Machine Learning Approach on Photoplethysmogram Morphology for Psychiatric Disorders Prediction

Azwani Awang
Dept. of Electrical, Electronic and Systems Engineering
Universiti Kebangsaan Malaysia
UKM Bangi, Malaysia
p106168@siswa.ukm.edu.my

Nazrul Anuar Nayan
Dept. of Electrical, Electronic and Systems Engineering
Universiti Kebangsaan Malaysia
UKM Bangi, Malaysia
nazrul@ukm.edu.my

Nik Ruzyanei Nik Jaafar
Dept. of Psychiatry
Hospital Canselor Tuanku Muhriz
WP Kuala Lumpur, Malaysia
njruzyanei@gmail.com

Mohd Zubir Suboh
Dept. of Electrical, Electronic and Systems Engineering
Universiti Kebangsaan Malaysia
UKM Bangi, Malaysia
mohdzubir@unikl.edu.my

Khairul Anuar A Rahman
Dept. of Electronics & Computer Engineering Technology
Universiti Tekninal Malaysia Melaka
Malacca, Malaysia
khairulanuar@utem.edu.my

Siti Nor Ashikin Ismail
Dept. of Electrical, Electronic and Systems Engineering
Universiti Kebangsaan Malaysia
UKM Bangi, Malaysia
p114762@siswa.ukm.edu.my

2022 5th International Conference on Communications, Signal Processing, and their Applications (ICCSPA) | 978-1-6654-8237-0/22/$31.00 ©2022 IEEE | DOI: 10.1109/ICCSPA55860.2022.10019188

***Abstract*—Psychiatric disorders (PDs) interfere with one's functioning and greatly affect a person's quality of life. Prompt diagnosis and intervention at the early stages of these illnesses are important. However, most people are oblivious or unaware of their mental health status as the symptoms may not be easily recognizable. Consequently, complications occur later in life. In this study, a machine learning (ML) approach that distinguishes between case (PD-diagnosed patients) and control (healthy) groups was developed using photoplethysmogram (PPG) morphology. 92 subjects with gender and age-matched PPG data were collected during two phases; baseline and stimulus state of a 10-min experiment. 60 features from PPG morphology were extracted from each phase, and another 30 were obtained from differences between the two phases. A total of 27 out of 90 features exhibited a significant difference. Twelve features extracted by heatmap based on the correlation analysis were fed to five types of ML algorithms: discrimination analysis, k-nearest neighbor, decision tree, support vector machine, and artificial neural network (ANN). The results showed the best performance of 92.86%, 100.00%, and 96.43% for sensitivity, specificity, and accuracy by ANN. Thus, a PD prediction model was developed using machine learning techniques from PPG morphology extraction.**

Keywords—Bio signals, Statistical analysis, Correlation analysis, Machine learning

I. INTRODUCTION

Psychiatric disorders (PDs) are characterized by the disturbance of emotions, behavior and cognition, cause distress, and affect one's functioning. PDs include major depression, bipolar disorder, schizophrenia, and anxiety disorder. Globally, the burden of PD continues to increase and has become a serious condition with a high prevalence (2018). Furthermore, in the wake of the SARS-CoV-2 pandemic in 2019, restrictions and social detention around the world have had a severe impact on people's mental health. Several studies have shown that the prevalence of community psychology and mental health increased during a pandemic compared to the previous year [1]. Until today, there was no consistent and reliable method to determine an individual's mental health status. The main problem is because mental health does not only involve individual physical aspect but also involves individual social factors, culture, surroundings, and genetics. Apart from that, there are other disruptions like pressure, nutrition, prenatal infections and exposure to environmental hazards. It was reported that repetitive exposure to pressure can affect one's physical and mental condition together with brain function [2]. People living in stressful environments are exposed to many disorders, such as anxiety or depression [3], [4]. Currently, screening questionnaires, such as the Depression Anxiety Stress Scale 21 (DASS-21) are used to measure depression, anxiety, and stress. However, in one study [5], DASS-21 was found to fail discriminate between symptoms associated with depression and anxiety in certain communities. The final diagnosis of PD is based on clinical assessment and evaluation by clinicians based on diagnostic guidelines such as the Diagnostic and Statistical Manual of Mental Disorders and the International Classification of Diseases. However, the initial screening approach relies on patient honesty in self-report, which is also subjected to the individual's self-awareness, memory, and sociocultural factors. In addition, most of the screening questionnaires used were for hospital-level research and studies only. There is no such mechanism for identifying individuals who with psychiatric disorders from a healthy population. Therefore, reliable diagnostic tools are needed that can easily and objectively assess and predict PD and that also take psychophysiology into account.

