Tornero, R.; Fácila, L.; Bertomeu-González, V.; Zangróniz, R.; Alcaraz, R.; Rieta, JJ. (2021). Is Short-Term Heart Rate Variability Good Enough to Predict Vascular Events in Hypertensive Patients?. IEEE. 1-4. https://doi.org/10.1109/EHB52898.2021.9657659

Is Short-Term Heart Rate Variability Good Enough to Predict Vascular Events in Hypertensive Patients?

Roberto Tornero[1], Lorenzo Fácila[2], Vicente Bertomeu-González[3], Roberto Zangróniz[4], Raúl Alcaraz[4], José J. Rieta[1*]

[1] BioMIT.org, Electronic Engineering Department, Universitat Politecnica de Valencia, Spain; {rotorcos, jjrieta}@upv.es

[2] Cardiology Department, Hospital General Universitario de Valencia, Spain; lfacila@gmail.com

[3] Clinical Medicine Department, Miguel Hernández University, Elche, Spain; vbertomeu@umh.es

[4] Research Group in Electronic, Biomedical and Telecomm. Eng., Univ. of Castilla-La Mancha, Spain; raul.alcaraz@uclm.es

***Abstract*—Vascular events are the main cause of premature death and disability in the developed countries, where there is great interest in the development of computational tools for their early detection. A very relevant variable for their study is the heart rate, that can be analyzed through heart rate variability (HRV). Furthermore, high blood pressure is an important risk factor for most cardiovascular diseases. In fact, small reductions in blood pressure are known to markedly reduce cardiovascular morbidity and mortality. This study evaluates the predictive value of short-term HRV (STHRV) by developing models based on data mining algorithms to stratify the risk of vascular events from hypertensive patients. For this specific framework, the performance of various machine learning models (Random Forest, Support Vector Machines, Gaussian Naive Bayes, K-N Nearest Neighbours and Logistic regression), trained with different time lengths of 5, 30 and 60 minutes of HRV features during sleep stage was compared. The analyzed HRV parameters were associated to time, frequency and nonlinear features. A total of 139 Holter recordings from hypertensive patients of whom 17 developed a vascular event were analyzed. Results indicated that classification models developed using STHRV, with only 5 minutes length, provided similar or even better results than those developed with longer time series. Furthermore, the STHRV models provided a higher sensitivity and a slightly higher F1 score. The best one, based on Support Vector Machines, yielded 88.2% sensitivity and 75% F1 score. Thus, this research suggests the feasibility of STHRV analysis for risk stratification of hypertensive patients to anticipate serious vascular events.**

***Keywords*—Heart rate variability; HRV; blood pressure; hypertension; machine learning; classification models; vascular event.**

I. Introduction

A report from the World Health Organization classifies cardiovascular diseases (CVD) as the main causes of mortalities in the last century, from the years 2000 to 2019, [1]. In this report, ischemic heart diseases and strokes are revealed as the two main factors of death around the world uninterruptedly since 2000. In fact, ischemic heart diseases led to more than 9 million of deaths only in 2019 and the 16% of deceases in this century. It is well known the impact of hypertension on the risk of many CVDs [2] and there is a clear link hypertension and the occurrence of serious cardiovascular events [3]. It has been suggested that hypertension may result in a reduced adaptation of the autonomic nervous system to haemodynamic changes.

Heart Rate Variability (HRV) analysis is used to investigate in a non-invasive way the influence of autonomic nervous system on the electric heart activity [4]. HRV has an incredible potential to provide insights into physiological and pathological conditions and to enhance risk stratification through the use of 24-hour electrocardiographic (ECG) monitoring technology. A low or depressed heart rate variability (HRV), where the sympathetic system dominates, may reduce this adaptation and anticipate a CVD [5], [6]. Therefore, HRV analysis could help to identify patients at higher risk of a potential vascular event and provide the possibility of acting accordingly in a preventive sense. The aim of this study is to validate the feasibility of using 5 minutes short-term HRV (STHRV) analysis to stratify vascular events' risk in hypertensive patients. Different time-lengths of HRV analysis will be compared to assess if performance of risk classification decreases with smaller lengths of HRV time series.

