Computers in Biology and Medicine 112 (2019) 103381

Contents lists available at ScienceDirect

Computers in Biology and Medicine

journal homepage: www.elsevier.com/locate/compbiomed

Detection of major depressive disorder from linear and nonlinear heart rate variability features during mental task protocol

Sangwon Byun[a,*], Ah Young Kim[b], Eun Hye Jang[b], Seunghwan Kim[b], Kwan Woo Choi[c,d], Han Young Yu[b,**], Hong Jin Jeon[c,***]

[a] Department of Electronics Engineering, Incheon National University, 22012, Incheon, South Korea
[b] Bio-Medical IT Convergence Research Division, Electronics and Telecommunications Research Institute (ETRI), 34129, Daejeon, South Korea
[c] Department of Psychiatry, Depression Center, Samsung Medical Center, Sungkyunkwan University School of Medicine, 06351, Seoul, South Korea
[d] Department of Psychiatry, Korea University Anam Hospital, Korea University College of Medicine, Seoul, 02841, South Korea

ARTICLE INFO

Keywords:
Heart rate variability (HRV)
Major depressive disorder (MDD)
Machine learning
Depression
Feature selection
Support vector machine (SVM)
Recursive feature elimination (RFE)
Mental task
Autonomic nervous system (ANS)

ABSTRACT

Background: Major depressive disorder (MDD) is one of the leading causes of disability; however, current MDD diagnosis methods lack an objective assessment of depressive symptoms. Here, a machine learning approach to separate MDD patients from healthy controls was developed based on linear and nonlinear heart rate variability (HRV), which reflects the autonomic cardiovascular regulation.

Methods: HRV data were collected from 37 MDD patients and 41 healthy controls during five 5-min experimental phases: the baseline, a mental stress task, stress recovery, a relaxation task, and relaxation task recovery. The experimental protocol was designed to assess the autonomic responses to stress and recovery. Twenty HRV indices were extracted from each phase, and a total of 100 features were used for classification using a support vector machine (SVM). SVM-recursive feature elimination (RFE) and statistical filter were employed to perform feature selection.

Results: We achieved 74.4% accuracy, 73% sensitivity, and 75.6% specificity with two optimal features selected by SVM-RFE, which were extracted from the stress task recovery and mental stress phases. Classification performance worsened when individual phases were used separately as input data, compared to when all phases were included. The SVM-RFE using nonlinear and Poincaré plot HRV features performed better than that using the linear indices and matched the best performance achieved by using all features.

Conclusions: We demonstrated the machine learning-based diagnosis of MDD using HRV analysis. Monitoring the changes in linear and nonlinear HRV features for various autonomic nervous system states can facilitate the more objective identification of MDD patients.

1. Introduction

Major depressive disorder (MDD) is characterized by persistent irritability, fatigue, altered appetite, insomnia, and physical aches [1,2]. Furthermore, the World Health Organization (WHO) predicts that depression will be the most common disease in the world by 2020 [1]. Untreated depression increases the mortality rate and may cause suicidal behavior, which is a serious public health problem [3]. Severe symptoms can also disrupt work performance, thereby generating significant economic and social burdens [4]. Therefore, it is important to accurately diagnose and treat patients in the early stages of depression.

To date, clinicians have been following the guidelines provided by the Diagnostic and Statistical Manual of Mental Disorders (DSM) to diagnose depression [5]. However, this conventional method mainly relies on clinical interviews and subjective self-reporting with test scales, and an objective and systematic method of assessing the biological aspects of depression is lacking [6]. Moreover, major depression is a heterogeneous syndrome, and patients may exhibit multiple sources of variance, rendering the correct diagnosis difficult. In previous studies, even highly trained clinicians agreed on MDD diagnoses in only 4–15% of the cases [7–9]. As a result, researchers have been working to develop a more reliable method of diagnosing depression.

* Corresponding author. Department of Electronics Engineering, Incheon National University, Incheon, South Korea.
** Corresponding author. Bio-Medical IT Convergence Research Division, Electronics and Telecommunications Research Institute (ETRI), Daejeon, South Korea.,
*** Corresponding author. Department of Psychiatry, Depression Center, Samsung Medical Center, Sungkyunkwan University School of Medicine, Seoul, South Korea.
E-mail addresses: swbyun@inu.ac.kr (S. Byun), uhan0@etri.re.kr (H.Y. Yu), jeonhj@skku.edu (H.J. Jeon).

https://doi.org/10.1016/j.compbiomed.2019.103381
Received 15 February 2019; Received in revised form 2 August 2019; Accepted 3 August 2019

Heart rate variability (HRV), which has been used to evaluate cardiac autonomic regulation, has attracted much attention in connection with depression. As previous studies have suggested, MDD is associated with a dysregulated autonomic nervous system (ANS) [10,11]. The HRV analysis can be used as a non-invasive tool to evaluate the autonomic modulation of the cardiovascular system and investigate various psychiatric disorders, such as bipolar disorder, anxiety disorder, schizophrenia, and posttraumatic stress disorder [11–14]. In general, the resting HRV is reduced in MDD patients, indicating a sympathovagal imbalance toward sympathetic dominance [10,11,15,16]. A recent meta-analysis also suggests that MDD is associated with reduced HRV, and that HRV decreases with increasing depression severity [17].

Based on these findings, recent studies have developed an automated diagnosis system for MDD using HRV features and machine learning methods [18–22]. For example, Zhang et al. [18] used a neuro-fuzzy approach to differentiate 10 MDD patients from 10 controls and achieved 95% accuracy. Matsui et al. [20] classified 13 MDD patients and 28 controls with 88% accuracy using linear discriminant analysis. Roh et al. [19] applied a support vector machine (SVM) algorithm to HRV features obtained from 23 subjects and classified the subjects' depression scale with 71% accuracy. However, these studies relied on a small number of samples [18,19], lack descriptions of the cross-validation method [18–21], or depended entirely on linear HRV features, which were calculated in the traditional time and frequency domains [18,20,21]. As heart rate fluctuation is a complex behavior originating from nonlinear regulatory processes, adaptation of nonlinear dynamics and information theory for HRV analysis has been suggested [23–25]. Notably, Schulz et al. [26] have shown that nonlinear HRV indices allow more reliable discrimination of MDD patients from controls than linear HRV features, as the latter exhibit high inter-subject variability. Kuang et al. [22] used a 10-fold cross-validation to evaluate the feature prediction performance and extracted 16 features, which included linear and nonlinear HRV features, but only female subjects participated in that study. To date, HRV-based automated recognition has not been comprehensively studied for major depression as it has for other diseases, such as cardiovascular diseases, bipolar disorder, and diabetes [27–29].

In this study, we aim to investigate the feasibility of automated MDD detection based on linear and nonlinear HRV features using a classification and feature selection algorithm. We employed an SVM as the classifier and a leave-one-out (LOO) procedure as the validation method. A support vector machine–recursive feature elimination (SVM-RFE) and a statistical filter were used as feature selection methods. Twenty HRV features—13 linear, five nonlinear, and two Poincaré plot—were extracted from electrocardiogram (ECG) recordings. In addition to the baseline measurement, we adopted laboratory stress testing to examine ANS reactivity and recovery to stimuli; this approach may improve the discrimination power of the HRV features [21]. For example, previous studies have reported impaired autonomic reactivity responses to external stimuli such as emotion and stress, manifesting as attenuated or exaggerated HRV reactivity [30–34]. In another study, MDD patients exhibited weaker stress recovery than control subjects [35].

In the present study, HRVs were measured during five experimental phases: the baseline, a mental stress task, stress recovery, a relaxation task, and relaxation task recovery (Fig. 1A), with the expectation that the incorporation of multiple alterations in the ANS activity would improve the prediction performance [36]. A total of 100 HRV features (20 features from five phases) were used as input data. We examined the effect of multiple autonomic alterations on the differentiation of depressive symptoms. Finally, we evaluated the performance of the SVM classifier when linear and nonlinear features were used separately as input data.

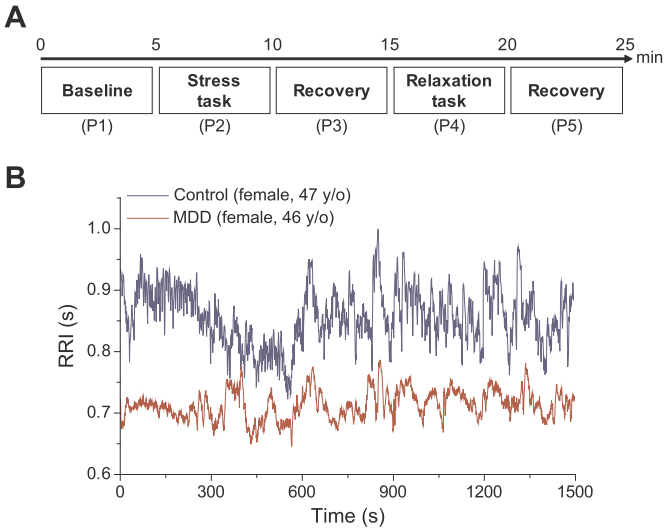

Fig. 1. A. Experimental protocol for ECG measurement. ECG readings were obtained during five consecutive phases, each of which had 5 min duration. **B. Sample RRI data from the measurement.** RRIs extracted from a healthy control (female, 47-years-old) and an MDD patient (female, 46-years-old) are shown.

2. Materials and methods

2.1. Participants

All subjects were recruited at the Samsung Medical Center in Seoul, Korea. The participants consisted of 37 MDD patients and 41 healthy controls, who were matched for age and gender. A board-certified psychiatrist evaluated the patients based on the DSM-IV criteria to verify those experiencing current major depressive episodes. The depressive episode severity was assessed based on the 17-item Hamilton depression rating score (HAMD) [37]. The exclusion criteria for both the control and MDD groups included any psychotic disorders, such as schizophrenia or delusional disorder; bipolar affective disorders; delirium; dementia; mental retardation; personality disorders; a serious risk of suicidal act; neurological illnesses or brain damage; alcohol/substance abuse or dependence; or serious medical conditions such as cardiovascular disease or cancer. Healthy subjects with no history of psychiatric disorder or significant medical conditions were recruited as the control group. The subjects were instructed to avoid substances affecting cardiovascular activity, such as alcohol and caffeine, before the examination. All the patients received standard pharmacotherapy using antidepressant medications, including selective serotonin reuptake inhibitors (SSRIs), serotonin norepinephrine reuptake inhibitors (SNRIs), norepinephrine dopamine reuptake inhibitors, and tricyclic antidepressants (TCAs) [38]. The experiment was explained to all subjects, who provided signed informed consent to participate in the study. In addition, $50 compensation was paid to each subject. This study was approved by the Institutional Review Board of the Samsung Medical Center, Seoul, Korea (No. 2015-07-151) and performed according to the relevant guidelines.

