MIPRO 2019, May 20-24, 2019, Opatija Croatia

Feature Extraction for Heartbeat Classification in Single-Lead ECG

J. Bogatinovski*, D. Kocev* and A. Rashkovska†
* Department of Knowledge Technologies, Jožef Stefan Institute, Ljubljana, Slovenia
† Department of Communication Systems, Jožef Stefan Institute, Ljubljana, Slovenia
E-mails:{jasmin.bogatinovski, dragi.kocev, aleksandra.rashkovska}@ijs.si

***Abstract*—The recent trends in ECG device development are heading towards wireless single-lead ECG sensors. The lightweight design of these wireless sensors allows the patients to wear it comfortably for a long period of time and during their ordinary everyday activities. Long-term ECG recordings are intended to help in detection or diagnosis of heart diseases. These measurements are significantly longer and more heterogeneous than the measurements performed at a controlled hospital environment. Consequently, their manual inspection is a tedious, hard and expensive job. An alternative is to use computational techniques for automatic classification of heartbeats and arrhythmia detection. In this paper, we investigate methods for feature extraction in single-lead ECG for the purpose of heartbeat classification. The used feature extraction methods originate from the field of time series analysis. The obtained features are then coupled with a classification algorithm to obtain predictive models. The usefulness of the proposed approach is demonstrated on the MIT-BIH arrhythmia database. Results show that features that emerge from different scientific areas can provide information for separation of different class distributions that appear in heartbeat classification problem.**

Keywords*—*ECG, single-lead, classification, feature, arrhythmia

I. Introduction

Unobtrusive wireless ECG measurements from devices employing smaller number of leads provide opportunity for continuous supervision for the patients with cardiovascular disorders. The main advantage of these ”off the person” measurements is obtaining longer and more heterogeneous measurements, meaning the measurements are acquired during different daily activities and, therefore, are more disturbed by noise. Nevertheless, these devices enable real-time tracking of the state of the patient, at a cost of more complicated tasks for automatic detection of specific type of arrhythmia.

Advancement of Medical Instrumation (AAMI) provides standard to which different methodologies for heartbeat classification are being tested. Similar recommendations are also part of the IEC 60601-2-47 standard. A part of the recommendation is to use the MIT-BIH database for benchmark [1]. The MIT-BIH database is well established standard database for testing different heartbeat classification methodologies. The database consists of 48 half-hour expert labeled two-channel ambulatory ECG recordings collected from 47 subjects. It includes heartbeats from 5 classes, including: nonectopic (N), supraventiricular ectopic beat (SVEB), ventricular ectopic beat (VEB), fusion beat (F) and unknown beat (Q).

Recent studies overview different sets of features being employed for the task of ECG arrhythmia classification [2] [3]. Most of the studies focus on feature extraction from measurements from at least two leads. However, it was shown that differential ECG leads provide different ECG signals from standard bipolar or unipolar ECG leads [4] [5]. Nevertheless, it was confirmed that the ECG from a differential lead is appropriate for hearth rhythm diagnostics [6]. Moreover, it has been shown that domain specific time series features might not be informative enough for solving a given time series classification task [7]. Furthermore, time series features derived from other domains showed to be competitive or better than the domain specific features. Combining global time series features under the AAMI inter-patient paradigm is the main focus on this work. The inter-patient paradigm refers to the process of discarding heartbeats from same patients in both training and test sets. The final goal is to find features that characterize different classes of heartbeats.

The remainder of the paper is organized as follows. Section II gives overview of the state-of-the-art methodologies for arrhythmia classification according to the AAMI standard and the domain specific global time series features derived for time-series classification. Section III presents the methodology. Section IV presents the results and discusses the usefulness of the features in distinguishing the different classes. Section V concludes the work and discusses further research directions.

II. Related work

Traditionally, the problem of heartbeat classification is recognized as a problem of classification of time series. Until recently, the main focus was on extraction of useful features from time series. However, with the recent advances in the area of deep learning, the focus is shifted towards automatic learning of features from time series [8] [9] [10].

