A multi-class approach to automatically detect Congestive Heart Failure in windowed ECG

Giovanni Rosa[a], Marco Russodivito[a], Gennaro Laudato[a], Simone Scalabrino[a], Angela Rita Colavita[b], Rocco Oliveto[a]

[a] *STAKE Lab, Department of Biosciences and Territory, University of Molise, Pesche (IS), Italy*
[b] *ASREM, Molise Region, Campobasso (CB), Italy*

Abstract

Congestive heart failure (CHF) is a chronic heart disease that causes debilitating symptoms and leads to higher mortality and morbidity. In this paper, we present *HARPER*, a novel automatic detector of CHF episodes able at distinguishing between Normal Sinus Rhythm (NSR), CHF and no-CHF. The main advantages of *HARPER* are its reliability and its capability of providing an early diagnosis. Indeed, the method is based on the evaluation of real-time features and on the observation of a brief segment of ECG signal. *HARPER* is an independent tool with the meaning that it does not need any ECG annotatotion or segmentation algorithms to provide a detection. The approach was submitted to a complete experimentation, by involving both the intra- and inter-patient validation schemes. The results are comparable to the state-of-art methods therefore highlighting the suitability of *HARPER* to be used in modern IoMT systems as a multi-class, fast and highly accurate detector of CHF. We also provide guidelines for the configuration of a temporal window to be used in automatic detection of CHF episodes.

Keywords:
CHF, Machine Learning, DSS, Wearable, IoMT

1 Introduction

We are living in the era of wearable devices, which are important components of human health for the prevention of diseases or pathological conditions. These devices have become so widespread that they play an important role in healthcare and tele-medicine systems [1]. The key aim of incorporating technology into healthcare systems is to increase the quality and usability of medical devices and facilities by providing improved interfacing capability between patients and caregivers [2, 3]. In remote healthcare monitoring, the Internet of Medical Things (IoMT) played a critical role. The Internet of Medical Things is primarily used to collect remote data for patients via wearable sensors/devices and store it in cloud databases. These data are made available to caregivers for real-time review and implementation but also to specific systems in order to provide automatic analysis [4, 5]. A recent example of tele-medicine system is ATTICUS [6, 7]. ATTICUS provides a bustier wearable [8] able to acquire at least 6-lead continuous electrocardiogram (ECG) signal and other vital parameters. A Decision Support System (DSS) [7] - structured as a distributed AI software - is in charge of providing the early diagnosis of several pathological conditions, which could possibly require immediate attention by doctors.

One of such conditions is Congestive Heart Failure (CHF), a common pathophysiological condition common, with around 26 million adults diagnosed with the disease worldwide in 2014 [9].

Thus, much effort was undertaken by the scientific community to the aim at automatically identify CHF. Most of the works propose a detector of CHF that provides a binary classification of a signal in CHF and Normal Sinus Rhythm (NSR) [10-13]. A minor part of the scientific community has modeled the CHF detection problem as a multi-class prediction problem to avoid that pathological signals different from CHF are erroneously classified as CHF [14]; in such cases, three classes are typically used: CHF, NSR, and no-CHF (i.e., pathology different from CHF). In both cases, a variety of temporal observations were adopted. For example, Xiong *et. al.* [13] experimented the use of an ECG segment with a fixed length in terms of samples, while Porumb *et. al.* [12] relied on observations at heartbeat level. The main drawback of such approaches is that they provide only binary classifications [10-13] or that they are dependent on other safety-critical algorithms to obtain the ECG segmentation.

In this paper, we introduce HARPER (*detector of congestive **H**eart f**A**ilu**R**e e**P**isodes for m**E**dical suppo**R**t*), an automatic method to detect CHF in a single-lead digital ECG. *HARPER* is a near real-time approach capable at providing a multi-class identification of a given windowed ECG signal in: CHF, NSR, and no-CHF (i.e., pathological rhythm, but different from CHF).

The main contributions of this paper are the followings:

- we introduce *HARPER*, a reliable and indipendent detector because it is capable of providing a multi-class identification (instead of a binary one) and it is not dependent on other algorithms (e.g. a R peak detector) for the ECG

segmentation because it involves a fixed length segmentation.

- due to the fact that CHF is a chronic condition [5], we conducted a complete study to assess the duration of the best temporal window in which to observe CHF. This was done within two scenarios: (i) *inter-patient* intended as the case when no personal data is available for a new subject to be monitored in ATTICUS and (ii) *intra-patients* when personal data are available in the training set.