2.2. Experimental procedure

All participants were welcomed into a room in which the temperature was maintained at 23 °C, and the humidity was below 50%. Because the temperature, humidity, and atmospheric pressure are known to affect the HRV measurements [39–41], the environmental

factors were maintained constant for all the participants. Therefore, we assume that the measurement artifact caused by the fluctuations in these environmental cues was negligible. The subjects were asked to sit in a comfortable armchair with a headrest. While attaching the electrodes for the physiological sensors, the experimenter explained the measurement procedure to the subjects in detail. Then, the subjects were allowed to adapt to the laboratory environment. The experimental protocol used in this study was designed to evaluate the autonomic responses to mental stress and relaxation tasks (Fig. 1A). The entire procedure consisted of five experimental phases: the baseline (P1), mental stress task (P2), stress task recovery (P3), relaxation task (P4), and relaxation task recovery (P5) phases. Each phase lasted for 5 min. During P1, the subjects were instructed to rest while minimizing their movement to facilitate the measurement of the ANS basal activity. During P2, a mental arithmetic test was conducted to induce stress in the subjects [42]. During this test, the subjects were required to subtract serial 7's beginning from 500 and to verbally report answers. If a subject made an error, the experimenter asked the subject to repeat that calculation. During P3, the subjects were asked to relax as much as possible to allow autonomic recovery from the mental stress of the previous task. Then, during P4, the subjects were asked to relax while watching 10 consecutive images of natural scenery, which were presented on a PC monitor. Each image lasted 30 s. In the fifth and final phase (P5), the image presentation ceased, and the subjects were asked to rest to recover from the relaxation task.

2.3. Physiological measurement

ECG data were recorded using the ProComp Infiniti system (SA7500, Thought Technology, Canada). Three disposable Ag/AgCl electrodes were positioned in a lead II configuration. The ECG signal was amplified, band-pass filtered, and sampled at 256 Hz. R-peak to R-peak (RR) interval data from the ECG signals were analyzed using Kubios HRV Premium software (Kubios, www.kubios.com), which detects R-peaks using an in-house-developed QRS detection algorithm based on the Pan-Tompkins method [43,44]. All the RR interval (RRI) data were visually inspected, and artifacts were corrected using a piecewise cubic spline interpolation method. The entire analysis was conducted by the same operator to ensure consistency. Then, the HRV features were separately calculated from the RRI data of the individual phases. Fig. 1B shows sample RRIs extracted from a healthy control (female, 47-years-old) and an MDD patient (female, 46-years-old).

2.4. HRV features

A standard HRV analysis based on international guidelines was conducted [45,46]. Twenty HRV features were extracted from the RRI data of each phase, using time, frequency, and nonlinear domain analyses (Table 1). A total of 100 HRV features (i.e., 20 features from five phases) were extracted from the entire ECG recording for a single subject.

Time-domain HRV features can be calculated directly from the time series of RRIs. We extracted six features using the time domain analysis: the mean of the RRIs, the standard deviation of the RRIs (SDNN), the root mean square of successive RR interval differences (RMSSD), the percentage of successive RRIs differing by more than 50 m s (pNN50), the integral of the histogram of the RRI divided by its height (TRI), and the baseline width of the RRI histogram (TINN).

Seven features were calculated based on the frequency domain analysis. The RRI data were converted to equidistantly sampled data using cubic spline interpolation (4 Hz). Then, a power spectral density (PSD) was estimated for the RRI data using a Welch's periodogram-based fast Fourier transform (FFT). Absolute powers in very low-frequency (VLF, 0–0.04 Hz), low-frequency (LF, 0.04–0.15 Hz), and high-frequency (HF, 0.15–0.4 Hz) bands were calculated. Relative powers of the LF and HF bands in normalized units and the LF/HF power ratio were also calculated. The absolute powers were expressed by their natural logarithms to reduce skewness in the distribution. A wavelet transform is another method for frequency domain analysis, which has been proposed to overcome the stationary assumption of the Fourier transform [47,48]. However, a previous study demonstrated that differences in HRV values calculated using the Fourier and wavelet transform methods were very small for both stationary and non-stationary data, suggesting that the wavelet method is only superior when additional analyses in the time-frequency domain are required [49]. In addition, since the Fourier transform has been primarily implemented for the frequency domain analysis in previous studies [45], we selected this method for the current study.

Table 1
HRV features used in the present study. Twenty HRV features were extracted from each phase.

Domain	Parameter	Unit	Description
Time	RRI	ms	Mean of RR intervals
	SDNN	ms	Standard deviation of RR intervals
	RMSSD	ms	Root mean square of successive RR interval differences
	pNN50	%	Percentage of successive RR intervals differing more than 50 m s
	TRI		Integral of the histogram of the RR interval divided by its height
	TINN	ms	Baseline width of the RR interval histogram
Frequency	logVLF	s^2	Log of absolute power in the VLF band (0–0.04 Hz)
	logLF	s^2	Log of absolute power in the LF band (0.04–0.15 Hz)
	LFnu	nu	Relative power of the LF band
	logHF	s^2	Log of absolute power in the HF band (0.15–0.4 Hz)
	HFnu	nu	Relative power of the HF band
	LF/HF		Ratio between LF and HF band powers
	logTot	s^2	Log of total power
Nonlinear	ApEn		Approximate entropy
	SampEn		Sample entropy
	α1		Short term fluctuation slope of DFA
	α2		Long term fluctuation slope of DFA
	CorDim		Correlation dimension
Poincaré plot	SD1	ms	Standard deviation of the Poincaré plot
	SD2	ms	perpendicular to (SD1) and along (SD2) the line of identity

Five nonlinear measures were extracted to evaluate nonlinear dynamics in heart rate signals. Approximate entropy (ApEn), first suggested by Pincus to measure irregularity in relatively short and noisy time-series data, does not involve any assumptions concerning the underlying system dynamics [24]. Higher values of ApEn indicate that the data are more irregular and unpredictable. In the present study, the embedding dimension and the tolerance value for the ApEn calculation were set to 2 and 0.2, respectively. Sample entropy (SampEn) was developed to reduce bias in ApEn, which is caused by self-comparison [50]. SampEn can be used and interpreted in the same manner as ApEn but is more reliable than ApEn for shorter time series. The parameters for SampEn were set to the same values as those for the ApEn calculation. Detrended fluctuation analysis (DFA) was proposed by Peng et al. [51] to assess fractal scaling properties of short-term RRI signals. In DFA, time-series data are integrated and detrended, and then a root-mean-square fluctuation is measured at a different time scale. The fluctuation is defined by α1 and α2, which represent short-range and long-range correlation, respectively. In this study, α1 was evaluated between the data length of 4 and 16, and α2 was evaluated between the length of 17 and 64. The correlation dimension (CorDim) was developed by Grassberger and Procaccia to estimate the number of independent variables required to model the underlying system in the signals [52]. CorDim is defined as the slope of the linear region of a correlation integral plot. Higher CorDim values indicate a greater number

of variables required to predict the signal and hence greater complexity [53].

Two features were extracted from the Poincaré plot analysis. The Poincaré plot graphically represents the correlation between successive RR intervals, in which each RR interval is plotted against the next interval. SD1 and SD2 represent the standard deviations of the Poincaré plot perpendicular to and along the line of identity, respectively.

Although SampEn is basically ApEn without the bias, ApEn was included in the feature set because previous studies on HRV in MDD patients have suggested that ApEn may serve as an important feature in classifying MDD and control groups [59,65,66]. In particular, ApEn was significantly reduced in MDD patients compared to healthy controls [59,65]. Valenza et al. [66] suggested that ApEn, not SampEn, was significantly reduced in MDD patients during an emotional task. These studies suggest that including ApEn in the feature set would be beneficial for classifying the MDD and control subjects. In addition, SampEn showed mixed results in differentiating MDD patients: SampEn showed a significant correlation with geriatric depression scale, but some studies suggested a non-significant relationship between SampEn and MDD [67–69]. Therefore, ApEn was included as a complementary role for SampEn because ApEn might provide additional information on depression symptoms.

SD1 and SD2 from the Poincaré plot can be obtained by combining the time-domain features, RMSSD and SDNN, suggesting that they may not provide indices that are independent of linear features. Indeed, SD1 is equivalent to RMSSD, and SD2 is strongly correlated with SDNN [54–56]. However, the metrics from the Poincaré plot have been primarily considered as nonlinear indices [46] and reported in previous studies on nonlinear biomarkers of psychiatric disorders [57]. For example, SD1 and SD2 have been reported as nonlinear HRV metrics in depressed patients [26,58,59]. In some studies, SD1 and SD2 were reported with their time-domain counterparts, RMSSD and SDNN [22,26]. SD1 and SD2 were also used as nonlinear input features in machine learning-based diagnoses of various medical conditions, including MDD, seizure, and cardiovascular diseases [22,60–64]. All these studies also included RMSSD and SDNN as linear input features. Therefore, despite the potential risk of redundant data usage, we included SD1 and SD2 as separate HRV metrics to compare our results to those in previous studies.

2.5. Statistical analyses

Statistical analyses were performed using MATLAB R2018a and SPSS 25 (SPSS Inc., Chicago, IL, USA). As the age, body mass index (BMI), and HAMD factors were not normally distributed, these factors were compared between the MDD and control groups using the Mann–Whitney *U* test. The sex and marital status were compared between the two groups using chi-square tests. The HRV features from the two groups were compared using the Mann–Whitney *U* test. A total of 100 HRV features were compared because there were 20 features extracted from each of the five phases. To control the type I error from multiple comparisons, *p*-values were adjusted for the false discovery rate (FDR) using the Benjamini–Hochberg method at the 0.05 significance level [70,71]. In all statistical tests, a *p*-value less than or equal to 0.05 was considered to indicate significance.

2.6. Leave-one-out procedure and feature selection

Fig. 2 shows an overview of the data processing pipeline. We adopted the LOO method to develop and evaluate a classifier. The entire dataset was split into two subsets: a test set (one subject) and a training set (the remaining 77 subjects). The training set was used to train the classifiers and assess feature rankings. The test set was used as unseen data to evaluate the performance of the classifier developed. During the entire LOO procedure, the HRV data were normalized by subtracting the median and dividing it by the mean absolute deviation (MAD).

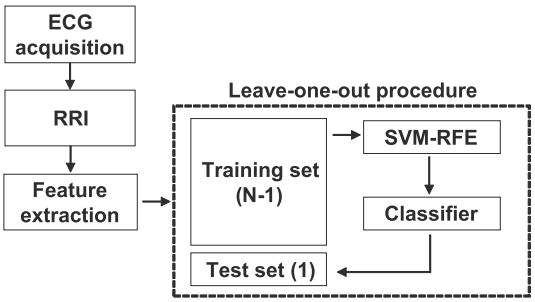

Fig. 2. Overview of data processing pipeline.

The LOO was chosen as a cross-validation method because it has been reported to perform well when the sample size is small (< 100), compared to other methods, such as 10-fold, 5-fold, and repeated *k*-fold [72]. Leave-*k*-out (LKO) cross-validation is the general case where *k* samples are selected as the test set. LKO has been shown to exhibit superior performance compared to LOO [73]. However, this method requires the evaluation of all possible $\binom{n}{k}$ subsets, where *n* is the total number of samples, leading to considerable computational complexity. For these reasons, LOO-CV was used in this study.

