# Ultra-short term HRV features as surrogates of short term HRV: a case study on mental stress detection in real life

R. Castaldo[1,2], L. Montesinos[1], P. Melillo[3], C. James[1] and L. Pecchia[1*]

## Abstract

**Background:** This paper suggests a method to assess the extent to which ultra-short Heart Rate Variability (HRV) features (less than 5 min) can be considered as valid surrogates of short HRV features (nominally 5 min). Short term HRV analysis has been widely investigated for mental stress assessment, whereas the validity of ultra-short HRV features remains unclear. Therefore, this study proposes a method to explore the extent to which HRV excerpts can be shortened without losing their ability to automatically detect mental stress.

**Methods:** ECGs were acquired from 42 healthy subjects during a university examination and resting condition. 23 features were extracted from HRV excerpts of different lengths (i.e., 30 s, 1 min, 2 min, 3 min, and 5 min). Significant differences between rest and stress phases were investigated using non-parametric statistical tests at different time-scales. Features extracted from each ultra-short length were compared with the standard short HRV features, assumed as the benchmark, via Spearman's rank correlation analysis and Bland-Altman plots during rest and stress phases. Using data-driven machine learning approaches, a model aiming to detect mental stress was trained, validated and tested using short HRV features, and assessed on the ultra-short HRV features.

**Results:** Six out of 23 ultra-short HRV features (MeanNN, StdNN, MeanHR, StdHR, HF, and SD2) displayed consistency across all of the excerpt lengths (i.e., from 5 to 1 min) and 3 out of those 6 ultra-short HRV features (MeanNN, StdHR, and HF) achieved good performance (accuracy above 88%) when employed in a well-dimensioned automatic classifier.

**Conclusion:** This study concluded that 6 ultra-short HRV features are valid surrogates of short HRV features for mental stress investigation.

**Keywords:** Heart rate variability (HRV), Ultra-short term HRV analysis, Mental stress detection, Data-driven machine learning

## Background

Stress is defined by the American Psychological Association as "the pattern of specific and nonspecific responses an organism makes to stimulus events that disturb its equilibrium and tax or exceed its ability to cope" [1]. In particular, mental stress has been defined by Lazarus and Folkman as a form of stress that occurs because of how events in one's external or internal environment are perceived, resulting in the psychological experience of distress and anxiety [2, 3]. In humans, mental stress has been investigated using several cognitive stressors in laboratory (e.g., computer work tasks, Stroop color word test, arithmetic tasks, game tasks) or in real-life scenarios (e.g., public speech tasks, academic examinations, during surgeries) [4]. Mental stress can manifest itself as many different symptoms and signs, ranging from physiological (i.e., increased heart rate, sweating) to psychological (i.e., anxiety) and behavioral (i.e., altered sleep patterns) manifestations. Moreover, subjects may experience these to varying degrees [5, 6]. In this study, mental stress is investigated during a verbal academic examination, which has shown to be a stressful situation resulting in accentuated sympathovagal antagonism [7–9]. Although different concerns have been raised about academic examination

* Correspondence: l.pecchia@warwick.ac.uk
[1]School of Engineering, University of Warwick, CV47AL, Coventry, UK
Full list of author information is available at the end of the article

as stressor due to arousal or other feelings, it has shown to be a reliable cognitive stressor [8–10]. In this study, the rest phase was acquired in a period where subjects were not under any academic pressure.

Mental stress has been investigated in various fields due to its detrimental effects on the daily routine [11]. In fact, whereas some kinds of stress may be beneficial by allowing humans to respond to threats in their environment, mental stress can also decrease attentional resources, impair working memory and memory retrieval, and overload cognitive systems [6, 12]. Stress influences judgment and decision-making, and has been shown to reduce human performance [4, 13]. There is a need to better understand the impact of stress on cognition and performance, especially in high-risk domains such as military, policing, surgery, aviation, driving, and elite-level sport, in which risks or threats are prevalent and they can result in devastating consequences [4, 14].

Although in the existing literature there are multiple physiological signals used to detect stress, such as galvanic skin response, blood pressure, electroencephalogram, respiration rate, and electrocardiogram (ECG), heart rate variability (HRV) is currently one of the most investigated methods for assessing mental stress [4]. Moreover, HRV is a more sensitive measure of stress than heart rate alone [15]. HRV describes the variations of the intervals between consecutive peaks of the R-waves in an ECG and it can be analyzed in the time, frequency and non-linear domains. HRV analysis can be performed on 24 h nominal recordings (defined as long term HRV analysis), 5 min recordings (defined as short term HRV analysis) or shorter recordings [16]. In this paper, ultra-short term HRV analysis is defined as the analysis performed on HRV excerpts shorter than 5 min.

During mental stress, there is an activation of the sympathetic nervous system and a withdrawal of the parasympathetic nervous system [5], which results in significant changes in many HRV features [4, 17]. Previous studies have shown that long and short HRV features change consistently during mental stress and that they are able to reliably capture stress in laboratory and real-life scenarios [4, 5, 18–20]. However, much less work has been done on real-life stress detection via ultra-short term HRV analysis. The demands of ultra-short term HRV analysis for monitoring individual's well-being status is increasing, due to the diffusion of wearable sensors in the healthcare and consumer devices such as mobile phones and smart watches [21, 22]. In e-health monitoring, in fact, the conventional 5 min recordings might be unsuitable, due to real-time requirements. Ultra-short term HRV analysis, especially in combination with wearable sensors, may allow continuous and real time monitoring of an individual' stress levels, which is important in some circumstances or jobs (e.g., surgeons, airplane pilots).

However, numerous challenges have arisen by shortening HRV excerpts below 5 min. In fact, a recent literature review highlights the lack of rigorous methods utilized to explore the extent of which ultra-short HRV features can be used to estimate short term ones [22]. In medicine, particularly in clinical trial designs, in order to cope with this kind of problems, the concept of surrogate endpoint (or marker) was introduced [23, 24]. However, proving whether a marker is a valid surrogate of a real clinical outcome can be quite difficult, and combination of appropriate statistical and correlation tests is required, as detailed elsewhere [22]. In a previous study [25] we explored the feasibility of using ultra-short HRV features for mental stress automatic detection basing on descriptive statistics and without developing a systematic method to identify reliable surrogates for short HRV features.

To the best of the authors' knowledge, none of the studies investigating ultra-short HRV features has proposed a robust methodology to assess if ultra-short HRV features are valid surrogates of short ones to detect stress [22]. There have been some attempts to investigate the reliability and accuracy of ultra-short term HRV analysis [15, 23, 24, 26–42], but only one study investigated the validity of ultra-short HRV features in a more rigorous way [39]. However, the authors in [39] only considered 2 time domain HRV features under one standard condition (i.e., rest phase).

Therefore, the current study is the first proposing a rigorous method to assess the validity of ultra-short HRV features for detecting mental stress. The current paper aims to show to what extent HRV features are reliable and accurate to automatically detect mental stress when moving from short (used as benchmark) to ultra-short term HRV analysis. Moreover, the proposed method could be suitable for other applications using ultra-short term HRV analysis to detect an adverse healthcare event.

## Methods
### Dataset
The data analyzed in the current study were acquired from 42 healthy students in the School of Biomedical Engineering of the University of Naples Federico II (Italy) as described in a previous study [7]. Melillo et al. assumed that verbal university examinations was a valid real-life stressor [7]. This hypothesis was previously proposed by other studies [8, 9] and confirmed by a systematic literature review proving that HRV features, computed during a verbal examination, presented a similar behavior as those acquired using other stressors [4]. As described in Melillo et al. [7], the data were acquired in two different days: the first recording was performed during an ongoing university verbal examination (i.e., stress phase) before Easter break (which in Italy last

less than 10 days), while the second one was taken in controlled resting condition (i.e., rest phase) after the vacations, far away from stress induced from study routines. During the rest phase, subjects were invited to sit on a comfortable chair and they were induced to talk, as they had done during the verbal examination, as talk has proven to alter respiration and therefore, HRV features [43]. Ethical permission was sought from the Local Ethic Committee. A commercial electrocardiograph (Easy ECG Pocket, manufactured by Ates Medical) was used to acquire 3-lead clinical research ECG signals, with a sampling frequency of 500 Hz and a resolution of 12 bits per sample. Kendall™ 530 series foam electrodes were used as they are designed for superior performance including adult stress, holter and diaphoretic applications. Electrodes were placed to the subjects' chest according to the standard guidelines [44]: one electrode was placed under right clavicle near right shoulder within the rib cage frame; the second electrode was placed under left clavicle near left shoulder within the rib cage frame; the third electrode was placed on the left side below pectoral muscles lower edge of left rib cage. Subjects were helped by expert staff of their gender to position the electrodes.