Photoplethysmogram (PPG) is one of the bio-signal methods used in the assessment of autonomic nervous system dysfunction (ANS) in PDs [6]. The signals are closely related to vital signs such as heart rate (HR), respiratory and blood pressure, which can be used to measure oxygen saturation and cardiac output and to assess autonomic function [7]. PPG uses infrared light to estimate the movement of blood under the skin, and a signal is divided into two main points: systolic and diastolic. A PPG waveform cycle consists of the onset, systolic peak, dicrotic notch, and diastolic peak [8]. PPG signals can provide information about pulse rate variability (PRV) and blood pressure modulated by ANS. Therefore, PPG features can indicate the psychological status of a person. Recent studies show that PPG signals can detect individual stress levels, anxiety, and major depression disorders (MDD) by PRV [9–11]. In general, PD patients showed a lower response to PRV than healthy subjects during a given mental task [12]. For example, in a study [13] using PPG sensors,

978-1-6654-8237-0/22/$31.00©2022 IEEE

patients with bipolar disorders were found to have lower PRV in the manic state than in the euthymic state. Another study [14] found that there were significant differences in cardiac function when studying PD patients and healthy controls during long-term continuous physical activity monitoring. In another study [15], the PPG signal was used to analyze PRV differences between schizophrenia patients and healthy individuals by an auditory stimulus test and found that PRV was lower in patients than in control subjects. Finally, studies by [16] found a positive theoretical relationship between mental health and the PPG pulse waveform. Preliminary results of the study [10] also suggest that PPG can be a promising emotion recognition tool and is suitable for human-machine interaction applications.

Despite all these findings, it remains difficult to treat patients with early-stage mental health problems because of a lack of reliable and consistent outcomes. The development of rapid and reliable automated screening methods is needed to refer at-risk patients to a psychiatrist or psychologist for further treatment. Machine learning (ML) is a very useful tool in healthcare, especially in implementing the internet of things [17]. Pattern recognition algorithms can be used to develop predictive models for certain diseases and can help with timely diagnoses, such as in studies [18]. Algorithms commonly used in ML are discriminant analysis (DA), k-nearest neighbors (KNNs), decision trees (DTs), support vector machines (SVMs), and artificial neural networks (ANNs), which predict and classify future events. In this paper, we aimed to predict PD from PPG signals by analyzing a case-control group on the basis of the reactivity of ANSs by using the mentioned ML algorithms.

II. Materials and Methods

The experimental paradigm consisted of five phases which are (1) data and PPG signal acquisition; (2) pre-processing and selection of a signal with good quality; (3) fiducial point detection from PPG signal followed by feature extraction; (4) feature selection based on significant *p*-value from statistical analysis and correlation analysis; and (5) the inputs of the most significant features to the ML process for the model prediction using a classifier with the best performance in terms of sensitivity (SN), specificity (SP) and accuracy (ACC).

A. Data Acquisition

The data acquisition for this study included participant demographic information and recordings of PPG signals. Sample size was calculated with the nomogram proposed by Gore and Altman, and the estimated prevalence rate of stress among Malaysians during the COVID-19 pandemic was 29% (P) [19]. This study targeted 95% for SN with a confidence interval (W) of 0.083. the estimated total number of samples was 92 and thus 46 subjects were included in each case and control group.

A total of 46 individuals met the criteria for the case group, currently having PDs and were recruited for this study. Their mean age was 24.76 (SD=5.43). All of them were patients admitted to the psychiatry clinic of Hospital Canselor Tuanku Muhriz (HCTM), Kuala Lumpur. They were clinically diagnosed with anxiety disorder (n = 13), depression (n = 20), both (n = 5), and schizophrenia (n = 8). The subjects were not assessed through DASS-21 as they had been clinically diagnosed. Then, 46 gender and age-matched normal subjects with a mean age of 23.80 (SD = 6.02) as the control participants were recruited from university students and members of the public who passed pre-screening through DASS-21. The age range for inclusion was 18-40 years. Informed written consent was obtained from all subjects, who were screened for any history of chronic medical conditions through verbal questioning. The exclusion criteria were currently having cardiovascular diseases and smoking, which may affect the PPG signals.