We used two feature selection methods, SVM-RFE and a statistical filter method, to reduce the dimensionality of the feature space. An RFE is a wrapper method for feature selection that evaluates a selected subset of features based on their usefulness in predicting the class [74]. In the present study, the feature importance was determined by the weights from a linear SVM, and the least relevant features were eliminated one by one during each iteration (backward sequential selection). A linear SVM-RFE has been known to perform well when the number of samples is smaller than the number of features and has been applied to various types of selection problems, such as genes and gas sensor data [74,75]. A statistical filter method filters out redundant features using a statistical test between features without involving a learning algorithm. In this study, the feature ranking was determined during each iteration by the statistical comparison of HRV features between the control and MDD groups, using the Mann–Whitney *U* test. The features were ranked in ascending order of the *p*-values [71]. All analyses were performed using MATLAB R2018a (MathWorks, Natick, MA). The SVM-RFE algorithm is available online from Mathworks File Exchange [75]. In addition, a Library for Support Vector Machines (LIBSVM) was used to perform the feature selection [76].

2.7. Classification and performance evaluation

We implemented an SVM algorithm to classify the control and MDD groups using the HRV features as input data. An SVM is based on the margin maximization principle [77]. It maps input data into a high-dimensional feature space. An optimal separating hyperplane is constructed in this space to maximize the interclass distances [78]. In the present study, a linear kernel function was used for the classifier model. The regularization parameter *C* was optimized by performing a grid search in the range of 0.1–10.

We selected an SVM as the classification method because SVMs exhibit good generalization performance with high-dimensional data [79]. In addition, SVMs can reduce the risk of overfitting by choosing appropriate regularization parameters. To further prevent overfitting, we implemented a linear SVM model. Nonlinear kernel methods tend to show higher risk of overfitting because these types of algorithms have more flexibility in building a model, which results in falsely high training accuracy from an unrealistic model. This can be prevented by using a linear algorithm, such as a linear SVM.

The accuracy, sensitivity, specificity, positive predictive value (PPV), and negative predictive value (NPV) were used as performance

Table 2
Performance measure formulae.

Performance measure	Formula
Accuracy	$\frac{TP + TN}{TP + TN + FP + FN}$
Sensitivity	$\frac{TP}{TP + FN}$
Specificity	$\frac{TN}{TN + FP}$
Positive predictive value	$\frac{TP}{TP + FP}$
Negative predictive value	$\frac{TN}{TN + FN}$

Abbreviations: TP, true positive; TN, true negative; FP, false positive; FN, false negative.

measures (Table 2). The prediction performance of the SVM classifier was evaluated using the unseen test set. This process was repeated $N = 78$ times, and the results from all folds were averaged to obtain performance measure estimates. Cross-validation was implemented externally to the feature selection process to accurately assess the prediction performance. We also computed the median rank for each feature over all folds to determine the relative importance of the feature.

Table 3
Demographic and clinical characteristics of control and MDD groups. The factors of the two groups were compared using the Mann–Whitney *U* test (*$p < 0.05$), except for the sex and marital status, which were compared using chi-square tests (χ^2). The data are expressed as median and MAD for continuous variables and as count and percentage for categorical variables.

Factors	Control ($N = 41$)	MDD ($N = 37$)	*p*-value
Sex (%)			0.800 (χ^2)
M	11 (27%)	9 (24%)	
F	30 (73%)	28 (76%)	
Age (MAD)	35 (12)	40 (14)	0.948
BMI (MAD)	22.3 (2.56)	22.4 (2.4)	0.980
Marital status (%)			0.064 (χ^2)
Single	20 (49%)	16 (43%)	
Married	21 (51%)	15 (41%)	
Divorced	0	3 (7%)	
Bereavement	0	3 (7%)	
HAMD (MAD)	1 (1)	18 (2)	< 0.001*

For all classifications, we fixed the threshold (or cutoff) value to 0.5 to determine the positive and negative classes. The best performance and the optimal feature sets were determined using the accuracy as an evaluation measure. Therefore, we compared the classifier outputs based on different feature selection methods (e.g., SVM-RFE vs. statistical filter) and different sets of input data (e.g., linear vs. nonlinear/Poincaré features) using the same fixed threshold. After obtaining the best feature set, we applied a receiver operator characteristic (ROC) curve analysis to this optimal model. The area under the ROC curve (AUC) was also evaluated for each case.

3. Results

3.1. Demographic and clinical characteristics of subjects

The participants in the present study included 41 healthy controls (30 females) and 37 MDD patients (28 females). Descriptive statistics of the demographic and clinical characteristics of the participants are summarized in Table 3. No significant differences were observed in sex, age, BMI, or marital status data among the subjects. As expected, the MDD group exhibited a significantly higher HAMD than the control group.

3.2. Statistical comparison of HRV features

HRV features were statistically compared between the control and MDD groups using the Mann–Whitney *U* test. Among the 100 HRV features considered, 20 exhibited significant differences between the two groups (Table 4 and Fig. 3). Note that all these features were significantly decreased in the MDD group compared to the control. SDNN and SD2 were significantly different between the two groups in all phases except P2 (stress task). Descriptive statistics and the *p*-values of the entire set of HRV features are summarized in the Supplementary Information (Table S1).

3.3. Classification of control and MDD participants using HRV features

An SVM learning algorithm was implemented to classify the healthy and MDD participants using HRV features. A total of 100 features (i.e., 20 features from five phases) were used as input data. The SVM-RFE and the statistical filter methods were used to perform feature selection,

Table 4
HRV features that differ significantly between the control and MDD groups. The features were compared using Mann–Whitney U tests, and *p*-values were corrected for the FDR (*p*-FDR) using the Benjamini–Hochberg method. The data are expressed as median and MAD (* *p*-FDR < 0.05). Statistical comparisons of the entire set of HRV features are summarized in Table S1.

Phase	Feature	Control	MDD	*p*	*p*-FDR
P1 (baseline)	SDNN	33.2 (12)	25.5 (8.76)	0.006	0.032*
	SD2	45.3 (17.4)	34.3 (11.6)	0.005	0.032*
P3 (recovery from the stress task)	SDNN	42.1 (10.6)	27.1 (9.48)	< 0.001	0.004*
	RMSSD	27.2 (7.53)	19.2 (8.41)	0.001	0.010*
	pNN50	6.69 (5.83)	0.568 (0.568)	0.001	0.010*
	TRI	10.2 (2.3)	7.22 (2.44)	< 0.001	0.004*
	TINN	203 (42)	126 (36)	< 0.001	0.010*
	SD1	19.3 (5.34)	13.6 (5.96)	0.001	0.010*
	SD2	58.3 (14.3)	36.4 (12)	< 0.001	0.004*
	CorDim	1.67 (1.14)	0.523 (0.399)	0.001	0.010*
P4 (relaxation task)	SDNN	36.2 (11.7)	30.7 (10.3)	0.004	0.032*
	TRI	9.44 (2.46)	6.74 (2.16)	0.003	0.030*
	SD2	48.3 (15.3)	35.8 (15.1)	0.004	0.032*
P5 (recovery from the relaxation task)	SDNN	45.9 (9.97)	26.5 (8.25)	0.005	0.032*
	RMSSD	26.7 (8.48)	19.9 (8.46)	0.005	0.032*
	TRI	10.6 (2.64)	7.24 (1.93)	0.001	0.010*
	TINN	185 (55)	125 (40)	0.007	0.039*
	SD1	18.9 (6)	14.1 (6)	0.005	0.032*
	SD2	61.7 (14.8)	36.5 (10.7)	0.005	0.032*
	SampEn	1.42 (0.122)	1.33 (0.156)	0.008	0.042*

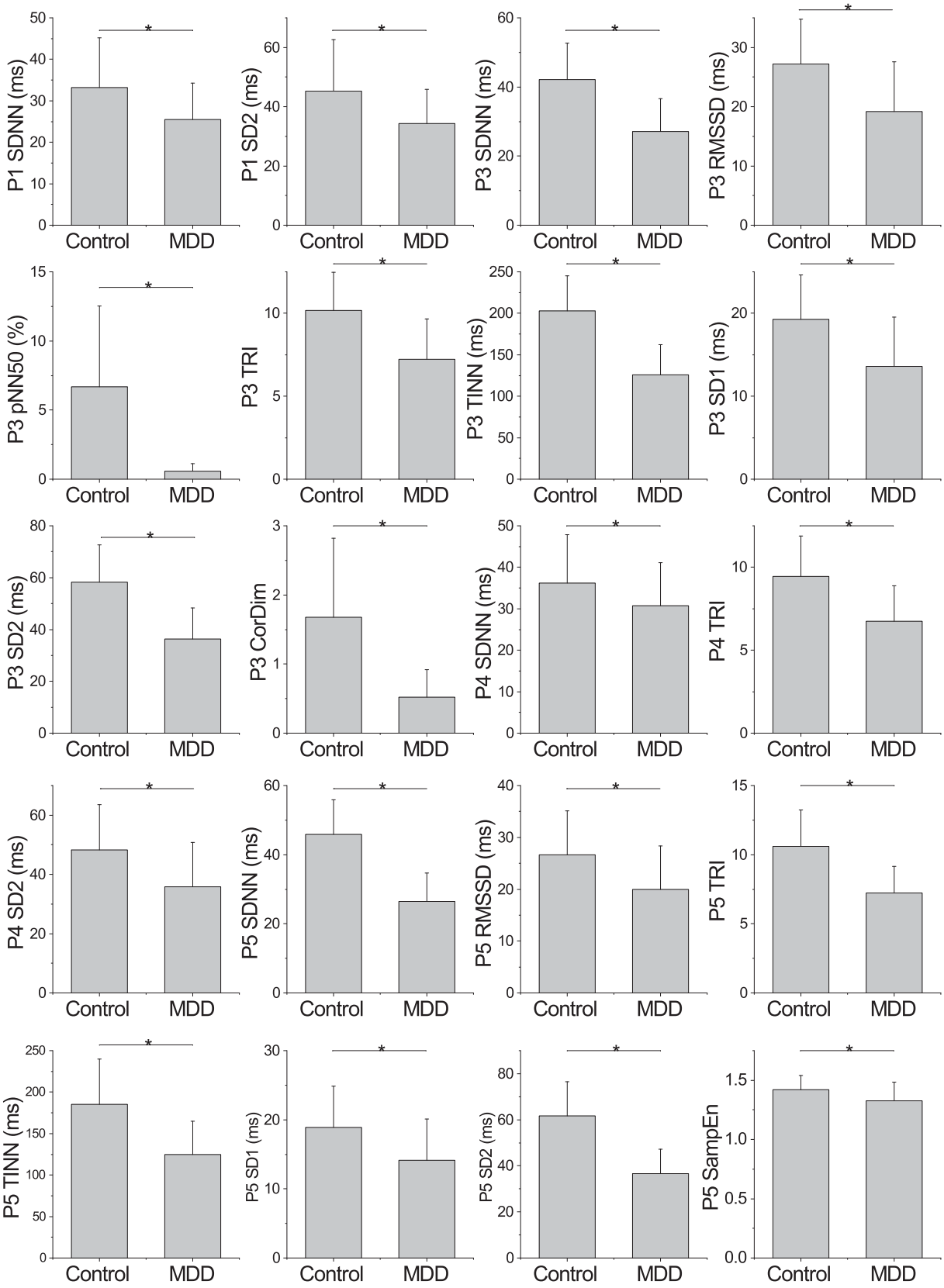

Fig. 3. Median ± MAD of HRV features listed in Table 4, which differ significantly between the control and MDD groups. Asterisks represent the statistical significance of the Mann–Whitney U tests (* p-FDR $<$ 0.05).

