Are we training our heartbeat classification algorithms properly?

Amalia Villa[1,2], Margot Deviaene[1,2], Rik Willems[3], Sabine Van Huffel[1,2] and Carolina Varon[1,2]

***Abstract*— Despite the multiple studies dealing with heartbeat classification, the accurate detection of Supraventricular heartbeats (SVEB) is still very challenging. Therefore, this study aims to question the current protocol followed to report heartbeat classification results, which impedes the improvement of the SVEB class without falling on over-fitting. In this study, a novel approach based on Variational Mode Decomposition (VMD) as source of features is proposed, and the impact of the use of the MIT-BIH Arrhythmia database is analyzed.**

The method proposed is based on single-lead ECG, and it characterizes heartbeats by a set of 45 features: 5 related to the time intervals between consecutive heartbeats, and the rest related to VMD. Each heartbeat is decomposed in their variational modes, which are, on their turn, characterized by their frequency content, morphology and higher order statistics. The 10 most relevant features are selected using a backwards wrapper feature selector, and they are fed into an LS-SVM classifier, which is trained to separate Normal (N), Supraventricular (SVEB), Ventricular (VEB) and Fusion (F) heartbeats. An inter-patient approach, using patient independent training, is considered as suggested in the literature.

The method achieved sensitivities above 80% for the three most important classes of the database (N, SVEB and VEB), and high specificities for the N and VEB classes. Given the challenges related to the SVEB and F class present in the literature, the composition of the MIT-BIH database is analyzed and alternatives are suggested in order to train heartbeat classification algorithms in a novel and more realistic way.

I. INTRODUCTION

The electrocardiogram (ECG) is commonly used for monitoring of the heart in several cardiac applications, such as the analysis of heart rate variability or the study of arrhythmias. The analysis of the ECG requires a step of heartbeat classification, which is very tedious and time consuming if done manually. Therefore, several approaches have been proposed to automatically perform this task.

Generally, the state-of-the-art makes use of a stage of feature extraction, in which interpretable characteristics of the heartbeats are described. The main difference between normal and irregular heartbeats is their temporal condition, since irregular heartbeats are characterized by arrhythmia, altering the normal sinus rhythm. Furthermore, the morphology must be taken into account to distinguish between the spatial origin of the different irregular heartbeats. Consequently, the most commonly used features are those related to the temporal differences between consecutive heartbeats [1], morphology [2], wavelets [3], higher order statistics (HOS) [4], or combinations of these [5] [6].

[1] KU Leuven, Dep. Electrical Engineering, ESAT-STADIUS. Leuven, BE
[2] IMEC, Leuven, BE
[3] Division of Experimental Cardiology, Dep. Cardiovascular Diseases, KU Leuven, BE

The MIT-BIH Arrhythmia database is the most used database to evaluate these algorithms [7]. Typically, this evaluation is performed following the protocol proposed in [2], in order to be able to compare results.

Even though several combinations of features and classifiers have been proposed, they often achieve very low performance in the accurate detection of Supraventricular heartbeats (SVEB), due to their difficult distinction from Normal (N) beats in terms of morphology, and the inter-patient variability within the SVEB class. One of the best methods uses features describing the temporal, morphological and HOS characteristics of the heartbeats, combined with an ensemble of SVM classifiers [6], achieving sensitivities and specificities above 93% for the N and Ventricular (VEB) class, and yet a sensitivity and specificity for the SVEB class of 78% and 50%, respectively.

This study aims to find features that characterize the different classes, improving the current SVEB state-of-the-art and making use of single-lead ECG. A combination of temporal features and characteristics extracted from Variational Mode Decomposition (VMD) is presented, together with new suggestions to better train heartbeat classification algorithms.

II. METHODOLOGY

A. ECG Dataset

This study used the first lead of the MIT-BIH Arrhythmia database from Physionet [7], corresponding to the modified-lead II (MLII) configuration. This database consists of 48 ECG signals recorded from 47 patients, sampled at 360Hz and with a duration of approximately 30 min.

The AAMI recommendations were followed to make the results comparable to the state-of-the-art [8]. Accordingly, 4 ECG signals were removed from the analysis since they correspond to patients with a pacemaker. In addition, the heartbeat annotations were divided in 5 classes, which are described in Table I.