The rest of the paper is structured as follows: Section 2 describes the planning of our study, highlighting the workflow of *HARPER* and the datasets used while Section 3 describes the experimental procedure and the validation schemes. In section 4 "Results" are described the results obtained in all the experimental settings and finally section 5 concludes the paper by reporting an analysis on the various outcomes and highlighting the future works.

2 Methods

In this section, we present the high-level worflow of *HARPER*. Then we describe the perfomed ECG processing, and the context of this study. Moreover, we describe the experimental procedure and the validation schemes.

2.1 *HARPER* Workflow

The high-level workflow of *HARPER* is described as follows. We consider as w the length of an observation window (in seconds). *HARPER* takes as input the ECG signal from the patient, having a length greater than or equal to $w \cdot f$, where f is the sampling frequency of the ECG signal. Next, the ECG signal is divided into several segments, based on the segmentation window. On each segment we perform a feature extraction step. The extracted features are submitted to a trained machine learning model which performs the signal classification. In this way, as output of the last step, *HARPER* provides a label for the most probable classification among NSR, CHF and no-CHF.

2.2 ECG Signal Processing and Features Extraction

The signal processing performed on the ECG starts with the detrend operation, where the signal mean is computed and subtracted from the input signal. Then the following literature features [15] are extracted:

- Energy of Maximal Overlap Discrete Wavelet Transform **(**using db2 as Daubechies wavelet transform and 15 levels of decomposition),
- Autoregressive Model of order 16. We computed the AR model coefficients using the Yule-Walker estimator,
- Multifractal Wavelet Leader (using db3 as wavelet transform);
- Fast Fourier Transform.

2.3 Study Design

The final goal of this study to assess the suitability of *HARPER*, as a detector of Congestive Heart Failure.

To achieve this, we designed two research questions:

- *RQ_1: What is the optimal value of the w parameter?* With the first research question, we aim to tune the *w* parameter to understand how many seconds of observation are necessary to perform the best prediction;
- *RQ_2: What is the classification effectiveness of HARPER?* With this second research question, we want to evaluate the accuracy of *HARPER* in the detection of CHF episodes.

We also want to evaluate *HARPER* in terms of different validation techniques such as intra-patient and inter-patient (where the data of a subject is not or partially considered in the training of the model, respectively) strategies.

2.4 Context of the Study

In our study we used a dataset of 162 ECG recordings extracted from PhysioNet database [16], provided by MathWorks[1]. In the final dataset, there are 36 recording of subjects with NSR, 96 with no-CHF anomalous episodes (*i.e.,* arrhythmia) and 30 recordings from subjects affected by CHF. The reason behind the use of this dataset is that (i) it is composed of signals from different datasets, providing a more heterogenous set of ECG recordings and (ii) it contains not only CHF recordings but also those no-CHF. In this way *HARPER* can distinguish abnormal recordings with CHF episodes from those that are abnormal but contains no-CHF heart diseases.

2.4.1 Baseline approach

We selected a recent approach from the literature, *i.e.* the approach proposed by Yang *et. al* [14] as reference baseline for the evaluation of *HARPER*. Beyond the high accuracy, we chose this work because the authors perform a three-class identification of ECG in: NSR, CHF and Coronary Artery Disease (CAD). Their approach combines an ECG fragment alignment (EFA) with principal component analysis (PCA) and a convolutional network (EFAP-Net) to ensure heartbeats consistency between subjects eliminating heart rate differences. Finally, they use a linear SVM as classifier. They also succesfully applied intra-patient and inter-patient validation techniques.

The main drawback of such an approach is that it involves the ECG segmentation at heartbeat level, which could lead to a less suitability in real-time scenarios because of the high computational cost of a robust R peak detector. Due to the consideration that CHF is a chronic condition,

[1] https://github.com/mathworks/physionet_ECG_data/

in *HARPER*, we tried to avoid any dependance on external algorithms and we only focused on a fixed length observation of the ECG. Indeed, in the context of ATTICUS, we needed to design a highly reliable *near* real-time approach of CHF episodes in two scenarios: when personal data are available in the monitoring and when a new ATTICUS user has to be continuously monitored.

3 Experimental Procedure

In the next sub-sections we describe the steps of our experimental procedure.

