Fall prediction in hypertensive patients via short-term HRV Analysis

R. Castaldo, *Student Member, IEEE,* P. Melillo, *Member, IEEE,* R. Izzo, N. De Luca, L. Pecchia, *Member, IEEE*

***Abstract*— Falls are a major problem of later life having severe consequences on quality of life and a significant burden in occidental countries. Many technological solutions have been proposed to assess the risk or to predict falls and the majority is based on accelerometers and gyroscopes. However, very little was done for identifying first time fallers, which are very difficult to recognise. This paper presents a meta-model predicting falls using short term Heart Rate Variability (HRV) analysis acquired at the baseline. 170 hypertensive patients (age: 72 ± 8 years, 56 female) were investigated, of which 34 fell once in the 3 months after the baseline assessment. This study is focused on hypertensive patients, which were considered as convenient pragmatic sample, as they undergo regular outpatient visits, during which short term ECG can be easily recorded without significant increase of healthcare costs. For each subject, 11 consecutive excerpts of 5 minutes each (55 min) were extracted from ECGs recorded between 10:30 and 12:30 and analysed. Linear and nonlinear HRV features were extracted and averaged among the 11 excerpts, which were, then, considered for the statistical and data mining analysis. The best predictive meta-model was based on Multinomial Naïve Bayes, which enabled to predict first-time fallers with sensitivity, specificity and accuracy rates of 72%, 61%, 68% respectively.**

***Index Terms*—Fall Prediction, Heart Rate Variability analysis, Multinomial Bayesian model, Accidental Falls**

I. INTRODUCTION

FALLS are a serious health problem among older citizens. In community dwelling old adults, the fall rate per year is around 30%; people aged 65 and older have higher risk of falling, and 50% of people older than 80 years fall at least once a year [1, 2]. In the UK, falls cost to the NHS more than -£2.3 billion per year [1]. Predicting falls is challenging, but could help designing more targeted and therefore sustainable fall prevention programs. Even defining a fall is a challenge itself. For example, the National Database of Nursing Quality Indicators defines a fall as “an unplanned descent to the floor with or without injuries”, whereas the World Health Organization defines a fall as “an event which results in a person coming to rest inadvertently on the ground or floor or some lower level”[3]. In this study we considered both the definitions, instructing patients and operators accordingly.

Regardless of the definition, a fall is often the result of complex and dynamic interactions between intrinsic (subject-specific) risk factors and extrinsic (environmental) risk factors. The former include, among others, age, history of recent fall, mobility impairments, urinary incontinence or frequency, certain medications and their combinations, postural hypertension, frailty, and other cardiovascular, neurological and visual concomitances [3]. Extrinsic include, among others, footwear, transient exposure to risky environments (i.e. unsupervised toileting) and so on [4].

Since falls in older citizens increase morbidity and mortality, and because of continuous ageing of occidental population, fall prevention has become an important priority in Europe and USA and the efforts about research and development of technologies aiming to screen the risk of falling and/or to detect and/or to predict falls are constantly increasing.

However, recent systematic reviews highlighted that many of the proposed technologies presented several limits including the elevated occurrence of false alarms, the obtrusiveness of those technologies and their costs-effectiveness [5]. Regarding costs-effectiveness, the majority of the proposed approaches require the use of additive sensors (mainly accelerometers, gyroscopes or ambient sensors) having no other direct utility for the older citizens’ health and therefore determining unsustainable additional costs [5]. Also, the mechanism that accelerometers, gyroscopes or ambient sensors uses, cannot detect all the risk factors for falls.

This paper presents the results of a study aiming to develop a method to assess the risk of falling using short-term HRV analysis in hypertensive patients. This is a particular subgroup of older citizens, because of drug prescription and prevalence

This paper was submitted for review on the 3rd of December 2015. P. Melillo was supported by the Project Smart Health and Artificial intelligence for Risk Estimation under Grant PON04a3_00139, funded in part by European Union, in the framework of 2007–2013 National Operational Programme for Research and Competitiveness.

P. Melillo is with the Multidisciplinary Department of Medical, Surgical and Dental Sciences of the Second University of Naples, Via Pansini, 5, Naples, Italy (e-mail: paolo.melillo@unina2.it).

R.Izzo and N. De Luca are with the Departments of Clinical Medicine, Cardiovascular and Immunological Sciences, Federico II University Hospital, Via Pansini, 5, Naples, Italy (email : raffaele.izzo@unina.it, nicola.deluca@unina.it)

R. Castaldo and L. Pecchia are with the School of Engineering, University of Warwick, Coventry CV4 7AL UK (email: r.castaldo@warwick.ac.uk, l.pecchia@warwick.ac.uk).

of cardiovascular risk factors for falls. However, this is a significant subgroup, given the hypertension incidence, which rise from the 60% in the 6th decade to the 70% in the 7th with a steep increase in the following decades of life[6]. Differently from the few previous studies investigating HRV in fallers [5, 7], which were focused on 24-hour HRV, this is the first paper describing the results obtained with short-term HRV analysis, which is much easier and cheaper to be translated in everyday outpatient clinical practice. This approach is based on the idea that it is possible to early detect constantly depressed autonomous nervous system states, which increase significantly the risk of falling. In fact, according to existing literature, 42% of falls among the community-based older population are due to transient problems, which are significantly related to cardiovascular system and autonomous nervous system conditions [8], including: gait/balance disorders, syncope, weakness, dizziness/vertigo, drop attacks and postural hypotension [2, 9, 10].