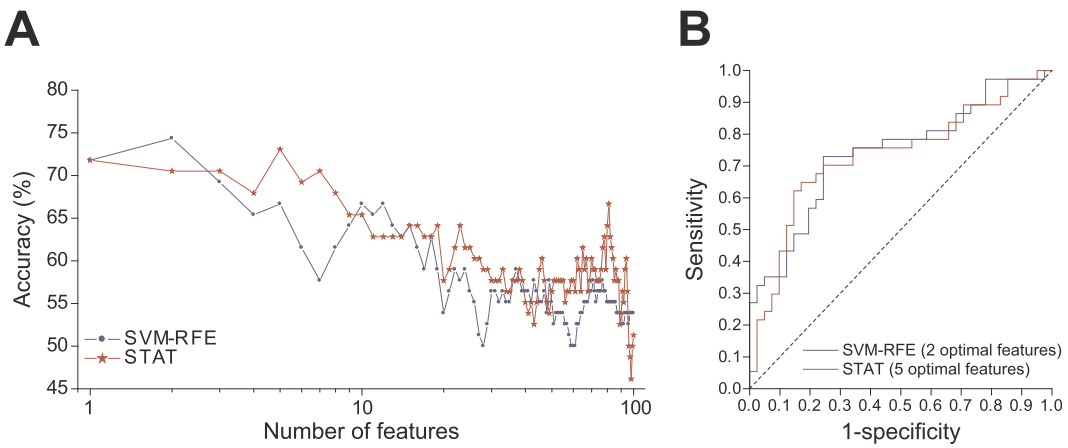

Fig. 4. A. Average accuracy as a function of the number of selected features. The features were ranked by SVM-RFE and statistical filter (STAT). **B. ROC curves of the SVM classifiers during the use of optimal feature sets selected by SVM-RFE and statistical filter.**

which ranked the features based on their relevance. The performance measures were assessed using the LOO procedure (Fig. 2).

Fig. 4A shows the average prediction accuracy as a function of the number of features selected by the SVM-RFE and the statistical filter methods. The best performance results for each feature selection method are summarized in Table 5. Training accuracy and test accuracy at the best performance were similar to each other, indicating that the SVM classifier did not overfit a model to the training set (Table S2). SVM-RFE marginally outperformed the statistical filter and achieved 74.4% accuracy, 73% sensitivity, and 75.6% specificity using only the two most relevant features. The best performance measures identified with the statistical filter used as the feature selection method were similar to those identified by SVM-RFE but were based on the five optimal features. When all 100 features were utilized for classification, the SVM classifier achieved only ~50% accuracy, demonstrating that the feature selection procedure substantially improved the classifier performance. Fig. 4B shows the ROC curves of the SVM classifiers during the use of optimal feature sets selected by the SVM-RFE and the statistical filter methods. The AUCs for SVM-RFE and statistical filter were 0.742 and 0.734, respectively (Table 5).

The top 10 features in the order of the median rank evaluated by the SVM-RFE and the statistical filter methods are listed in Table 6. In addition, the features that differ significantly between the control and MDD groups, which are listed in Table 4, are marked with asterisks. For both selection methods, the most relevant feature was SDNN from P3 (stress task recovery). SD2 from P3 (stress task recovery) was selected as the third and the second most relevant feature by the SVM-RFE and the statistical filter methods, respectively. As expected, all 10 most relevant features selected by the statistical filter method exhibited significant group differences. In contrast, only two features from SVM-RFE ranking exhibited significant group differences.

To investigate the role of autonomic arousal and recovery during the measurement in detecting MDD, we performed a classification using features from an individual phase. For this classification, we only used SVM-RFE for feature selection because SVM-RFE outperformed the statistical filter method. Table 7 shows the classification results for different sets of input features. All performance measures decreased when individual phases were used as input data compared to when all phases were included, except for the specificity achieved with the features from P4 (relaxation task). Using the P2 features (stress task) resulted in the lowest accuracy (56.4%). In addition, the number of features required to achieve the best accuracy increased substantially when only one phase was used, except for the results from P3 (stress task recovery). Table 8 lists the top five features ranked by SVM-RFE when a single phase was used as input data. Note that a Poincaré plot feature, SD2, was selected as the most relevant feature in all five individual phases, suggesting that SD2 played an important role in differentiating the MDD patients from the healthy controls. The ROC curves of the SVM classifiers using optimal features from individual phases are shown in Fig. S1. The AUC values ordered in the same way as the accuracy, were the lowest with the P2 features (stress task) and highest with the P5 features (relaxation task recovery) (Table 7).

Based on these results, we tested the performance of the classifier using the linear and nonlinear/Poincaré features separately and using SVM-RFE as the feature selection method. Thirteen linear, five nonlinear, two Poincaré plot features were extracted from each phase (Table 1); therefore, a total of 65 linear features and 35 nonlinear/Poincaré features were used separately as input data. Using the nonlinear/Poincaré features led to 74.4% accuracy, 73% sensitivity, and 75.6% specificity, which was the same as the best performance achieved with all the features included (Table 7). In contrast, when the

Table 5
Best performance of the SVM classifier achieved with the SVM-RFE and statistical filter as feature selection methods.

Feature selection	NF	ACC (%)	SEN (%)	SPE (%)	PPV (%)	NPV (%)	AUC
SVM-RFE	2	74.4	73.0	75.6	73.0	75.6	0.742
Statistical filter	5	73.1	67.6	78.0	73.5	72.7	0.734

Abbreviations: NF, number of features; ACC, accuracy; SEN, sensitivity; SPE, specificity.

Table 6
Top 10 HRV features in order of median rank as evaluated by the SVM-RFE and statistical filter methods. The features with asterisks are those that differ significantly between the control and MDD groups, as shown in Table 4.

Rank	SVM-RFE		Statistical filter	
	Phase	Feature	Phase	Feature
1	P3	SDNN*	P3	SDNN*
2	P2	TINN	P3	SD2*
3	P3	SD2*	P3	TRI*
4	P3	α1	P3	TINN*
5	P5	LFnu	P3	CorDim*
6	P5	LF/HF	P5	TRI*
7	P5	logHF	P3	RMSSD*
8	P2	RRI	P3	SD1*
9	P5	logVLF	P3	pNN50*
10	P1	HFnu	P4	TRI*

Abbreviations: P1, baseline; P2, stress task; P3, stress task recovery; P4, relaxation task; P5, relaxation task recovery.

Table 7
Classification results based on different sets of input data. The best performance for each case is listed. Feature selection was performed by the SVM-RFE.

Data used	NF/TF	ACC (%)	SEN (%)	SPE (%)	PPV (%)	NPV (%)	AUC
All (P1–P5)	2/100	74.4	73.0	75.6	73.0	75.6	0.742
P1	15/20	66.7	62.2	70.7	65.7	67.4	0.621
P2	7/20	56.4	56.8	56.1	53.8	59.0	0.561
P3	1/20	69.2	67.6	70.7	67.6	70.7	0.732
P4	15/20	67.9	56.8	78.0	70.0	66.7	0.666
P5	11/20	71.8	70.3	73.2	70.3	73.2	0.755
Linear features	23/65	70.5	73.0	68.3	67.5	73.7	0.753
Nonlinear/Poincaré	4/35	74.4	73.0	75.6	73.0	75.6	0.730

Abbreviations: P1, baseline; P2, stress task; P3, stress task recovery; P4, relaxation task; P5, relaxation task recovery; NF, number of features used for the best performance; TF, total number of features; ACC, accuracy; SEN, sensitivity; SPE, specificity.

Table 8
Top five HRV features in the order of median rank from the classifications based on different sets of input data. Feature ranking was evaluated by the SVM-RFE.

Data used	Rank				
	1	2	3	4	5
All (P1–P5)	P3 SDNN	P2 TINN	P3 SD2	P3 α1	P5 LFnu
P1	SD2	HFnu	LFnu	SDNN	RRI
P2	SD2	RRI	SampEn	SDNN	HFnu
P3	SD2	SDNN	α1	RRI	TRI
P4	SD2	SampEn	CorDim	SDNN	LF/HF
P5	SD2	logHF	logVLF	SampEn	RRI
Linear features	P3 SDNN	P2 TINN	P3 TRI	P1 HFnu	P5 logHF
Nonlinear/Poincaré	P3 SD2	P3 SD1	P2 SD2	P2 SD1	P3 α1

Abbreviations: P1, baseline; P2, stress task; P3, stress task recovery; P4, relaxation task; P5, relaxation task recovery.

linear features were used, the best accuracy was 70.5%, which was achieved with 23 features (Table 7 and Fig. 5A). The most relevant features selected from the linear and nonlinear/Poincaré datasets were SDNN and SD2 from P3 (stress task recovery), respectively (Table 8). Fig. 5B shows the ROC curves of the SVM classifiers using optimal features selected from linear and nonlinear/Poincaré feature pools separately. The AUCs for the linear and nonlinear/Poincaré feature sets were 0.753 and 0.730, respectively (Table 7).

4. Discussion

We demonstrated that patients with MDD can be differentiated from healthy controls with 74.4% accuracy, 73% sensitivity, and 75.6% specificity using the two most relevant HRV features. Two feature selection methods, SVM-RFE and a statistical filter, selected SDNN from the P3 phase (stress task recovery) as the most important feature in the classification. The value of the performance measures decreased when individual phases were considered separately as input data, suggesting that combining HRV features measured during autonomic arousal and recovery may enhance the discriminative power of this technique.

We compared the HRV features between the MDD and control groups. The values of 20 of the 100 tested HRV features significantly decreased in the MDD group compared to the control (Table 4, Fig. 3). These results are consistent with those of the previous studies that have reported reduced HRV in MDD patients [10,11,15–17]. Two features, SDNN and SD2, were significantly different between the two groups in all phases except for P2 (stress task), suggesting that these two features may represent altered autonomic cardiac regulation in MDD patients.