Given the unbalance in the class distribution and the heterogeneity of the recordings, which differ in the number of classes presented, a protocol based on inter-patient training and testing was proposed by Chazal et al [2]. This protocol suggests the sorting of the recordings of the database in a training and a test set, DS1 and DS2 respectively, in such a way that patients are not mixed in these processes. Furthermore, both groups were designed to present a comparable number of heartbeats of each class. The number of heartbeats of each class, in both groups, are presented in Table II.

TABLE I

TYPES OF BEATS FROM THE MIT-BIH ARRHYTHMIA DATASET INCLUDED IN CLASSES NORMAL,SVEB,VEB,F AND Q

	Type of heartbeat	Symbol
N	Normal beat	N
	Left Bundle Branch Block (LBBB)	L
	Right Bundle Branch Block (RBBB)	R
SVEB	Atrial premature beat (APB)	A
	Aberrated APB	a
	Nodal/Junctional prem. beat	J
	Prem./Ectopic supraventr. beat	S
	Atrial escape beat	e
	Junctional escape beat	j
VEB	Ventricular premature beat	V
	Ventricular escape beat	E
F	Fusion ventricular and normal beat	F
Q	Paced beat	P o /
	Fusion of paced and normal beat	f
	Unclassifiable beat	U

TABLE II

DESCRIPTION OF THE MIT-BIH ARRHYTHMIA DATABASE

	N	SVEB	VEB	Fusion	Unknown
DS1	45850	944	3788	414	8
DS2	44240	1837	3220	388	7

B. Pre-processing and segmentation

All ECG signals were pre-processed to remove the baseline wander and high frequency noise affecting further steps of feature extraction. The method proposed in [2], based on a two-step application of median filters, was used to remove the baseline wander. Firstly, a median filter of 200ms was used to extract the QRS complexes and the P waves, followed by a 600ms median filter to remove the T waves. The remaining signal was then subtracted from the original in order to remove the baseline wander.

Regarding high frequency noise, a Butterworth low-pass filter with cut-off frequency at 50Hz and order 6 was applied.

The Rpeak annotations provided by the MIT-BIH database were used for segmentation, applying a window of 500ms around the Rpeak, as suggested in [6].

C. Feature extraction

Continuing the work presented in [9], this study proposes new morphological, HOS and frequency features based on VDM, added to a set of temporal characteristics. The complete list of features is indicated in Table III.

1) Temporal features: The four standard temporal features previously used in [9] were extracted, normalized per patient as suggested in [1] to reduce the inter-patient variability. Furthermore, the angle between post and pre RR intervals was added as a new temporal feature, aiming to distinguish between an ideal $\sim 45^{\circ}$ angle characteristic of normal heartbeats, and other values coming from irregular heartbeats.

2) Morphological, frequency and HOS features: The morphological, HOS and frequency features of the heartbeats were extracted in this study making use of VMD, which adaptively decomposes a signal into its principal modes [10]. VMD has been proven to be a valid signal decomposition method when applied to ECG signals, given the quasi-stationary properties of these and the robustness of VMD against noise. Furthermore, the band-limited characteristics of the principal modes can be related to the different waveforms present in the ECG, making this method able to detect Rpeaks [11] or classify heartbeats [9].

TABLE III

LIST OF FEATURES EXTRACTED

Time features	Pre RR
	Post RR
	|Pre RR - post RR|
	mean(previous 10 RR) - preRR
	angle post RR, pre RR
VMD features	Bandwidth of the mode
	Power in the band
	Symmetry around the R peak
	Number of zero crossings
	Difference between minima around R peak
	Central Frequency
	Skewness
	Variance

The same tuning scheme considered in [9] was used to select the number of modes k and the bandwidth control parameter α of the VMD decomposition, which were set to 5 and 30, respectively. VMD was applied to all heartbeats and 8 features were extracted from each of the 5 modes, resulting in a set of 40 features.