II. Materials

The database Smart Health for Assessing the Risk of Events via ECG, collected at the Centre of Hypertension of the University Hospital Federico II was retrieved from Physionet [7]. It consisted of samples of 24-h ECG Holter recordings from 139 patients, plus other details such as vascular characteristics evaluated by cardiac and carotid ultrasonography.

Patients aged over 55 years (49 female and 90 male) were followed up for 12 months after the recordings and labelled whether they suffered major cardiovascular or cerebrovascular events. 17 patients suffered a major event (11 myocardial infarctions, 3 strokes, 3 syncopal events). This database has already being tested [8], but due to the low proportion of patients who suffered an event and patients who did not (17 and 122), the authors altered the dataset with oversampling techniques. In the present work, the dataset was neither altered nor oversampled.

III. METHODS

A. Extraction and preprocessing of RR series

Samples were obtained during patient's sleep at night, which is assumed to be between 00:00 A.M. and 06:00 A.M.. HRV analysis during sleep stages is preferred and highly valued, because HRV is relatively high, whereas the presence of different cardiovascular diseases can make HRV lower due to a loss in activation capacity of vagus nerves [9]. Furthermore, ECG contamination due to patient's muscular activity is reduced at night. Excerpts of 5 minutes, 30 minutes and 1 hour in length were randomly extracted from the ECG recordings of the each patient. The global process is illustrated in Fig. 1. Since recordings were sampled at 128 Hz and guidelines about HRV analysis suggest a minimum sampling frequency of 250 Hz [4], all the signals were accordingly upsampled to 256 Hz.

R peaks from the ECG were detected by the Pan-Tompkins algorithm [10]. Then, RR time series were obtained calculating the difference between adjacent R points. However, the series must be corrected from outliers due to noise, missing values or ectopic beats. This was mostly made by visual inspection.

Possible outliers were identified first by some physiological decision rules such as longer than 2000 ms or shorter than 345 ms. Also, a simple detection algorithm was developed based on targeting values out of $\pm$ 3 times the standard division of the detrended series. Finally, a tool to detect premature ventricular contractions was used [11]. It was based on the application of a convolutional neural network to the wavelet transform of the raw ECG channel. The final decision to label a value as outlier was based on visual inspection.

B. HRV analysis features

Several types of HRV analysis were performed to study the randomly chosen signal sections. HRV analysis could be classified as linear analysis in time and frequency domains as well as non linear analysis. In total, 32 HRV features [4] were calculated to later develop machine learning models.

- *Time domain features*. Statistical methods were calculated such as the mean (AVNN), standard division (SDNN), the square root of the mean squared of differences of successive NN intervals (RMSSD), number of successive differences of NN intervals greater than 50 ms (NN50) and the proportion of the total (pNN50). Also, geometrical methods were included such as the HRV triangular index (HRVTi) and the triangular interpolation of NN interval histogram (TINN).
- *Frequency domain features*. Power spectral density (PSD) was computed with Lomb-Scargle periodogram. Frequency was divided in three components: very low frequency (VLF: $0 - 0.04$ Hz), low frequency (LF: $0.04 - 0.15$ Hz) and high frequency (HF: $0.15 - 0.4$ Hz). Absolute power, relative power and peak frequency was calculated for every frequency band. Also, spectral power for LF and HF was normalized. Finally, LF/HF ratio was included as well as the total PSD.

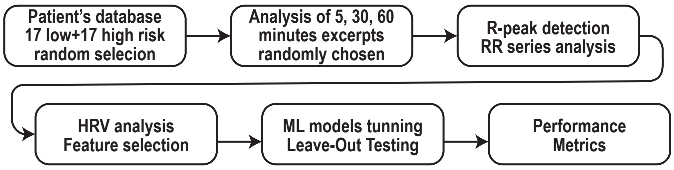


Fig. 1. General flowchart applied illustrating the steps carried out.