The state-of-the-art employs 9-layered end-to-end convolution neural network with batch-weight loss to tackle the imbalance of the problem [10]. Namely, due to the nature of

Authors J. Bogatinovski and D. Kocev acknowledge the financial support of the Slovenian Research Agency under the grants P2-0103 and J2-9230. Author AR acknowledges the financial support of the Slovenian Research Agency under the grant P2-0095 and the EU H2020 project SAAM under grant No. 769661.

the process, it is expected that most of the heartbeats will belong to the N class. Arrhythmia and its different forms are usually present with far less heartbeat examples. In data science literature, presence of such rear categories among majority of others is known as imbalance of the problem. The data from the MIT-BIH database from a single lead (MLII) are used. The impact of neighbouring heartbeats is reported to have great significance on improving the performance. Specificity or True Negative Rate is reported instead of False Positive Rate, as recommended by AAMI. Under the inter-patient paradigm, it is reported that the results from the studies in [11] and [12] are outperformed on all of the 4 measures (accuracy, positive predictivity, sensitivity and specificity) as recommended by AAMI, except for sensitivity in [11] and accuracy in [12].

In [11], an end-to-end system for automatic arrhythmia classification is proposed. It consists of 4 standard processes as described in [3]: preprocessing, noise removal, feature extraction and classification. Total of 15 features separated in 3 groups (R-R intervals, beat intervals and morphological features) are used. These features are domain specific. It is important to note that features are extracted on both leads from the MIT-BIH database. The final prediction is given as a combination of the output of two weighted linear discriminant classifiers. This work employs filtering of the signal, which might result in discarding important artifacts of the heartbeats and increase the prediction time.

Despite wavelets and R-R based features, other popular features include: duration of QRS complex, distance between fiducial points of the heartbeat, PCA (principle component analysis) features on raw signal, ICA sources of the raw signal, GDA (Generalized Discriminant Analysis), Random Projections, high order accumulative features, correlation dimensions, highest Lyapunov exponents, Hermite transformations, fractal dimension features, Fourier transformation, vectrograms and others. A detailed overview on the domain-specific features used for arrhythmia classification are given [2] and [3].

The work presented in [13] is one of the first that utilizes feature selection techniques for selecting the most relevant domain-specific features for heartbeat classification. Incremental wrapper and filtering approaches based on mutual information are utilized for selecting the top performing features from more than 200 features available. Their results show that it is possible to use single-lead ECG for the task of classification of heartbeats. Although they follow the AAMI procedure for splitting the heartbeats, they do not evaluate the performance as the standard suggest, neither provide the confusion matrix. So, direct comparison with their results according to the standard is not possible.

III. Materials and Methods

Recent findings show that domain specific features alone might not be enough for solving a given time series classification task [7]. In [14], a library for extracting global features from time series data, named HCTSA (Highly Comparative Time Series Analysis), is proposed. The features originate from interdisciplinary studies interested in dynamical modeling. Such global features quantify patterns in time series across the full time interval. The experiments in [14] show that features facilitate interpretable insights, are often selected from unexpected literature (drawing attention to novel features for specific application) and best performing classifiers are often constructed using a novel combination of interdisciplinary features (e.g., combining features from economics and biomedical signal processing). Motivated by these conclusions, we employ this library to the problem of heartbeat classification on a single-lead ECG (MLII) from the MIT-BIH database, following the AAMI recommendations.

The HCTSA library combines time series global feature operators derived across various scientific areas during the years. Roughly, it can be organized into 14 groups: statistical, measures of distribution, correlation, basic function representation, stationary, scaling, entropy, non-linear time series analysis, non-linearity, time domain transformations, model fitting and forecasting, domain specific operators, fanciful operations and others. Each of these groups has its own subgroups of operators. Setting different parameters for operations yields different features. Thus, instead of 1064 basic operations, one can easily finish with few thousand features. Increasing the number of features comes with a cost expressed in time needed for computation and increasing the correlation between the generated features. The later can increase the noise in the data and consequently make harder to distinguish the right features for performing the classification task. Regarding the time complexity, calculating features is expensive. However, selecting the right features will result in low cost for obtaining predictions by a pre-trained model. This makes the approach useful in scenario where a fast prediction is needed. The implementation of the library is in MATLAB and is freely available for non-commercial purposes[1].