Differently from other technological approaches used in previous studies, HRV can be extracted from Electrocardiogram (ECG), largely used to monitor/screen patients over 60 years old. In fact, ECG monitoring is beneficial for several cardiovascular diseases, and the application of ECG monitoring during real-life activities are under investigation for several purposes and particularly because of its effectiveness as early detector of cardiovascular diseases worsening [5, 11, 12]. Accordingly, most of the wearable and ambient sensing technologies aiming to monitor older subjects in real life include ECG or HRV monitoring.

Therefore, while older citizens could be sceptical of wearing technologies embedding accelerometers and gyroscopes “only” for falls prevention, it is expected that the same users would be less sceptical of adopting technologies that have been already proven effective for other cardiovascular diseases. In other words, enriching those technologies today under exploration with an ECG sensor could be convenient combination in order to predict/detect a fall, while being used to monitor cardiovascular problems. For these reasons, in this study, we focused on hypertensive patients undergoing regular outpatient visits, for which ECG recordings were already going to be prescribed in order to monitor the risk of other cardiovascular events [13]. Moreover, other well known risk factors for falls (e.g. multiple-prescriptions) are also systematically monitored and recorded in hypertensive patients undergoing regular outpatient visits, facilitating this study.

Differently from other methodologies used in previous studies, this paper presents a meta-model to automatically identify subjects at higher risk of falling via HRV features using advanced data mining methods.

II. Methods

A. Dataset

This study was carried in the outpatient clinic for hypertension at the University Hospital of Naples “Federico II”, and therefore it was approved by the Local Ethic Committee and all the participants signed specific informed consent to allow the use of their data for this study. Hypertensive patients were enrolled in this study if they met the following inclusion criteria: home dwelling autonomous above 55 years old, without cognitive impairments and without history of falls in the previous years. At the baseline, a nominal 24h ECG Holter registration was performed, together with the other periodic controls for hypertension management. ECGs were recorded using Holter ECG Cardioscan DMS 300-3A and downloaded for analysis using Cardioscan software (V12.0; DMS Holter, Stateside, NV, USA). Further details on the clinical protocol for hypertension management, other clinical outcomes (non-falls) and the ECG recording specifications could be found in [13]. Falls were self-reported by patients. The following definitions for accidental falls were used in order to instruct patients and operators: “an unplanned descent to the floor with or without injuries” and/or “an event which results in a person coming to rest inadvertently on the ground or floor or some lower level”.

B. HRV Processing

The series RR beat intervals were obtained from ECG recordings using an automatic QRS detector based on nonlinearly scaled ECG curve length feature[14]. The QRS detection was performed through the WQRS implementation[14] , freely available from PhysioNet.

All the Holter recordings started in early morning (i.e. from 8:30am to 9:30am). In order to avoid the white coat effect, and to maximally standardize the protocol (i.e., minimize heterogeneity due to circadian cycle), the second and third hours of each recording were considered (approximately between 10:30 and 12:30). From these two hours the first 11 consecutive 5-minutes excerpts were used for the analysis. The two hours were initially selected as a quality check was performed using the NN/RR ratio, and each excerpt was included among the consecutive 11 only if the NN/RR ratio resulted more than 90%. According to the protocol, a subject would have been excluded if 11 consecutive excerpts would have been not identifiable in those two hours. This did not happen in the current study.

Standard linear HRV analysis according to International Guidelines was performed [15]. Moreover, nonlinear features were computed according to recent literature [16]. The HRV analysis was performed using an ad hoc developed HRV software based on MATLAB version R2013a (The MathWorks Inc., Natick, MA) implementation [17].

As shown in Table I, time-domain HRV features, reliable in 5-min HRV analysis, were calculated: Average of all RR intervals (AVNN), standard deviation of all NN intervals (SDNN), square root of the mean of the sum of the squares of differences between adjacent NN intervals (RMSSD), number and percentage of differences between adjacent NN intervals that are longer than 50 ms (NN50 and pNN50).

The frequency-domain HRV features rely on the estimation of power spectral density, computed with Lomb-Scargle periodogram. The generalized frequency bands in case of short-term HRV recordings were low frequency (LF, 0.04-0.15 Hz), and high frequency (HF, 0.15-0.4 Hz). The included frequency-domain features were absolute for each band, LF, HF, and the LF/HF power ratio (see Table I).

Nonlinear HRV was analyzed with the following methods: Poincaré plot, Approximate entropy, Correlation dimension, Detrended fluctuation analysis, and Recurrence Plot (RP) [13, 18](see Table I).

TABLE I
HRV FEATURES

Features	Units	Description
Time Domain		
AVNN	ms	The mean of RR intervals
SDNN	ms	Standard deviation of RR intervals
RMSSD	ms	Square root of the mean squared differences between successive RR intervals
NN50	/	Number of successive RR interval pairs that differ more than 50 ms
pNN50	ms	NN50 divided by the total number of RR intervals
Frequency Domain		
Absolute power	ms^2	Absolute power of LF and HF bands
LF/HF	/	Ratio between LF and HF band powers
Non-linear		
SD1,SD2	ms	The standard deviation of the Poincare plot perpendicular to the line-of-identity (SD1) and along the line-of-identity (SD2)
ApEn		Approximate Entropy
SampEn		Sample Entropy
D2		Correlation Dimension
DFA:		Detrended fluctuation analysis
α_1	/	Short-term fluctuation slope
α_2	/	Long-term fluctuation slope
RPA:		Recurrence plot analysis:
Lmean	Beats	Mean line length
Lmax	Beats	Maximum line length
REC	%	Recurrence rate
DET	%	Determinism
ShanEn	/	Shannon entropy
DIV		Divergence

C. *Statistical Analysis*

Median, standard deviation, 25^{th} and 75^{th} percentiles were calculated to describe distribution of HRV features for fallers and no-fallers. The non-parametric Wilcoxon Signed-Rank Test was used to investigate the statistical significances of feature variation between fallers and no-fallers. The Wilcoxon test was chosen as several HRV features, as expected, were not normally distributed. Baseline continuous and categorical variables were presented as median (± standard deviation) or as count (percentage), respectively. Wilcoxon test and Chi-square test were adopted to compare continuous and categorical variables, respectively, between those who experienced a fall and those who did not. The statistical analysis was performed using IBM SPSS statistics 22.

