considered.

**3.7. Classification without the AMI class: C-database**

The confusion matrix of classification with neural network in the complete C-database is shown (Table 6), since this algorithm performed best overall. Compared to other metrics, precision allows easier visualization of system performance in determining the percentage of predicted classes belonging to each class, i.e. positives that are correct [51]. Given that a 90% precision was obtained, the AMI classification, with only 60% precision, was notably poor.

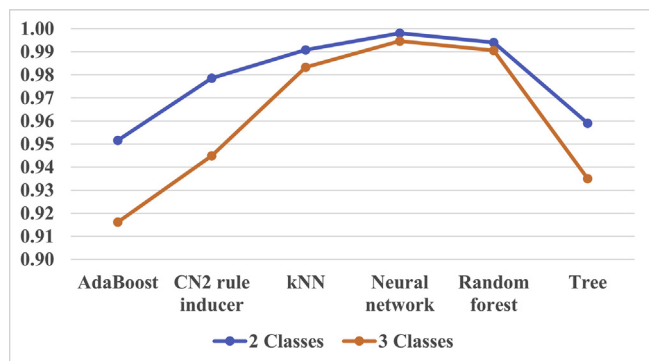
Taking into consideration that anterior myocardial infarction is not usually recognized in lead II of ECG, a new classification process was performed without the AMI class. This classification was made using MIR features and neural network (Figure 11). Performance in this configuration was improved significantly with respect to that of Figure 10, with values of AUC and accuracy observed to be around 0.90 and 0.80 respectively.

**4. Discussion**

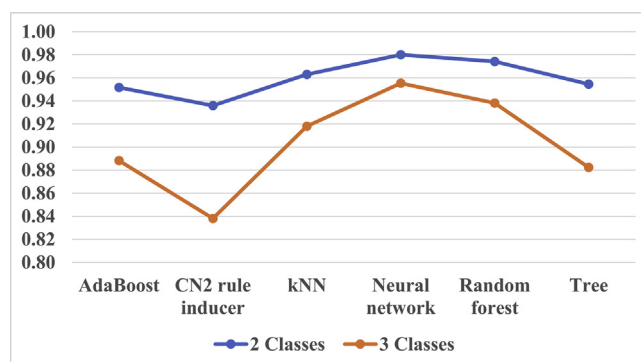
Comparing the performance level of the present study with previous work, the results with MFCC features were not able to match the correct

classification of 99% reported by [13] in the use of neural network classifiers and MFCC. In that study, the high performance was achieved in a binary classification considering normal and abnormal classes, in which the authors took segments of ECG signals with a one second duration, whereas in our study, in addition to normal signals, three more classes were examined, thereby significantly increasing the complexity in the classification task; likewise, the analysis was carried out with different subjects in the training and test sets. Using the same MIT-BIH database and considering MIR features, an accuracy of 0.98 and 0.96 was obtained for 2 and 3 classes, respectively (Figure 7). As with the C-database, neural network was the best algorithm overall (Figures 6 and 7); CN2 rule inducer was the only algorithm with an accuracy below 0.85 (Figure 7) and all algorithms performed well, with AUC higher than 0.90 and accuracy above 0.85. It is not clear whether periods of one second are representative or not in the context of physiology and signal processing. Clearly, it is difficult to determine if a condition of arrhythmia is present in signals of one second duration. In the present study, therefore, ten-second segments were used.

Anterior myocardial infarction consistently produced the weakest predictions due to the lack of information present in the original lead II signal (Table 6). Lead II is most able to recognize elements in the inferior wall of the heart. For the detection of infarction in the anterior wall, precordial leads V3 and V4 are recommended, as well as leads V1 and V2 [27], highlighting the need to establish which diseases or conditions it is possible to recognize in each electrocardiogram lead. Classification exempting the AMI class consequently gave a better performance than with all four classes (Figure 11), with AUC and accuracy around 0.9 and 0.8 respectively, stabilizing from the 12<sup>th</sup> best-ranked features. These metrics would thus appear to ratify the potential capacity of the selected