The resting condition was measured at the same time slot as for the stress phase, in order to minimize circadian cycle effects on the HRV, and in the same menstrual cycle for women, as this is also a relevant measure for HRV features [7, 45]. The participants were examined under standard conditions during rest and stress phases: in the same quiet room, at a comfortable temperature, while sitting. Data acquisitions were carried out in the morning for both rest and stress phases between 9 a.m. and 12 p.m.
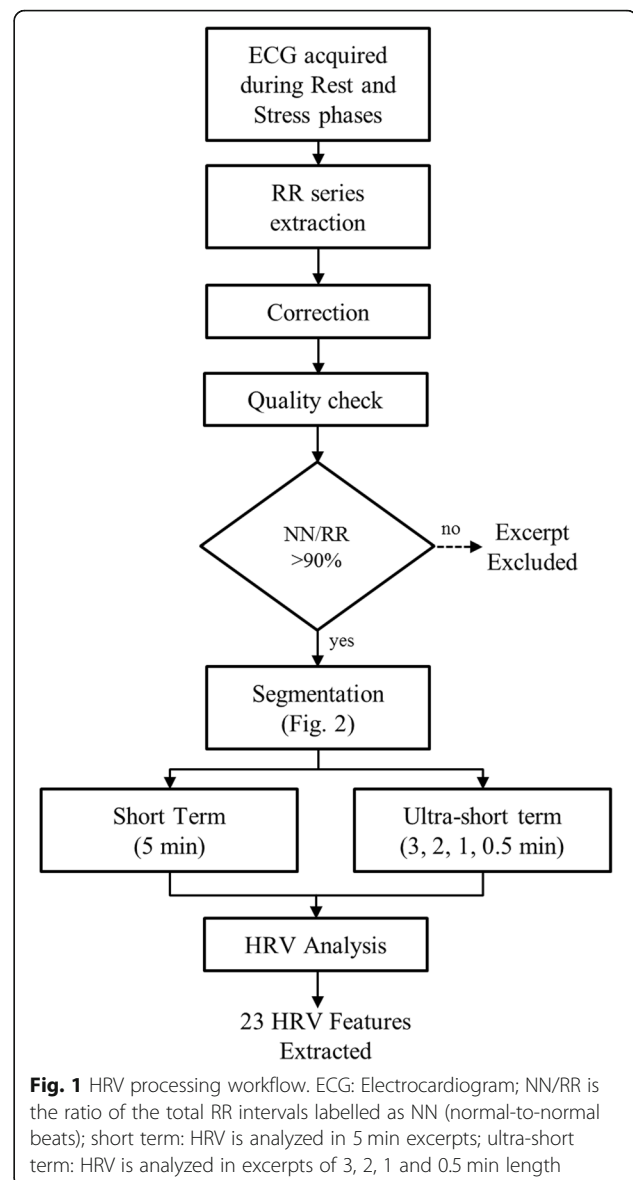
Three-lead ECG was recorded for at least 30 min. The first 15 min (i.e., adaptation time) were excluded and one ECG excerpt of 5 min was extracted and analyzed.

Participants were invited to refrain from drinking alcohol and to limit their tea and coffee intake (2 cups max.) in the 24 h prior to the acquisitions, as alcohol, caffeine and tea have shown to alter HRV features [46].

The participants enrolled in the study had no history of heart disease, systemic hypertension, metabolic disorders or other diseases potentially influencing HRV. They were not obese and did not consume medication, drugs or alcohol in the 24 h preceding the experiments. All of the participants signed specific informed consent form before the acquisitions. More details on the protocol can be found in [7].

### HRV analysis
As shown in Fig. 1, the RR interval time series were extracted from ECG recordings using an automatic QRS detector, WQRS, available in the PhysioNet's toolkit [47], based on nonlinearly scaled ECG curve length



**Fig. 1** HRV processing workflow. ECG: Electrocardiogram; NN/RR is the ratio of the total RR intervals labelled as NN (normal-to-normal beats); short term: HRV is analyzed in 5 min excerpts; ultra-short term: HRV is analyzed in excerpts of 3, 2, 1 and 0.5 min length

features. Details on ECGs pre-processing used in the PhysioNet's toolkit can be read in Zong et al. [48]. The automatic QRS detection was followed by visual inspection and manual correction. QRS review and correction was performed using WAVE, which is the graphical user interface to visualize biomedical signals provided by PhysioNet and includes facilities for interactive annotation editing. Additional details can be found elsewhere [49].

The fraction of total RR intervals labelled as normal-to-normal (NN) intervals was computed as NN/RR ratio. When ectopic beat correction methods are not adopted, and more than one RR excerpt is available per each subject, the NN/RR ratio is used to identify a window of time of sufficient quality, excluding those windows of time in which this ratio is lower than a threshold.

Thresholds of 80% [47] and 90% [50] have been proposed. In studies enrolling only healthy and young subjects, a lower NN/RR ratio is associated with movement artifacts. In the current study, in which subjects were healthy and young, sitting in a comfortable position, a threshold of 90% was chosen and still no records were excluded. Therefore, short HRV features were computed from the first 5 min after the adaptation time for all of the participants.

The same 5-min excerpts were later used to extract shorter NN excerpts (Fig. 2, left-hand side) from which the ultra-short HRV features were computed. The initial choice of extracting the central excerpts was arbitrary. Therefore, we decided to assess this choice by repeating the extraction from different locations within the 5-min excerpt (Fig. 2, right-hand side). This was done only with the shortest significant time length excerpt, resulting from the statistical significance and correlation analysis.

The HRV analysis was performed using Kubios software [51]. Time and frequency features were analyzed according to international guidelines [16], whereas non-linear features were analyzed as described in [7]. Spectral analysis can be performed by different methods, which are classified as non-parametric, such as the Fast Fourier Transform (FFT), and the autoregressive method (AR). The FFT has the advantage of the low computational cost. However, it has limitations such as a poor spectrum resolution, mainly when short data excerpts are used, and leakages. On the other hand, the AR method became popular because it produces a spectrum with better resolution when short data excerpts are used, and the spectrum can be divided into independent components. Therefore, frequency domain features were extracted from power spectrum estimated with autoregressive (AR) model methods. As reported in Additional file 1: Table S1[1] 23 HRV features were extracted from

5-min, 3-min, 2-min, 1-min, and 30-s excerpts and subsequently analyzed.

However, some HRV features were not computable in ultra-short time excerpts. In fact, it is recommended that spectral analyses are performed on recordings at least 10 times longer than the wavelength of the lower frequency limit that is at least 2 min for the Low Frequency power (LF). Therefore, LF was not computed for excerpts below 2 min along with LF/HF ratio and total power. Additionally, High Frequency power (HF) was not computed for excerpts below 1 min [16]. As far as non-linear HRV features are concerned, less has been explored in the existing literature. However, approximate entropy (ApEn) values were excluded for length below 3 min, since they have shown to be unreliable due to the small number of samples presented in the RR series [31, 52, 53]. Moreover, when the length of the data was reduced to 30 s, most of the non-linear features became non-computable, due to the lack of samples. The entire list of features calculated for the different excerpt lengths is reported in Table 1.

## Multiscale HRV comparison: Short vs ultra-short
Median (MD), standard deviation (SD), 25th and 75th percentiles were calculated for all of the subjects to describe the distribution of HRV features during rest and stress phases at 5 min, 3 min, 2 min, 1 min, and 30 s (see Additional file 1:Table S2-S6). Moreover, a non-parametric statistical significance test and a correlation analysis were performed in parallel, as shown in Fig. 3, in order to select the subset of ultra-short HRV features that were good surrogates of short HRV features.