D. *Model training, validation and testing procedure*

According to [19], the whole dataset was split per subject in three folders (Fig. 1): folder 1 (34%) was used for feature selection; folder 2 (39%) was used for training the classification models; finally folder 3 (27%) was adopted to evaluate the performance of the developed classification models.

The subjects not included in folder 1, were randomly assigned to folder 2 or folder 3 according to a 2:3 ratio. The reason of this asymmetric splitting was that the folder 2 was further split in 3 subsamples because of the 3-fold cross-validation technique (as detailed in subsection II.D.3).

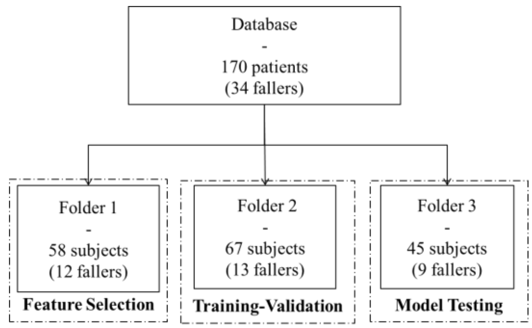

Fig. 1. Splitting of the dataset in three folders

1) *HRV Feature Selection*

As recalled also in [19], the number of features used in a machine learning model should be strongly limited by the number of subjects presenting the event to detect (falls) in each folder, in order to minimize the risk of over-fitting. Moreover, a smaller set of significant features strongly simplifies the medical interpretation of the achieved results, by directing attention only on the most important informative part of the utilized signal [19]. However, selecting the minimum set of features using the same folder utilized to train the machine-learning model can reduce the generalizability of the final decisional model. Therefore, the HRV features were minimized using only the folder 1 (58 patients, of which 12 fallers). The feature selection was based on two main stages: the relevance analysis performed by Wilcoxon Signed-Rank Test and redundancy analysis in term of feature correlation (Fig. 2) [20].

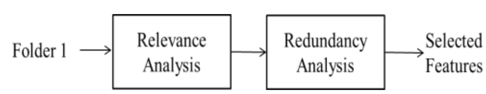

Fig. 2. Framework of Feature Selection

The relevance analysis aimed to identify the HRV features changing more significantly among fallers and non-fallers, according to the Wilcoxon Signed-Rank Test. Since not all the HRV features were normally distributed, (i.e., frequency features have non-symmetric distributions) a non-parametric test was adopted. All the HRV features changing significantly between fallers and non-fallers (p-value less than 0.05) were selected at this stage.

All the relevant HRV features ($p<0.05$) were then further minimized with the redundancy analysis aiming to exclude highly correlated features. Notions of measure redundancy are normally explored in terms of feature correlation. It is widely accepted that two features are redundant to each other if their values are strongly correlated. The features with a Pearson's coefficient above 0.7 in absolute magnitude and with a significant p-value (less than 0.05) were excluded. In this final stage, only the HRV features relevant and not redundant were considered for the next steps.

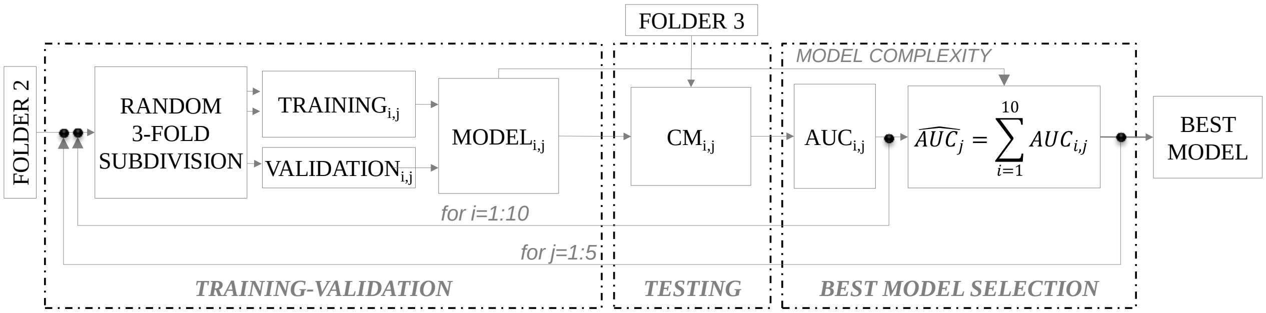

Fig. 3. Model training validation and testing. For each of the 5 learning-machine methods used (*j=1…5*), the training-validation procedure was repeated 10 times (*i=1…10*). For each iteration, the Confusion Matrix ($CM_{i,j}$) and the $AUC_{i,j}$ were calculated. The best method was the one with $\max_j\{\widehat{AUC_j}\}$.