In [9], the modes extracted from each heartbeat were arranged in terms of their central frequency, extracting features only from those modes related to the characteristic frequencies of the ECG (from 1 to approx 25 Hz). However, in this work this restriction was eliminated in order to exploit the frequency differences between each of the classes. Furthermore, by not sorting the modes, the relevance of each frequency band in the heartbeats is preserved. This new condition is described by the central frequency of each mode as a new feature. Additionally, the skewness and the variance of each mode were also used as features, motivated by the relevance of HOS in heartbeat classification.

D. Feature selection

Given the high number of features used to describe heartbeats, a stage of feature selection was introduced in this study. A backwards wrapper feature selector was used, based on random forest as classifier and the change in the Cohen's Kappa κ value as performance evaluation metric, given the unbalance in the class distribution. The method performs classification using the entire feature set and evaluates the difference in performance using the κ index when the values of the feature of interest are randomly permuted. Those features affecting less the performance are assumed to be redundant or irrelevant, and they are removed from the feature set. This process is repeated until only one feature remains for classification.

E. Classifier and performance metrics

The selected features were fed into an LS-SVM classifier using an RBF kernel, using 10-fold cross-validation for tuning the parameters of the kernel.

The classification results were derived from the confusion matrix as suggested by the AAMI standards, with the peculiarity of not rewarding or penalizing the classification of Fusion beats as VEBs. In addition to the sensitivity and specificity of each class, general performance metrics that take into account the high imbalance of the classes must be used. 3 indexes were considered for this aim: (1) the *j* index, which adds the sensitivities and specificities of the most important arrhythmias (SVEB and VEB) [3]; (2) the Cohen's Kappa κ, as a measure of agreement between the ground truth and the classifier; and (3) the *jk* index [4], which is a weighted sum of the *j* and κ indices.

III. RESULTS AND DISCUSSION

A. Selected features

The Cohen's Kappa κ value extracted from the feature selector is depicted in Figure 1. This error provides information about the relevance of the features. Since the drop of the classification error starts around 35 features removed, the last 10 features were kept for further classification. The features selected are listed in Table IV in order of relevance.

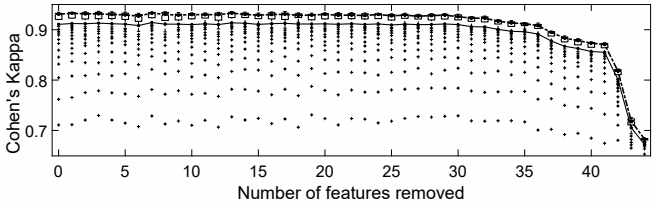

Fig. 1. Classification error vs. number of features removed

TABLE IV
SELECTED FEATURES IN ORDER OF RELEVANCE

Pre RR	Time
Mode 1 Variance	**VMD**
Mode 1 Central Freq	**VMD**
angle postRR,preRR	**Time**
PostRR	Time
Mode 1 Power	VMD
mean(previous 10 RR)-preRR	Time
Mode 2 Central Freq	**VMD**
Mode 2 Skewness	**VMD**
Mode 1 Symmetry mode around the Rpeak	VMD

As it can be seen, 4 out of the 5 proposed temporal features were selected, given the relevance of rhythm in the detection of irregular heartbeats. The angle between the post and pre RR intervals, is the second most important temporal feature. Furthermore, the comparison of the previous RR interval with the tendency of the previous 10 intervals shows more relevance than the difference between previous and posterior RR intervals. Regarding the VMD features, it is clear that the most relevant modes are the lowest ones, generally related to the lowest frequencies. The new features, related to HOS and to the central frequency of the first and second modes, are among those selected, representing the distinction in frequency content of the different classes.

B. Classification results

Table V shows the results obtained in this study and it compares them with the state-of-the-art. As it can be seen, the sensitivity of SVEB has increased in a 15% with respect to [9], where VMD was also used. The addition of the new temporal feature, and the new VMD descriptors, together with the selection of 10 features out of the entire feature set improves the detection of SVEB, at the expense of a 4% drop in specificity. The *j*,*k* and *jk* indices show all their best performance when feature selection is used, compared to [9] where 29 features were used for classification.

The sensitivity of the SVEB class is at the level of Mar et al. [4], being the best among the state-of-the-art. Regarding the sensitivity of VEB, this study presents the second best results, only below [6]. However, in terms of specificity, neither SVEB nor VEB class reach the best results of the literature, affected by the miss-classification of some of the N beats. The F class, on the other hand, presents very low sensitivity. This can be explained by the design of the features for this analysis, which originally aimed to tackle specifically SVEB and VEB.