The 44 records from lead MLII of the MIT-BIH database are first segmented. The cut-off time for the segments is 200 ms before the R peak time – the approximate duration of the PR interval. The segmented heartbeats are given as input to the HCTSA library and the result are 7873 produced features per heartbeat. Some of the operations showed to be not suited for small number of samples and resulted in errors that are tracked by the library. The processing of the calculated features is performed with removing the features where these errors appeared, which resulted in 3324 total number of features.

Since many of the features are product of one particular operator instantiated with different parameter settings, high correlation between the features could be expected. To assess the relevance of the features, we perform feature ranking using random forests impurity scores [15]. The feature ranking has two goals. First, it provides additional insights into the feature relevance/importance for the classification. Second, it can help in reduction of the number of features used fo classification. To select the features, a threshold on the relevance is imposed.

[1] https://hctsa-users.gitbook.io/hctsa-manual/

In such a way, the features that have higher relevance over the specified threshold, are preserved.

The last step in the workflow is using algorithms to build predictive models. As a set of predictive modeling algorithms, AdaBoost and Gradient Boosting, implemented under the sckitlearn library, were used[2]. According to the AAMI guidelines, we calculate four performance metrics: accuracy, positive predictivity (pp), sensitivity and specificity. Their definitions are as follows:

$$accuracy = \frac{TP + TN}{TP + TN + FP + FN}, \quad (1)$$

$$pp = \frac{TP}{TP + FP}, \quad (2)$$

$$sensitivity = \frac{TP}{TP + FN}, \quad (3)$$

$$specificity = \frac{TN}{TN + FP}, \quad (4)$$

where TP, TN, FP and FN are the numbers of true positives, true negatives, false positives and false negatives, respectively.

IV. Results and Discussion

The time required for computation of all features for 100 733 time series was 240 hours on a single machine with 32GB of RAM and Intel(R) Core(TM) i7-8700K CPU @ 3.70GHz processor. This is a big overhead, but the used post-processing methods aim at reducing the number of features thus reducing the time overhead.

The average features rankings with gini impurity criteria method for all the classes are given in Table I. The confusion matrix for the best performing model AdaBoost is given in Table II. In the following, we discuss the obtained results.

A. Dataset representation

Fig. 1 represents visualization of the training dataset obtained with t-distributed stochastic neighbor embedding (TSNE) from the MIT-BIH database, as discussed in [11]. A clear observation of the classes can be distinguished, with the VEB class belonging to the top head of the predominantly N class spherical shape, with partially localization of SVEB and F classes. Given the spread of the SVEB classes, it can be concluded that algorithms will have hard time in distinguishing the SVEB class from the N class. However, comparing the VEB and the N classes, the separation is more clear.

In Fig. 1, a dozen groups of points are observed. With close inspection of these groups, it can be concluded that they belong to the same patient (Fig. 2). It is important to note that, although the majority of the heartbeats for one patient will happen to belong to one group, there exist heartbeats that do not belong to the same group, if the heartbeats are of different class. For example, beats for the patients 208 and 209, as observed in Fig. 1. This is encouraging findings since it is a sign of existing good features in separation of at least some of the classes.