2) Machine learning methods

Five different machine-learning methods were used to develop models aiming to automatically detect future fallers based on HRV features: *Naïve Bayes (NB)*, which uses the naive Bayes formula to calculate the probability of each class given the values of all the attributes and assuming the conditional independence and the Gaussian distribution of the attributes [21]; *Multinomial Naïve Bayes(MNB)*, which is based on Bayes' theorem (Bayes rule), with the additional incorporation of frequency information and multinomial distribution for each of the features [22]; *Support Vector Machine(SVM)*, which belongs to a general field of kernel-based machine learning methods and are used to efficiently classify both linearly and non-linearly separable data[23]; *Multilayer Perceptron (MLP)*consisting of an artificial neural network of nodes (processing elements) arranged in layers [24]; *Neighbor Search(IBK)*, which finds a group of k objects in the training set that are closest to the test object, and bases the assignment of a label on the predominance of a particular class in this neighborhood [25]. Regarding model parameters, for multilayer perceptron classifier, the learning rate was varied from 0.3 to 0. 9, the momentum from 0.2 to 1 and the number of epoch from 100 to 2000; for support vector machine, basis function kernel was used, varying gamma from 10^{-5} to 10. As regards neighbor search, it was trained by varying K from 1 to 5. Each of those methods was used with all the possible combinations of N out the D selected features (with *D* equal to the number of features selected and N spanning from 3 to *D*). The Weka platform for knowledge discovery (version 3.6.10), issued by the University of Waikato as an open source software under the GNU General Public License[26], was used to train, validate and test the classification models.

3) Training and validation

The training of the machine-learning models was performed on the folder 2 (67 patients, of which 13 patients experienced a fall). Folder 2 was further divided in 3 equal sized subsamples, according to the 3-fold person-independent cross-validation approach. Of these 3 subsamples, 2 subsamples are used as training data and the remaining one is retained for validating the model. The process is then repeated 3 times, with each of the 3 subsamples used exactly once as the validation data. Finally, the cross-validated estimations are computed by averaging the performances over the 3 validation subsamples. Binary classification measures were adopted according to the standard formulae reported in Table II [27, 28].

TABLE II
BINARY CLASSIFICATION PERFORMANCE FEATURES

MEASURE	FORMULA
SENSITIVITY (TPR)	SEN = TP / (TP + FN)
SPECIFICITY (1-FPR)	SPE = TN / (FP + TN)
ACCURACY	ACC = (TP+TN) / (TP+TN+FP+FN)
AREA UNDER CURVE	AUC = AREA UNDER ROC CURVE

TP: the number of fallers correctly classified; TPR: TP rate; TN: the number of non-fallers correctly classified; FP: the number of non-fallers incorrectly classified as fallers; FPR: FP rate; FN: the number of fallers incorrectly classified as non-fallers

Given the relatively small and unbalanced number of events (falls) in each subsample; the random allocation of one subject to one of the three subsamples can significantly alter the cross-validation estimates. Therefore, we repeated 10 times the cross-validation procedure and averaged over those 10 iterations the cross-validation estimates. This procedure was performed 5 times: one for each machine-learning method used to develop predictive models (see Fig. 3).

4) Testing of predictive model and best model selection

Testing a classifier involves analyzing its performances on a set of subjects that is independent from the training and validation set [19]. Accordingly, folder 3 (45 patients) was used to test the trained models. Finally, the best performing model was selected as the one achieving the highest averaged AUC, which is a reliable estimator of both sensitivity and specificity rates and, in case of equal AUC average, the model with minimal complexity (i.e. minor mean number of features employed).

E. Final model generation

For the best performing method, a meta-model was produced by averaging the coefficients of the hyperplanes separating fallers by non-fallers for each of the 10 models generated during the validation. Coefficients were assumed to be zero for features that were not selected. The performances of this final model were computed, according to the formulae introduced in Table II, using folder 3. In addition, the Diagnostic Odds Ratio (DOR) was computed and ROC curve of the best model was constructed.

III. RESULTS

The current study was performed enrolling 170 hypertensive patients (including 56 female and 114 male), age 72 ±8 years, of which 34 subjects experienced a fall within 3 months from the baseline assessment. Other clinical and metabolic information are reported in Table III. According to the baseline data, no statistically significant differences were observed between fallers and non-fallers.

TABLE III
PATIENT BASELINE CHARACTERISTICS

Clinical features	Non-Fallers MD±SD	Fallers MD±SD	p-val
Age (Years)	71.85(±7.046)	70.33(±9.6)	0.22
Gender (Female)	45(26.7%)	12(7.14%)	0.93
History of Hypertension	46(27.8%)	12(7.2%)	0.90
History of stroke	13(7.8%)	2(1.2%)	0.43
Diabetes	22(13.1%)	5(3%)	0.68
Diastolic BP(mmHg)	76.00(±8.97)	75.92(±11.75)	0.69
Systolic BP (mmHg)	136.76(±20.15)	144.44(±21.31)	0.06
Total cholesterol	176.96(±36.34)	188.86(±40.99)	0.13
LDL(mg/dl)	101.56(±30.012)	113.67(±35.16)	0.11
HDL(mg/dl)	52.25(±13.33)	51.33(±13.61)	1.00
BMI(kg/m2)	27.76(±4.06)	27.27(±4.13)	0.43
BSA(m2)	1.89(±0.16)	1.9(±0.22)	0.84
Alpha-blockers	21(12.6%)	7(4.2%)	0.64
Beta-blockers	56(33.7%)	13(7.8%)	0.45
ACE inhibitor	45(27.1%)	14(8.43%)	0.64
Dihydropyridine	35(21.08%)	9(5.4%)	0.82
IMT Mean(mm)	1.57(±0.45)	1.41(±0.36)	0.07
IMT Max(mm)	2.35(±0.75)	2.23(±0.89)	0.19
LVMi(g/m2)	131.84(±26.32)	133.62(±22.99)	0.68
EF(%)	58.90(±11.24)	63.47(±6.51)	0.05

BP: blood pressure; IMT: intima media thickness; LVMi: left ventricular mass index; EF: ejection fraction