**Figure 6.** Area under the ROC curve (AUC), MIT-BIH database: classification with MIR features.



**Figure 7.** Accuracy, MIT-BIH database: classification with MIR features.

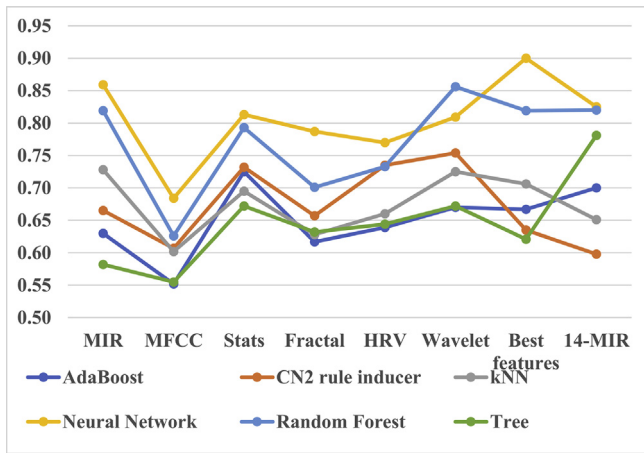


Figure 8. Area under the ROC curve (AUC), C-database: classification analysis of features and classifier algorithms.

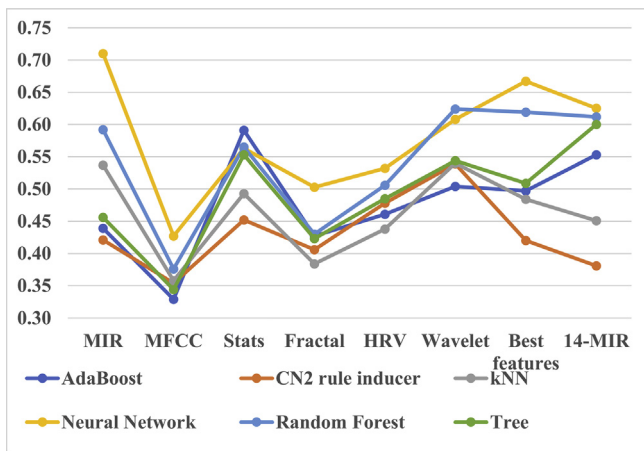


Figure 9. Accuracy, C-database: classification analysis of features and classifier algorithms.

MIR features to represent the ECG signals studied. Moreover, ranking of the MIR features in the C-database revealed tempo, pulse clarity, root

mean square and spectral elements (skewness, roll-off, centroid, and kurtosis) as the strongest contributors to the classification task (Table 4). From the 13<sup>th</sup> best-ranked features, classification with neural network stabilized around 0.85 in AUC and 0.65 in accuracy.

The neural network classifier, almost without exception, performed best in both AUC and accuracy for all feature sets of the C-database. Within the features themselves, the best performance involved the MIR, statistics, and wavelet groups. The MFCC and fractal features generally produced the lowest performances. However, combining the five best-ranked features from each of the MIR, Statistics, HRV and Wavelet groups also gave good results, noticeably using the neural network and random forest classifiers (Figures 8 and 9).

From the correlation analysis, a direct relationship was observed between a number of features extracted from the ECG signal and certain MIR features. In particular, tempo was strongly linked to mean heart rate and R-R interval, while root mean square had a close association with energy, variance, maximum, and mean of the wavelet coefficient zero (cfs0). In the musical context, tempo, slower rhythms, and low-pitched notes are known to be linked generally with low frequencies and the energy of the selected ECG signals was found to be concentrated in the lower frequencies (cfs0). In terms of the rhythmic function of the human heart, it would seem to be more coherent to associate the low frequencies with tempo or with more ponderous rhythms (i.e. the normal heart rate of between 60 and 90 beats per minute). Given that tempo describes the speed of music [42], a close link would be anticipated with heart rate over a given length of time, as described by the R-R interval and heart rate average.

