Automatic prediction of falls via Heart Rate Variability and data mining in hypertensive patients: the SHARE project experience.

P. Melillo[1,2], A. Jovic[3], N. De Luca[4], and L. Pecchia[2,5]

[1] Multidisciplinary Department of Medical, Surgical and Dental Sciences, Second University of Naples, Naples, Italy
[2] SHARE Project, Italian Ministry of Education, Scientific Research and University, Rome, Italy
[3] Faculty of Electrical Engineering and Computing, University of Zagreb, Croatia
[4] Department of Translational Medical Sciences, University of Naples Federico II, Naples, Italy
[5] School of Engineering, University of Warwick, Coventry, UK

Abstract— **Accidental falls in elderly is a major problem. This paper presents the preliminary results of a retrospective study investigating association between Heart Rate Variability (HRV) measures and risk of falling, analyzing 168 clinical 24-hour ECG recording from hypertensive patients, 47 of them experienced at least one fall in the three months before/after the registration. Several HRV patterns, based on 68 linear and non-linear HRV measures, were analyzed in relation to falls using advanced statistical and data mining methods.**

The results demonstrated that there is a significant association between a depressed HRV and the risk of falling, suggesting that a depressed HRV could be a new independent risk factor for falls with an odds ratio of 5.12 (CI 95% 1.42-18.41; p<0.01).

Keywords— **heart rate variability (HRV), accidental falls, fall risk factors, falls prediction, data mining.**

I. Introduction

FALLS represent a major problem for modern societies given its burden and implication on quality of life and autonomy of elderly and their informal caretakers [1]. The mean and median costs for a fall are about 9,000 and 11,000 euro [2]. Falls are caused by complex and dynamic interactions between intrinsic (subject-based) and extrinsic (environmental) factors [3]. Over 400 risk factors have been identified [4] and their prioritization remains unclear [5]. Moreover, the applicability, sensitivity and particularly, the specificity of subject-specific assessment of falls' risks remain imprecise [6].

Recent systematic reviews investigated the independent capability of different technologies to prevent falls [7-11], highlighting their limits. For instance, false alarms' rate is too high to maintain full attention of the nursing staff [7]. Moreover, the majority of these technologies have no other direct benefit for the elderly health problems (i.e. cardiovascular disease), and may not be cost-effective and sustainable. Moreover, older people are used to be monitored for health issues, but are not willing to be monitored in order to eventually prevent 'just' a fall [12].

Differently, this study investigated association between Heart Rate Variability (HRV) and falls, as HRV is proved to be effective in monitoring other conditions that affect later life. This is a relevant consideration for falls prevention, as the most frequent co-morbidities of patients hospitalized for a fall are cardiovascular diseases [2] that significantly benefit from HRV monitoring. These include: hypertension (63%), coronary atrial fibrillation (30%), artery disease (25%), and congestive heart failure (20%).

This paper presents the preliminary results of a retrospective study investigating the relationship between HRV patterns and the risk of falling. The hypothesis that we explored in this study is that a reduced HRV complexity is associated with an increased risk of falling, because it reflects the deteriorating state of the Autonomous Nervous System (ANS).

II. Methods and Materials

A. Study design

According to conventions used for retrospective cross-sectional studies, the following was assumed: HRV pattern corresponding to the rules identified with data mining methods was the "risk factor" under investigation; fallers were considered as "cases" and non-fallers as "controls"; patients with a HRV positive to these rules for at least the 10% of the day (nominally 24 hours) where considered "exposed" to the risk factor under investigation.

The current study analyzed clinical 24h ECG Holter recordings of 168 hypertensive patients (72±8years, 60 female) and among them 47 subjects experienced a fall within the 3 months before or after the registration. The database was collected in the framework of the Smart Health and Artificial intelligence for Risk Estimation (SHARE) project and details about the database can be found elsewhere [13].

B. HRV processing

The series of RR beat intervals were obtained from ECG recordings using an open-source software [14]. HRV was analyzed concurrently in segments (excerpts) of 30 minutes, excluding the ones with less than 600 valid beats [15]. Standard linear HRV analysis, according to International Guidelines, was performed [16]. Additionally, nonlinear features were computed according to recent literature [17]. The HRV analysis was performed using open-source HRV software [18, 19].