- *Non-linear features*. SD1 and SD2 measures from Poincaré Plot were included. Approximate Entropy and Sample Entropy were calculated with an embedding dimension of $m = 2$ and tolerance of $r = 0.2$ times standard deviation of the data. Correlation Dimension was calculated with the embedding dimension set at 10 and time delay at 5. Detrended Fluctuation Analysis was also performed with α_1 for short-term within range $4-16$ beats and α_2 for longer fluctuations within range $16-64$ beats. Finally, several recurrence plots measures were obtained such as recurrence rate, maximal length of lines, mean of length, determinism, and Shannon Entropy.

C. Machine Learning for classification

A total of 5 types of machine learning models were tested for this framework: Random Forest, Support Vector Machines, Gaussian Naive Bayes, K-N Nearest Neighbours and Logistic regression. Due to imbalance of the dataset and the tiny number of samples for the target class, with only 17 patients, we opted for using leave-one-out cross-validation. We did not choose to oversample the dataset, as other previous studies [8], due to the risk of overfitting with such a few samples. Instead, classes were balanced with a severe undersampling, randomly keeping 17 samples from the low-risk class to use a smaller subset of 34 samples, composed of 17 samples from each class.

Because the goal of this study is to validate if 5 minutes of HRV analysis are able to provide similar results than larger HRV time series, accurate classification models for 5, 30 minutes and 1 hour were developed. To do so, the models were hyperparameter-tuned with a cross-validation of 17-fold and later tested with the leave-one-out strategy. Classification performance was computed using confusion matrix measures: accuracy, sensitivity, specificity and F1-score.

D. Feature selection

Due to the small number of samples compared to the number of measures and analysis, three feature selection algorithms were applied to select the best six predictors for training classification models in 5, 30 and 60 minutes segments: χ^2 statistics method [12], Minimum redundancy maximum relevance (mRMR) method [13] and Predictor Importance Estimates by permutation using Random Forest. By using relevant features, classification algorithms can generally improve their predictive accuracy, shorten the learning period and avoid possible overfitting of models, thus increasing the generalization of their predictions.

TABLE I

PERFORMANCE MEASUREMENTS ESTIMATED BY LEAVE-ONE-OUT FOR 5 MINUTES HRV ANALYSIS. FS: FEATURE SELECTION ALGORITHM, ACC: ACCURACY, SEN: SENSITIVITY, SPE: SPECIFICITY, F1: F1-SCORE

Model	FS	ACC	SEN	SPE	F1
RF	RF-FS(6)	61.8%	70.6%	52.9%	64.9%
SVM	**X2-FS(6)**	**70.6%**	**88.2%**	**52.9%**	**75.0%**
NB	**X2-FS(6)**	**70.6%**	**82.4%**	**58.8%**	**73.7%**
KNN	X2-FS(6)	52.9%	47.1%	58.8%	50.0%
LR	RF-FS(6)	64.7%	70.6%	58.8%	66.7%

TABLE II

LEAVE-ONE-OUT RESULTS FOR 30 MINUTES HRV ANALYSIS.

Model	FS	ACC	SEN	SPE	F1
RF	X2-FS(6)	64.7%	47.1%	82.4%	57.1%
SVM	**X2-FS(6)**	**76.5%**	**58.8%**	**94.1%**	**71.4%**
NB	mRMR(6)	67.6%	52.9%	82.4%	62.1%
KNN	RF-FS(6)	61.8%	52.9%	70.6%	58.1%
LR	RF-FS(6)	70.6%	58.8%	82.4%	66.7%

TABLE III

LEAVE-ONE-OUT RESULTS FOR 60 MINUTES HRV ANALYSIS.