RQ_1: ECG Segmentation

As we want to evaluate the best ECG temporal window to provide an accurate multi-class detection (therefore a better observation of CHF episodes), we need to (i) split the ECG recordings in segments and (ii) assess what is the best classifier that it is possible to define on the top of that data. First, we defined a set of possible durations of time windows. Furthermore, we performed the ECG segmentation based on the defined temporal windows and the extraction of the previously described features. We selected a set of time windows ranging from 5 to 120 seconds.

Next, for the intra-patients scenario, we built different machine learning pipelines, testing out different models in combination with pre-processing and sampling techniques. In this way, we aimed at assessing which is the best temporal window and which could be the best model to use with for the specific intra-patients scenario. For example, for RandomForest we only applied a min-max scaling, but for SVM we applied the standardization. We also evaluated the impact of data balancing techniques, such as SMOTE [17]. In detail, we first removed higly-correlated features, removing those having a Pearson correlation coefficient *r* greater than 0.95. Then, we applied a tree-based estimator feature selection technique, where the impurity-based feature importance is computed. In this way, we discarded the irrelevant features according to their importance. The final step of our classification pipeline consisted in a combination of random split of training and test set (*i.e.* 80-20), data sampling (*i.e.* SMOTE), data pre-processing (*i.e.* scaling, standardization) and a classification algorithm.

Finally, we opted for using the best model obtained from the intra-patients scenario to assess the best window duration also in the inter-patients scenario. The preprocesssing scheme used was the same.

We evaluated the classification performance using widely-used metrics for classification tasks, namely *Sensitivity, Specificity, Precision, F1 score.*

Table 2 - Dataset overview after ECG segmentation

Temporal window (seconds)	ARR	NSR	CHF	Total
5	9792	3672	3060	16524
10	4896	1836	1530	8262
15	3264	1224	1020	5508
20	2400	900	750	4050
25	1920	720	600	3240
30	1632	612	510	2754
60	768	288	240	1296
120	384	144	120	648

RQ_2: HARPER Classification and Validation

Taking into account the results of RQ_1, to assess the classification performance of *HARPER*, we conducted its validation considering the best ECG time window combined with the best performing classification pipeline. Moreover, we compared our approach with the selected baseline [14] to assess if *HARPER* has comparable performance to a state-of-the-art approach for CHF detection.

As we used the random split technique to split the dataset in train e test set, we decided to perform 1000 executions to reduce a possible bias due to the randomness.

On the other hand, for the intra-patient protocol, we performed *n* executions where *n* corresponds to the number of patients in our dataset (*i.e.* 162). For each execution, a specific patient is selected as test set and the remaining are used as training set. In this way the model observes a brand-new set of ECG recordings not evaluated before. We used the same classification metrics used in RQ_1 (*i.e. Sensitivity, Specificity, Precision, F1 score.*) taking the average value across all patients.

Table 1 - Evaluation of the best temporal ECG window and classifiers in the intra-patients scenario.

Model	Sampling	Preprocess	5s				10s				15s				20s			
			Se %	Sp %	Pre %	f1 %	Se %	Sp %	Pre %	f1 %	Se %	Sp %	Pre %	f1 %	Se %	Sp %	Pre %	f1 %
RandomForestClassifier	SMOTE	MinMaxScaler	97.73	**98.76**	97.76	97.74	98.67	**99.35**	98.69	98.67	98.46	**99.39**	98.49	98.46	98.64	**99.51**	98.67	98.65
RandomForestClassifier	-	-	**97.94**	98.08	**97.95**	**97.94**	**98.85**	99.07	**98.85**	**98.85**	**99.18**	99.24	**99.19**	**99.18**	**99.01**	98.98	**99.01**	**99.01**
SVM_Classifier	SMOTE	StandardScaler	97.43	98.69	97.48	97.44	97.76	98.82	97.80	97.77	97.91	99.35	97.99	97.92	97.90	98.85	97.94	97.91
SVM_Classifier	-	StandardScaler	97.43	98.21	97.44	97.43	98.25	98.68	98.25	98.25	97.91	98.71	97.93	97.91	98.15	98.45	98.15	98.15
KNeighborsClassifier	SMOTE	-	49.74	71.66	55.50	51.39	57.11	76.33	63.04	58.86	51.81	73.89	57.85	53.50	54.20	73.41	59.33	55.64
KNeighborsClassifier	-	-	58.79	57.82	55.30	55.68	63.04	61.83	60.30	60.88	60.25	60.90	57.35	57.81	61.85	61.80	59.87	60.01
LogisticRegressionClassifier	-	StandardScaler	94.49	95.44	94.49	94.49	96.85	97.54	96.87	96.86	96.82	97.68	96.84	96.83	97.16	98.04	97.19	97.17