Standard time-domain HRV measures were calculated: average of all RR intervals, standard deviation of all NN intervals (SDNN), squares of differences between adjacent NN intervals, number and percentage of differences between adjacent NN intervals that are longer than 50 ms (NN50 and pNN50, respectively), standard deviation of the averages of NN intervals in all 5-min segments, mean of the standard deviations of NN intervals in all 5-min segments, maximum of RR intervals (RR_{MAX}), minimum of RR intervals (RR_{MIN}), median of RR intervals, HRV triangular index, i.e. the proportion of all accepted RR intervals to their modal measurement at a discrete scale of 1/128 s bins, triangular interpolation of RR interval histogram, i.e. the baseline width of the distribution measured as a base of a triangle, approximating the NN interval distribution by using the minimum square difference.

The frequency-domain HRV measures rely on the estimation of power spectral density (PSD) computed, in this work, with three different methods: Welch periodogram, Auto-Regressive (AR) method and Lomb-Scargle periodogram. For the Welch's periodogram, the NN interval was first interpolated with cubic spline interpolation at 4 Hz. The interpolated series was then divided into overlapping segments of length 256 points and each segment was Hamming windowed. The overlap was chosen to be 128 points. AR model order was 16. The generalized frequency bands in case of short-term HRV recordings are the 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). The frequency-domain measures extracted from the PSD estimate for each frequency band include absolute and relative powers of VLF, LF, and HF bands, LF and HF band powers in normalized units, the LF/HF power ratio, and peak frequencies for each band. Further, in the paper, we will refer to Welch-based, AR-based and the Lomb-Scargle-based measures by using the subscript WE, AR and LS, respectively. For instance, TP_{WE} refers to the estimation of the total power computed by using the Welch periodgram, while LF_{LS} refers to LF computed by using the Lomb-Scargle periodogram.

Nonlinear properties of HRV were analysed by the following methods: Poincaré Plot [20], Approximate Entropy [21], Correlation Dimension [22], Detrended Fluctuation Analysis [23], and Recurrence Plot [24]. The **Poincaré Plot** estimates the correlation between successive RR intervals [25]. **Approximate entropy** measures the complexity or irregularity of the RR series [21] as described in [26, 27], with the values of parameters *r* and *m*, respectively 2 and 20% of SDNN, chosen according to [28]. The **correlation dimension** measures the complexity used for the HRV time series [22], with a parameter m=10 as described in [28]. **Detrended Fluctuation Analysis**, which measures the correlation within the signal, by two parameters: short-term fluctuations (α_1) and long-term fluctuations (α_2) [23]. **Recurrence Plot** (RP), which measures the complexity of a time-series, with the values of parameters chosen according to [29, 28]. The following measures of RP were computed: recurrence rate (REC); maximal length of lines (L_{MAX}); mean length of lines (L_{MEAN}); the determinism (DET); the Shannon Entropy (ShanEn).

C. Statistics and data mining methods

Differences between HRV features of fallers and non-fallers subjects were assessed by repeated measure regression model estimated using Generalized Estimating Equaions (GEE) [30].

Two well-known and complementary approaches of data mining were used: divide-and-conquer decision tree algorithms such as C4.5, CART [31], or M5P and covering rule induction algorithms such as RIPPER [32] or PART. Particularly, a combination of Naïve Bayes [33], lift chart and the C4.5 [34], CART [31], or RIPPER [32] methods was employed, using the Weka platform for knowledge discovery [35]. C4.5 is the landmark decision tree algorithm developed by [34]. Binary splits at internal nodes are made for numerical features, while multi-way splits are done for the categorical ones. All of the features are considered for splits in each node of the tree and a gain ratio is used for evaluating the contribution of each feature. CART is a decision tree algorithm developed by [31]. There are some differences between CART and C4.5, when CART is used for classification, including the criterion used for branching in internal nodes and the type of pruning. CART uses the Gini index as impurity measure for determining the split at the internal node. The CART method also uses cost-complexity based sub-tree post-pruning, which is more conservative than empirically based sub-tree replacement and raising of C4.5 and leads to construction of smaller trees. The idea is to first prune those sub-trees that, relative to their size, lead to the smallest increase in error on the training data. The pruning is usually performed by an internal cross-validation procedure. The RIPPER (*Repeated Incremental Pruning to Produce Error Reduction*) algorithm splits the training set into two distinctive sets: a growing set and a pruning set. On the growing set, it greedily constructs rules with perfect coverage. A Naïve Bayes (NB) classifier is used to obtain a probabilistic model of the dataset with respect to the health event investigated in this study. It outputs posterior probabilities as a result of the classification procedure. Excerpts were ranked according to posterior Bayesian probability for falling. Highest 10%, 20% and 30% excerpts were then analyzed by C4.5, CART and RIPPER.