[2]https://scikit-learn.org/stable/

TABLE I
Top 5 average ranks for the features for the classes N, SVEB, VEB, F and Q

vs all	average rank	feature name
N	1.7272	CO Add Noise 1 gaussian firstUnder75
	2.2727	CO TranslateShape circle 25 pts fives
	6.5000	EX MovingThreshold 01 01 maxq
	8.4545	EN CID minCE1
	8.5909	SP Summaries pgram hamm w10 90
SVEB	4.6818	CP ML StepDetect l1pwc 005 medianstepint
	4.7727	FC Surprise T2 50 3 udq 500 tstat
	4.9090	PP Compare resample 1 2 gauss1 kd resAC2
	9.31818	CP ML StepDetect l1pwc 02 minstepint
	11.5454	length
VEB	2.5454	CO AddNoise 1 gaussian firstUnder75
	3.6363	SP Summaries pgram hamm w10 90
	4.5454	EN CID minCE1
	6.5454	CO StickAngles y std p
	7.0454	PH Walker momentum 2 w std
F	15.4545	DN SimpleFit sin1 resruns
	19.9545	CO TranslateShape circle 15 pts threes
	21.0000	PH Walker biasprop 01 05 res runstest
	25.3181	PP Compare spline44 gauss1 kd resruns
	30.1818	HT DistributionTest lillie ev
Q	2.9545	FC LocalSimple mean1 ac1
	3.5454	FC LocalSimple mean3 ac1
	7.0000	MF steps ahead ar 2 6 rmserr 1
	7.6818	FC LocalSimple mean2 ac1
	10.9545	MF AR arcov 2 a3

TABLE II
Confusion matrix for our method. *number of trees 500, learning rate 0.1*

	Class	N	SVEB	VEB	F	Q
	N	43832	25	396	3	0
True	SVEB	1721	15	100	1	0
Label	VEB	708	76	2432	4	0
	F	281	1	106	0	0
	Q	3	0	0	4	0

B. Feature ranking

From the results presented in Table I, according to interclass feature importance, it can be conducted that the two most common classes, N and VEB, have features that occur frequently as top ranked. Moreover, the same top ranked features are shared between the N and VEB classes. For example, *Co add noise* – the best ranked feature comes from chaos theory. It is a measure of chaos, calculated by adding Gaussian noise to the time series in increasing manner across some range of values and then measuring the mutual information at each point by calculating histograms. The larger the measure, the greater the relative intensity of the chaos is. Predominantly, the values for the N beats tend to be positively valued. According to this measure, positive values indicate chaotic behaviour. Conversely,the VEB heartbeats tend to have negative value for this feature. For both N and VEB, this is the most important feature.

The *normalized length* feature is closely related with one of the most exploited features in heartbeat classification literature – normalized R-R interval. *Normalized length* measures the length of the time series normalized with respect to the average length of the patient-specific heartbeat. Fig. 3 depicts the *normalized length* feature vs. the *CO add noise* feature. It can be seen that the *CO add noise* feature makes separation between classes VEB (blue) and N (red) with observation that VEB heartbeats tend to have larger duration with respect to normal ones. Inspection of the N heartbeats with negative value for the *CO add noise* feature suggests that they belong to patient that experiences VEB heartbeats. Combining this novelty feature together with previously known features enables clearer separation of classes N and VEB. The same discussion for the SVEB and the VEB classes can be drawn in parallel. In Fig. 4, one can also observe a clearer separation of the VEB vs all other classes, which the second ranked feature *CO TranslateShape circle 25 pts fives* imposes. This feature is instantiate from the CO basicrecurf function operator. It calculates the number of points that are close to certain geometric shapes in a plots generated from the time series with lag *t*. As such, those represent measures for point density estimation.

Regarding the categorization of the SVEB class, such calculated global features do not preserve the local deviation of the SVEB and the N classes. These two types are distinct only in the PR interval of the beat. The global time series features diminish the effect of this local property and thus it is harder to separate the classes N and SVEB. Therefore, a more suitable way would be to threat that part of the heartbeat as separate time series. Extracting features on that part, hopefully will bring additional discriminate power.

For class F, the average feature rankings suggest that there is no strong feature for classification of this class. The time series for class Q are varying by large margin in patients, and are of little support. Although, *FC LocalSimple mean1 ac1* appears quite frequently as top ranked feature, its impact can not be well determined due to overlapping of class distributions by large extend.