From the same analysis, but in the time domain, zero-crossing allowed conclusions to be drawn concerning the higher frequencies. Zero-crossing is a measurement of sign changes in the signal. It is a noisiness indicator that could explain its association with the Higuchi fractal dimension, which measures irregularity in a signal or time series [52]. Where more zero crosses abound, the higher the fractal dimension will be. The fractal nature of the selected ECG signals was clearly associated with the presence of high frequencies, as revealed by features both of time and of frequency (e.g. spectral spread, spectral flatness). The direct relationships between the Higuchi fractal dimension and these spectral features imply that the wider the frequency range of the ECG signal, the higher the self-similarity may be, as observed particularly in the relationship with spectral spread. The correlation with spectral flatness confirms a fractal behaviour with the frequency content in all frequency bands, but especially in the higher frequencies. Thus, given the influence of frequency content in the fractal dimension of the electrocardiographic signals, through the sound and music property of timbre a link can be contemplated between the ECG signals and musical signals.

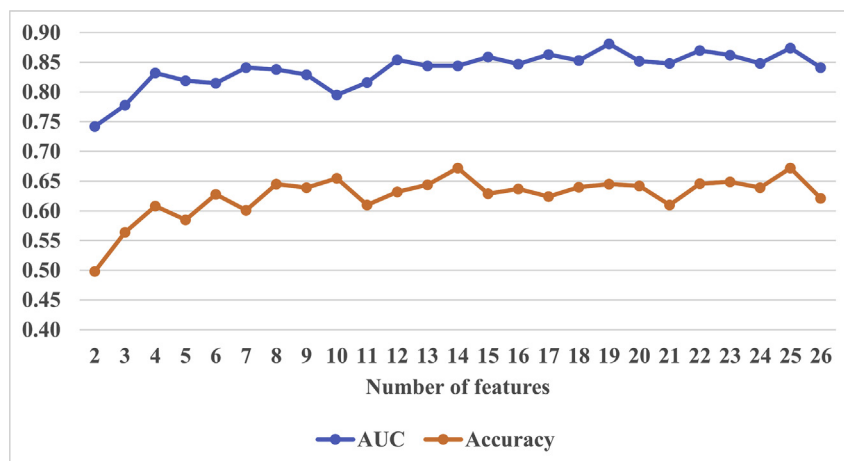


Figure 10. Neural network classification with ranked MIR features (C-database).

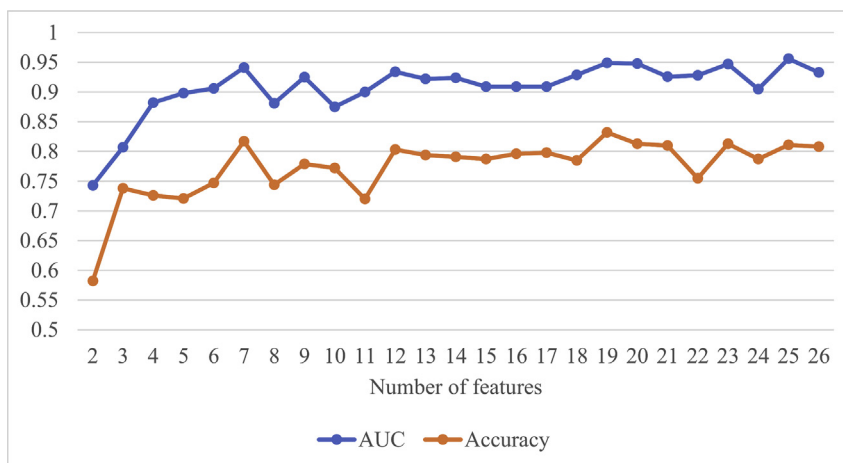
**Table 5.** Correlation coefficients (Corr) between selected MIR features and other groups of features (ECG features).