All of the classifiers were evaluated by 10-fold cross-validation on the whole dataset. The minimum number of instances in leaf nodes (or minimum rule coverage) was varied for C4.5, CART, and RIPPER in order to obtain relevant rules. The rules can be read from the C4.5 and CART tree classifier structure once the tree is built.

III. Results

The GEE statistical analysis revealed that the following 10 HRV measures changes significantly ($p<0.01$) in fallers:

1. RR_{MIN}, L_{MAX}, L_{MEAN}, ShanEn increased;
2. VLF_{WE}, TP_{WE}, LF_{LS}, RP_{LS}, DIV, DET decreased.

The data mining methods revealed several risk factors for fallers. The most accurate rule was: $\alpha_2 \leq 0.947$ and $pNN50 \leq 26.7$ and $RR_{MAX} \geq 2265.6$. This pattern was found in 132 excerpts out of which 94 were from fallers, therefore with an accuracy of 71% among excerpts. The odds ratio (OR) of falling being positive to this pattern was significant: OR=5.12 (CI 95% 1.42-18.41; $p<0.01$).

IV. Discussion

The retrospective cross-sectional study presented in this paper investigated the odds ratio of falling among those subjects having an abnormal HRV pattern. The results demonstrated that there is a significant association between a depressed HRV pattern and risk of falling in hypotensive patients. In fact, both the statistical analysis and the data mining methods highlighted that features extracted with linear and nonlinear methods, indicate consistently a recurrent depressed HRV in fallers. In fact, a consistently depressed PSD is recurrent in fallers and nonlinear features (e.g. L_{MAX}) exhibited a significant increase, indicating a less 'chaotic' behavior of HRV in fallers. These results suggest that fallers may present an abnormal HRV compared to non-fallers.To the best of authors' knowledge, only another retrospective study [39] investigated the relation between HRV and falls, finding no significant differences in HRV between the two groups. However, the authors [39] adopted only standard linear methods to investigate HRV, such as SDNN, RMSSD, pNN50, which did not show statistical differences also in the current study. Regarding the frequency domain measures, Isik et al. [39] did not provide information about the PSD methods, while the current study compared three different methods to estimate PSD, reporting that the Welch and Lomb-Scargle periodgram provided more significant information compared to AR methods, which is consistent with previous studies [40]. Moreover, our findings reinforce the importance of nonlinear analysis of HRV, which has been shown to improve the discrimination power between different patho-physiological conditions [27, 41]. Finally, the comparison of our method based on HRV with functional mobility tests [37] and computer-based tests [38] to discriminate between fallers and non-fallers, reported in Table 1, showed higher OR and/or relative risk values of our method.

Table 1 Comparison with other tests for fallers' identification

Test	Odds ratio (95% CI)	Relative risk (95% CI)
Sit-to-stand once [37]	-	1.3 (0.8, 2.1)
Sit-to-stand five times [37]	-	2.0 (1.3, 3.0)
Pick-up-weight test [37]	-	1.5 (0.8, 2.6)
Half-turn test [37]	-	1.3 (0.8, 2.0)
Alternate-step test [37]	-	2.3 (1.4, 3.5)
Six-metre walk [37]	-	1.8 (1.2, 2.6)
Stair ascent [37]	-	1.4 (1.0, 2.1)
Stair descent [37]	-	1.7 (1.2, 2.6)
Stroop Stepping Test [38]	1.7 (1.2, 2.3)	-
Abnormal HRV	5.1 (1.4, 18.4)	2.5 (1.5, 4.2)

Acknowledgment

The current study was partially supported by "the 2007-2013 NOP for Research and Competitiveness for the Convergence Regions (Calabria, Campania, Puglia and Sicilia)" with code PON04a3_00139 - Project Smart Health and Artificial intelligence for Risk Estimation.

Conflict of Interest

The authors declare that they have no conflict of interest.