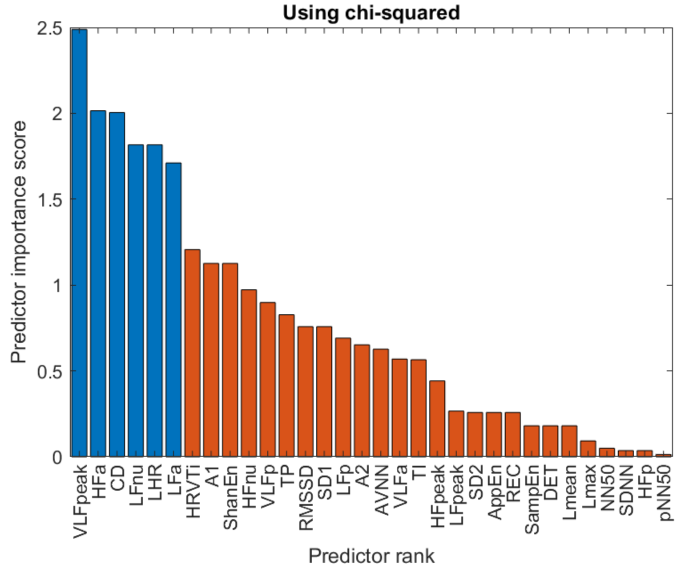
Model	FS	ACC	SEN	SPE	F1
RF	RF-FS(6)	67.6%	64.7%	70.6%	66.7%
SVM	**RF-FS(6)**	**70.6%**	**70.6%**	**70.6%**	**70.6%**
NB	X2-FS(6)	47.1%	52.9%	41.2%	50.0%
KNN	X2-FS(6)	64.7%	70.6%	58.8%	66.7%
LR	RF-FS(6)	64.7%	58.8%	70.6%	62.5%

IV. RESULTS

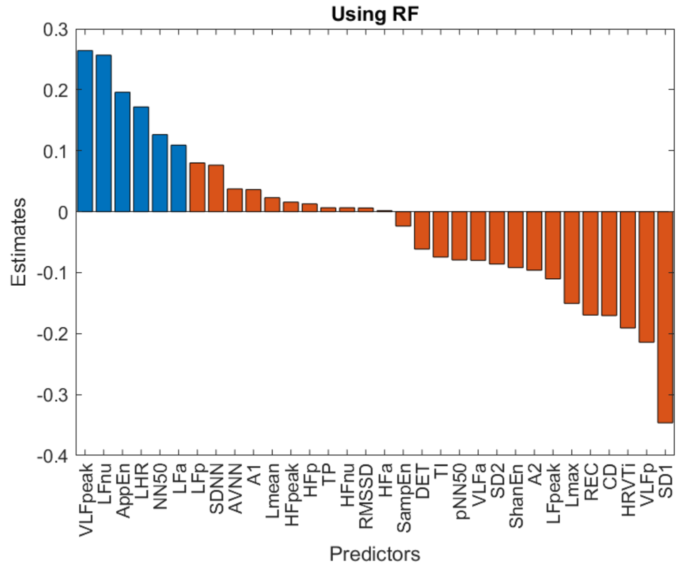
Tables I, II and III show results of the five classification models applied. The best model in each case has been bolded. Figure 2 shows an example of the outcomes of feature selection methods for the 5 minutes HRV analysis.

In accordance with the general metrics, accuracy (ACC) and F1-Score, results are similar for the three HRV lengths. With the exception of KNN, classification models trained with only 5 minutes of HRV analysis showed a F1 Score of 65% - 75%, with higher sensitivity values than specificity. In this case, the target class "*high-risk*" was better classified while the "*low-risk*" class tended to be misclassified more often. Although linear SVM model showed the best sensitivity, Gaussian NB was also provided as the best performance due to a smaller imbalance between sensitivity and specificity. NB classification model showed slightly good accuracy of 70.6% and a good ability to detect high-risk of 82.4%. F1 score of this model was 73.7% while SVM showed 88.2% sensitivity and F1 score of 75%. Although SVM and NB models were highly good in selecting the target class, specificity was worser than models in Tables II and III.