A pulse oximeter (CMS50E, Contec Medical Systems Co., Ltd., China) with a sampling frequency of 100 Hz was put on the index finger of each subject's left hand for PPG signal recording. The study protocol consisted of two sessions: the baseline (T1) and stimulus state (T2). The sessions were designed to evaluate differences between the case and control groups concerning the reactivity of their ANSs. In the T1 session, subjects' PPG signals were recorded under rested and relaxed conditions. In the second session, subjects underwent a Stroop color test. According to a 1935 study, the Stroop test is useful as a cognitive stress test and can elicit high levels of physiological stimulation [20]. Subjects took longer to read words printed in a different color than they did to read the same words printed in black. This approach is similar to previous research that used stimuli to detect changes in the ANS [21–22]. The total time for data recording was 10 min. Once PPG signal recording was completed, data was stored using the Smart Device Assistant, SpO_2 software, for further analysis.

B. Pre-processing

A second order butterworth high-pass filter with a cut-off frequency of 0.5 Hz, and a first order butterworth low-pass filter with a cut-off frequency of 6 Hz were used [23]. Following that, signal quality indexing (SQI) is used to establish trustworthy signal stability. The length of a quality PPG signal segment is determined by the SQI procedure. The SQI algorithm employed is based on the raw signal quality of PPG and employs a template-matching technique as proposed by prior works [24–25]. The SQI process is governed by three conditions: the extrapolated 10-second PPG signal must be between 40 bpm and 180 bpm; the PPG pulse-peak gap must not exceed 3 seconds to avoid missing more than one beat; and the maximum and minimum beat-to-beat interval within a sample must be less than 2.2. For feature extraction, only high-quality signals will be processed. This process is performed using MATLAB (Mathworks Inc., USA).

C. Feature Extraction

Peak position, mean of peak-to-peak intervals calculated over the analyzed segment, beats per minute (HR algorithms), and common heart rate variability (HRV) are the four variables involved to measure the performance of PRV features. For example, HR is calculated using the time interval between two successive peaks and a common PRV measure using the standard deviation of the successive differences. The calculated measures in the frequency domain are the low-frequency (LF) frequency spectrum between 0.05-0.15 Hz and the high-frequency (HF) frequency spectrum between 0.15-0.5 Hz. The measures are calculated from the power spectral density, which in turn is estimated using welch-based methods. The high performance of the algorithm has been reported in a study [26].

In addition, the association between other features of the PPG waveform was investigates, as this could lead to more diagnostic features and improve the of PDs because of the

great potential of this signal. The delineator [27] and bp_annotate [28] methods algorithms have been applied to detect the fiducial points in the PPG signals. The determination of pulse onset was related to the zero-crossing point before maximal diffraction was used to determine pulse start and the systolic peak was defined as the zero-crossing point following diffraction [27]. MATLAB was used in detecting these fiducial points in the PPG signal. As shown in Fig. 1, the raw PPG signal with the fiducial point where O is onset, S is systolic, N is a notch, D is diastolic, and *i* is the sequence of a cycle.

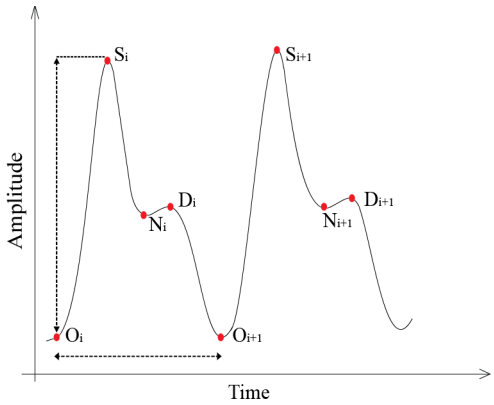


Fig. 1. Fiducial points in PPG or points of interest are identified on the pulse wave.