## Non-parametric statistical significance and inter-group assessment
The non-parametric Wilcoxon signed-rank test was used to investigate the statistical significances ($p$-value< 0.05)
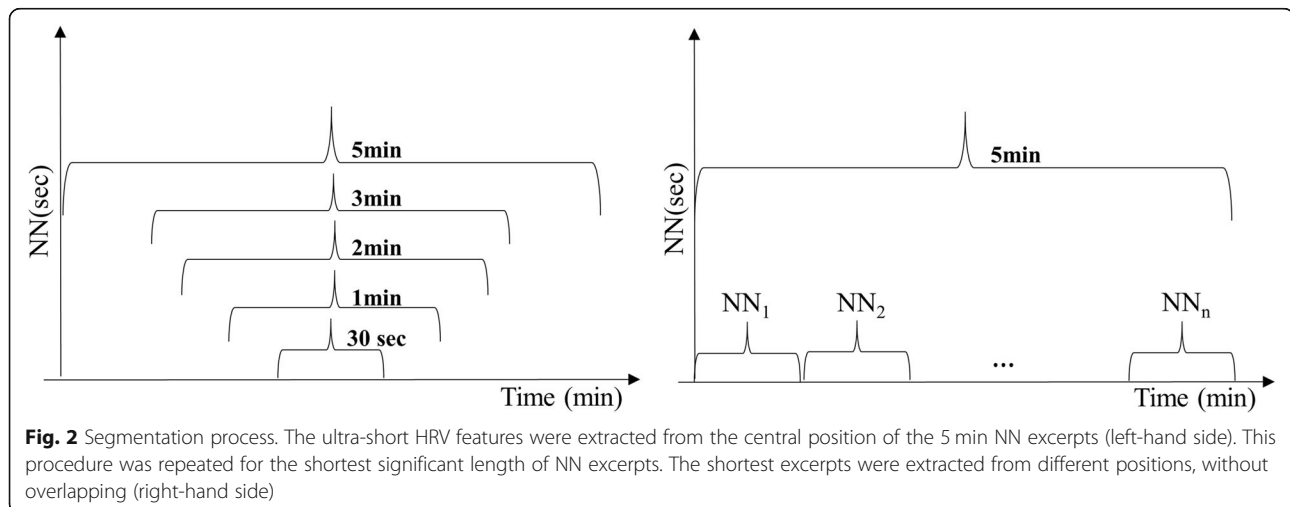


**Fig. 2** Segmentation process. The ultra-short HRV features were extracted from the central position of the 5 min NN excerpts (left-hand side). This procedure was repeated for the shortest significant length of NN excerpts. The shortest excerpts were extracted from different positions, without overlapping (right-hand side)

**Table 1** HRV feature trends

| HRV Features | 5 min | 3 min | 2 min | 1 min | 30 sec |
|---|---|---|---|---|---|
| MeanNN | ↓↓ | ↓↓ | ↓↓ | ↓↓ | ↓↓ |
| StdNN | ↓↓ | ↓↓ | ↓↓ | ↓↓ | ↓↓ |
| MeanHR | ↑↑ | ↑↑ | ↑↑ | ↑↑ | ↑↑ |
| Std HR | ↑↑ | ↑↑ | ↑↑ | ↑↑ | ↑↑ |
| RMSSD | ↑ | ↑ | ↑ | ↑ | ↓ |
| NN50 | ↑ | ↑ | ↑ | ↑ | ↑ |
| pNN50 | ↓ | ↓ | ↑ | ↓ | ↓ |
| LF | ↓↓ | ↓↓ | ↓↓ | – | – |
| HF | ↓↓ | ↓↓ | ↓↓ | ↓↓ | – |
| LF/HF | ↓↓ | ↓↓ | ↓↓ | – | – |
| TotPow | ↓↓ | ↓↓ | ↓↓ | – | – |
| SD1 | ↑ | ↑ | ↑ | ↑ | ↓ |
| SD2 | ↓↓ | ↓↓ | ↓↓ | ↓↓ | ↓↓ |
| ApEn | ↓↓ | ↓↓ | – | – | – |
| SampEn | ↓↓ | ↓↓ | ↓↓ | ↓↓ | – |
| D2 | ↓↓ | ↓↓ | ↓↓ | ↓↓ | – |
| dfa1 | ↓↓ | ↓↓ | ↓↓ | ↓↓ | – |
| dfa2 | ↑ | ↑ | ↑ | ↓ | – |
| RPlmean | ↑↑ | ↑↑ | ↑↑ | ↑ | – |
| RPlmax | ↓↓ | ↓↓ | ↓ | ↑ | – |
| REC | ↑↑ | ↑↑ | ↑↑ | ↑ | – |
| RPadet | ↑↑ | ↑ | ↑ | ↓ | – |
| ShanEn | ↑↑ | ↑↑ | ↑↑ | ↑↑ | – |

Trend ↓↓ (↑↑): significantly lower (higher) under stress ($p < 0.05$), ↓(↑) lower (higher) under stress ($p > 0.05$), – not computable

of the HRV feature variations between stress and rest phases for each excerpt length (i.e., 5 min, 3 min, 2 min, 1 min, and 30 s).

The increase or decrease of the HRV feature median between rest and stress phases was reported, referred as median trend or feature trend, using the following convention [4, 54, 55]:

- Two arrows, ↓↓ (or ↑↑) were used to report a significant (*p*-value< 0.05) decrease (or increase) of a feature median during the stress phase;
- One arrow was used for non-significant variations: ↓ (or ↑) indicated a non-significant (p-value> 0.05) decrease (or increase) of a feature median during the stress phase.

Trend analysis consisted in inspecting inter-group changes (i.e., increase or decrease) of HRV feature median across time scales.

An HRV feature was assumed to maintain the same behavior across the 5 different time-scales (5 min, 3 min, 2 min, 1 min, and 30 s) if:

- the Wilcoxon's test *p*-value was less than 0.05 between rest and stress phases over all time scales;
- the ultra-short HRV feature's median trend was changing between rest and stress phases consistently with the equivalent short HRV feature.

### Correlation analysis and Bland-Altman plots as intra-group assessment

An intra-group assessment was carried out using the Spearman's rank correlation and Bland-Altman plots to investigate to what extent an ultra-short HRV feature was correlated with the equivalent short HRV feature during rest and stress conditions. For instances, Mean-$NN_{3min}$ (i.e., computed at 3 min during rest) VS Mean-$NN_{5min}$ (i.e., computed at 5 min during rest).

The statistical significance of this association was demonstrated by a p-value ($p_{rho}$) lower than 0.05. As first screening, each ultra-short HRV feature was investigated against the equivalent short HRV feature during resting condition (Fig. 3). In addition, each ultra-short HRV feature was also explored during stress condition. However, a correlation coefficient is blind to the possibility of bias caused by the difference in the mean or standard deviation between two measurements; in other words, a strong correlation does not necessarily imply a close agreement. Therefore, Bland-Altman procedure was used to calculate 95% LoA (Limits of Agreement) [56]. In contrast to the traditional Bland-Altman plots, the measurements of the 5 min was plotted on the x-axis [39]. The bias was calculated as the median difference between the HRV features at 5 min and the ultra-short HRV features.

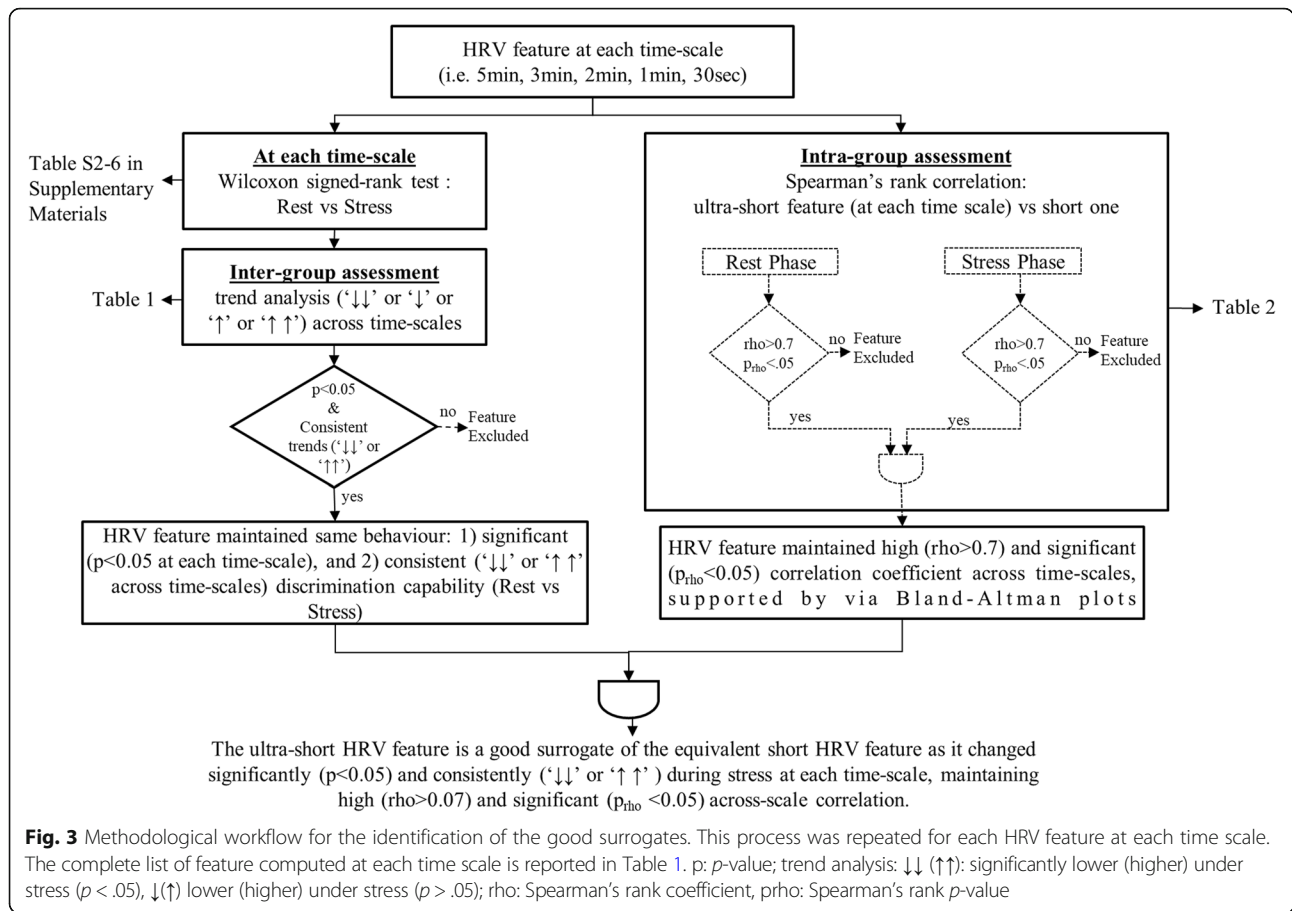### Surrogate feature subset selection

At this stage, it was assumed that an ultra-short HRV feature was a valid surrogate of the equivalent short one, only if:

- the feature maintained the same behavior between rest and stress conditions at each time scale;
- the ultra-short HRV feature was highly and significantly correlated (i.e. rho> 0.7 and $p_{rho} < 0.05$), with the equivalent short feature, over all of the time scales in both rest and stress phases.

### Classification and performance evaluation

Short HRV features (benchmark) were used to train, validate and test an automatic classifier to detect mental stress. The performance of this classifier was then tested inputting ultra-short HRV features in order to assess the discriminant power of ultra-short term HRV analysis.

To reduce overfitting problems and bias in the overall accuracy of the classifier, the whole dataset was randomly split per subject into two folders: folder 1 (60%) was used for feature selection, training and validation of

**Fig. 3** Methodological workflow for the identification of the good surrogates. This process was repeated for each HRV feature at each time scale. The complete list of feature computed at each time scale is reported in Table 1. p: *p*-value; trend analysis: ↓↓ (↑↑): significantly lower (higher) under stress (*p* < .05), ↓(↑) lower (higher) under stress (*p* > .05); rho: Spearman's rank coefficient, prho: Spearman's rank *p*-value

the classifiers; folder 2 (40%) for testing the classifier. The reasoning behind this split is that a classifier should be tested on a set of data that is independent of the training data [55, 57]. Moreover, although the best approach is to select the minimum set of features using a different folder from the one adopted to train the machine learning classifier [55, 57], due to the small number of subjects, feature selection, training and validation were performed on the same folder (folder 1).

### HRV feature selection for modelling

In order to minimize the over-fitting risk in a machine learning model, the number of features used in the model and its cardinality should be limited by the number of subjects presenting the event to detect (i.e., stress) in each folder [57]. Moreover, a significant small set of

clinical features strongly simplifies the physiological interpretation of results, by directing attention only on the most informative features [57]. In addition, although the best approach is to select the minimum set of features using a different folder from the one adopted to train the machine learning model [55], due to the small number of subjects in this study, feature selection and model training were performed on the same folder (folder 1: 25 subjects). Because 5 min is defined as standard length for short term HRV analysis, the feature selection was performed using 5-min HRV features. Among the 23 HRV features initially computed, only those that showed to be good surrogates in the ultra-short time excerpts entered the feature selection process to build a classifier.

The feature selection was based on two main stages (Fig. 4): the relevance analysis and the redundancy



**Fig. 4** Framework of feature selection
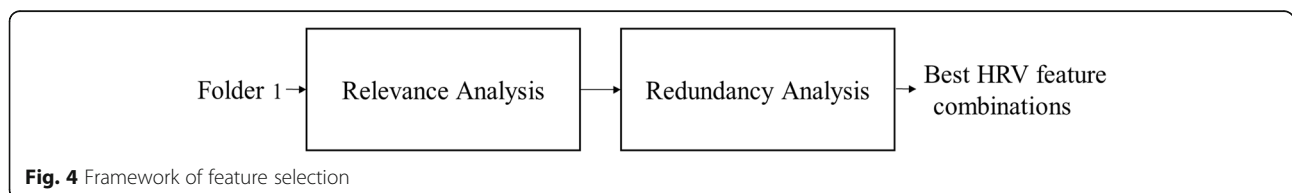
analysis. The former was performed using the Wilcoxon signed-rank test. The latter consisted in selecting only one feature from each cluster of features mutually correlated using Spearman's rank correlation to reduce multicollinearity in the models. More details can be found in [55].

### Machine learning methods

Five different machine-learning methods were used to develop classifiers aiming to automatically detect mental stress based on short HRV features: 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 [58]; Multilayer Perceptron (MLP) consists of an artificial neural network of nodes (processing elements) arranged in layers [59]; Neighbor Search (IBK), which finds a group of K objects in the training set that are the closest to the test object, bases the assignment of a label on the predominance of a particular class in the neighborhood [60]; C4.5 builds decision trees from a set of training data, using the concept of information entropy [61]; Linear Discriminant Analysis (LDA) aims to find linear combinations of the input features that can provide an adequate separation between two classes [62].

Regarding model parameters, for the MLP classifier, the learning rate was varied from 0.3 to 0.9, the momentum from 0.2 to 1, the number of excerpts from 100 to 2000 and the number of hidden layers was set to 1 with 3 units [63, 64]; for the SVM, polynomial kernel function was used, varying the degree from 1 to 5. As regards IBK, it was trained with K equals to 1, 3 and 5 due to the binary nature of the classification problem [65]. C4.5 trees were developed by varying confidence factor (CF) for pruning from 0.05 to 0.5, minimum number of instances (ML) per leaf from 2 to 20. The algorithm parameters were tuned during training on folder 1.

Each of those methods was used with all of the combinations of relevant and non-redundant HRV features.

The Weka platform for knowledge discovery (version 3.6.10) (University of Waikato) was used to train, validate and test the classification models. This platform was issued as an open source software under the GNU General Public License [63].

### Training, validation and testing

The training of the machine-learning models (including feature selection and algorithm parameter tuning) was performed on the folder 1 (25 subjects) and using 5-min HRV features (benchmark). Folder 1 was also used to validate the classifier using a k-fold cross-validation technique. The choice of the k-value is crucial in order to achieve high overall accuracy and reduce bias in the model. As a rule of thumb, the k-value should allow each of the k-folds to have at least 10 occurrences

presenting the events to detect. Therefore, the 3-fold person-independent cross-validation approach was used to validate the models in folder 1 [55].

Binary classification performance measures were adopted according to the standard formulae reported in [66, 67].

Among the five different machine-learning methods used to train, validate and test the classifiers (SVM, MLP, IBK, C4.5, and LDA), the best-performing model was chosen as the classifier achieving the highest Area under the Curve (AUC), which is a reliable estimator of both sensitivity and specificity rates.

The model was then tested on folder 2 (17 subjects) using the short and ultra-short HRV features, in order to assess their efficacy in automatically detecting mental stress. Binary performances were computed again, but this time to observe how the performance of the model changed across different time-scales.

## Results

ECGs recorded from 42 healthy subjects (19 female, 23 male) were analyzed in the current study. Subjects were aged 18 to 25 years old (age: 21.5 ± 3.5), were no obese (BMI 22.3 ± 2.7) and were not taking any medication for the duration of the study. HRV features median (MD), standard deviation (SD), 25th and 75th percentiles calculated on 5-min, 3-min, 2-min, 1-min, and 30-s NN data series are given in the Additional file 1: Table S2-S6, respectively.[2]

### Multiscale HRV comparison: Short vs ultra-short

Table 1 summarizes the results of the significance and trend analysis, presenting the HRV features' median trend at each time-scale. Table 1 also reports the HRV features calculated for the different excerpt lengths (i.e., features indicated with '-' were not computable).

As shown in Table 1, from 5-min excerpts of NN data series, 18 out of the 23 selected HRV features showed significant changes from resting to stress conditions. Twelve out of these 18 features decreased significantly during stress phase, while the remaining 6 features (MeanHR, StdHR, RPlmean, REC, RPadet and ShanEn) showed a significant increase.

The second column in Table 1 demonstrates that from 3-min excerpts of NN data series all of the 23 features were computable, and 12 features decreased significantly during stress, while 5 (MeanHR, StdHR, RPlmean, REC, and ShanEn) increased significantly. However, RPladet which showed significant increase during 5 min, failed to show any significant change when the data length was shortened below 5 min.