Model	Sampling	Preprocess	25s				30s				60s				120s			
			Se %	Sp %	Pre %	f1 %	Se %	Sp %	Pre %	f1 %	Se %	Sp %	Pre %	f1 %	Se %	Sp %	Pre %	f1 %
RandomForestClassifier	SMOTE	MinMaxScaler	99.54	**99.70**	99.54	99.54	**99.27**	**99.36**	**99.28**	**99.27**	98.08	98.40	98.09	98.07	96.15	98.88	96.40	96.19
RandomForestClassifier	-	-	**99.69**	99.55	**99.69**	**99.69**	98.73	98.28	98.74	98.73	**99.23**	**99.26**	**99.23**	**99.23**	96.92	98.45	96.93	96.92
SVM_Classifier	SMOTE	StandardScaler	97.38	98.03	97.39	97.38	98.91	99.03	98.91	98.91	96.92	98.17	97.00	96.94	95.38	96.73	95.38	95.38
SVM_Classifier	-	StandardScaler	97.99	98.37	98.00	97.99	98.37	98.19	98.38	98.35	98.08	98.44	98.10	98.09	93.85	94.94	93.86	93.83
KNeighborsClassifier	SMOTE	-	50.93	73.08	57.02	52.69	46.46	68.32	52.63	48.18	46.54	67.72	54.51	48.71	44.62	73.56	51.71	46.51
KNeighborsClassifier	-	-	60.34	60.55	57.43	57.92	56.08	52.83	51.02	52.31	58.46	55.84	56.17	56.89	54.62	64.81	51.29	52.21
LogisticRegressionClassifier	-	StandardScaler	96.91	97.53	96.92	96.91	97.82	97.83	97.84	97.82	97.69	98.30	97.74	97.69	**97.69**	**99.42**	**97.93**	**97.74**

4 Results

In the next sub-section we describe the results achieved to answer our research questions.

RQ₁: ECG Segmentation

In Table 1 the results of our experiment in the intra-patients scenario are reported. For each time window, the percentage scores are displayed related to the classification metrics corresponding to each classification pipeline, where we described the used model and the data preprocessing techniques.

The main outcome of this experimentation is that an observation of 25 seconds allowed to obtain the best overall metrics with a pipeline composed only of a RandomForest model. In this case, the results exceeded 0.99 on all the evaluated classification metrics. In Figure 1, the boxplots of the classification metrics obtained by experimenting each time window are shown. In particular, from this distribution of data it is also possible to derive how the time observation with the highest median of the classification performances is the 25 seconds window. However, in terms of robustness the 20 and 25 time windows are the best, because even the outliers are approximately above 0.96. On the contrary, this happens for the 120s time window. For example, the Specificity score is below 0.88 considering the outliers below the first quartile.

The results of the intra-patient valildation protocol showed that in this specific condition the best time window is defined by 60 s and not 25 s as previously obtained in the other scenario (93.13 vs 91.64 in terms of F1 score).

The reason behind this could be that in the case of a patient that is never examined before, a larger time window (*i.e.* a longer ECG buffering) is needed to achieve an accurate classification.

RQ₂: HARPER Classification and Validation

Taking into account the results from RQ_1, we used as reference time window 25 seconds and the *raw* RandomForest as reference classification pipeline. We also executed the validation on the other temporal windows to compare and verify if the previously selected time window is the best also in a inter-patient scenario. In Table 3 Table 3 the average percentage classification metrics of the validation protocol are reported, for the inter-patients scheme. We achieved a score sligthly worse than the one achieved in RQ_1 (where we reported the results of a single execution) but we aimed at preserve the repeatibility of our experiment by avoiding the contribution of randomness.

We compared our two validation results with the baseline approach where we only considered the NSR and CHF classes from the multi-class detection, because one of our classes (ARR) differs from the one proposed by Yang et al. [14] (CAD). In Table 4 we reported our results compared to the baseline approach. On the left side, there are the results of the inter-patient validation where the baseline approach slightly outperforms *HARPER* for a few percentage points. This can due to their perfectly balanced dataset. For the intra-patient validation (right side), in some cases HARPER outperforms the baseline, mainly on NSR class. For the CHF class, we have a better Sensitivity and Specificity values but much lower values of Precision and F1-score. This means that our approach have more false positives than the baseline.

Figure 1 - Metrics distribution for the intra-patient scheme.