In total, ten HR/PRV features and twenty PPG pulse fiducial point detection features were obtained (see Table 1).

TABLE I. FEATURES AND LABELS COVERED BY THE STUDY

Features	Label	Details
HR	BPM, IBI	beats per minute, mean inter-beat interval
PRV	SDNN, SDSD, RMSSD, PNN20, PNN50, MAD, HF, LF	standard deviation of intervals between heart beats, standard deviation of successive differences between neighboring heart beats intervals, the root mean square of successive differences between neighboring heart beat intervals, proportion of differences between successive heart beats greater than 20ms, proportion of differences between successive heart beats greater than 50ms, and median absolute deviation of intervals between heart beats, high frequency, low frequency.
Onset	O_iO_{i+1}, O_iS_i, O_iN_i, O_iD_i, O_i–S_i, O_i–N_i, O_i–D_i	Time domain: $Onset_i$ to $Onset_{i+1}$, $Onset_i$ to $Systolic_i$, $Onset_i$ to $Notch_i$, $Onset_i$ to $Diastolic_i$, Apmplitude Domain: $Onset_i$–$Systolic_i$, $Onset_i$–$Notch_i$, $Onset_i$–$Diastolic_i$
Systolic	S_iS_{i+1}, S_iO_{i+1}, S_iN_i, S_iD_i, S_i–N_i	Time domain: $Systolic_i$ to $Systolic_{i+1}$, $Systolic_i$ to $Onset_{i+1}$, $Systolic_i$ to $Notch_i$, $Systolic_i$ to $Diastolic_i$, Amplitude Domain: Systolic –$Notch_i$
Notch	N_iN_{i+1}, N_iS_{i+1}, N_iO_{i+1}, N_iD_i	Time domain: $Notch_i$ to $Notch_{i+1}$, $Notch_i$ to $Systolic_{i+1}$, $Notch_i$ to $Onset_{i+1}$, $Notch_i$ to $Diastolic_i$
Diastolic	D_iD_{i+1}, D_iO_{i+1}, D_iS_{i+1}, D_iN_{i+1}	Time domain: $Diastolic_i$ to $Diastolic_{i+1}$, $Diastolic_i$ to $Onset_{i+1}$, $Diastolic_i$ to $Systolic_{i+1}$, $Diastolic_i$ to $Notch_{i+1}$

D. Statistical Analysis and Feature Selection

Statistical analysis was performed using IBM SPSS Statistical Version 26. Shapiro-Wilk test (p>0.05) and skewness and kurtosis (2.0<p<2.0) were performed for normality assumption test [29]. Independent t-test was performed for the normal distribution feature, while the Mann-Whitney U test was performed for the abnormal distribution feature. The features were then sorted according to the smallest to highest p-value, with a 95% confidence interval. Interquartile range assessment was used on potential features for isolated variability assessment. Significant features based on (p<0.05) were compared for the correlation analysis. Heat maps were used in finding features that were highly connected to a target and ensuring that no connections occurred between features. Feature selection primarily focused removing uninformed or redundant predictors from a model [30].

E. Classification

Five conventional ML models, DA, KNN, DT, SVM, and ANN have been utilized to differentiate case and control group data. Linear and quadratic types from a DA algorithm are used because DA is fast and can provide accurate discrimination between groups [31]. A KNN algorithm can determine the minimum distance (euclidean) of test and training data and categorize features with different numbers of neighbors, k value is set from 1 to 100. A DT classifier is used to produce "true" or "false" responses to training data with Gini's diversity index splitting criterion, and the number of splits is set from 1 to 100. An SVM algorithm is used to create decision boundaries (hyperplane) that can differentiate two groups using the four different kernel functions of linear, radial basis, and third- and fourth-order polynomial functions. In ANNs, a multilayer perceptron network is developed with two hidden layers and one output layer. The number of hidden nodes varied from 1 node to 15 nodes at each hidden layer. The network is trained with Levenberg-Marquardt training algorithm, the 'log sig' transfer function (hidden layers), the 'purelin' transfer function (output layer), and a training goal of 1 x 10^{-7} mean squared error (MSE).