MIR Features		ECG Features	Corr
tempo	-	mean heart rate	+0.979
tempo	-	mean R-R interval	-0.958
root mean square	-	energy (cfs0)	+0.951
root mean square	-	variance	+0.951
root mean square	-	variance (cfs0)	+0.950
root mean square	-	energy	+0.950
root mean square	-	maximum (cfs0)	+0.946
root mean square	-	maximum	+0.935
zero-crossing	-	Higuchi fractal dimension	+0.842
root mean square	-	mean	+0.834
spectral spread	-	Higuchi fractal dimension	+0.829
root mean square	-	mean (cfs0)	+0.829
spectral flatness	-	Higuchi fractal dimension	+0.819

signals in the study was achieved in most cases by the MIR features. The features extracted from the statistics and wavelet groups also had a significant level of description. In the correlation phase, a solid association was established between the musical features and a number of features of interest extracted from ECG signals, most evident in the strong relationships found between tempo, heart rate, and Higuchi fractal dimension. Given that tempo relates to the speed of music, a close link could therefore be expected with heart rate over a given length of time, as described by the R-R interval and heart rate average. The fractal behaviour of the heart was further clearly associated with the frequency content in the ECG signals across all bands. When the energy was predominantly in the lower frequencies, this fractal nature was less pronounced. Conversely, the greatest fractal quality was observed with frequency content predominantly at the higher frequencies. Given the influence of frequency content in the fractal dimension of the electrocardiographic signals, a link might thus be considered between the ECG signals and musical signals through and music property of timbre. This study constitutes an initial approach to relating MIR features with ECG

**Table 6.** Confusion matrix: classification with MIR features and neural network.

		Predicted heart condition				Σ
		N	T	IMI	AMI	
Actual heart condition	N	189 (90%)	28 (10%)	42 (15%)	65 (20%)	324
	T	0 (0%)	248 (84%)	40 (14%)	0 (0%)	288
	IMI	0 (0%)	9 (3%)	180 (62%)	63 (20%)	252
	AMI	22 (10%)	9 (3%)	27 (9%)	194 (60%)	252
	Σ	211	294	289	322	1116



**Figure 11.** Neural network classification with ranked MIR features (C-database without AMI class).

MIR features were found to be descriptors for ECG signals of inferior and anterior myocardial infarction, T-wave alternating, congestive heart failure, arrhythmia and of the normal condition. Finally, because these particular ECG signals can be represented through MIR descriptors, they might be considered as analogues to musical signals, since MIR features are often used in musical signal processing.

**5. Conclusions and future work**

The use of music information retrieval (MIR) features as electrocardiographic signal descriptors was explored. By means of AI classification techniques and correlation analysis, a relationship was established between MIR features and ECG signals. The best representation of the ECG

signals. The MIR features selected were shown to be capable of discriminating ECG signals in the study, making them potential candidates for use as electrocardiographic signal descriptors and contributing to the development of feature extraction of these signals. This approach could be expanded to discover other possible applications. This would necessitate the study of MIR features in more depth; the inclusion of new MIR features; the incorporation of other ECG referrals and different pathologies within a larger database; and eventually combining MIR features with other classic features of ECG analysis to improve classification performance. Moreover, a new classification metric might be developed, adjusted to these types of feature. Finally, a logical development of the present study would seek, through music information retrieval (MIR) features, to relate sound stimuli and ECG captured in subjects as they



listen to these stimuli, adding to the science of music perception. It would also give a physiological meaning to MIR features associated with ECG signals, and thereby contribute greatly to the science of music therapy.

## Declarations

### Author contribution statement

Ennio Idrobo-Ávila, R. Vargas-Cañas: Conceived and designed the experiments; Performed the experiments; Analyzed and interpreted the data; Contributed reagents, materials, analysis tools or data; Wrote the paper.

H. Loaiza-Correa, F. Muñoz-Bolaños, L. van Noorden: Analyzed and interpreted the data; Wrote the paper.

### Funding statement

This work was supported by Universidad del Valle, Universidad del Cauca, and Colciencias (Funding call No. 727 of 2015), Colombia.

### Data availability statement

Data associated with this study has been deposited at PhysioNet physiological signals database.

### Declaration of interests statement

The authors declare no conflict of interest.

### Additional information

No additional information is available for this paper.