The highest accuracy, 74.4%, was achieved using SVM-RFE in the present study. SVM-RFE outperformed the statistical filter as a feature selection method, but the differences in performance measures between the two methods were not substantial (Table 5). Because the performance in feature selection and classification depends on input data, this result may only be applicable to our study. Nonetheless, a previous study on the HRV-based detection of congestive heart failure showed that a backward elimination method yielded better performance than a statistical significance method [63]. Adopting feature selection substantially improved the performance of the SVM classifier. The accuracy was only ~50% when all the features were used as input data but increased to > 73% when relevant features were selectively used (Fig. 4A). In general, the accuracy of classification algorithms deteriorates with high-dimensional data due to the curse of dimensionality; as the number of features increases, the amount of data required for accurate classification grows exponentially [80]. Therefore, with the fixed number of samples, the ability of a classifier to find a true model deteriorates as the dimensionality increases.

SVM-RFE achieved the best performance using two features, SDNN from P3 (stress task recovery) and TINN from P2 (stress task). The statistical filter reached its maximum accuracy using five features from the P3 phase (stress task recovery), SDNN, SD2, TRI, TINN, and CorDim (Table 6). These results suggest that the observation of the baseline ANS activity (the P1 phase) may not be ideal for differentiating MDD patients from the healthy population. The analysis of the resting HRV only assesses cardiac vagal modulation and reveals little information on cardiac sympathetic modulation [81]. A previous study suggested that MDD is associated with reduced cardiovascular recovery from laboratory stress, which may explain why the features from the P3 phase (stress task recovery) were selected as key relevant features [35]. Therefore, our feature selection results suggest that evaluating the responses to autonomic stimuli and subsequent recovery from those stimuli may reveal the differences in ANS activity between depressed and healthy subjects. In previous studies on depression classification, various ANS stimuli were also employed during HRV measurements, such as emotion elicitation [18], relaxation [20], random number generation [21], and Ewing tests [22]. This approach has allowed the monitoring of autonomic reactivity related to disrupted arousal control in depressed patients.

Classification based on SVM-RFE performed more poorly when individual phases were used separately as input data than when all phases were utilized (Table 7). The highest accuracy achieved using the individual phase was 71.8% with P5 (relaxation task recovery), and the lowest was 56.4% with P2 (stress task). However, the best performance with the P5 phase (relaxation task recovery) was achieved with 11 features, which is quite a high number, considering that two features were required when all phases were included as input data. With the P3 phase (stress task recovery), only one feature was required for the best performance, but the accuracy, sensitivity, and PPV were below 70%. These results suggest that combining autonomic responses from various stimuli can be complementary and improve decision-making based on HRV.

We conducted ROC curve analyses for classifier outputs listed in Tables 5 and 7. Since ROC curves are obtained by varying the threshold from 0 to 1, higher AUC does not necessarily reflect higher accuracy evaluated based on the fixed threshold, 0.5. For example, although SVM-RFE showed higher accuracy than the statistical filter as the

feature selection method, AUC was lower in SVM-RFE than in statistical filter (Table 5). In addition, the highest accuracy was achieved when all features or only nonlinear/Poincaré features were used as input data. However, the highest AUC was achieved when only the features from the last phase were used (Table 7).

Our results indicate that SDNN and SD2 may represent the distinctive difference between MDD patients and healthy controls. Both feature selection methods selected SDNN from P3 (stress task recovery) as the most important feature and ranked SD2 from P3 (stress task recovery) within the top three relevant features (Table 6). These results are similar to those of the statistical comparison, in which these two features from P3 (stress task recovery) exhibited the lowest adjusted *p*-values (Table 4). When the individual phases were used separately as input data, SVM-RFE selected SD2 as the most relevant feature for all cases (Table 8). SDNN represents overall HRV and correlates strongly with total power [45,46]. SD2 from the Poincaré plot represents the long-term variation in RRIs and is affected by both parasympathetic and sympathetic activity [82,83]. SDNN and SD2 are also strongly correlated [82], which partially explains why these two indices were selected together as top-ranked features by various methods. Previous studies have reported reduced SDNN in MDD patients and also a significant negative correlation between the severity of depression and SDNN [58,59,84,85]. Reduced SD2 in MDD patients has also been detected in previous studies [58,86]. Chen et al. [58] measured HRV from healthy controls and MDD patients undergoing a series of autonomic tests, i.e., rest, deep breathing, the Valsalva test, and standing up. They found that SDNN and SD2 were significantly lower in the MDD group than in the control group during these tests, except for the resting state. These results suggest that SDNN and SD2 can represent dysregulated autonomic reactivity to sympathetic and parasympathetic stimuli in MDD patients and therefore can be used as important factors in distinguishing MDD patients from healthy controls.

Since SDNN reflects all frequency bands of HRV, it is affected by both vagal and sympathetic activities of the autonomic nervous system [45]. The HF component of HRV is primarily determined by vagal activity, but the LF component is known to be mediated by both vagal and sympathetic activities [46]. A strong correlation between SDNN and pharmacologically controlled vagal tone was reported, suggesting that SDNN can represent the changes in vagal activity [87]. Moreover, a moderate correlation between the LF component and the vagal tone was detected, partially explaining why SDNN, which encompasses LF and HF bands, showed a good correlation with the vagal tone [87]. Therefore, significantly decreased SDNN in the MDD group in our study may represent reduced vagal activity, which is a typical characteristic of HRV in depressed patients [88].

To describe the neurological association between major depression and autonomic disruption, a neurovisceral integration model has been suggested, which is referred to as the central autonomic network (CAN) [89]. In this model, the CAN is involved in an internal regulation through which the brain mediates the visceromotor, neuroendocrine, and behavioral responses in adaptation to environmental, cognitive, and emotional influences. Structurally, the CAN includes the prefrontal cortex, cingulate cortex, insular cortex, amygdala, hypothalamus, and brainstem autonomic nuclei, which form neurological interconnection that enables the bidirectional flow of information between the lower and higher layers of the central nervous system. The output of the CAN is connected to the sinoatrial node of the heart through the stellate ganglia and the vagus nerve, suggesting that cortical activity modulates cardiac activity. Therefore, HRV can reflect brain-heart interactions and the integrity of the CAN. Previous neuroimaging studies have demonstrated that reduced HRV is associated with decreased activation in various brain regions, including the right superior prefrontal, left rostral anterior cingulate, right dorsolateral prefrontal, and right parietal cortices [15]. These results suggest that decreased HRV and dysfunctional cardiac adaptation observed in MDD patients are related to hypoactivation within the CAN.

Using the nonlinear/Poincaré HRV features exclusively, we achieved 74.4% accuracy, 73% sensitivity, and 75.6% specificity, which were the same as the best performance measures achieved when all the features were considered as input data (Table 7). The importance of the nonlinear/Poincaré HRV features in differentiating the pathological ANS dynamics of clinical diseases has been previously suggested. Valenza et al. extensively examined the potential applications of nonlinear/Poincaré HRV features to the algorithm-based diagnosis of various diseases, including bipolar disorder [28], congestive heart failure [90], and Parkinson's disease [91]. Recently, subclinical depression (dysphoria) was differentiated from healthy controls using linear and nonlinear/Poincaré HRV features [71]. We also demonstrated that the model using the nonlinear/Poincaré HRV features performed better in classification than that using the linear features (Table 7, Fig. 5A), which is consistent with the results of the previous studies on Parkinson's disease [91] and dysphoria [71]. In addition, Schulz et al. [26] demonstrated that nonlinear/Poincaré HRV features are more reliable in discriminating MDD patients from controls than linear HRV features, as the latter exhibit high inter-subject variability.

Previous studies in which HRV features were used to discriminate MDD patients have also yielded promising results obtained via various machine learning methods, which are summarized in Table 9. Compared to those previous results, the highest accuracy achieved in the present study (74.4%) is relatively low. However, some of the previous

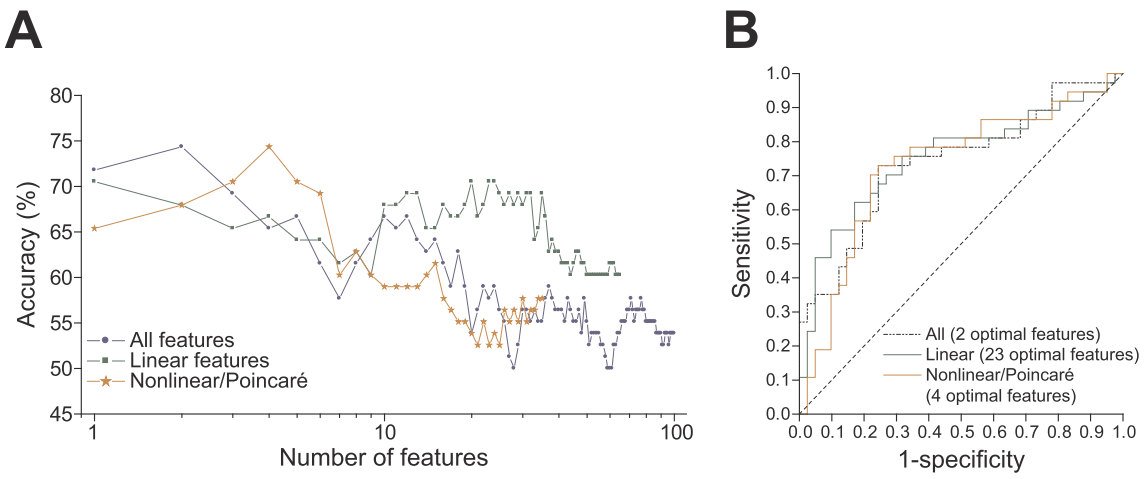

Fig. 5. A. Average accuracy as a function of the number of selected features for three different datasets. The features were ranked by SVM-RFE. **B. ROC curves of the SVM classifiers using optimal features selected from linear and nonlinear feature pools separately.**

Table 9
Summary of previous HRV-based MDD detection studies.