The changes in the features extracted from 2-min excerpts, shown in the third column of Table 1, present the same significant trends as the 3-min features, apart from ApEn, which is not computable, and RPlmax, which is no longer significant (*p*-value< 0.05).

The changes in the features extracted from 1-min excerpts, shown in the fourth column of Table 1, present the same significant trends as the 2-min features, except for 3 HRV features (LF, LF/HF ratio, TotPow), which are not computable, and 2 HRV features (RPlmean and REC), which are no more significant (p-value< 0.05).

The changes in the features extracted from 30-s excerpts, shown in the fifth column of Table 1, present the same significant trends as the 1-min features, apart from those features that are not computable.

Table 2 shows the results of the correlation analysis used to select the subset of ultra-short HRV features that were good surrogates of short HRV features. The correlation analysis was run between ultra-short HRV features and the equivalent short ones. This analysis was not used to eliminate multicollinearity between features but to investigate the interdependence between an ultra-short HRV feature and its equivalent in 5-min excerpt.

Time domain HRV features maintained a significantly high correlation coefficient at 3 min, 2 min, and 1 min. Conversely, from 30-s excerpts, StdNN showed a

Spearman coefficient above 0.70 at rest and below 0.70 during stress, while StdHR showed a Spearman coefficient below 0.70 during both rest and stress phases. Regarding frequency-domain HRV features, they showed to be highly correlated with the equivalent short HRV features at each time-scale (i.e., from 3 min to 1 min) during both resting and stress phases. As far as non-linear features are concerned, SD1 maintained a constant behavior between short and ultra-short term during rest and stress phases while SD2 was less correlated at 30 s during stress. ApEn, SampEn, D2, RPlmean, RPlmax, REC, RPadet and ShanEn showed to be highly correlated with short HRV features over 3-min excerpts during resting and stress conditions, while they resulted less correlated in shorter time-scales. In general, HRV features resulted less correlated in resting than during stress conditions. This is most likely due to the fact that HRV showed a more depressed dynamic during stress phase. Similar behaviors have been observed in other studies [68].

Due to this first analysis, the HRV features computed on 30-s excerpts were at this point excluded from the

**Table 2** Correlation analysis of ultra-short HRV features vs equivalent short ones

| HRV Features | Rest Phase | | | | Stress Phase | | | |
|---|---|---|---|---|---|---|---|---|
| | 3 vs 5 min | 2 vs 5 min | 1 vs 5 min | 30 s vs 5 min | 3 vs 5 min | 2 vs 5 min | 1 vs 5 min | 30 s vs 5 min |
| **MeanNN** | **0.984** | **0.890** | **0.975** | **0.936** | **0.985** | **0.937** | **0.955** | **0.964** |
| **StdNN** | **0.954** | **0.875** | **0.905** | **0.749** | **0.962** | **0.912** | **0.791** | 0.640 |
| **MeanHR** | **0.984** | **0.891** | **0.975** | **0.947** | **0.985** | **0.938** | **0.954** | **0.964** |
| **StdHR** | **0.914** | **0.789** | **0.796** | 0.635 | **0.971** | **0.904** | **0.784** | 0.696 |
| **RMSSD** | **0.961** | **0.914** | **0.946** | **0.859** | **0.983** | **0.928** | **0.915** | **0.852** |
| **NN50** | **0.972** | **0.883** | **0.949** | **0.822** | **0.971** | **0.920** | **0.905** | **0.894** |
| **pNN50** | **0.967** | **0.882** | **0.943** | **0.818** | **0.969** | **0.915** | **0.913** | **0.881** |
| **LF** | **0.894** | **0.886** | – | – | **0.921** | **0.916** | – | – |
| **HF** | **0.915** | **0.906** | **0.901** | – | **0.925** | **0.915** | **0.798** | – |
| **LF/HF** | **0.830** | **0.839** | – | – | **0.846** | **0.807** | – | – |
| **TotPow** | **0.897** | **0.882** | – | – | **0.900** | **0.905** | – | – |
| SD1 | **0.961** | **0.914** | **0.945** | **0.862** | **0.983** | **0.928** | **0.915** | **0.852** |
| SD2 | **0.956** | **0.865** | **0.876** | **0.707** | **0.941** | **0.898** | **0.755** | 0.694 |
| ApEn | **0.771** | 0.169 | – | – | **0.918** | **0.790** | – | – |
| SampEn | **0.855** | 0.666 | 0.681 | – | **0.931** | **0.826** | 0.599 | – |
| D2 | **0.922** | 0.674 | 0.330 | – | **0.967** | **0.876** | **0.816** | – |
| dfa1 | 0.661 | 0.687 | 0.637 | – | **0.927** | **0.908** | **0.799** | – |
| dfa2 | 0.633 | 0.611 | 0.673 | – | **0.767** | 0.563 | 0.485 | – |
| **RPlmean** | **0.837** | **0.708** | 0.645 | – | **0.901** | **0.730** | 0.503 | – |
| **RPlmax** | **0.738** | 0.588 | 0.583 | – | **0.896** | **0.737** | 0.678 | – |
| **REC** | **0.880** | 0.643 | 0.608 | – | **0.892** | 0.689 | 0.513 | – |
| **RPadet** | **0.852** | 0.645 | 0.495 | – | **0.948** | **0.817** | 0.642 | – |
| **ShanEn** | **0.795** | 0.661 | 0.614 | – | **0.907** | **0.720** | 0.463 | – |

All the correlations resulted significant ($p_{rho}$ < 0.05); in bold Spearman's correlation coefficient (rho) greater than 0.7; –: not computable

rest of the study due to the low number of HRV features behaving coherently with the benchmark. The results from the correlation analysis were supported by the visual inspection of the Bland-Altman plots. A decrease in bias and in width of the 95% LoA was observed as the excerpts length increased for all of the HRV features. A representative example is shown in Additional file 2: Figure S1 and Additional file 3: Figure S2.

As a result, MeanNN, StdNN, MeanHR, StdHR, HF and SD2 were selected as valid surrogates of short HRV features to investigate mental stress, as they displayed consistency across all of the excerpt lengths (i.e., from 5 to 1 min). Moreover, the discrimination power to automatically detect stress of these features across all of the excerpt lengths (i.e., from 5 to 1 min) was also corroborated as detailed in the section below.

## Classification and performance measurement

Regarding the feature selection process, all of the six HRV features (MeanNN, StdNN, MeanHR, StdHR, HF, and SD2), selected as valid surrogates of short HRV features resulted also relevant in folder 1. This was not a trivial result given the lower number of subjects included in folder 1. In fact, a reduction in the number of subjects may result in an increase of $p$-values. As result of the redundancy analysis the minimum set of relevant but mutually non-correlated features resulted to be: MeanNN, StdHR, and HF.

Each machine learning method was trained and validated with this combination of short HRV features (MeanNN, StdHR, and HF) using folder 1. The classifiers were then tested on short HRV features using folder 2 as shown in Table 3.

According to the criteria defined above, the IBK classifier showed the highest AUC with 88% sensitivity, 100% specificity, 94% accuracy, and 99% AUC, using MeanNN, StdHR and HF as HRV features. Therefore, the IBK was chosen as model to automatically detect mental stress.

**Table 3** Model performance measurements estimated on the test set (Folder 2) on 5 min excerpts

| Method | Parameters | AUC | SEN | SPE | ACC |
|--------|-----------|-----|-----|-----|-----|
| MLP | LR = 0.3; ML = 0.2; NE = 500 | 98% | 100% | 88% | 94% |
| SVM | PolyKernel, E = 1.0 | 88% | 88% | 88% | 88% |
| C4.5 | CF = 0.25; ML = 2 | 94% | 88% | 100% | 94% |
| IBK | K = 3 | 99% | 88% | 100% | 94% |
| LDA | – | 98% | 88% | 100% | 94% |

*MLP* Multilayer Perceptron, *SVM* Support Vector Machine, *C4.5* decision trees, *IBK* Neighbor Search, *LDA* Linear Discriminate Analysis, *AUC* area under the curve, *SEN* sensitivity, *SPE* specificity, *ACC* accuracy

The IBK model was then tested using ultra-short HRV features in folder 2 to further evaluate their capability to automatically detect mental stress (Table 4).

The length of data seemed to slightly affect the performance of the model. However, as shown in Table 4, the model outperformed in 3-min time-scale with 97% AUC.

Compared to the short term performances, sensitivity increased by 6% and specificity decreased by 6% respectively using 3-min excerpts. Nevertheless, the model achieved good performances also using 1-min HRV excerpts. After observing these results, the model was also assessed on consecutive 1-min excerpts (as shown in Fig. 2, right-hand side) within the 5-min NN data series in order to understand if the performances were changing significantly according to the extracted excerpts. In fact, the performances using 1-min HRV features showed to be constantly good with 86 ± 4.1% sensitivity, 95 ± 4.4% specificity and 92 ± 3.75% accuracy.