C. Feature analysis

The analysis of the results presented in Table V highlight the general problem of the detection of SVEB in all classifiers. This problem has commonly been explained by the similar morphology of N and SVEB, even though temporal features should clearly differ from the N class.

Figure 2 shows the kernel function estimate of (a) the previous and (b) posterior RR interval for the SVEB class in DS1 and DS2. These temporal features were normalized as suggested in [1], and they were selected because they are often used in literature for heartbeat classification.

As it can be seen, the DS1 and DS2 distributions for the previous RR interval do not overlap, which explains why a classifier trained with DS1 is not able to properly classify DS2. Furthermore, there is a clear peak in DS2, corresponding to the record 232. 78% of the heartbeats in this record correspond to SVEB, which represent 50% of the SVEB in the entire database, and 77% of the SVEB in the test set.

93% of the SVEBs present in DS1 come only from 5 patients, and 40% of these belong to one single patient. Similarly, regarding the F class, 91.5% of these beats in the entire database belong to 2 patients, one used for train and the other for test in the given protocol.

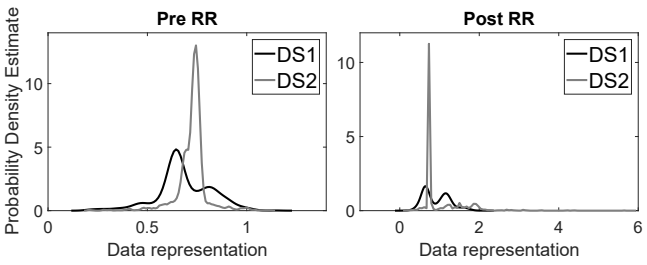

Fig. 2. Kernel function estimate for RR intervals of the SVEB class

TABLE V

RESULTS ON MIT-BIH (DS2) COMPARING THIS STUDY TO STATE-OF-THE-ART METHODS.

	N		SVEB		VEB		F		Average					
Classifier	**Se**	**P+**	**Se**	**P+**	**Se**	**P+**	**Se**	**P+**	**Se**	**P+**	**Acc**	**j index**	**k index**	**jk index**
Our method	*0.82*	*0.98*	*0.83*	*0.29*	*0.88*	*0.84*	*0.11*	*0.08*	*0.63*	*0.53*	*0.81*	*2.83*	*0.50*	*0.61*
Villa et al. [9]	0.81	0.98	0.68	0.32	0.88	0.82	0.23	0.01	0.65	0.53	0.80	2.69	0.48	0.57
Chazal et al. [2]	0.87	0.99	0.76	0.39	0.80	0.87	0.89	0.09	0.83	0.57	0.86	2.77	0.53	0.61
Mar et al. [4]	0.90	0.99	0.83	0.34	0.87	0.76	0.61	0.17	0.80	0.56	0.90	2.80	0.60	0.65
Zhang et al. [5]	0.89	0.99	0.79	0.36	0.86	0.93	0.94	0.14	0.87	0.60	0.88	2.93	0.59	0.66
Mondéjar et al. [6]	0.96	0.98	0.78	0.50	0.95	0.94	0.12	0.24	0.70	0.66	0.95	3.17	0.76	0.77

The described differences between patients and the unbalance of the class distribution emphasize the role of the database in the poor results obtained generally in the state-of-the-art for the SVEB and F classes. Given the lack of well-known substitute databases, a solution could be a different selection of patients as training set, which could allow to better train heartbeats classification algorithms.

TABLE VI

CLASSIFICATION RESULTS FOR DIFFERENT TRAINING-TEST SETS

	N		SVEB		VEB	
	Se	**P+**	**Se**	**P+**	**Se**	**P+**
(1)	0.82	0.98	0.83	0.29	0.88	0.84
(2)	0.67	0.99	0.87	0.35	0.78	0.87
(3)	0.67	1.00	0.83	0.41	0.89	0.66

D. Alternative results

A new study was performed, using the same 45 features, but dividing the recordings in new training and test set. Nevertheless, this distribution was done taking into account the amount of recordings present in each group, and the presence of all classes approximately comparable in both datasets. After dividing the files, feature selection was applied to the new training set and those features were used to describe the heartbeats. The results obtained for the N, SVEB and VEB classes are shown in Table VI for 3 cases: (1) DS1 and DS2 as recommended in [2], (2) using DS2 as training and DS1 as test and (3) a new random setting suggested.