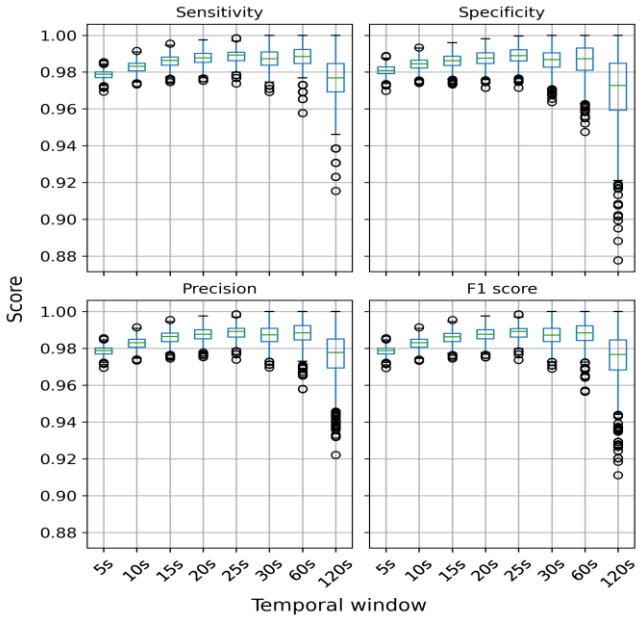

Table 3 – Classification metrics of HARPER using a window of 25 seconds and inter-patient validation protocol.

Class	Se (%)	Sp (%)	Pre (%)	f1 (%)
NSR	98.74	98.3	99.23	98.98
CHF	97.63	99.67	98.55	98.08
ARR	99.31	99.78	98.83	99.07
Avg.	**98.88**	**98.87**	**98.88**	**98.87**

Table 4 - Classification performance compared with the baseline approach [14] of intra-patient validation (left) and inter-patient validation (right).

Class NSR

Metric	Our approach	Yang *et. al* [25]	Delta
Se (%)	98.74	99.89	-1.15
Sp (%)	98.3	99.86	-1.56
Pre (%)	99.23	99.82	-0.59
f1 (%)	98.98	99.93	-0.95

Class CHF

Metric	Our approach	Yang *et. al* [25]	Delta
Se (%)	97.63	99.87	-2.24
Sp (%)	99.67	99.81	-0.14
Pre (%)	98.55	99.82	-1.27
f1 (%)	98.08	99.90	-1.82

Class NSR

Metric	Our approach	Yang *et. al* [25]	Delta
Se (%)	98.95	96.57	+ **2.38**
Sp (%)	99.60	98.79	+ **0.81**
Pre (%)	98.61	92.44	+ **6.17**
f1 (%)	98.78	99.26	-0.48

Class CHF

Metric	Our approach	Yang *et. al* [25]	Delta
Se (%)	91.94	89.61	+ **2.33**
Sp (%)	93.78	85.78	+ **8.00**
Pre (%)	71.25	83.85	-12.60
f1 (%)	80.28	92.65	-12.37

5 Conclusions & Future Works

The results of this work clearly highlight that a longer (60 s) observation of the ECG is needed to best detect CHF episodes when monitoring in real-time a new user of an IoMT system. Such a duration can be reduced (25 s) once enough data points are made available to the ML pipeline.

HARPER is highly accurate and it showed a great potential to be embedded in scenarios of continuous monitoring due to its high accuracy in the detection and to the technological independence: indeed, no other algorithms are needed to obtain the ECG segmentation.

HARPER can be considered reliable because it concerns also the classification of pathological rhythm different from CHF.

As part of our future agenda, we plan to validate the accuracy of *HARPER (i)* when common ECG noises (electrode movement or motion artefact) are spread in the signal, and *(ii)* within the signals directly acquired by the ATTICUS smart vest.

Acknowledgements

This work is supported by the project PON-ARS01_00860 titled *Ambient-intelligent Tele-monitoring and Telemetry for Incepting and Catering over hUman Sustainability - ATTICUS* funded by the Italian Ministry of Education and Research - RNA/COR 576347.

References

[1] J.J. Rodrigues, D.B.D.R. Segundo, H.A. Junqueira, M.H. Sabino, R.M. Prince, J. Al-Muhtadi, and V.H.C. De Albuquerque, Enabling technologies for the internet of health things, *Ieee Access* **6** (2018), 13129-13141.