Finally, another representation of the data is presented in Fig. 5. This figure illustrates the difficulty of the prediction problem at hand. Namely, only 68.58 % of the variance in the dataset is explained with the first three components.

C. Predictive model

The confusion matrix on the test set, reported in Table II, shows that the proposed features are good enough for separation of VEB and other classes. The accuracy, sensitivity, positive predictivity and specificity for the VEB class are: 0.97, 0.76, 0.80 and 0.99 accordingly, which are quite good given the imbalance of the problem. All SVEB heartbeats tend to be predicted as N due to the previously discussed reasons. The accuracy, sensitivity, positive predictivity and specificity for the SVEB class are: 0.96, 0.01, 0.13, 0.99. The low value for the sensitivity is unsatisfactory. The F and Q beats as expected, are not distinguishable and thus are miss-classified.

Compared to the state-of-the-art methods that extract features from a single-lead ECG, our approach is poorer in performance and solving the task. One of the explanations

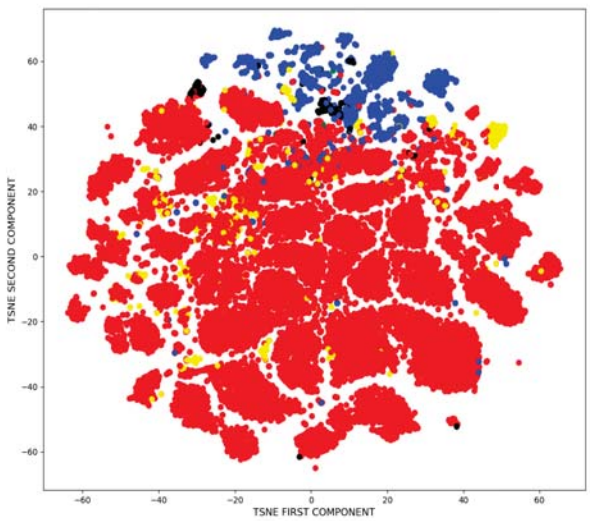

Fig. 1. TSNE representation of the training MIT-BIH dataset represented in two dimensions. Classes: N=Red, VEB=Blue, SVEB=Yellow, F=Black, Unknown=Green.

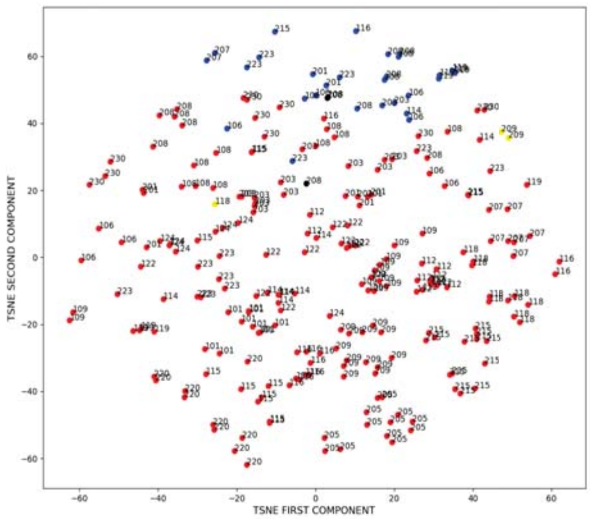

Fig. 2. TSNE representation of the training MIT-BIH dataset represented in two dimensions along with patients IDs to present the clustering groups depicted in Fig. 1. Red=N, Blue=VEB, Yellow=SVEB, Black=F, Green=Unknown.

is that the work perfomed in [10] extracts patient specific features. As depicted in Fig. 1, the heartbeats of same patient tend to cluster themselves in groups. This might reflect the belief of unique anatomy of a patients' cardiovascular system, which implies different characteristic features of the types of