Authors	Subjects	Stimulus (number of test phases)	Features	Feature selection	Classification method	ACC
Zhang et al. [18]	MDD 10 HC 10	Multimodal affective contents (1 phase)	6	N/A	Neuro-fuzzy network	95%
Roh et al. [19]	23[a]	N/A	10	N/A	SVM	71%
Matsui et al. [20]	MDD 13 HC 28	Relaxation (2 phases)	4 (2 per phase)	N/A	Linear discrimination analysis	88%
Sun et al. [21]	MDD 44 HC 47	Random number generation (3 phases)	9 (3 per phase)	N/A	Logistic regression	79%
Kuang et al. [22]	MDD 38 HC 38 (female)	Ewing test (4 phases)	64 (16 per phase)	Correlation-based method (10 selected)	Bayesian networks 10-fold CV	86%
This study	MDD 37 HC 41	Mental arithmetic test (5 phases)	100 (20 per phase)	SVM-RFE (2 selected)	SVM LOO CV	74%

Abbreviations: HC, healthy control; N/A, not applicable; CV, cross-validation; ACC, accuracy.
[a] Total number of subjects.

studies reported the results based on a small number of samples [18,19] or they lacked the description of the prediction accuracy validation method [18,20,21]. Without cross-validation, predictive performance can be over-estimated. Kuang et al. [22] used a 10-fold cross-validation method to evaluate their prediction performance. They achieved 86% accuracy with 10 selected features, but only female subjects participated in the study [22], suggesting that future studies may need to investigate the potential improvement of the model performance by excluding male subjects. The other possible reason for the highest accuracy being relatively low was that we used a linear kernel function for both SVM-RFE and SVM classifier models. We may achieve higher accuracy if the linear kernel is replaced with nonlinear kernel functions, but this change may also increase the risk of overfitting. Note that the number of features required to achieve the best performance in our study was two using SVM-RFE, which is relatively small compared to the numbers of features indicated in previous studies. Since the classification was based on a linear kernel SVM, our results demonstrated that the two optimal features separated the two groups successfully using a simple linear function. Even accounting for the relatively low accuracy, the features were very efficient in the classification. The RFE method becomes more beneficial in the case of a large amount of potentially correlated features (e.g., this study) because the nested feature subsets contain complementary features, which are not necessarily most relevant individually [74]. Therefore, the top-ranked feature subset selected from the RFE algorithm can have a strong discriminative power with high accuracy in classification. We also demonstrated that the algorithm using nonlinear/Poincaré HRV features performed better than that using linear indices and matched the best performance achieved by all features (Table 7). To the best of our knowledge, classification of MDD solely based on nonlinear/Poincaré features has not been extensively investigated. Kuang et al. [22] included nonlinear/Poincaré features as part of the input data but the performance based on nonlinear/Poincaré data has not been evaluated separately.

Recent advances in healthcare technology have facilitated the clinical applications of wearable monitoring systems for psychiatric disorders [92]. For example, a wearable textile-based sensor system has been used to measure long-term ambulatory ECG of patients with bipolar disorders. Longitudinal HRV dynamics from these ECG data have been analyzed to predict the clinical states of individual bipolar patients [93–95]. Furthermore, an adhesive patch sensor worn on the chest has been used to monitor ECG and locomotor activity in patients with schizophrenia; by using features derived from these data, these patients were differentiated from healthy controls [96]. A similar approach can be applied to the management and treatment of major depression. As a result of the heterogeneous nature of MDD, such patients would benefit from a personalized HRV assessment to track their depressive symptoms in association with abnormal changes in ANS activity. Electroencephalography (EEG) has also been widely studied as an objective marker for the detection of depression using machine learning methods [97]. However, EEG is more prone to noise than ECG and easily contaminated by electrooculogram or electromyogram artifacts, which are the disadvantages of a wearable system. Therefore, HRV analysis of ECG measurements may play a more important role in the development of mobile healthcare services for mental illness management.

4.1. Limitations

Antidepressant medications have been associated with changes in HRV although a definite conclusion has not yet been reached. For example, a meta-analysis of antidepressant treatment found that TCAs significantly reduced HRV, but other antidepressants had a minimal impact on HRV [17]. In contrast, a large cohort study (2114 subjects) by Licht et al. [84] revealed that MDD is not associated with HRV but MDD patients using SSRIs, SNRIs, and TCAs have significantly reduced HRV. These findings suggest that antidepressants might be responsible for the outcomes in this study because all the patients were receiving antidepressant treatment. However, direct comparisons with previous studies are not straightforward because of the methodological issues, and we may need to include unmedicated depressed patients in a future study to investigate the effect of antidepressants [98].

Our results are based on a total of 78 subjects. Previous studies have demonstrated methods to estimate the sample size required to perform binary classification based on high-dimensional data, such as DNA microarray data [99,100]. Mukherjee et al. [100] used a learning curve fitting to estimate the dataset size for classifying cancer morphology and treatment outcome based on microarray data. They used a statistical test to determine the minimum number of samples to produce a classifier that performs significantly better than a random classifier. In addition, learning curves estimated the empirical error rate as a function of dataset size, which provided the sample size to build a good predictor. They suggested that sample sizes of 30–75 were adequate for achieving a reasonably accurate classification. Dobbin and Simon [99] used a different method, a model-based approach, to estimate the sample size for classifying DNA microarray data and suggested that at least 40–60 samples are required to produce small error rates. These results suggest that we have a sufficient number of samples to build a significant classifier while achieving relatively low errors. Nonetheless, an increase in the sample size generally leads to an improvement in performance until the performance approaches its plateau [101]. Currently, we are recruiting more subjects, and a future study of this expanded cohort will allow us to develop a more robust screening method for major depression.

The limitation of classifier models used in this study is that SVM algorithms were implemented only with a linear kernel function. Performance of SVMs is substantially affected by the kernel function and its parameters. In general, a linear kernel can perform well with linearly separable data, but if the relationship between the labels and features is nonlinear, it is outperformed by nonlinear kernel functions,

such as radial basis function, sigmoid, and polynomial [102]. Previous studies have demonstrated that nonlinear kernel functions outperform linear models in the classification of image data and acoustic signals [103,104]. Nonlinear kernel functions were further modified for specific data types, and these tailor-made kernels improved the performance of SVM classifiers [105,106]. For future studies, we will perform additional tests with different kernel functions to evaluate their performances.

5. Conclusion

We demonstrated the machine learning-based automated detection of depression using linear and nonlinear HRV measures. We found that ANS stimulation during measurements was crucial to revealing abnormal reactivity and recovery of the heartbeat dynamics of the depressed subjects because these behaviors were not detectable during the baseline activity. In addition, nonlinear/Poincaré HRV features played a crucial role in differentiating MDD patients. These findings suggest that using linear and nonlinear HRV features measured during various states of ANS has the potential to more objectively identify patients with depressive symptoms.

Ethics approval and consent to participate

This study was approved by the Institutional Review Board of the Samsung Medical Center, Seoul, Korea (No. 2015-07-151), and performed according to the relevant guidelines. The experiment was explained to all subjects, who provided signed informed consent to participate in the study. In addition, $50 compensation was paid to each subject.

Conflicts of interest

None Declared.

Funding

This work was supported by the National Research Foundation of Korea (NRF) grant funded by the Ministry of Science and ICT of the South Korea government (MSIT) (No. 2017R1C1B5017730) and the Institute for Information and Communications Technology Promotion (IITP) grant funded by the MSIT (No. 2015-0-377 00062, The development of skin adhesive patches for the monitoring and prediction of mental disorders). This research was also supported by the Original Technology Research Program for Brain Science of the NRF grant funded by the MSIT (No. NRF-2016M3C7A1947307; PI HJJ), and the Bio & Medical Technology Development Program of the NRF grant funded by the MSIT (No. NRF-2017M3A9F1027323; PI HJJ).

Author contributions

S.B., A.Y.K., E.H.J., H.J.J., H.Y.Y., and S.K. designed research. S.B., A.Y.K., E.H.J., H.Y.Y., H.J.J. and K.W.C. performed research. S.B., A.Y.K., and E.H.J. analyzed the data. S.B. wrote the paper. All authors commented on the manuscript.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.compbiomed.2019.103381.

References

[1] World Health Organization, The Global Burden of Disease: 2004 Update, (2008), https://doi.org/10.1038/npp.2011.85.
[2] M. Luppa, S. Heinrich, M.C. Angermeyer, H.-H. König, S.G. Riedel-Heller, Cost-of-illness studies of depression, J. Affect. Disord. 98 (2007) 29–43, https://doi.org/10.1016/j.jad.2006.07.017.
[3] J.C. Franklin, J.D. Ribeiro, K.R. Fox, K.H. Bentley, E.M. Kleiman, A.C.A.C. Jaroszewski, et al., Risk factors for suicidal thoughts and behaviors: a meta-analysis of 50 Years of research, Psychol. Bull. (2016) 187–232, https://doi.org/10.1037/bul0000084.
[4] P.S. Wang, G. Simon, R.C. Kessler, The economic burden of depression and the cost-effectiveness of treatment, Int. J. Methods Psychiatr. Res. 12 (2003) 22–33.
[5] A.P. Association, Diagnostic and Statistical Manual of Mental Disorders, fifth ed., American Psychiatric Publishing, Arlington, VA, 2013.
[6] A. Jacob, Limitations of clinical psychiatric diagnostic measurements ayden, J. Neurol. Disord. 1 (2013), https://doi.org/10.4172/2329-6895.1000122.
[7] S.M. Lieblich, D.J. Castle, C. Pantelis, M. Hopwood, A.H. Young, I.P. Everall, High heterogeneity and low reliability in the diagnosis of major depression will impair the development of new drugs, Br. J. Psychiatry Open 1 (2015), https://doi.org/10.1192/bjpo.bp.115.000786 e5–7.
[8] D.A. Regier, W.E. Narrow, D.E. Clarke, H.C. Kraemer, S.J. Kuramoto, E.A. Kuhl, et al., DSM-5 field trials in the United States and Canada, part II: test-retest reliability of selected categorical diagnoses, Am. J. Psychiatry 170 (2013) 59–70.
[9] N. Craddock, L. Mynors-Wallis, Psychiatric diagnosis: impersonal, imperfect and important, Br. J. Psychiatry 204 (2014) 93–95.
[10] D.L. Musselman, D.L. Evans, C.B. Nemeroff, The relationship of depression to cardiovascular disease, J. Am. Med. Assoc. 55 (July 1998) 580–592 2014.
[11] J.M. Gorman, R.P. Sloan, Heart rate variability in depressive and anxiety disorders, Am. Heart J. 140 (4 Suppl) (2000) 77–83.
[12] G. Valenza, M. Nardelli, A. Lanatà, C. Gentili, G. Bertschy, R. Paradiso, et al., Wearable monitoring for mood recognition in bipolar disorder based on history-dependent long-term heart rate variability analysis, IEEE J. Biomed Heal Informatics 18 (2014) 1625–1635.
[13] S. Boettger, D. Hoyer, K. Falkenhahn, M. Kaatz, V.K. Yeragani, K.J. Bär, Altered diurnal autonomic variation and reduced vagal information flow in acute schizophrenia, Clin. Neurophysiol. 117 (2006) 2715–2722.
[14] H. Cohen, M. Kotler, M a Matar, Z. Kaplan, U. Loewenthal, H. Miodownik, et al., Analysis of heart rate variability in posttraumatic stress disorder patients in response to a trauma-related reminder, Biol. Psychiatry 44 (1998) 1054–1059 S0006-3223(97)00475-7 [pii].
[15] A.H. Kemp, D.S. Quintana, K.L. Felmingham, S. Matthews, H.F. Jelinek, Depression, comorbid anxiety disorders, and heart rate variability in physically healthy, unmedicated patients: implications for cardiovascular risk, PLoS One 7 (2012) 1–8.
[16] R.M. Carney, J.A. Blumenthal, P.K. Stein, L. Watkins, D. Catellier, L.F. Berkman, et al., Depression, heart rate variability, and acute myocardial infarction, Circulation 104 (2001) 2024–2029.
[17] A.H. Kemp, D.S. Quintana, M.A. Gray, K.L. Felmingham, K. Brown, J.M. Gatt, Impact of depression and antidepressant treatment on heart rate variability: a review and meta-analysis, Biol. Psychiatry 67 (2010) 1067–1074, https://doi.org/10.1016/j.biopsych.2009.12.012.
[18] Z.X. Zhang, X.W. Tian, J.S. Lim, New algorithm for the depression diagnosis using HRV: a neuro-fuzzy approach, Proc. 2011 Int. Symp. Bioelectron. Bioinform. 2011 (2011) 283–286 ISBB.
[19] T. Roh, S. Hong, H.J. Yoo, Wearable depression monitoring system with heart-rate variability, Conf. Proc. Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. IEEE Eng. Med. Biol. Soc. Annu. Conf. 2014 (2014) 562–565.
[20] T. Matsui, K. Kakisaka, T. Shinba, Impaired parasympathetic augmentation under relaxation in patients with depression as assessed by a novel non-contact microwave radar system, J. Med. Eng. Technol. 40 (2016) 15–19.
[21] G. Sun, T. Shinba, T. Kirimoto, T. Matsui, An objective screening method for major depressive disorder using logistic regression analysis of heart rate variability data obtained in a mental task paradigm, Front. Psychiatry 7 (NOV) (2016) 1–7.
[22] D. Kuang, R. Yang, X. Chen, G. Lao, F. Wu, X. Huang, et al., Depression recognition according to heart rate variability using Bayesian Networks, J. Psychiatr. Res. 95 (2017) 282–287, https://doi.org/10.1016/j.jpsychires.2017.09.012.
[23] M. Costa, A.L. Goldberger, C. Peng, Multiscale entropy analysis of complex physiologic time series, Phys. Rev. Lett. 89 (2002) 6–9.
[24] S.M. Pincus, Approximate entropy as a measure of system complexity, Proc. Natl. Acad. Sci. U. S. A. 88 (1991) 2297–2301, https://doi.org/10.1073/pnas.88.6.2297.
[25] A.L. Goldberger, L.A.N. Amaral, J.M. Hausdorff, P.C. Ivanov, C.-K. Peng, H.E. Stanley, Fractal dynamics in physiology: alterations with disease and aging, Proc. Natl. Acad. Sci. 99 (Supplement 1) (2002) 2466–2472, https://doi.org/10.1073/pnas.012579499.
[26] S. Schulz, M. Koschke, K.-J.J. Bär, A. Voss, The altered complexity of cardiovascular regulation in depressed patients, Physiol. Meas. 31 (2010) 303–321, https://doi.org/10.1088/0967-3334/31/3/003.
[27] 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 10 (2015) 1–14.
[28] G. Valenza, M. Nardelli, A. Lanata, C. Gentili, G. Bertschy, M. Kosel, et al., Predicting mood changes in bipolar disorder through heartbeat nonlinear dynamics, IEEE J. Biomed Heal Informatics 20 (2016) 1034–1043.
[29] U. Rajendra Acharya, O. Faust, N. Adib Kadri, J.S. Suri, W. Yu, Automated identification of normal and diabetes heart rate signals using nonlinear measures, Comput. Biol. Med. 43 (2013) 1523–1529, https://doi.org/10.1016/j.compbiomed.2013.05.024.
[30] J.L. Kibler, M. Ma, Depressive symptoms and cardiovascular reactivity to laboratory, Behav. Stress 11 (2004) 81–87.