## Acknowledgements

We are especially grateful to Colin McLachlan for suggestions relating to the English text.

## References

- [1] R. Rangayyan, *Biomedical Signal Analysis*, second ed., John Wiley & Sons, 2015.
- [2] R. Northrop, *Analysis and Application of Analog Electronic Circuits to Biomedical Instrumentation*, second ed., CRC press, 2012.
- [3] X.-L. Yang, G.-Z. Liu, Y.-H. Tong, H. Yan, Z. Xu, Q. Chen, X. Liu, H.-H. Zhang, H.-B. Wang, S.-H. Tan, The history, hotspots, and trends of electrocardiogram, *J. Geriatr. Cardiol.* 12 (2015) 448–456.
- [4] WHO, *Cardiovascular Diseases (CVDs)*, WHO, 2017. [https://www.who.int/news-room/fact-sheets/detail/cardiovascular-diseases-\(cvds\)](https://www.who.int/news-room/fact-sheets/detail/cardiovascular-diseases-(cvds)). (Accessed 9 May 2019).
- [5] L.J. Laslett, P. Alagona, B.A. Clark, J.P. Drozda, F. Saldivar, S.R. Wilson, C. Poe, M. Hart, The worldwide environment of cardiovascular disease: prevalence, diagnosis, therapy, and policy issues, *J. Am. Coll. Cardiol.* 60 (2012) S1–S49.
- [6] G.D. Clifford, J. Behar, Q. Li, I. Rezek, Signal quality indices and data fusion for determining clinical acceptability of electrocardiograms, *Physiol. Meas.* 33 (2012) 1419–1433.
- [7] J. Behar, J. Oster, Q. Li, G.D. Clifford, ECG signal quality during arrhythmia and its application to false alarm reduction, *IEEE Trans. Biomed. Eng.* 60 (2013) 1660–1666.
- [8] Q. Li, C. Rajagopalan, G.D. Clifford, Ventricular fibrillation and tachycardia classification using a machine learning approach, *IEEE Trans. Biomed. Eng.* 61 (2014) 1607–1613.
- [9] U. Parlitz, S. Berg, S. Luther, A. Schirdewan, J. Kurths, N. Wessel, Classifying cardiac biosignals using ordinal pattern statistics and symbolic dynamics, *Comput. Biol. Med.* 42 (2012) 319–327.
- [10] C. Ye, B.V.K. Vijaya Kumar, M.T. Coimbra, Heartbeat classification using morphological and dynamic features of ECG signals, *IEEE Trans. Biomed. Eng.* 59 (2012) 2930–2941.
- [11] V. Abvelj, Auditory display of biomedical signals through a sonic representation: ECG and EEG sonification, in: 2012 Proc. 35th Int. Conv. MIPRO, 2012, pp. 474–475.
- [12] J.N. Kather, T. Hermann, Y. Bukschat, T. Kramer, L.R. Schad, F.G. Zöllner, Polyphonic sonification of electrocardiography signals for diagnosis of cardiac pathologies, *Sci. Rep.* 7 (2017) 44549.
- [13] M. Boussaa, I. Atouf, M. Atibi, A. Bennis, ECG signals classification using MFCC coefficients and ANN classifier, in: 2016 Int. Conf. Electr. Inf. Technol., 2016, pp. 480–484.
- [14] M.S. Nagawade, V.R. Ratnaparkhe, Musical instrument identification using MFCC, in: 2017 2nd IEEE Int. Conf. Recent Trends Electron. Inf. Commun. Technol., 2017, pp. 2198–2202.
- [15] K. Siedenburg, I. Fujinaga, S. McAdams, A comparison of approaches to timbre descriptors in music information retrieval and music psychology, *J. N. Music Res.* 45 (2016) 27–41.
- [16] B.L. Sturm, Alexander Lerch, An introduction to audio content analysis: applications in signal processing and music informatics, *Comput. Music J.* 37 (2013) 90–91.