[2] F. Firouzi, A.M. Rahmani, K. Mankodiya, M. Badaroglu, G.V. Merrett, P. Wong, and B. Farahani, Internet-of-Things and big data for smarter healthcare: From device to architecture, applications and analytics, in, Elsevier, 2018.

[3] A. Gatouillat, Y. Badr, B. Massot, and E. Sejdić, Internet of medical things: A review of recent contributions dealing with cyber-physical systems in medicine, *IEEE internet of things journal* **5** (2018), 3810-3822.

[4] M. Cornacchia, K. Ozcan, Y. Zheng, and S. Velipasalar, A survey on activity detection and classification using wearable sensors, *IEEE Sensors Journal* **17** (2016), 386-403.

[5] G.J. Joyia, R.M. Liaqat, A. Farooq, and S. Rehman, Internet of Medical Things (IOMT): applications, benefits and future challenges in healthcare domain, *J Commun* **12** (2017), 240-247.

[6] E. Balestrieri, F. Boldi, A.R. Colavita, L. De Vito, G. Laudato, R. Oliveto, F. Picariello, S. Rivaldi, S. Scalabrino, and P. Torchitti, The architecture of an innovative smart T-shirt based on the Internet of Medical Things paradigm, in: *2019 IEEE International Symposium on Medical Measurements and Applications (MeMeA)*, IEEE, 2019, pp. 1-6.

[7] R. Oliveto, G. LAUDATO, S. Scalabrino, A.R. Colavita, Q. Chiacchiari, R. D'orazio, R. Donadelli, L. De Vito, F. Picariello, and I. Tudosa, ATTICUS: Ambient-intelligent Tele-monitoring and Telemetry for Incepting and Catering over hUman Sustainability, *Frontiers in Human Dynamics* **3** (2021), 19.

[8] L. De Vito, E. Picariello, F. Picariello, I. Tudosa, D. Loprevite, D. Avicolli, G. LAUDATO, and R. Oliveto, An undershirt for monitoring of multi-lead ECG and respiration wave signals, in: *International workshop on Metrology for Industry 4.0 and IoT* IEEE, 2021, p. To Appear.

[9] P. Ponikowski, S.D. Anker, K.F. AlHabib, M.R. Cowie, T.L. Force, S. Hu, T. Jaarsma, H. Krum, V. Rastogi, and L.E. Rohde, Heart failure: preventing disease and death worldwide, *ESC heart failure* **1** (2014), 4-25.

[10] U.R. Acharya, H. Fujita, S.L. Oh, Y. Hagiwara, J.H. Tan, M. Adam, and R. San Tan, Deep convolutional neural network for the automated diagnosis of congestive heart failure using ECG signals, *Applied Intelligence* **49** (2019), 16-27.

[11] Z. Masetic and A. Subasi, Congestive heart failure detection using random forest classifier, *Computer methods and programs in biomedicine* **130** (2016), 54-64.

[12] M. Porumb, E. Iadanza, S. Massaro, and L. Pecchia, A convolutional neural network approach to detect congestive heart failure, *Biomedical Signal Processing and Control* **55** (2020), 101597.

[13] J. Xiong, X. Liang, L. Zhao, B. Lo, J. Li, and C. Liu, Improving Accuracy of Heart Failure Detection Using Data Refinement, *Entropy* **22** (2020), 520.

[14] W. Yang, Y. Si, D. Wang, G. Zhang, X. Liu, and L. Li, Automated intra-patient and inter-patient coronary artery disease and congestive heart failure detection using EFAP-Net, *Knowledge-Based Systems* **201** (2020), 106083.

[15] G. Rosa, G. Laudato, A.R. Colavita, S. Scalabrino, and R. Oliveto, Automatic Real-time Beat-to-beat Detection of Arrhythmia Conditions, in: *14th International Conference on Health Informatics*, 2021.

[16] A.L. Goldberger, L.A. Amaral, L. Glass, J.M. Hausdorff, P.C. Ivanov, R.G. Mark, J.E. Mietus, G.B. Moody, C.-K. Peng, and H.E. Stanley, PhysioBank, PhysioToolkit, and PhysioNet: components of a new research resource for complex physiologic signals, *circulation* **101** (2000), e215-e220.

[17] N.V. Chawla, K.W. Bowyer, L.O. Hall, and W.P. Kegelmeyer, SMOTE: synthetic minority over-sampling technique, Journal of artificial intelligence research 16 (2002), 321-357.

Address for correspondence

Rocco Oliveto
rocco.oliveto@unimol.it