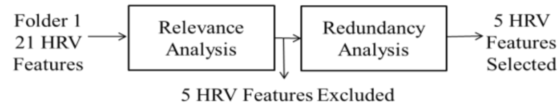

Fig. 4. HRV Feature Selection

Table IV reports the median (MD), the standard deviation (SD), the 25th and the 75th percentiles for the 21 HRV features extracted from faller and no-faller subjects for all the dataset. The last column of Table IV shows the Wilcoxon Signed-Rank Test p-values of feature variations between fallers and no-fallers. As shown in Table IV, 20 out of 21 HRV features changed significantly between fallers and no-fallers. Particularly, in fallers lower value of all time-domain features except AVNN was observed. Moreover, in fallers both LF and HF power were lower, while LF/HF increased. Furthermore, the statistical analysis showed that fallers had significantly lower SD1, SD2, D2 and DIV, and significantly higher ApEn, SampEn, α1, α2, D2, REC, Lmax, Lmean, DET, and ShanEn.

HRV feature selection, performed on the first folder, showed that 16 HRV features were relevant (changed significantly also in this folder), of which 11 were excluded with the redundancy analysis as strongly and significantly correlated (Fig. 4). Finally, 5 HRV features (pNN50, HF, SD1, Lmax, ShanEn) resulted not statistically correlated ($|\rho|< 0.7$ and p-value>0.05).

Table V reports the performance measurements (mean and standard deviation) estimated on the independent test set of the 5 models, averaged over the 10 iterations. According to the criteria defined in sub-section II.D.4, the Multinomial Naïve Bayes model outperformed the other data-mining methods achieving the best mean value of performance measures over 10 iterations: 72% sensitivity, 61% specificity and 68% accuracy. This method achieved the best average AUC and it

TABLE IV
HRV FEATURES IN NO-FALLERS AND FALLERS

HRV Features	Non-Fallers MD±SD	25th	75th	Fallers MD±SD	25th	75th	p-val	Trend
AVNN	773.75±244.9	640.6	899.3	782.9±185.5	676.5	901.7	0.162	↑
SDNN	**57.35±64.9**	**35.60**	**91.9**	**46.2±70.11**	**30.8**	**73.8**	**<0.01**	↓↓
RMSSD	**48.65±64.7**	**26.25**	**86.4**	**29.25±83.4**	**19.9**	**50.4**	**<0.01**	↓↓
NN50	**30±39.9**	**11.00**	**62.0**	**16±35.5**	**6.00**	**28.0**	**<0.01**	↓↓
pNN50	**9.4±16.5**	**3.50**	**22.6**	**4.85±12.78**	**1.60**	**9.10**	**<0.01**	↓↓
LF	**0.010±0.009**	**0.01**	**0.02**	**0.007±0.009**	**0.00**	**0.02**	**<0.01**	↓↓
HF	**0.016±0.09**	**0.00**	**0.03**	**0.006±0.02**	**0.00**	**0.02**	**<0.01**	↓↓
LF/HF	**0.666±1.53**	**0.46**	**1.10**	**0.97±2.01**	**0.57**	**2.11**	**<0.01**	↑↑
SD1	**34.44±45.9**	**18.58**	**61.2**	**20.71±59.15**	**14.1**	**35.61**	**<0.01**	↓↓
SD2	**71.72±79.7**	**43.26**	**115.5**	**60.59±72.6**	**40.5**	**91.43**	**<0.01**	↓↓
ApEn	**0.94±0.21**	**0.76**	**1.05**	**0.96±0.23**	**0.77**	**1.07**	**<0.01**	↑↑
SampEn	**1.06±0.51**	**0.70**	**1.45**	**1.23±0.57**	**0.75**	**1.61**	**<0.01**	↑↑
α1	**0.9±0.28**	**0.71**	**1.10**	**1.03±0.3**	**0.78**	**1.26**	**<0.01**	↑↑
α2	**0.87±0.29**	**0.66**	**1.07**	**0.97±0.32**	**0.75**	**1.15**	**<0.01**	↑↑
D2	**0.80±1.39**	**0.06**	**2.37**	**0.46±1.34**	**0.04**	**1.90**	**<0.05**	↓↓
REC	**0.44±0.16**	**0.32**	**0.52**	**0.45±0.15**	**0.36**	**0.53**	**<0.05**	↑↑
Lmax	**126±106.4**	**67.00**	**212**	**184±109.9**	**111**	**289**	**<0.01**	↑↑
Lmean	**15.2±14.7**	**9.73**	**23.04**	**16.86±14.9**	**11.8**	**24.91**	**<0.01**	↑↑
DIV	**0.008±0.01**	**0.00**	**0.01**	**0.006±0.008**	**0.00**	**0.01**	**<0.01**	↓↓
DET	**0.99±0.02**	**0.98**	**1.00**	**0.99±0.01**	**0.99**	**1.00**	**<0.01**	↑↑
ShanEn	**3.34±0.58**	**2.97**	**3.78**	**3.42±0.58**	**3.21**	**3.88**	**<0.01**	↑↑

was also the less complex employing a less mean number of features (Table V).