[31] D. Carroll, A.C. Phillips, K. Hunt, G. Der, Symptoms of depression and cardiovascular reactions to acute psychological stress: evidence from a population study, Biol. Psychol. 75 (2007) 68–74.
[32] B.M. Appelhans, L.J. Luecken, Heart rate variability as an index of regulated emotional responding, Rev. Gen. Psychol. 10 (2006) 229–240.
[33] A.C. Nugent, E.E. Bain, J.F. Thayer, J.J. Sollers, W.C. Drevets, Heart rate variability during motor and cognitive tasks in females with major depressive disorder, Psychiatry Res. Neuroimaging 191 (2011) 1–8, https://doi.org/10.1016/j.pscychresns.2010.08.013.
[34] T. Shinba, Altered autonomic activity and reactivity in depression revealed by heart-rate variability measurement during rest and task conditions, Psychiatry Clin. Neurosci. 68 (2014) 225–233.
[35] K. Salomon, A. Clift, J. Rottenberg, Major Depressive Disorder Is Associated with Attenuated Cardiovascular Reactivity and Impaired Recovery Among Those Free of Cardiovascular Disease vol. 28, (2009), pp. 157–165.
[36] A.Y. Kim, E.H. Jang, S. Kim, K.W. Choi, H.J. Jeon, H.Y. Yu, et al., Automatic detection of major depressive disorder using electrodermal activity, Sci. Rep. 8 (2018) 17030, https://doi.org/10.1038/s41598-018-35147-3.
[37] M.A.X. Hamilton, Development of a rating scale for primary depressive illness, Br. J. Clin. Psychol. 6 (1967) 278–296.
[38] K.W. Choi, E.H. Jang, A.Y. Kim, M. Fava, D. Mischoulon, G. Papakostas, et al., Heart rate variability for treatment response between patients with major depressive disorder versus panic disorder: a 12-week follow-up study, J. Affect. Disord. 246 (2019) 157–165, https://doi.org/10.1016/J.JAD.2018.12.048.
[39] J.J. Sollers, T.A. Sanford, R. Nabors-Oberg, C.A. Anderson, J.F. Thayer, Examining changes in HRV in response to varying ambient temperature, IEEE Eng. Med. Biol. Mag. 21 (2002) 30–34.
[40] S. Yamamoto, M. Iwamoto, M. Inoue, N. Harada, Evaluation of the effect of heat exposure on the autonomic nervous system by heart rate variability and urinary catecholamines, J. Occup. Health 49 (2007) 199–204.
[41] E. Barbosa, J.M. García-Manso, J.M. Martín-González, S. Sarmiento, F.J. Calderón, M.E. Da Silva-Grigoletto, Effect of hyperbaric pressure during scuba diving on autonomic modulation of the cardiac response: application of the continuous wavelet transform to the analysis of heart rate variability, Mil. Med. 175 (2013) 61–64.
[42] P. Zarjam, J. Epps, F. Chen, N.H. Lovell, Estimating cognitive workload using wavelet entropy-based features during an arithmetic task, Comput. Biol. Med. 43 (2013), https://doi.org/10.1016/j.compbiomed.2013.08.021 2186–1295.
[43] M.P. Tarvainen, J.P. Niskanen, J.A. Lipponen, P.O. Ranta-aho, P.A. Karjalainen, Kubios HRV - heart rate variability analysis software, Comput. Methods Progr. Biomed. 113 (2014) 210–220, https://doi.org/10.1016/j.cmpb.2013.07.024.
[44] J. Pan, W.J. Tompkins, A real-time QRS detection algorithm, IEEE Trans. Biomed. Eng. 32 (1985) 230–236, https://doi.org/10.1109/TBME.1985.325532.
[45] M. Malik, 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. 17 (1996) 354–381.
[46] F. Shaffer, J.P. Ginsberg, An overview of heart rate variability metrics and norms, Front Public Heal 5 (September) (2017) 1–17, https://doi.org/10.3389/fpubh.2017.00258.
[47] V. Pichot, J. Gaspoz, S. Molliex, A. Antoniadis, R.I.C. Roche, R.I.C. Costes, et al., Wavelet transform to quantify heart rate variability and to assess its instantaneous changes, J. Appl. Physiol. (1999) 1081–1091.
[48] U. Wiklund, M. Akay, U. Niklasson, Short-term analysis of heart-rate variability by adapted wavelet transforms, IEEE Eng. Med. Biol. Mag. 16 (1997).
[49] J.H. Houtveen, P.C.M. Molenaar, Comparison between the Fourier and Wavelet methods of spectral analysis applied to stationary and nonstationary heart period data, Psychophysiology 38 (2001) 729–735.
[50] J.S. Richman, J.R. Moorman, Physiological time-series analysis using approximate entropy and sample entropy, Am. J. Physiol. Heart Circ. Physiol. 278 (2000) H2039–H2049.
[51] C.K. Peng, S. Havlin, H.E. Stanley, A.L. Goldberger, Quantification of scaling exponents and crossover phenomena in nonstationary heartbeat time series, Chaos 5 (1995) 82–87, https://doi.org/10.1063/1.166141.
[52] P. Grassberger, I. Procaccia, Characterization of strange attractors, Phys. Rev. Lett. 50 (1983).
[53] M.I. Owis, A.H. Abou-Zied, a.BM. Youssef, Y.M. Kadah, Study of features based on nonlinear dynamical modeling in ECG arrhythmia detection and classification, IEEE Trans. Biomed. Eng. 49 (2002) 733–736.
[54] M. Brennan, M. Palaniswami, P. Kamen, Do existing measures of Poincareé plot geometry reflect nonlinear features of heart rate variability? IEEE Trans. Biomed. Eng. 48 (2001) 1342–1347.
[55] A.B. Ciccone, J.A. Siedlik, J.M. Wecht, J.A. Deckert, N.D. Nguyen, J.P. Weir, Reminder: RMSSD and SD1 are identical heart rate variability metrics anthony, Muscle Nerve 56 (2017) 674–678.
[56] P. Guzik, J. Piskorski, T. Krauze, R. Schneider, K.H. Wesseling, A. Wykretowicz, et al., Correlations between the Poincaré plot and conventional heart rate variability parameters assessed during paced breathing, J. Physiol. Sci. 57 (2007) 63–71.
[57] A. de la Torre-Luque, X. Bornas, M. Balle, A. Fiol-Veny, Complexity and nonlinear biomarkers in emotional disorders: a meta-analytic study, Neurosci. Biobehav. Rev. 68 (2016) 410–422, https://doi.org/10.1016/j.neubiorev.2016.05.023.
[58] X. Chen, R. Yang, D. Kuang, L. Zhang, R. Lv, X. Huang, et al., Heart rate variability in patients with major depression disorder during a clinical autonomic test, Psychiatry Res. 256 (2017), https://doi.org/10.1016/j.psychres.2017.06.041 June:207–11.
[59] D.E. Vigo, L.N. Siri, M.S. Ladrón De Guevara, J.A. Martínez-Martínez, R.D. Fahrer, D.P. Cardinali, et al., Relation of depression to heart rate nonlinear dynamics in patients ≥60 years of age with recent unstable angina pectoris or acute myocardial infarction, Am. J. Cardiol. 93 (2004) 756–760, https://doi.org/10.1016/j.amjcard.2003.11.056.
[60] A. Davari Dolatabadi, S.E.Z. Khadem, B.M. Asl, Automated diagnosis of coronary artery disease (CAD) patients using optimized SVM, Comput. Methods Progr. Biomed. 138 (2017) 117–126, https://doi.org/10.1016/j.cmpb.2016.10.011.
[61] O.M. Doyle, A. Temko, W. Marnane, G. Lightbody, G.B. Boylan, Heart rate based automatic seizure detection in the newborn, Med. Eng. Phys. 32 (2010) 829–839, https://doi.org/10.1016/j.medengphy.2010.05.010.
[62] Y. Işler, M. Kuntalp, Combining classical HRV indices with wavelet entropy measures improves to performance in diagnosing congestive heart failure, Comput. Biol. Med. 37 (2007) 1502–1510.
[63] A. Narin, Y. Isler, M. Ozer, Investigating the performance improvement of HRV Indices in CHF using feature selection methods based on backward elimination and statistical significance, Comput. Biol. Med. 45 (2014) 72–79, https://doi.org/10.1016/j.compbiomed.2013.11.016.
[64] J.F. Ramirez-Villegas, E. Lam-Espinosa, D.F. Ramirez-Moreno, P.C. Calvo-Echeverry, W. Agredo-Rodriguez, Heart rate variability dynamics for the prognosis of cardiovascular risk, PLoS One 6 (2011).
[65] S. Berger, A. Kliem, V. Yeragani, K.J. Bär, Cardio-respiratory coupling in untreated patients with major depression, J. Affect. Disord. 139 (2012) 166–171, https://doi.org/10.1016/j.jad.2012.01.035.
[66] G. Valenza, R.G. Garcia, L. Citi, E.P. Scilingo, C.A. Tomaz, R. Barbieri, Nonlinear digital signal processing in mental health: characterization of major depression using instantaneous entropy measures of heartbeat dynamics, Front. Physiol. 6 (2015) MAR:1–8.
[67] M. Baumert, G.W. Lambert, T. Dawood, E a Lambert, M.D. Esler, M. McGrane, et al., Short-term heart rate variability and cardiac norepinephrine spillover in patients with depression and panic disorder, Am. J. Physiol. Heart Circ. Physiol. 297 (2009) H674–H679.
[68] C. Blasco-Lafarga, I. Martinez-Navarro, M.E. Sisamon, N. Caus, E. Yanguez, P. Llorens-Soriano, Linear and nonlinear heart rate dynamics in elderly inpatients. Relations with comorbidity and depression, Med 46 (2010) 393–400 1006-04 [pii].
[69] J.S. Chang, C.S. Yoo, S.H. Yi, J.Y. Her, H.M. Choi, T.H. Ha, et al., An integrative assessment of the psychophysiologic alterations in young women with recurrent major depressive disorder, Psychosom. Med. 74 (2012) 495–500.
[70] Y. Benjamini, Y. Hochberg, Y. Benjamini, Y. Hochberg, Controlling the false discovery rate: a practical and powerful approach to multiple testing, J. R. Stat. Soc. B 57 (1995) 289–300, https://doi.org/10.2307/2346101.
[71] A. Greco, Assessment of linear and nonlinear/complex heartbeat dynamics in subclinical depression (dysphoria), Physiol. Meas. (2018) 0–7.
[72] A.M. Molinaro, R. Simon, R.M. Pfeiffer, Prediction error estimation: a comparison of resampling methods, Bioinformatics 21 (2005) 3301–3307.
[73] P. Zhang, Model selection via multifold cross validation, Ann. Stat. 21 (1993) 299–313.
[74] I. Guyon, Gene Selection for Cancer Classification, (2002), pp. 389–422.
[75] K. Yan, D. Zhang, Feature selection and analysis on correlated gas sensor data with recursive feature elimination, Sens. Actuators B Chem. 212 (2015) 353–363, https://doi.org/10.1016/j.snb.2015.02.025.
[76] R. Fan, P. Chen, C. Lin, Working Set Selection Using Second Order Information for Training Support, Vector Machines vol. 6, (2005) 1889–1918.
[77] V.N. Vapnik, Statistical learning theory, Adapt Learn Syst. Signal Process Commun. Control 2 (1998) 1–740, https://doi.org/10.2307/1271368.
[78] V.N. Vapnik, The Nature of Statistical Learning Theory, Springer, 2000.
[79] C.J.C. Burges, A tutorial on support vector machines for pattern recognition, Data Min. Knowl. Discov. 2 (1998) 121–167.
[80] G.V. Trunk, A problem of dimensionality: a simple example, IEEE Trans. Pattern Anal. Mach. Intell. 1 (1979) 306–307.
[81] A. Schumann, C. Andrack, K.J. Bär, Differences of sympathetic and parasympathetic modulation in major depression, Prog. Neuro-Psychopharmacol. Biol. Psychiatry 79 (2017) 324–331, https://doi.org/10.1016/j.pnpbp.2017.07.009.
[82] R.A. Hoshi, C.M. Pastre, L.C.M. Vanderlei, M.F. Godoy, Poincaré plot indexes of heart rate variability: relationships with other nonlinear variables, Auton Neurosci. Basic Clin. 177 (2013) 271–274, https://doi.org/10.1016/j.autneu.2013.05.004.
[83] G. De Vito, S.D.R. Galloway, M.A. Nimmo, P. Maas, J.J.V. McMurray, Effects of central sympathetic inhibition on heart rate variability during steady-state exercise in healthy humans, Clin. Physiol. Funct. Imaging 22 (2002) 32–38, https://doi.org/10.1046/j.1475-097X.2002.00395.x.
[84] C.M.M. Licht, E.J.C. De Geus, R. Van Dyck, B.W.J.H. Penninx, Longitudinal evidence for unfavorable effects of antidepressants on heart rate variability, Biol. Psychiatry 68 (2010) 861–868, https://doi.org/10.1016/j.biopsych.2010.06.032.
[85] W. Y, Z. X, O. A, T. A, L. X, B. M, Altered cardiac autonomic nervous function in depression, BMC Psychiatry 13 (2013) 187, https://doi.org/10.1186/1471-244X-13-187.
[86] H.F. Jelinek, A.H. Khandoker, D.S. Quintana, M.H. Imam, a H. Kemp, Complex Correlation Measure as a sensitive indicator of risk for sudden cardiac death in patients with depression, Comput. Cardiol. CinC 2011 (2011) 809–12 http://www.cinc.org/2011/preprints/109.pdf.
[87] J. Hayano, Y. Sakakibara, A. Yamada, M. Yamada, S. Mukai, T. Fujinami, et al., Accuracy of assessment of cardiac vagal tone by heart rate variability in normal subjects, Am. J. Cardiol. 67 (1991) 199–204.
[88] D. Bassett, A literature review of heart rate variability in depressive and bipolar disorders, Aust. N. Z. J. Psychiatr. 50 (2015) 1–9, https://doi.org/10.1177/