## Discussion

The current study aimed to investigate if ultra-short HRV features are valid surrogates of short ones to automatically detect mental stress. This is a topic of growing interest. In fact, the continued rise of consumer wearable devices able to instantaneously assess mental stress level is raising the attention of the scientific community around the use of HRV features computed over excerpts shorter than 5 min [22].

Differently from Melillo et al. [7], this study explored the validity of ultra-short HRV features as surrogates of short HRV features to detect mental stress in real-life scenario. Moreover, in Melillo et al. [7] only non-linear HRV features were analyzed in 5-min excerpts and employed to develop a linear classifier.

Regarding the methodology, this study presents an innovative method to assess the minimum length of HRV excerpts to detect mental stress in healthy young subjects. In fact, to the best of the authors' knowledge [22], only two studies evaluated the reliability of ultra-short HRV features during stress condition, but only using statistical significance tests, which as demonstrated in this study are not sufficient to draw any conclusion [31, 69]. In fact, differently from the methods described in the current paper, Pereira et al. used only a

**Table 4** Model performance measurements on different time-scale excerpts

| Duration | AUC | SEN | SPE | ACC |
|----------|-----|-----|-----|-----|
| 3 min | 97% | 94% | 94% | 94% |
| 2 min | 93% | 94% | 88% | 91% |
| 1 min | 93% | 82% | 94% | 88% |

*AUC* area under the curve, *SEN* sensitivity, *SPE* specificity, *ACC* accuracy

parametric statistical test (one-way ANOVA) to determine which HRV features (i.e., 220, 150, 100 and 50 s) could discriminate between rest and stress sessions ($p <$ 0.05) with small windows of analysis [69]. For their part, Salahuddin et al. used the non-parametric Kruskal–Wallis test to assess that ultra-short term analysis was not significantly different to the short term analysis if the *p*-value was greater than 0.05 and Wilcoxon sign-rank test ($p < 0.05$) to find the shortest duration that distinguished between rest and stress phases. However, no correlation or machine learning methods were utilized to validate their findings. Moreover, if the p-value is greater than 0.05 then the null hypothesis cannot be neither rejected nor accepted [70]. Therefore, no conclusion can be drawn using only the statistical significance tests, which make the results reported in [31] not sufficiently reliable. Hence, it is difficult to compare the results reported in [31] with the one reported in the current paper. Unfortunately, this study has been used to support the majority of works related to mental stress detection using ultra-short HRV features [31]. In fact, many wearable systems [28, 41, 42] and scientific studies [26–28, 30, 32, 41, 42] monitoring stress via ultra-short term HRV analysis have based their feature selection on Salahuddin et al. [31] results, which should be read more carefully.

Other studies have investigated the reliability and accuracy of ultra-short HRV features in different conditions (e.g., athletic performance, acoustic sounds and controlled resting conditions) [33–36, 38, 39]. However, none of these studies employed rigorous statistical approaches to identify ultra-short HRV features as good surrogates of short term ones [22].

Differently from our study, few studies only employed correlation tests to prove that ultra-short term HRV features behaved as good surrogates of short-term ones, concluding that ultra-short HRV features were good surrogates of short-term ones if significantly correlated with their equivalent short HRV features [34, 35, 38]. This result is arguable because, as demonstrated in our study, although an appropriate correlation test is the first step for the identification of a good surrogate, a much stronger condition than correlation is required to identify a surrogate [22, 71].

Other studies performed both statistical significance test and correlation analysis in alignment with our study, but they presented various methodological ambiguities [33, 36, 39]. For more details refer to [22]. Only one study investigated HRV features in time domain in 10, 30 and 120 s compared to 5 min, using a more rigorous method [39]. In fact, they used Pearson correlation, after log transforming HRV features, Bland-Altman plots and Cohen's d. However, although the approach used in Munoz et al. [39] to assess the validity of ultra-short

HRV features seems more rigorous than other studies [22], they only investigated 2 time domain HRV features (SDNN and RMSSD) in resting condition.

Regarding our results, the statistical analysis in the short term showed a significant depressed HRV during stress, in agreement with the previously published literature [4]. Ultra-short term HRV features also resulted in being significantly depressed during mental stress over each time-scale. Concerning the HRV features in time domain, all of them maintain the same behavior across the 5 different time-scales (i.e., 5 min, 3 min, 2 min, 1 min, and 30 s). Moreover, four of them (MeanNN, StdNN, MeanHR and StdHR) were also significantly different between rest and stress phases and were significantly correlated (Spearman's rank rho> 0.7) across time-scales (i.e., each ultra-short vs short time-scale per each feature). These results, achieved with a more robust method, confirm the findings of Baek et al. [36], McNames and Aboy [35], Nussinovitch et al. [34], Pereira et al. [69] and Munoz et al. [39], which showed that MeanNN, StdNN, MeanHR are reliable for length from 5 to 1 min in a controlled resting condition. However, some HRV features that showed to be good surrogates in the existing literature, failed to show good results in the present study. Our interpretation of this result is that the method used in the present study is based on more stringent and reliable requirements, compared to other studies, which demonstrated significant methodological limitations [22]. Concerning the HRV features in frequency domain, it is well-known that a minimum of 1 min is required to estimate HF and a minimum of 2 min is required to estimate LF component [16, 36]. Accordingly, the present study showed that for HRV features in frequency domain such as LF, the minimum length is 2 min. However, HF component could be extracted from 1-min excerpts, as confirmed by the fact that in this study HF resulted to be a good surrogate of the 5 min equivalent. In fact, as also proved by Baek et al. [36], LF had a very low Pearson coefficient below 2 min whilst HF below 1 min. In relation to non-linear HRV features, no study has investigated their reliability in excerpts shorter than 5 min. The current study empirically demonstrated that they lose their utility for excerpts below 3 min due to computational problems. In fact, non-linear HRV features require a high number of samples in order to appreciate the dynamics of the heartbeat series over time. Only two HRV non-linear features (SD1 and SD2) showed to be good surrogates over 3, 2 and 1-min lengths as also shown by Nardelli et al. [38].

Although our study employed only 42 healthy subjects to develop a model to automatically detect stress, it is able to detect stress with higher accuracy than the models presented in the existing literature [7, 15, 20, 23, 24, 27–29, 32, 72].

Three studies proposed a model to detect mental stress using short term HRV analysis [7, 20, 72], whilst seven studies developed a model for the detection of mental stress using ultra-short HRV features [15, 23, 24, 27–29, 32]. Melillo et al. [7] adopted the same dataset as in this study and proposed a model based on LDA, employing only three HRV non-linear features: SD1, SD2 and ApEn in short term HRV analysis. The model proposed in their study, achieved sensitivity, specificity and accuracy, of 86, 95 and 90%, respectively, which are lower than the ones achieved by the model developed in this study. Whereas Traina et al. [72] studied the Pearson correlation between frequency domain measures before and after the stress session, demonstrating that those correlations were significant. However, as discussed above, the Pearson correlation lays on the assumption that the HRV measures are normally distributed, yet HRV frequency measures are not. In 2015, Munla et al. [20] used an SVM-RBF classifier using time and non-linear HRV features, with only 16 different individuals, to predict drivers' stress with an accuracy of 83%. However, no validation or testing was applied in that study.

Mayya et al. [28] proposed a method for automatically detecting mental stress using smartphone and focusing on 1-min HRV features. The model was built on the assumption that ultra-short HRV features were relevant according to the available literature [31], which has been proved to lack of a robust method to identify ultra-short HRV features that are good surrogates of short HRV features [22]. They used a multinomial logistic regression applied to 2 features, RMSSD and dfa1, which were excluded in our study, and achieved 80.5% accuracy, which is lower than the accuracy achieved in the present study, supporting the idea that an suboptimal ultra-short feature selection generates low performances. Choi et al. [24], Brisinda et al. [27] and Sun et al. [32] also proposed a method to automatically detect mental stress focusing on 4-min, 2-min and 1-min HRV features respectively. Also in these studies, the models were built on the assumption that ultra-short HRV features were relevant according to the available literature, although Brisinda et al. [27] confirmed their findings using only ICC analysis. These studies used linear classifiers achieving accuracy lower than the one achieved in the current study. Other models were developed using ultra-short term HRV analysis along with other physiological measurements but they are not discussed here [15, 29, 32]. To conclude, none of those papers achieved better results than the one presented in this study. This also supports our convincement that a reliable identification of good surrogates is important to identify a good set of features aiming to detect mental stress. However, it is important to highlight that these studies employed protocols and sample sizes different from our study and therefore, a strict comparison of the classifiers' performance may be equivocal [15, 23, 24, 27–29, 32].