TABLE V
PERFORMANCE MEASUREMENT (MEAN±SD) AND MODEL COMPLEXITIES (COMP) ESTIMATED ON THE TEST SET (FOLDER 3)

Method	AUC	SEN	SPE	ACC	COMP
NB	68±7.5	55.6±24.7	75.0±13.2	72.2±8.7	4.2
MNB	70.0±7.8	72.2±10.9	61.1±7.4	67.8±5.9	3.1
SVM	58.0±10.2	22.2±11.6	81.9±9.9	68.9±7.8	3.9
MLP	60.0±6.2	5.6±16.5	84.7±9.3	71.1±6.0	4.3
IBK	54.0±10.3	22.2±12.2	79.2±6.4	67.8±6.1	4.3

As described in section II.E a meta-model was generated by averaging the coefficients of the hyperplanes separating fallers and non-fallers obtained at each iteration of the validation process. In log-space, the *Multinomial Naïve Bayes* meta-model equation was:

$$.09\ HF - .02\ pNN50 - .20\ SD1 + .05\ L_{max} - .05\ ShanEn - .59 \cong 0 \quad (1)$$

Because 3 features (SD1, Lmax, ShanEn) out of the 5 selected were always employed in all the 10 models, they significantly dominate this meta-model. In fact, the remaining HF and pNN50 were employed respectively only 2 times and 1 time out of 10 iterations. Therefore, the reduced meta-model employing only the 3 non-linear features, achieved the same performance:

$$-.20\ SD1 + .05\ L_{max} - .05\ ShanEn - .59 \cong 0 \quad (2)$$

The interpretation of equation (2) could be the following: “a subject is classified as faller if it lies above the hyperplane”. In other words, a subject is identified at high risk of falling if:

$$-.20\ SD1 + .05\ L_{max} - .05\ ShanEn - .59 > 0 \quad (3)$$

Using this interpretation, the Diagnostic Odds Ratio (DOR) for this meta-model was 4.9 (CI 95%: 1.49 - 11.7). The ROC curve for the meta-model estimated on the independent test set is shown in Fig. 5.

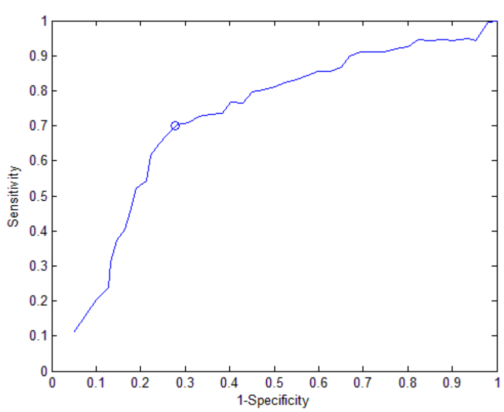

Fig. 5. ROC curve for the multinomial naïve Bayes final meta-model

IV. DISCUSSION

The current study proposed a mathematical model to automatic assess the risk of first-time falls in hypertensive patients, based on few HRV features. These features were extracted from 11 consecutive 5 minutes HRV excerpts extrapolated from the second and third hour of Holter registrations (approximately between 10:30 and 12.30).

The statistical analysis showed that fallers presented a generally depressed HRV and a decline of non-linear heartbeat dynamics. It is known that HRV depression can be due to drug therapy or ageing. However, our results suggested that this difference is not due to those factors, because, as reported in Table III, no significant statistical differences were observed in drug therapy or age between fallers and non-fallers.

These results confirmed our previous finding on long-term HRV analysis [5]. Differently, a precedent study investigating HRV feature changes between fallers and non-fallers [7] did not find significant differences, although it demonstrated the same features trends for AVNN, SDNN, pNN50, LF. This difference could be due to several reasons, including the following: in [7] only linear long-term HRV analysis was performed; it enrolled a smaller sample size (about 60 patients); they used history of falls and not future falls to classify the subjects.

In our study, significant differences in both linear and nonlinear HRV features between the two groups emerged. However, nonlinear ones appeared to have better discrimination ability: the 3 nonlinear features selected during the feature selection phase (SD1, Lmax, ShanEN) were then utilized independently by each method.

During the testing, the best performances were achieved by Multinomial Naïve Bayes, with relative high sensitivity (72%), specificity (61%) and accuracy (68%). The other models achieved high specificity and accuracy but quite low sensitivity during the testing. Therefore, Multinomial Naïve Bayes model was the best model according to our selection criteria, based on AUC and model complexity. For other applications (e.g., screening), other criteria (e.g., highest sensitivity rate) could represent a better choice.

The results presented in this paper reinforce the idea that dysfunctions between cardiovascular system and autonomous nervous system are associated with higher risk of falling in the next few weeks and can be used to predict the risk of falling. According to previous findings [5], the reason could be that a depressed HRV reflects a reduced capability to react to extrinsic risk factors avoiding falls. In our previous study [5], an automatic classifier based on 24 hours HRV features was proposed achieving a DOR of 4.2 (CI 95% 2.0-8.7). Differently from this previous study, the current paper achieved better results by using 1-hour recording and analyzing the HRV on 5-min excerpts (short-term analysis).

The models presented in the current paper were developed through a rigorous training, validation and testing procedure following recent recommendations for machine-learning in biomedical engineering research [19], using three independent subsets of data for feature selection, model training and testing and averaging the performances by repeating the procedure ten times. Differently from machine learning algorithms, the statistical tests were applied on the whole dataset (n=170), not requiring further evaluation or testing, in order to achieve the highest statistical power.

The method proposed in the present paper is clinically feasible, since it only requires a 1-hour ECG recording, which is often performed in cardiovascular patients also through wearable devices[29]. For instance, the method proposed does not require the use of other technologies as wearable accelerometers or pressure matrices, which are not used in everyday clinical practices. For that reason, the method proposed could be easily integrated with other clinical tools for estimating the risk of falling and used widely in outpatient settings to identify high-risk patients who need further assessment and could benefit from fall prevention programs or fall detection systems [30-34]. This is important as falls depend from hundreds of risk factors and integration of complimentary approaches could be more effective in predicting falls. For instance, the mechanisms that accelerometers or gyroscopes use could perform better in a population in which other intrinsic risk factors are prevalent. In fact, focusing on hypertensive patients may have narrowed our study to a population where risk factors for falls due to cardiovascular problems were prevalent. Therefore, although hypertension affects the 60% of people in the 6th decade of life, 70% in the 7th and so far, future studies on a different population and combining different approaches seems to be needed.