For 30 minutes HRV series, F1 scores were slightly lower than for 5 minutes, providing F1 values between 57% and 71%. The best model was a square polynomial SVM. Its accuracy was 76.5%, with a specificity for low-risk samples of 94.1%. However, sensitivity for the target class was poor, being only 58.8%. Surprisingly, classification models for this time period performed much better when it came to dismissing "*low-risk*" class, but their general ability to detect "*high-risk*" samples was poor. The period of 1 hour provided F1 values similar to models for 30 minutes, but with classification ability more balanced. The best classification model was a square polynomial SVM again, with sensitivity and specificity of 70.6%, therefore, accuracy and F1 score showed 70.6% as well. In general, the average F1 score of all models was similar to 30 minutes models, but with a more balanced predictions, exchanging ability to dismiss "*low-risk*" samples for improving detection of "*high-risk*".



(a) χ^2 feature selection



(b) Out-of-bag RF permutation

Fig. 2. Example of ranking of predictors by two feature selection algorithms for 5-minutes HRV dataset. (a) χ^2 feature selection. (b) Out-of-bag RF permutation. In Blue are the top 6 ranked predictors, whereas in orange are the discarded predictors.

V. Discussion

Although the goal of this research was to obtain high-performance classification models able to anticipate serious vascular events in hypertensive patients, the significant unbalance of the dataset did not allow to do so. Hence, the purpose of the study was to assess if using only 5 minutes of HRV time series allowed to train a classification models with similar performance as classification models trained with significantly longer time periods. This topic had already been discussed in the SHAREE project [8], but the authors applied an oversampling technique to an extremely small subset of samples for the target class. Besides, the authors only compared the HRV analysis with vascular echographic analysis and there was never an attempt to evaluate longer HRV time series.

As a consequence, this research proposes a new approach without any oversampling technique, making use of only 34 samples, but with a robust leave-one-out cross-validation strategy. Obviously, the 34 samples dataset can be considered as not enough to obtain accurate real performance, but could be enough to compare the suitability of 5 minutes length HRV time series with other time lengths. Because of the small number of samples, a feature selection of the best predictors was mandatory to avoid overfitting of the classification models.

Although, the average performance was not incredibly high due to the unbalanced dataset and its small number of samples, sensitivity in classification models trained with 5 minutes short-term HRV analysis have provided a high performance to alert of a possibly important risk of developing serious vascular events in the incoming year. This outcome is quite adequate for this topic, where misclassifying sometimes a low-risk patient is preferable rather than failing in the detection of high-risk patients. Hence, our results suggest that machine learning models could be very useful to monitor hypertensive patients with STHRV excerpts aimed at alerting of potential cardiac risks. In fact, similar works in the context of sudden cardiac death have already been done analyzing STHRV time series with promising results [14].

Finally, sleep stage has been introduced in several studies as a good condition to perform HRV analysis, due to an absence of sympathetic activity burst and a more stationary heart rate [15]. Feature selection algorithms showed that the most important ones to predict the risk of vascular events were frequency domain features. Thus, the variation of spectral power in frequency bands of the inter-beat intervals during sleep could show the adaptation of the autonomous nervous system and the competition between sympathetic and parasympathetic branches, which is reduced in hypertensive patients. Next steps in this research should be collecting a larger number of samples of hypertensive patients with vascular events to make sure of representing the real distribution of HRV metrics.

VI. Conclusions

The present study has demonstrated that five minutes short-term Heart Rate Variability is a feasible tool to anticipate serious vascular events in hypertensive patients. Classification models showed similar performance for the different HRV series lengths analyzed, but those trained with STHRV showed a higher sensitivity for 5 minutes analysis and a slightly higher F1 score metric. Support Vector Machines classifiers provided the highest performance.

Acknowledgments

Research supported by grants DPI2017–83952–C3 from MINECO/AEI/FEDER UE, SBPLY/17/180501/000411 from JCCLM and AICO/2021/286 from GVA.

References

[1] World Health Organization, "Leading causes of death and disability 2000-2019: A visual summary," WHO, Tech. Rep., 2019.