References

1. Tinetti ME, Kumar C. The patient who falls: "It's always a trade-off". JAMA. 2010;303(3):258-66
2. Siracuse JJ, Odell DD, Gondek SP, Odom SR, Kasper EM, Hauser CJ et al. Health care and socioeconomic impact of falls in the elderly. American journal of surgery. 2012;203(3):335-8; discussion 8
3. Guideline for the prevention of falls in older persons. American Geriatrics Society, British Geriatrics Society, and American Academy of Orthopaedic Surgeons Panel on Falls Prevention. J Am Geriatr Soc. 2001;49(5):664-72
4. Peel NM. Epidemiology of falls in older age. Can J Aging. 2011;30(1):7-19
5. Pecchia L, Bath PA, Pendleton N, Bracale M. Analytic Hierarchy Process (AHP) for examining healthcare professionals' assessments of risk factors. The relative importance of risk factors for falls in community-dwelling older people. Methods Inf Med. 2011;50(5):435-44
6. Perell KL, Nelson A, Goldman RL, Luther SL, Prieto-Lewis N, Rubenstein LZ. Fall risk assessment measures: an analytic review. J Gerontol A Biol Sci Med Sci. 2001;56(12):M761-6

7. Kosse N, Brands K, Bauer J, Hortobagyi T, Lamoth C. Sensor technologies aiming at fall prevention in institutionalized old adults: A synthesis of current knowledge. International journal of medical informatics. 2013;82(9):743-52

8. Hamacher D, Singh NB, Van Dieen JH, Heller MO, Taylor WR. Kinematic measures for assessing gait stability in elderly individuals: a systematic review. Journal of the Royal Society, Interface / the Royal Society. 2011;8(65):1682-98

9. Yang CC, Hsu YL. A Review of Accelerometry-Based Wearable Motion Detectors for Physical Activity Monitoring. Sensors-Basel. 2010;10(8):7772-88

10. Allet L, Knols RH, Shirato K, de Bruin ED. Wearable Systems for Monitoring Mobility-Related Activities in Chronic Disease: A Systematic Review. Sensors-Basel. 2010;10(10):9026-52

11. Gates S, Smith LA, Fisher JD, Lamb SE. Systematic review of accuracy of screening instruments for predicting fall risk among independently living older adults. Journal of rehabilitation research and development. 2008;45(8):1105-16

12. Ganyo M, Dunn M, Hope T. Ethical issues in the use of fall detectors. Ageing Soc. 2011;31:1350-67

13. Melillo P, Izzo R, Luca N, Pecchia L. Heart rate variability and target organ damage in hypertensive patients. BMC Cardiovasc Disord. 2012;12(1):105

14. Zong W, Moody G, Jiang D, editors. A robust open-source algorithm to detect onset and duration of QRS complexes. Computers in Cardiology, 2003; 2003: IEEE.

15. Hautala AJ, Karjalainen J, Kiviniemi AM, Kinnunen H, Mäkikallio TH, Huikuri HV et al. Physical activity and heart rate variability measured simultaneously during waking hours. Am J Physiol-Cell Ph. 2010;298(3):H874

16. Malik M, Bigger JT, Camm AJ, Kleiger RE, Malliani A, Moss AJ et al. Heart rate variability: Standards of measurement, physiological interpretation, and clinical use. Eur Heart J. 1996;17(3):354-81

17. Acharya R, Joseph P, Kannathal N, Lim CM, Suri JS. Heart rate variability: a review. Med Biol Eng Comput. 2006;44(12):1031-51

18. J R. Design, Evaluation and application of Heart rate variability software. 2010.

19. Melillo P, De Luca N, Bracale M, Pecchia L. Classification Tree for Risk Assessment in Patients Suffering From Congestive Heart Failure via Long-Term Heart Rate Variability. IEEE J Biomed Health Inform. 2013;17(3):727-33

20. Melillo P, Fusco R, Sansone M, Bracale M, Pecchia L. Discrimination power of long-term heart rate variability measures for chronic heart failure detection. Medical and Biological Engineering and Computing. 2011;49(1):67-74

21. Richman JS, Moorman JR. Physiological time-series analysis using approximate entropy and sample entropy. American Journal of Physiology-Heart and Circulatory Physiology. 2000;278(6):H2039-H49