However, this study has some limitations that should be considered before adopting these methods in other contexts. This study focused on hypertensive patients, which represent special population with distinguished characteristics, different from the population of community-dwelling older citizens. Future study should explore how this method can be adapted to a more general population recruiting participants from a geriatric outpatient clinic. The patients were enrolled in an outpatient clinic for hypertension and not in a fall clinic. Therefore, important information, such as the exposure to other independent intrinsic/extrinsic risk factors for falls could not be accessed or used to verify independently the results. Moreover, the fall recordings were based on patient self-reports and potentially relevant characteristics of the recorded falls were not systematically recorded.

V. Conclusion

The current study proposed a method based on short-term HRV analysis to identify automatically future fallers among hypertensive patients aged 55 or over. The presented classifier achieved satisfactory results through a rigorous validation procedure, enabling to predict fallers with a sensitivity rate of 72% and a specificity rate of 61%. The proposed method is based on analysis that is already in use in many outpatient clinics and could be used widely in outpatient settings to identify high-risk patients who need further assessment and could benefit from fall prevention programs.

Acknowledgment

PM thanks all the participants to the SHARE project. We thank our colleague Luis Montesinos Silva who provided insight and expertise that greatly assisted the research.

References

[1] N. Kosse, K. Brands, J. Bauer, T. Hortobagyi, and C. Lamoth, "Sensor technologies aiming at fall prevention in institutionalized old adults: A synthesis of current knowledge," *International journal of medical informatics,* vol. 82, pp. 743-752, 2013.

[2] N. I. f. C. Excellence, *Clinical Practice Guideline for the Assessment and Prevention of Falls in Older People: Guidelines Commissioned by the National Institute for Clinical Excellence (NICE)*: Royal College of Nursing, 2013.

[3] I. M. Miake-Lye, S. Hempel, D. A. Ganz, and P. G. Shekelle, "Inpatient Fall Prevention Programs as a Patient Safety StrategyA Systematic Review," *Annals of internal medicine,* vol. 158, pp. 390-396, 2013.

[4] L. Pecchia, P. A. Bath, N. Pendleton, and M. Bracale, "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,* vol. 50, pp. 435-444, 2011.

[5] P. Melillo, A. Jovic, N. De Luca, and L. Pecchia, "Automatic classifier based on heart rate variability to identify fallers among hypertensive subjects," *Healthcare Technology Letters,* vol. 2, pp. 89-94, 2015.

[6] G. Mancia, et al., "2013 ESH/ESC guidelines for the management of arterial hypertension: the Task Force for the Management of Arterial Hypertension of the European Society of Hypertension (ESH) and of the European Society of Cardiology (ESC)." *Blood pressure,* vol. 22, p. 193, 2013.

[7] M. Isik, M. Cankurtaran, B. Yavuz, A. Deniz, B. Yavuz, M. Halil, *et al.*, "Blunted baroreflex sensitivity: An underestimated cause of falls in the elderly?," *European Geriatric Medicine,* vol. 3, pp. 9-13, 2012.

[8] L. Z. Rubenstein, "Falls in older people: epidemiology, risk factors and strategies for prevention," *Age Ageing,* vol. 35 Suppl 2, pp. ii37-ii41, Sep 2006.

[9] P. C. Fletcher and J. P. Hirdes, "Risk factors for falling among community-based seniors using home care services," *The Journals of Gerontology Series A: Biological Sciences and Medical Sciences,* vol. 57, pp. M504-M510, 2002.

[10] P. Kannus, H. Sievänen, M. Palvanen, T. Järvinen, and J. Parkkari, "Prevention of falls and consequent injuries in elderly people," *The Lancet,* vol. 366, pp. 1885-1893, 2005.

[11] L. Pecchia, P. Melillo, M. Sansone, and M. Bracale, "Discrimination Power of Short-Term Heart Rate Variability Measures for CHF Assessment," *IEEE Transactions on Information Technology in Biomedicine,* vol. 15, pp. 40-46, Jan 2011.

[12] L. Pecchia, P. Melillo, and M. Bracale, "Remote Health Monitoring of Heart Failure With Data Mining via CART Method on HRV Features," *IEEE Transactions on Biomedical Engineering,* vol. 58, pp. 800-804, Mar 2011.

[13] P. Melillo, R. Izzo, A. Orrico, P. Scala, M. Attanasio, M. Mirra, *et al.*, "Automatic Prediction of Cardiovascular and Cerebrovascular Events Using Heart Rate Variability Analysis," *PLoS ONE,* vol. 10, p. e0118504, 2015.

[14] W. Zong, G. Moody, and D. Jiang, "A robust open-source algorithm to detect onset and duration of QRS complexes," in *Computers in Cardiology,* pp. 737-740, 2003.

[15] M. Malik, J. T. Bigger, A. J. Camm, R. E. Kleiger, A. Malliani, A. J. Moss, *et al.*, "Heart rate variability: Standards of measurement, physiological interpretation, and clinical use," *Eur Heart J,* vol. 17, pp. 354-381, March, 1996.

[16] R. Acharya, P. Joseph, N. Kannathal, C. M. Lim, and J. S. Suri, "Heart rate variability: a review," *Med Biol Eng Comput,* vol. 44, pp. 1031-51, Dec 2006.

[17] J. Ramshur, "Design, Evaluation and application of Heart rate variability software," ed, 2010.

[18] J. P. Zbilut, N. Thomasson, and C. L. Webber, "Recurrence quantification analysis as a tool for nonlinear exploration of

nonstationary cardiac signals," *Medical Engineering & Physics,* vol. 24, pp. 53-60, Jan 2002.