The current study showed that IBK was able to detect stressed subjects with 88, 100, 94% of sensitivity, specificity and accuracy respectively, using short HRV features (MeanNN, StdHR and HF). IBK was the most recurrent machine learning used among the papers identified in the existing literature [23, 24, 29].

Finally, it is useful to mention that the proposed methodology could be used in any application aiming to automatically detect a condition using ultra-short HRV features. In particular, the proposed method can improve the identification of the minimal length of HRV excerpts enabling the detection of an anomaly in real time.

## Conclusion

Currently, 5-min recordings are regarded as being an appropriate option for HRV analysis to detect mental stress in healthy subjects. However, the continued rise in the interest of everyday wearable devices being able to instantaneously assess mental stress level is rising the attention of the scientific community around the use of RR interval shorter than 5 min.

This study demonstrates that not all the ultra-short HRV features are good surrogates of short term ones. In fact, only six ultra-short HRV features resulted to be good surrogates of short term ones: MeanNN, StdNN, MeanHR, StdHR, HF, and SD2. Those six features displayed consistency across all the excerpt lengths (i.e., from 5 to 1 min) and MeanNN, StdHR and HF showed good performance if employed in a well-dimensioned automatic classifier.

Moreover, an automatic classifier based on IBK is able to detect stressed subjects with very high performances, using 3-min HRV analysis, and relatively good performances using 1-min HRV excerpts. The former achieved sensitivity, specificity and accuracy of 94, 94 and 94% respectively and the latter achieved sensitivity, specificity and accuracy of 82, 94 and 88% respectively.

Therefore, we conclude that it is possible to automatically detect mental stress using ultra-short HRV features with excerpts not shorter than 1 min. According to the specific application, 3- or 2-min excerpts could be preferable, because features having a clear physiological significance (e.g., HF and LF) remain computable.

## Endnotes

[1]Supplementary materials are available in the supporting document

[2]Supplementary materials are available in the supporting document

## Additional files

**Additional file 1: Table S1-S6.** HRV features and their distributions. HRV feature distributions at rest and stress and different time scales (i.e. 5 min, 3 min, 2 min, 1 min, and 30s). (PDF 307 kb)

**Additional file 2: Figure S1.** Bland-Altman Plot of MeanNN during Rest. Bland-Altman Plot of MeanNN during Rest. (PDF 136 kb)

**Additional file 3: Figure S2.** Bland-Altman Plot of MeanNN during Stress. Bland-Altman Plot of MeanNN during Stress. (PDF 207 kb)

### Abbreviations
ACC: Accuracy; AUC: Area under the curve; C4.5: Decision trees classifier; ECG: Electrocardiogram; HRV: Heart rate variability; IBK: K-nearest neighbors classifier; LDA: Linear discriminant analysis classifier; MLP: Multilayer perceptron classifier; Rho: Spearman's correlation coefficient; SEN: Sensitivity; SPE: Specificity; SVM: Support vector machine classifier

### Availability of data and materials
RR raw data are available on request.

### Authors' contributions
RC and LP conceived this study. RC and LM drafted methods and presented the results. PM and LP previously acquired the signals. RC, PM, CJ and LP discussed the results considering the state of the art of the literature. RC, CJ and LP made substantial contributions to interpretation of the results. RC, LM, PM, CJ and LP drafted and reviewed the manuscript critically. All the authors contributed to the paper. All authors read and approved the final manuscript.

### Ethics approval and consent to participate
This study was approved by the local Ethic Committee at the University of Naples Federico II. All of the participants signed specific informed consent form before the acquisitions.

### Consent for publication
All of the participants gave written informed consent for publication.

### Competing interests
The authors declare that they have no competing interests.

## Publisher's Note
Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

### Author details
[1]School of Engineering, University of Warwick, CV47AL, Coventry, UK.
[2]Institute of Advanced Studies, University of Warwick, CV47AL, Coventry, UK.
[3]Multidisciplinary Department of Medical, Surgical and Dental Sciences, University of Campania Luigi Vanvitelli, Naples, Italy.

## References
1.  R. J. Gerrig and P. G. Zimbardo, American Psychological Association: Glossary of Psychological Terms: Pearson Education, Education, Incorporated (COR), 2002.
2.  Lazarus RS, Folkman S, Stress A. Coping. In: Stress and emotion: A new synthesis, vol. 5; 1984. p. 101–26.
3.  Folkman S. Stress: appraisal and coping: Springer; 2013.
4.  Castaldo R, Melillo P, Bracale U, Caserta M, Triassi M, Pecchia L. Acute mental stress assessment via short term HRV analysis in healthy adults: A systematic review with meta-analysis. Biomedical Signal Processing and Control. 2015;18:370–7.
5.  Hernando A, Lázaro J, Gil E, Arza A, Garzón JM, López-Antón R, et al. Inclusion of respiratory frequency information in heart rate variability analysis for stress assessment. IEEE journal of biomedical and health informatics. 2016;20:1016–25.
6.  Teisala T, Mutikainen S, Tolvanen A, Rottensteiner M, Leskinen T, Kaprio J, et al. Associations of physical activity, fitness, and body composition with heart rate variability–based indicators of stress and recovery on workdays: a cross-sectional study. Journal of Occupational Medicine and Toxicology. 2014;9:16.
7.  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).
8.  Lacey K, Zaharia M, Griffiths J, Ravindran A, Merali Z, Anisman HJP. A prospective study of neuroendocrine and immune alterations associated with the stress of an oral academic examination among graduate students, vol. 25; 2000. p. 339–56.
9.  Tharion E, Parthasarathy S, Neelakantan NJNMJI. Short-term heart rate variability measures in students during examinations. Natl Med J India. 2009; 22:63–6.
10. 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," Biomed Eng Online, vol. 10, pp. 1–13, 2011/11/07 2011.
11. Minakuchi E, Ohnishi E, Ohnishi J, Sakamoto S, Hori M, Motomura M, et al. Evaluation of mental stress by physiological indices derived from finger plethysmography. J Physiol Anthropol. 2013;32:17.
12. Yang Z, Jia W, Liu G, Sun M. Quantifying mental arousal levels in daily living using additional heart rate. Biomedical Signal Processing and Control. 2017; 33:368–78.
13. Castaldo R, Montesinos L, Wan TS, Serban A, Massaro S, Pecchia L. Heart rate variability analysis and performance during a repeated mental workload task. In: EMBEC & NBC 2017: Springer; 2017. p. 69–72.
14. Sun Y, Yu XB. An innovative nonintrusive driver assistance system for vital signal monitoring. IEEE journal of biomedical and health informatics. 2014; 18:1932–9.
15. Xu Q, Nwe TL, Guan C. Cluster-based analysis for personalized stress evaluation using physiological signals. IEEE journal of biomedical and health informatics. 2015;19:275–81.
16. Force T. Heart rate variability guidelines: standards of measurement, physiological interpretation, and clinical use. Eur Heart J. 1996;17:354–81.
17. Wang Z, Parnandi A, Gutierrez-Osuna R. BioPad: leveraging off-the-shelf video games for Stress self-regulation. IEEE Journal of Biomedical and Health Informatics. 2017;22:47–55.
18. Delaney J, Brodie D. Effects of short-term psychological stress on the time and frequency domains of heart-rate variability. Percept Mot Skills. 2000;91: 515–24.
19. Salai M, Vassányi I, Kósa I. Stress detection using low cost heart rate sensors. Journal of Healthcare Engineering. 2016;2016.
20. Munla N, Khalil M, Shahin A, Mourad A. Driver stress level detection using HRV analysis. In: Advances in biomedical engineering (ICABME), 2015 international conference on; 2015. p. 61–4.
21. Athavale Y, Krishnan S. Biosignal monitoring using wearables: observations and opportunities. Biomedical Signal Processing and Control. 2017;38:22–33.
22. Pecchia L, Castaldo R, Montesinos L, Melillo P. Are ultra-short heart rate variability features good surrogates of short term ones? Literature review and method recommendations. Healthcare technology letters. 2018;5:94–100.
23. Pandey P, Lee EK, Pompili D. A distributed computing framework for real-time detection of Stress and of its propagation in a team. IEEE journal of biomedical and health informatics. 2016;20:1502–12.
24. Choi J, Gutierrez-Osuna R. "Using heart rate monitors to detect mental stress," in Wearable and implantable body sensor networks, 2009. In: BSN 2009. Sixth international workshop on; 2009. p. 219–23.
25. Castaldo R, Montesinos L, Melillo P, Massaro S, Pecchia L. To what extent can we shorten HRV analysis in wearable sensing? A case study on mental Stress detection. In: EMBEC & NBC 2017: Springer; 2017. p. 643–6.

26. Boonnithi S, Phongsuphap S. Comparison of heart rate variability measures for mental stress detection. In: Computing in cardiology, 2011; 2011. p. 85–8.