These ML models are trained based on training and testing data groups, which are randomly divided and stratified according to a group with a division ratio of 70:30 (training and testing, respectively). Each ML model is fed with increasing number of features, from the least p-value to the highest one. The target is set in a binary of "0" and "1" for the control and case groups, respectively. The training is validated using a five-fold cross-validation algorithm. Then, model performance is compared in terms of MSE, SN, SP, ACC and receiver operating characteristic curves (ROC) analysis. In ROC, the false positive rate and true positive rate of the trained models were recorded with discrimination thresholds varying from 0 to 1. The area under the ROC curve was compared, and an area under the curve (AUC) value closer to 1 indicated a good predictor model [32]. The details of the performance evaluators are shown in (1-4):

$$SN = TP/(TP + FN) \quad (1)$$

$$SP = TN/(TN + FP) \quad (2)$$

$$ACC = (TP + TN)/(TP + FP + FN + TN) \quad (3)$$

where *TP* = number of PD accurately detected as the case group, *TN* = number of healthy controls accurately identified as the control group, *FN* = number of PD inaccurately detected as the control group, and *FP* = number of healthy controls inaccurately detected as the case group. While the *AUC* of the *ROC* curve is given by,

$$AUC = \int_0^1 ROC(t)\, dt \tag{4}$$

where t = (1 – specificity) and the $ROC(t)$ is sensitivity.

The study was approved by the Research and Ethics Committee of HCTM, with a registration number of UKM.PPI/111/8/JEP-2021-322.

III. Results

A. Data Acquisition

Table 2 shows a summary of participant characteristics, including sex, age, height, weight, BMI, and Stroop test accuracy. The control group included 16 males and 30 females (34.8% and 65.2%, respectively) with an average age of 23.8 years. The case group included 16 males and 30 females, with a mean age of 24.76 years. The number of female participants was higher than that of males because the prevalence of mental illness was higher in women than in men [33]. The statistical characteristics using Levene's Test for Equality of Variances showed no significant differences between the two groups regarding gender (p = 1.00), age (p = 0.491), BMI (p = 0.366), and Stroop test accuracy (p = 0.640).

TABLE II. SUBJECT DEMOGRAPHIC AND CHARACTERISTICS (N = 92)

Characteristics	Label		*P*
	Features control (*N*=46)	Case (*N*=46)	
Gender, male	16	16	1.00
Age	23.80 (6.02)	24.76 (5.43)	0.491
BMI (kg/m2)	23.49 (5.29)	24.20 (7.79)	0.366
ST Accuracy	96.10 (14.47)	93.99 (11.71)	0.640

B. Pre-processing

Once the PPG signal is filtered, the signal is subjected to a SQI process to determine a reliable signal. In Fig. 2, poor-quality PPG signal is highlighted in red color, and good-quality signal is indicated in blue color. The SQI is one of the artefact detection algorithms aimed at assessing the quality of physiological data [34].

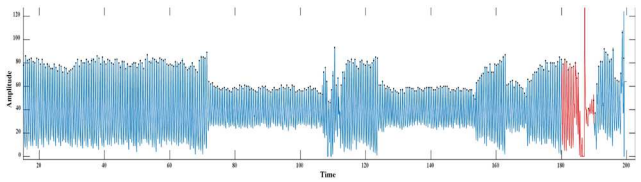


Fig. 2. PPG signals with low quality are red, while those with high quality are blue.

C. Feature Extraction

In this study, 60 features were extracted with the PPG signal from the subject's left hand based on the T1 and T2 tasks. HR and PRV features were extracted using Python, whereas other features from fiducial point detection (onset, systolic, notch, and diastolic) were extracted with MATLAB. These features were divided as follows: 48 from the time domain, four from the frequency domain, and eight from the amplitude domain. The features would be evaluated according to T1 and T2 for the discrimination between the case and control groups. Another 30 features were obtained from changes in features between T1 and T2. These features were extracted and used in assessing the effects of changes before and during subjects experienced a stimulus state, which is described as T3.

D. Statistical Analysis and Feature Selection

The rank of p-value in descriptive statistics was applied by analyzing the mean values for each feature. Table 3 shows that among the 90 features, 27 features with significant differences (p-values<0.05) were identified. The significant features were as follows: 13 of T1, 6 of T2, and 8 of T3. These features were ranked from the smallest to the largest according to the tasks.