[19] K. R. Foster, R. Koprowski, and J. D. Skufca, "Machine learning, medical diagnosis, and biomedical engineering research-commentary," *Biomedical engineering online,* vol. 13, p. 94, 2014.

[20] L. Yu and H. Liu, "Efficient feature selection via analysis of relevance and redundancy," *The Journal of Machine Learning Research,* vol. 5, pp. 1205-1224, 2004.

[21] G. H. John and P. Langley, "Estimating continuous distributions in Bayesian classifiers," in *Proceedings of the Eleventh conference on Uncertainty in artificial intelligence*, pp. 338-345, 1995.

[22] D. J. Hand, H. Mannila, and P. Smyth, *Principles of data mining*: MIT press, 2001.

[23] V. N. Vapnik, *Statistical learning theory*. New York: Wiley, 1998.

[24] C. M. Bishop, *Neural networks for pattern recognition*. Oxford: Clarendon Press ; Oxford University Press, 1995.

[25] X. Wu, V. Kumar, J. R. Quinlan, J. Ghosh, Q. Yang, H. Motoda, *et al.*, "Top 10 algorithms in data mining," *Knowledge and Information Systems,* vol. 14, pp. 1-37, 2008.

[26] M. Hall, E. Frank, G. Holmes, B. Pfahringer, P. Reutemann, and I. H. Witten, "The WEKA data mining software: an update," *ACM SIGKDD explorations newsletter,* vol. 11, pp. 10-18, 2009.

[27] M. Kohl, "Performance measures in binary classification," *International Journal of Statistics in Medical Research,* vol. 1, pp. 79-81, 2012.

[28] P. Melillo, M. Bracale, and L. Pecchia, "Nonlinear Heart Rate Variability features for real-life stress detection. Case study: students under stress due to university examination," *BioMedical Engineering OnLine,* vol. 10, pp. 1-13, 2011.

[29] M. Baig, H. Gholamhosseini, and M. Connolly, "A comprehensive survey of wearable and wireless ECG monitoring systems for older adults," *Medical & Biological Engineering & Computing,* vol. 51, pp. 485-495, 2013.

[30] B. Mirmahboub, S. Samavi, N. Karimi, and S. Shirani, "Automatic Monocular System for Human Fall Detection Based on Variations in Silhouette Area," *IEEE Transactions on Biomedical Engineering,* vol. 60, pp. 427-436, 2013.

[31] L. Yun, K. C. Ho, and M. Popescu, "Efficient Source Separation Algorithms for Acoustic Fall Detection Using a Microsoft Kinect," *IEEE Transactions on Biomedical Engineering,* vol. 61, pp. 745-755, 2014.

[32] E. E. Stone and M. Skubic, "Fall detection in homes of older adults using the microsoft kinect," *Biomedical and Health Informatics, IEEE Journal of,* vol. 19, pp. 290-301, 2015.

[33] J. Cheng, X. Chen, and M. Shen, "A framework for daily activity monitoring and fall detection based on surface electromyography and accelerometer signals," *Biomedical and Health Informatics, IEEE Journal of,* vol. 17, pp. 38-45, 2013.

[34] X. Ma, H. Wang, B. Xue, M. Zhou, B. Ji, and Y. Li, "Depth-based human fall detection via shape features and improved extreme learning machine," *Biomedical and Health Informatics, IEEE Journal of,* vol. 18, pp. 1915-1922, 2014.

Rossana Castaldo (SM'15) was born in Torre del Greco, Naples, Italy on August 20, 1990. She received the M.Sc. degree in biomedical engineering at University of Warwick in 2014. She is currently Ph.D. student in Biomedical Engineering at University of Warwick, UK, where she has been an associate of the *Applied Biomedical Signal Processing and Intelligent eHealth Lab (ABSPIE)* since 2013. Her main areas of expertise are biomedical signal processing, machine learning and data analytics. Rossana is member of the Italian Scientific society on Medical and Biomedical Engineering and member of the International Women Committee (Women in MBE) of the International Federation of Medical and Biomedical Engineering (IFMBE).

Paolo Melillo (M'12) was born in Naples, Italy, on June 29, 1985. He received the M.Sc. degree (Hons.) in 2008 in biomedical engineering and the Ph.D. degree in 2012 from the University of Naples "Federico II" Naples. He is Assistant Professor at the Multidisciplinary Department of Medical, Surgical and Dental Sciences of the Second University of Naples, Naples, Italy. He has authored or coauthored about 40 journal and conference papers in the fields of data mining applied to health information, health management, telemedicine, and signal and image processing. Dr. Melillo is a member of the IEEE Engineering in Medicine and Biology Society, the Italian Association of Medical and Biological Society, the International Federation of Medical and Biological Society, and the Italian Mathematical Union.

Leandro Pecchia (M'13) was born in Naples, Italy, on September 13, 1975. He received the Ph.D. degree in Biomedical Engineering, Economy and Management of Healthcare Services and Organizations in 2019 from the University of Naples "Federico II". He is Assistant Professor of Biomedical Engineering at The University of Warwick, Coventry, UK, where he directs the *Applied Biomedical Signal Processing and Intelligent eHealth Lab (ABSPIE)*. He has authored or coauthored about 90 journal, book and conference papers in the fields of machine learning applied to health information, health management, telemedicine, and signal and image processing.

Dr. Pecchia is the Chairman of the Health Technology Assessment Division of the International Federation of Medical and Biomedical Engineering (IFMBE), AC member of the European Alliance of Medical and Biological Engineering and Science and member of the IEEE UK/IR Chapter.