0004867415622689.
[89] J.F. Thayer, A.L. Hansen, E. Saus-Rose, B.H. Johnsen, Heart rate variability, prefrontal neural function, and cognitive performance: the neurovisceral integration perspective on self-regulation, adaptation, and health, Ann. Behav. Med. 37 (2009) 141–153.
[90] G. Valenza, H. Wendt, K. Kiyono, J. Hayano, E. Watanabe, Y. Yamamoto, et al., Mortality prediction in severe congestive heart failure patients with multifractal point-process modeling of heartbeat dynamics, IEEE Trans. Biomed. Eng. 9294 (2018) 1–10.
[91] G. Valenza, S. Orsolini, S. Diciotti, L. Citi, E.P. Scilingo, M. Guerrisi, et al., Assessment of spontaneous cardiovascular oscillations in Parkinson's disease, Biomed. Signal Process. Control 26 (2016) 80–89, https://doi.org/10.1016/j.bspc.2015.12.001.
[92] E. Reinertsen, G.D. Clifford, A review of physiological and behavioral monitoring with digital sensors for neuropsychiatric illnesses, Physiol. Meas. 39 (2018).
[93] G. Valenza, C. Gentili, A. Lanatà, E.P. Scilingo, Mood recognition in bipolar patients through the PSYCHE platform: preliminary evaluations and perspectives, Artif. Intell. Med. 57 (2013) 49–58, https://doi.org/10.1016/j.artmed.2012.12.001.
[94] C. Gentili, G. Valenza, M. Nardelli, A. Lanat, G. Bertschy, L. Weiner, et al., Longitudinal monitoring of heartbeat dynamics predicts mood changes in bipolar patients: a pilot study, J. Affect. Disord. 209 (2017) 30–38, https://doi.org/10.1016/j.jad.2016.11.008 November 2016.
[95] A. Lanata, G. Valenza, M. Nardelli, C. Gentili, E.P. Scilingo, Complexity index from a personalized wearable monitoring system for assessing remission in mental health, IEEE J. Biomed Heal Informatics 19 (2015) 132–139.
[96] E. Reinertsen, M. Osipov, C.Y. Liu, J.M. Kane, G. Petrides, G.D. Clifford, Continuous assessment of schizophrenia using heart rate and accelerometer data, Physiol. Meas. 38 (2017) 1456–1471.
[97] U.R. Acharya, V.K. Sudarshan, H. Adeli, J. Santhosh, J.E.W. Koh, A. Adeli, Computer-aided diagnosis of depression using EEG signals, Eur. Neurol. 73 (2015) 329–336.
[98] A.H. Kemp, D.S. Quintana, G.S. Malhi, Effects of serotonin reuptake inhibitors on heart rate variability: methodological issues, medical comorbidity, and clinical relevance, Biol. Psychiatry 69 (2011) e25–e26, https://doi.org/10.1016/j.biopsych.2010.10.035.
[99] K.K. Dobbin, R.M. Simon, Sample size planning for developing classifiers using high-dimensional DNA microarray data, Biostatistics 8 (2007) 101–117.
[100] S. Mukherjee, P. Tamayo, S. Rogers, R. Rifkin, A. Engle, C. Campbell, et al., Estimating dataset size requirements for classifying DNA microarray data, J. Comput. Biol. 10 (2003) 119–142.
[101] X. Zhu, C. Vondrick, C.C. Fowlkes, D. Ramanan, Do we need more training data? Int. J. Comput. Vis. 119 (2016) 76–92.
[102] S.S. Keerthi, C.J. Lin, Asymptotic behaviors of support vector machines with Gaussian kernel, Neural Comput. 15 (2003) 1667–1689.
[103] B. Yekkehkhany, A. Safari, S. Homayouni, M. Hasanlou, A comparison study of different kernel functions for SVM-based classification of multi-temporal polarimetry SAR data, Int. Arch. Photogram Remote Sens Spat. Inf. Sci. - ISPRS Arch. 40 (2014) 281–285.
[104] M.A. Nanda, K.B. Seminar, D. Nandika, A. Maddu, A comparison study of kernel functions in the support vector machine and its application for termite detection, OR Inf. 9 (2018).
[105] S. Amari, S. Wu, Improving support vector machine classifiers by modifying kernel functions, Neural Netw. 12 (1999) 783–789.
[106] B. Guo, S.R. Gunn, R.I. Damper, J.D.B. Nelson, Customizing Kernel functions for SVM-based hyperspectral image classification, IEEE Trans. Image Process. 17 (2008) 622–629.