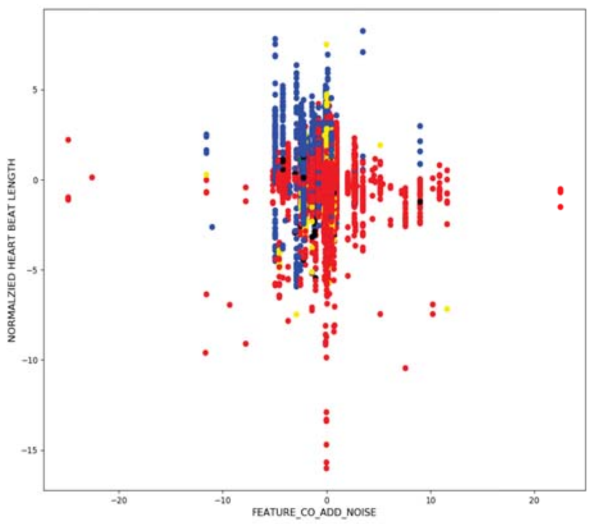

Fig. 3. Feature *CO add noise* – x axis and Feature *normalized time series length* – y axis. Depicts the separation of VEB versus SVEB and N.

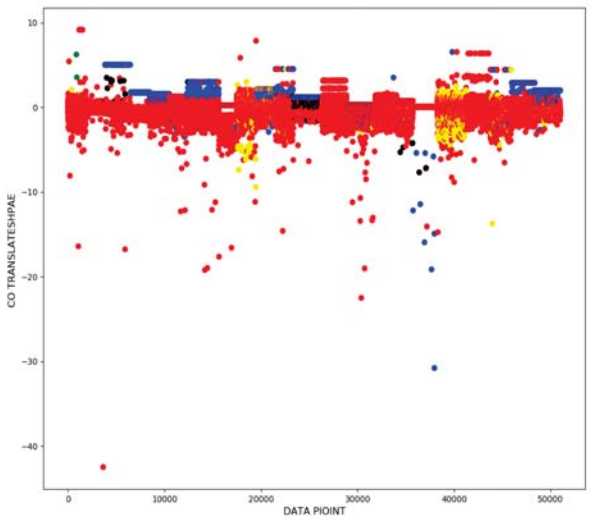

Fig. 4. *Sample ID* – x axis and Feature *CO TranslateShape circle 25 pts fives* – y axis. Depicts the separation of VEB and N.

arrhythmia a patient might experience. Also, in their work, an information for the neighbouring heartbeat is used, which seems useful, since the heartbeats are not independent.

V. Conclusion

It this paper, we propose methods for feature extraction in single-lead ECG for the purpose of heartbeat classification. The extracted features belong to the domain of global time series features derived from various technical and scientific

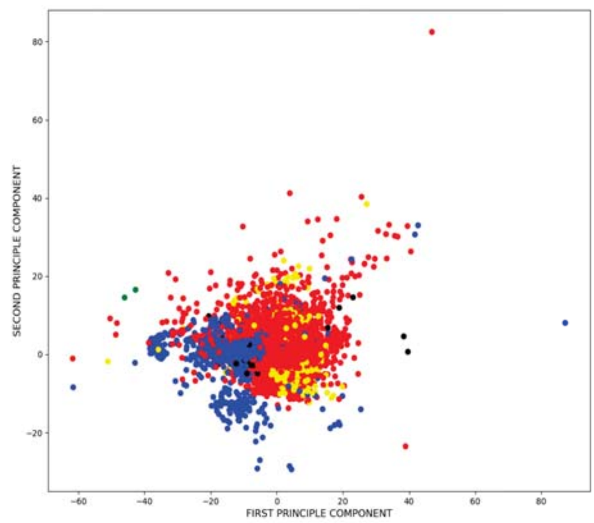

Fig. 5. PCA representation of the training MIT-BIH dataset represented in two dimensions. The first 3 components express 68.58 % of the variance in the data.

areas. Results show that features that emerge from different scientific areas can provide information for separation of different class distributions that appear in the heartbeat classification problem.