- [17] B. Kostek, Ł. Kania, Music information analysis and retrieval - a review, in: G. T. (Ed.), *Politechnika Wroclawska, Multimedia Systems Department, Gdańsk University of Technology, ul. Narutowicza 11/12, Gdańsk, 80-952, Poland*, 2019, pp. 27–42. <https://www.scopus.com/inward/record.uri?eid2-s2.0-85071973358&partnerID40&md535310d1b6b12d86c29be69f427ca35f8>.
- [18] Y. Ospitia Medina, S. Baldassarri, J.R. Beltrán, High-level libraries for emotion recognition in music: A review 847 (2019) 158–168.
- [19] 13th International Conference on Intelligent Information Hiding and Multimedia Signal Processing, IHH-MSP 2017, in: P. J.-S., T. P.-W., J. L.C., W. J. (Eds.) 81, 2018, pp. 1–448. <https://www.scopus.com/inward/record.uri?eid2-s2.0-85026780681&partnerID40&md521fb3338ef18ece9875cc684237f4be4>.
- [20] Y.-L. Hsu, J.-S. Wang, W.-C. Chiang, C.-H. Hung, Automatic ECG-based emotion recognition in music listening, *IEEE Trans. Affect. Comput.* 11 (2020) 85–99.
- [21] C. Stolojescu, ECG signals classification using statistical and time-frequency features, *Appl. Med. Informatics* 30 (No 1) (2012). <http://ami.info.umfcluj.ro/inde x.php/AMI/article/view/351>.
- [22] R.J. Martis, U.R. Acharya, L.C. Min, ECG beat classification using PCA, LDA, ICA and Discrete Wavelet Transform, *Biomed. Signal Process Contr.* 8 (2013) 437–448.
- [23] B.M. Asl, S.K. Setarehdan, M. Mohebbi, Support vector machine-based arrhythmia classification using reduced features of heart rate variability signal, *Artif. Intell. Med.* 44 (2008) 51–64.
- [24] A.M.A. Molina, R.I.R. Jiménez, A.M. Diosdado, Multifractal analysis of ECG time series of stress tests in healthy subjects, *AIP Conf. Proc.* 2090 (2019) 50001.
- [25] A.L. Goldberger, L.A.N. Amaral, L. Glass, J.M. Hausdorff, P.C. Ivanov, R.G. Mark, J.E. Mietus, G.B. Moody, C.-K. Peng, H.E. Stanley, PhysioBank, PhysioToolkit, and PhysioNet, *Circulation* 101 (2000) e215–e220. [http://circ.ahajournals.org/co ntent/101/23/e215.abstract](http://circ.ahajournals.org/content/101/23/e215.abstract).
- [26] G.B. Moody, R.G. Mark, The impact of the MIT-BIH arrhythmia database, *IEEE Eng. Med. Biol. Mag.* (2001).
- [27] L.W., Wilkins, *ECG Interpretation Made Incredibly Easy*, fifth ed., Wolters Kluwer/Lippincott Williams & Wilkins Health, 2011.
- [28] PTB, The PTB diagnostic ECG database signals, PhysioBank ATM, 2017 (n.d.), <https://physionet.org/cgi-bin/atm/ATM>. (Accessed 12 July 2017).
- [29] R.-D. Bousseljot, D. Kreiseler, A. Schnabel, The PTB Diagnostic ECG Database, (n.d.). doi
- [30] G.B. Moody, The physionet/computers in cardiology challenge 2008: T-wave Alternans, *Comput. Cardiol.* 35 (2008) 505–508.
- [31] PhysioBank ATM, T-wave Alternans Challenge Database Signals, PhysioBank ATM, 2017 (n.d.), <https://physionet.org/cgi-bin/atm/ATM>. (Accessed 12 July 2017).
- [32] X. Wang, Y. Zhou, M. Shu, Y. Wang, A. Dong, ECG baseline wander correction and denoising based on sparsity, *IEEE Access* 7 (2019) 31573–31585.
- [33] M. Misiti, Y. Misiti, G. Oppenheim, J.-M. Poggi, *Wavelets and Their Applications*, John Wiley & Sons, 2013.
- [34] G.D. Clifford, F. Azuaje, P. McSharry, *Advanced Methods and Tools for ECG Data Analysis*, Artech House, Inc., Norwood, MA, USA, 2006.