[2] R. S. Vasan, M. G. Larson, E. P. Leip, J. C. Evans, C. J. O'Donnell, W. B. Kannel, and D. Levy, "Impact of high-normal blood pressure on the risk of cardiovascular disease," *New England journal of medicine*, vol. 345, no. 18, pp. 1291–1297, 2001.

[3] T. Harris, E. Cook, W. Kannel, A. Schatzkin, and L. Goldman, "Blood pressure experience and risk of cardiovascular disease in the elderly." *Hypertension*, vol. 7, no. 1, pp. 118–124, 1985.

[4] M. Malik, J. T. Bigger, A. J. Camm, R. E. Kleiger, A. Malliani, A. J. Moss, and P. J. Schwartz, "Heart rate variability: Standards of Measurement, Physiological Interpretation and Clinical Use," *European Heart Journal*, vol. 17, pp. 354–381, 1996.

[5] S. Hillebrand, K. B. Gast, R. de Mutsert, C. A. Swenne, J. W. Jukema, S. Middeldorp, F. R. Rosendaal, and O. M. Dekkers, "Heart rate variability and first cardiovascular event in populations without known cardiovascular disease: meta-analysis and dose–response meta-regression," *Europace*, vol. 15, no. 5, pp. 742–749, 2013.

[6] V. Bertomeu, A. Cordero, J. Quiles, P. Mazón, J. Aznar, and H. Bueno, "Control de los factores de riesgo y tratamiento de los pacientes con cardiopatía isquémica: registro trece," *Revista española de cardiología*, vol. 62, no. 7, pp. 807–811, 2009.

[7] 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, and H. E. Stanley, "PhysioBank, PhysioToolkit, and PhysioNet: Components of a new research resource for complex physiologic signals," *Circulation*, vol. 101, no. 23, pp. e215–e220, 2000 (June 13).

[8] P. Melillo, R. Izzo, A. Orrico, P. Scala, M. Attanasio, M. Mirra, N. De Luca, and L. Pecchia, "Automatic prediction of cardiovascular and cerebrovascular events using heart rate variability analysis," *PloS one*, vol. 10, no. 3, p. e0118504, 2015.

[9] E. Vanoli, P. B. Adamson, Ba-Lin, G. D. Pinna, R. Lazzara, and W. C. Orr, "Heart rate variability during specific sleep stages: a comparison of healthy subjects with patients after myocardial infarction," *Circulation*, vol. 91, no. 7, pp. 1918–1922, 1995.

[10] J. Pan and W. J. Tompkins, "A real-time QRS detection algorithm." *IEEE Transactions on Biomedical Engineering*, vol. 32, no. 3, pp. 230–236, Mar. 1985.

[11] Q. Li, C. Liu, Q. Li, S. P. Shashikumar, S. Nemati, Z. Shen, and G. D. Clifford, "Ventricular ectopic beat detection using a wavelet transform and a convolutional neural network," *Physiological measurement*, vol. 40, no. 5, p. 055002, 2019.

[12] H. Liu and R. Setiono, "Chi2: Feature selection and discretization of numeric attributes," in *Proceedings of 7th IEEE International Conference on Tools with Artificial Intelligence*. IEEE, 1995, pp. 388–391.

[13] H. Peng, F. Long, and C. Ding, "Feature selection based on mutual information criteria of max-dependency, max-relevance, and min-redundancy," *IEEE Transactions on pattern analysis and machine intelligence*, vol. 27, no. 8, pp. 1226–1238, 2005.

[14] E. Ebrahimzadeh, M. Pooyan, and A. Bijar, "A novel approach to predict sudden cardiac death (scd) using nonlinear and time-frequency analyses from hrv signals," *PloS one*, vol. 9, no. 2, p. e81896, 2014.

[15] G. Brandenberger, M. Buchheit, J. Ehrhart, C. Simon, and F. Piquard, "Is slow wave sleep an appropriate recording condition for heart rate variability analysis?" *Autonomic Neuroscience*, vol. 121, no. 1-2, pp. 81–86, 2005.