The extracted features are disriminative for the VEB class versus the remaining 4 classes. However, a clear distinction between nonectopic and SVEB heartbeats is absent. This is due to the focus of the employed time series features to extract the global properties - nonectopic and SVEB heartbeats are different mostly in the PR interval. Thus such global approach might not be suitable. As future work, we would investigate if adding extracted information from different wave forms of the ECG provides greater diversity in the problem expression.

Since the literature shows that the neighbouring heartbeats can help in expressing the differences between the classes, for future work, global time series features on two heartbeats can also be extracted. This will preserve the local correlation between the heartbeats. The focus on this work is primarily on weather features from various domains are helpful in describing the problem of heartbeat classification. However, in order to exploit the information these features provide, using different algorithms with different parametarization should also be conducted. Any other further work should take this in consideration.

References

[1] G. Moody and R. Mark, “The impact of the mit-bih arrhythmia database,” *IEEE Engineering in Medicine and Biology Magazine*, vol. 20, no. 3, pp. 45–50, 2001.

[2] S. K. Berkaya, A. K. Uysal, E. S. Gunal, S. Ergin, S. Gunal, and M. B. Gulmezoglu, “A survey on ecg analysis,” *Biomedical Signal Processing and Control*, vol. 43, pp. 216 – 235, 2018.

[3] E. J. da S. Luz, W. R. Schwartz, G. Cámara-Chávez, and D. Menotti, "Ecg-based heartbeat classification for arrhythmia detection: A survey," *Computer Methods and Programs in Biomedicine*, vol. 127, pp. 144 – 164, 2016.
[4] I. Tomašić and R. Trobec, "Electrocardiographic systems with reduced numbers of leads—synthesis of the 12-lead ecg," *IEEE reviews in biomedical engineering*, vol. 7, 05 2013.
[5] T. Roman, T. Ivan, K. Aleksandra, Rashkovska, D. Matjaž, and A. Viktor, *Body Sensors and Electrocardiography*. Springer International Publishing, 2018.
[6] I. Tomašić, S. Frljak, and R. Trobec, "Estimating the universal positions of wireless body electrodes for measuring cardiac electrical activity," *IEEE Transactions on Biomedical Engineering*, vol. 60, no. 12, pp. 3368–3374, 2013.
[7] G. Dong and H. Liu, *Feature Engineering for Machine Learning and Data Analytics*, 2018.
[8] U. R. Acharya, H. Fujita, O. S. Lih, Y. Hagiwara, J. H. Tan, and M. Adam, "Automated detection of arrhythmias using different intervals of tachycardia ecg segments with convolutional neural network," *Information Sciences*, vol. 405, pp. 81 – 90, 2017.
[9] S. Kiranyaz, T. Ince, and M. Gabbouj, "Real-time patient-specific ecg classification by 1-d convolutional neural networks," *IEEE Transactions on Biomedical Engineering*, vol. 63, no. 3, pp. 664–675, March 2016.
[10] A. Sellami and H. Hwang, "A robust deep convolutional neural network with batch-weighted loss for heartbeat classification," *Expert Systems with Applications*, vol. 122, pp. 75 – 84, 2019.
[11] P. de Chazal, M. O'Dwyer, and R. B. Reilly, "Automatic classification of heartbeats using ecg morphology and heartbeat interval features," *IEEE Transactions on Biomedical Engineering*, vol. 51, no. 7, pp. 1196–1206, 2004.
[12] T. Li and M. Zhou, "Ecg classification using wavelet packet entropy and random forests," *Entropy*, vol. 18, p. 285, 2016.
[13] G. Doquire, G. de Lannoy, D. François, and M. Verleysen, "Feature selection for interpatient supervised heart beat classification," *Computational intelligence and neuroscience*, vol. 2011, 2011.
[14] B. D. Fulcher and N. S. Jones, "Highly comparative feature-based time-series classification," *IEEE Transactions on Knowledge and Data Engineering*, vol. 26, pp. 3026–3037, 2014.
[15] R. Díaz-Uriarte and S. Alvarez de Andrés, "Gene selection and classification of microarray data using random forest," vol. 7, p. 3.