These results show that there can be a wide variability in the performance, especially for the SVEB class, if the recordings are shuffled differently to the well-known protocol. Even thought the amount of SVEB samples and the diversity of their recordings are not enough to properly characterize this phenomena, the distribution of the recordings in train and test set seem to additionally play a major role. Given the amount of combinations that could be considered for the redistribution of recordings, there is a need for methods that can automatically select the most interesting patients of a database, in order to obtain an optimal training set, where the majority of the examples can be presented to the classifier. This would be useful for this and many other biomedical applications which require the inter-patient paradigm.

IV. CONCLUSION

A new combination of Variational Mode Decomposition and time features was proposed for heartbeat classification based on the MIT-BIH arrhythmia database. The relevance of the features related to the lowest variational modes and the temporal features was identified by a backwards wrapper feature selector. The selected features were classified using LS-SVM to distinguish 4 classes suggested by AAMI: Normal, Supraventricular, Ventricular and Fusion. This setting achieved results aligned with the state-of-the-art, which yet presents poor results for the SVEB class.

Future work should consider the addition of specific features to distinguish the F and SVEB classes. Furthermore, the cascading of classifiers, as suggested in [6], should be studied, aiming to focus on rhythmic heartbeats vs arrhythmia and subsequently on the distinction between arrhythmias.

An analysis of the temporal features and of the patients present in the database were provided, emphasizing the role of the database in the results obtained in the literature.

Methods that allow the automatic selection of the most interesting patients for training classifiers are required, or as an alternative, new heartbeat databases that provide more representative samples of the critical heartbeat classes.

ACKNOWLEDGMENT

Agentschap Innoveren & Ondernemen (VLAIO): STW 150466 OSA+. imec funds 2017. C. Varon is a postdoctoral fellow of the Research Foundation-Flanders (FWO). R. Willems is supported as postdoctoral clinical researcher by the Fund for Scientific Research Flanders.

REFERENCES

[1] C. Lin "Heartbeat classification using normalized RR intervals and morphological features", Mathematical Problems in Engineering 2014.
[2] P. de Chazal "Automatic Classification of Heartbeats Using ECG Morphology and Heartbeat Interval Features", IEEE Trans. Biomed. Eng., vol. 51, no. 7, pp. 1196–1206, 2004.
[3] M. Llamedo "An ECG classification model based on multilead wavelet transform features," Comput. Cardiol., vol. 34, pp. 105–108, 2007.
[4] T. Mar "Optimization of ECG classification by means of feature selection," IEEE Trans. Biomed. Eng., vol. 58, no. 8, 2011.
[5] Z. Zhang "Heartbeat classification using disease-specific feature selection", Comput. Biol. Med., vol. 46, no. 1, pp. 79–89, 2014.
[6] V. Mondéjar-Guerra "Heartbeat classification fusing temporal and morphological information of ECGs via ensemble of classifiers", Biomed. Signal Process. Control, vol. 47, pp. 41–48, Jan. 2019.
[7] G. B. Moody "The impact of the MIT-BIH Arrhythmia Database", IEEE Eng. Med. Biol. Mag., vol. 20, no. 3, pp. 45–50, 2001.
[8] AAMI, "Testing and reporting performance results of cardiac rhythm and ST segment measurement algorithms", ANSI/AAMI EC38, 1998.
[9] A. Villa "Variational mode decomposition features for heartbeat classification," Comput. Cardiol., 2018, pp. 1–4.
[10] K. Dragomiretskiy "Two-Dimensional Variational Mode Decomposition," IEEE Trans. Signal Process., vol. 62, no. 3, pp. 531–544, 2014.
[11] G. Goovaerts "A Machine Learning Approach for Detection and Quantification of QRS Fragmentation", IEEE J. Biomed.Heal.Inf., 2018.