27. D. Brisinda, A. Venuti, C. Cataldi, K. Efremov, E. Intorno, and R. Fenici, "Real-time imaging of Stress-induced cardiac autonomic adaptation during realistic Force-on-force police scenarios," Journal of Police and Criminal Psychology, pp. 1–16, 2014/01/31 2014.

28. Mayya S, Jilla V, Tiwari VN, Nayak MM, Narayanan R. Continuous monitoring of stress on smartphone using heart rate variability. In: Bioinformatics and bioengineering (BIBE), 2015 IEEE 15th international conference on; 2015. p. 1–5.

29. Wijsman J, Grundlehner B, Liu H, Hermens H, Penders J. Towards mental stress detection using wearable physiological sensors. In: Engineering in Medicine and Biology Society, EMBC. Annual International Conference of the IEEE; 2011. p. 1798–801.

30. Kim D, Seo Y, Cho J, Cho C-H. Detection of subjects with higher self-reporting stress scores using heart rate variability patterns during the day. In: Engineering in medicine and biology society EMBS 2008. 30th annual international conference of the IEEE; 2008. p. 682–5.

31. Salahuddin L, Cho J, Jeong MG, Kim D. Ultra short term analysis of heart rate variability for monitoring mental stress in mobile settings. Conf Proc IEEE Eng Med Biol Soc. 2007;2007:4656–9.

32. Sun F-T, Kuo C, Cheng H-T, Buthpitiya S, Collins P, Griss M. Activity-aware mental stress detection using physiological sensors. In: International conference on Mobile computing, applications, and services; 2010. p. 211–30.

33. Esco MR, Flatt AA. Ultra-short-term heart rate variability indexes at rest and post-exercise in athletes: evaluating the agreement with accepted recommendations. J Sports Sci Med. 2014;13:535–41.

34. Nussinovitch U, Elishkevitz KP, Katz K, Nussinovitch M, Segev S, Volovitz B, et al. Reliability of ultra-short ECG indices for heart rate variability. Ann Noninvasive Electrocardiol. 2011;16:117–22.

35. McNames J, Aboy M. Reliability and accuracy of heart rate variability metrics versus ECG segment duration. Med Biol Eng Comput. 2006;44:747–56.

36. Baek HJ, Cho C-H, Cho J, Woo J-M. Reliability of ultra-short-term analysis as a surrogate of standard 5-min analysis of heart rate variability. Telemedicine and e-Health. 2015;21:404–14.

37. Thong T, Li K, McNames J, Aboy M, Goldstein B. Accuracy of ultra-short heart rate variability measures. In: Engineering in medicine and biology society, 2003. Proceedings of the 25th annual international conference of the IEEE; 2003. p. 2424–7.

38. Nardelli M, Greco A, Bolea J, Valenza G, Scilingo EP, Bailon R. Reliability of lagged Poincaré plot parameters in ultra-short heart rate variability series: application on affective sounds. IEEE Journal of Biomedical and Health Informatics. 2018;22:741–9.

39. Munoz ML, van Roon A, Riese H, Thio C, Oostenbroek E, Westrik I, et al. Validity of (ultra-) short recordings for heart rate variability measurements. PLoS One. 2015;10:e0138921.

40. Schroeder EB, Whitsel EA, Evans GW, Prineas RJ, Chambless LE, Heiss G. Repeatability of heart rate variability measures. J Electrocardiol. 2004;37:163–72.

41. Flatt AA, Esco MR. Validity of the ithleteTM smart phone application for determining ultra-short-term heart rate variability. Journal of human kinetics. 2013;39:85–92.

42. Kwon S, Lee D, Kim J, Lee Y, Kang S, Seo S, et al. Sinabro: A smartphone-integrated opportunistic electrocardiogram monitoring system. Sensors. 2016;16:361.

43. Bernardi L, Wdowczyk-Szulc J, Valenti C, Castoldi S, Passino C, Spadacini G, et al. Effects of controlled breathing, mental activity and mental stress with or without verbalization on heart rate variability. J Am Coll Cardiol. 2000;35:1462–9.

44. Schreck DM, J RD, A T, Fishberg JOEM. Derivation of the 12-lead electrocardiogram and 3-lead vectorcardiogram. Am J Emerg Med. 2013;31:1183–90.

45. Sato N, Miyake S, Akatsu J i, Kumashiro M. Power spectral analysis of heart rate variability in healthy young women during the normal menstrual cycle. Psychosom Med. 1995;57:331–5.

46. Ferré S, O'Brien MC. Alcohol and caffeine: the perfect storm. Journal of caffeine research. 2011;1:153–62.

47. Goldberger AL, Amaral LA, Glass L, Hausdorff JM, Ivanov PC, Mark RG, et al. Physiobank, physiotoolkit, and physionet components of a new research resource for complex physiologic signals. Circulation. 2000;101:e215–20.

48. Zong W, Moody GB, Jiang D. A robust open-source algorithm to detect onset and duration of QRS complexes. presented at the Computers in Cardiology. 2003;2003.

49. G. B. Moody, "Evaluating ECG analyzers," WFDB Applications Guide, 2003.

50. Asyali M. Discrimination power of long-term heart rate variability measures. In: Engineering in medicine and biology society, 2003. Proceedings of the 25th annual international conference of the IEEE; 2003. p. 200–3.

51. Tarvainen MP, Niskanen J-P, Lipponen JA, Ranta-Aho PO, Karjalainen PA. Kubios HRV–heart rate variability analysis software. Comput Methods Prog Biomed. 2014;113(1):210–20.

52. Yentes JM, Hunt N, Schmid KK, Kaipust JP, McGrath D, Stergiou N. The appropriate use of approximate entropy and sample entropy with short data sets. Ann Biomed Eng. 2013;41:349–65.

53. Castaldo R, Montesinos L, Pecchia L. Ultra-short entropy for mental Stress detection. In: World Congress on Medical Physics and Biomedical Engineering, vol. 2019; 2018. p. 287–91.

54. Castaldo R, Melillo P, Pecchia L. Acute mental Stress detection via ultra-short term HRV analysis. In: World Congress on Medical Physics and Biomedical Engineering. Toronto: Canada; 2015. p. 1068–71.

55. Castaldo R, Melillo P, Izzo R, Luca ND, Pecchia L. Fall prediction in hypertensive patients via short-term HRV analysis. IEEE Journal of Biomedical and Health Informatics. 2017;21:399–406.

56. Bland JM, Altman D. Statistical methods for assessing agreement between two methods of clinical measurement. Lancet. 1986;327:307–10.

57. Foster KR, Koprowski R, Skufca JD. Machine learning, medical diagnosis, and biomedical engineering research-commentary. Biomed Eng Online. 2014;13:10.1186.

58. Vapnik VN. Statistical learning theory. New York: Wiley; 1998.

59. Bishop CM. Neural networks for pattern recognition. In: Oxford: Clarendon press ; Oxford University press; 1995.

60. Wu X, Kumar V, Quinlan JR, Ghosh J, Yang Q, Motoda H, et al. Top 10 algorithms in data mining. Knowl Inf Syst. 2008;14:1–37.

61. Quinlan JR. C4.5: programs for machine learning. San Mateo, Calif: Morgan Kaufmann Publishers; 1993.

62. Krzanowski W. Principles of multivariate analysis. Oxford: OUP; 2000.

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

64. S. J. I. J. O. E. T. Karsoliya and Technology. Approximating number of hidden layer neurons in multiple hidden layer BPNN architecture. 2012; 3:714–7.

65. A. B. Hassanat, M. A. Abbadi, G. A. Altarawneh, and A. A. J. A. P. A. Alhasanat, "Solving the problem of the K parameter in the KNN classifier using an ensemble learning approach," 2014.

66. Kohl M. Performance measures in binary classification. International Journal of Statistics in Medical Research. 2012;1:79–81.

67. 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;17:727–33.

68. Pecchia L, Melillo P, Sansone M, Bracale M. Heart rate variability in healthy people compared with patients with congestive heart failure. In: 2009 9th international conference on information technology and applications in biomedicine; 2009. p. 1–4.

69. Pereira T, Almeida PR, Cunha JP, Aguiar A. Heart rate variability metrics for fine-grained Stress level assessment. Comput Methods Prog Biomed. 2017; 148:71–80.

70. Rice WR. Analyzing tables of statistical tests. Evolution. 1989;43:223–5.

71. Fleming TR, DeMets DL. Surrogate end points in clinical trials: are we being misled? Ann Intern Med. 1996;125:605–13.

72. Traina M, Galullo F, Russo G. Effects of anxiety due to mental stress on heart rate variability in healthy subjects. Minerva Psichiatr. 2011;52:227–31.