22. Carvajal R, Wessel N, Vallverdú M, Caminal P, Voss A. Correlation dimension analysis of heart rate variability in patients with dilated cardiomyopathy. Computer Methods and Programs in Biomedicine. 2005;78(2):133-40

23. Penzel T, Kantelhardt JW, Grote L, Peter JH, Bunde A. Comparison of detrended fluctuation analysis and spectral analysis for heart rate variability in sleep and sleep apnea. IEEE Trans Bio Med Eng. 2003;50(10):1143-51

24. Zbilut JP, Thomasson N, Webber CL. Recurrence quantification analysis as a tool for nonlinear exploration of nonstationary cardiac signals. Medical Engineering & Physics. 2002;24(1):53-60

25. Brennan M, Palaniswami M, Kamen P. Do existing measures of Poincare plot geometry reflect nonlinear features of heart rate variability? IEEE Trans Bio Med Eng. 2001;48(11):1342-7

26. Pincus S. Approximate entropy as a measure of system complexity. Proc Natl Acad Sci USA. 1991;88:2297 - 301

27. Melillo P, Bracale M, Pecchia L. Nonlinear Heart Rate Variability features for real-life stress detection. Case study: students under stress due to university examination. Biomed Eng Online. 2011;10(1):96

28. Niskanen J-P, Tarvainen MP, Ranta-aho PO, Karjalainen PA. Software for advanced HRV analysis. Computer Methods and Programs in Biomedicine. 2004;76(1):73-81

29. Dabire H, Mestivier D, Jarnet J, Safar ME, Chau NP. Quantification of sympathetic and parasympathetic tones by nonlinear indexes in normotensive rats. American Journal of Physiology-Heart and Circulatory Physiology. 1998;275(4):H1290-H7

30. Zeger SL, Liang K-Y, Albert PS. Models for longitudinal data: a generalized estimating equation approach. Biometrics. 1988:1049-60

31. Breiman L, Friedman JH, Olshen RA, Stone CJ. Classification and regression trees. Belmont, Calif.: Wadsworth International Group; 1984.

32. Cohen W, editor. Fast effective rule induction. Twelfth International Conference on Machine Learning; 1995.

33. Friedman N, Geiger D, Goldszmidt M. Bayesian Network Classifiers. Machine Learning. 1997;29(2):131-63

34. Quinlan JR. C4.5: programs for machine learning. Morgan kaufmann; 1993.

35. Hall M, Frank E, Holmes G, Pfahringer B, Reutemann P, Witten IH. The WEKA data mining software: an update. ACM SIGKDD explorations newsletter. 2009;11(1):10-8

36. Breiman L. Random forests. Machine Learning. 2001;45:5-32

37. Tiedemann A, Shimada H, Sherrington C, Murray S, Lord S. The comparative ability of eight functional mobility tests for predicting falls in community-dwelling older people. Age Ageing. 2008;37(4):430-5

38. Schoene D, Smith ST, Davies TA, Delbaere K, Lord SR. A Stroop Stepping Test (SST) using low-cost computer game technology discriminates between older fallers and non-fallers. Age Ageing. 2014;43(2):285-9

39. Isik M, Cankurtaran M, Yavuz BB, Deniz A, Yavuz B, Halil M et al. Blunted baroreflex sensitivity: An underestimated cause of falls in the elderly? European Geriatric Medicine. 2012;3(1):9-13

40. Clifford GD, Tarassenko L. Quantifying errors in spectral estimates of HRV due to beat, replacement and resampling. IEEE Trans Bio Med Eng. 2005;52(4):630-8

41. Khandoker AH, Jelinek HF, Palaniswami M. Identifying diabetic patients with cardiac autonomic neuropathy by heart rate complexity analysis. Biomed Eng Online. 2009;8

Author:	Paolo Melillo
Institute:	SHARE Project, Italian Ministry of Education. Scientific Research and University (PI)
Street:	via Pansini, 5
City:	Naples
Country:	Italy
Email:	paolomelillo85@gmail.com, paolo.melillo@unina2.it

Author:	Alan Jovic
Institute:	Faculty of Electrical Engineering and Computing, University of Zagreb
Street:	Unska 3
City:	Zagreb
Country:	Croatia
Email:	alan.jovic@fer.hr

Author:	Leandro Pecchia
Institute:	University of Warwick
Street:	Library Road
City:	Coventry
Country:	UK
Email:	l.pecchia@warwick.ac.uk