- [35] C. Wang, S. Yang, X. Tang, B. Li, A 12-Lead ECG Arrhythmia Classification Method Based on 1D Densely Connected CNN, 11794 LNCS, 2019, pp. 72–79.
- [36] A. Patané, M. Kwiatkowska, Calibrating the classifier: siamese neural network architecture for end-to-end arousal recognition from ECG, in: *Lect. Notes Comput. Sci. (Including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics)*, 2019.
- [37] G.K. Pal, Resting heart rate is the index of cardiovascular health, *Int. J. Clin. Exp. Pathol.* 1 (2014). <http://ijcep.org/index.php/ijcep/article/view/135>.
- [38] L. Munoz, A. Roon, H. Riese, C. Thio, E. Oostenbroek, I. Westrik, E. Geus, R. Gansevoort, J. Lefrandt, I. Nolte, H. Snieder, Validity of (Ultra-)Short recordings for heart rate variability measurements, *PLoS One* 10 (2015), e0138921.
- [39] S. Laborde, E. Mosley, J.F. Thayer, Heart rate variability and cardiac vagal tone in psychophysiological research – Recommendations for experiment planning, data analysis, and data reporting, *Front. Psychol.* 8 (2017) 213. <https://www.frontiers in.org/article/10.3389/fpsyg.2017.00213>.
- [40] S. Krishnan, Y. Athavale, Trends in biomedical signal feature extraction, *Biomed. Signal Process Contr.* 43 (2018) 41–63.
- [41] B.A. de Castro, D. de M. Brunini, F.G. Baptista, A.L. Andreoli, J.A.C. Ulson, Assessment of Macro fiber composite sensors for measurement of acoustic partial discharge signals in power transformers, *IEEE Sensor. J.* 17 (2017) 6090–6099.
- [42] B. Benward, M. Saker, *Music in Theory and Practice*, eighth ed., I, McGraw-Hill, New York, 2009.
- [43] S. Dubnov, Generalization of spectral flatness measure for non-Gaussian linear processes, *IEEE Signal Process. Lett.* 11 (2004) 698–701.
- [44] O. Lartillot, P. Toivainen, T. Eerola, in: C. Preisach, H. Burkhardt, L. Schmidt-Thieme, R. Decker (Eds.), *A Matlab Toolbox for Music Information Retrieval BT - Data Analysis, Machine Learning and Applications*, Springer Berlin Heidelberg, Berlin, Heidelberg, 2008, pp. 261–268.
- [45] J. Pan, W.J. Tompkins, A real-time QRS detection algorithm, *IEEE Trans. Biomed. Eng.* (1985).

- [46] MathWorks, Makers of MATLAB and simulink. <https://www.mathworks.com/>, 2016. (Accessed 15 April 2019).
- [47] U. of Jyväskylä, MIRtoolbox — Humanities and Social Sciences (n.d.), <https://www.jyu.fi/hytk/fi/laitokset/mutku/en/research/materials/mirtoolbox>. (Accessed 15 April 2019).
- [48] A. Goshvarpour, A. Goshvarpour, Human identification using information theory-based indices of ECG characteristic points, *Expert Syst. Appl.* 127 (2019) 25–34.
- [49] N. Jatupaiboon, S. Pan-Ngum, P. Israsena, Subject-dependent and subject-independent emotion classification using unimodal and multimodal physiological signals, *J. Med. Imaging Heal. Informatics.* 5 (2015) 1020–1027.
- [50] J.R. Landis, G.G. Koch, The measurement of observer agreement for Categorical data, *Biometrics* 33 (1977) 159–174.
- [51] T. Alaa, Classification assessment methods, *Appl. Comput. Informatics* (2020) ahead-of-.
- [52] H. Ahammer, Higuchi dimension of digital images, *PloS One* 6 (2011), e24796.