TABLE III. SIGNIFICANT FEATURES ARE RANKED BY *P*-VALUE IN STATISTICAL ANALYSES FROM SMALLEST TO LARGEST

No	Feature	*p*-value	Task
1	N_iO_{i+1}_T1	<0.05	T1
2	HF_T1	<0.05	
3	MAD_T1	<0.05	
4	PNN50_T1	<0.05	
5	LF_T1	<0.05	
6	S_iO_{i+1}_T1	<0.05	
7	PNN20_T1	<0.05	
8	O_iO_{i+1}_T1	<0.05	
9	SDNN_T1	<0.05	
10	BPM_T1	0.001	
11	IBI_T1	0.001	
12	N_iN_{i+1}_T1	0.003	
13	N_iS_{i+1}_T1	0.007	
14	N_iO_{i+1}_T2	<0.05	T2
15	MAD_T2	0.001	
16	PNN50_T2	0.002	
17	HF_T2	0.012	
18	PNN20_T2	0.014	
19	SDNN_T2	0.023	
20	LF_T3	0.001	T4
21	N_iD_i_T3	0.005	
22	S_iO_{i+1}_T3	0.006	
23	HF_T3	0.011	
24	IBI_T3	0.018	
25	D_iN_i_T3	0.024	

26	BPM_T3	0.042	
27	PNN20_T3	0.044	

Next, the significant features were validated through correlation analysis. Fig. 3 through Fig. 5 shows the heat map of the correlation matrix and correlation strength (absolute value of the correlation coefficient, |r|) for the features from T1, T2, and T3. Based on the heat map, it was found that there are features that are highly correlated with each other and have similar information. If both are selected as ML inputs, the complexity of ML training increases, and ML performance is affected. Therefore, features that are highly correlated with each other will be discarded. Only features that are related to the group will be included in the ML.

In T1, PNN20 was removed as it had a high correlation (correlation $|r| > 0.8$) with PNN50 and MAD. S_iO_{i+1}, BPM, IBI, and N_iS_{i+1} were highly correlated to one another given that most measurements were close and consistent to a pulse measurement. Only the S_iO_{i+1} feature was selected as it had the least *p*-value and highest correlation strength. O_iO_{i+1} BPM, IBI, and N_iS_{i+1} features were removed. Correlation strength for T1 was in accordance with the *p*-value, and the least *p*-value feature had the highest correlation strength. This finding was also found in T2.

In T2, most features had a low correlation strength to their groups, except for the N_iO_{i+1} feature. PNN20 and SDNN were removed in the analysis as they were highly correlated to PNN50. Compared with the features in T1 and T2, the features in T3 had the least correlation strength to the groups, with a mean of 0.24 ± 0.0639. This result suggested that changes in the parameters measured from T1 to T2 had negligible differences from those in the control and case groups, as their correlations were weak. Thus, the features from T3 were excluded from the study.

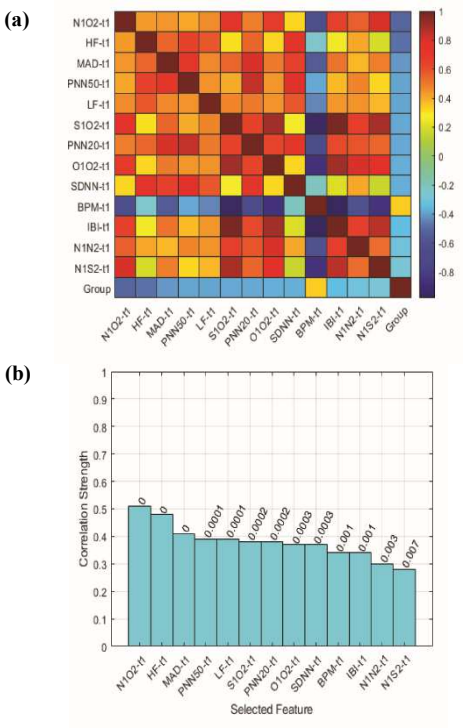


Fig. 3. (a) Correlation matrix of T1. (b) Correlation strength of T1.

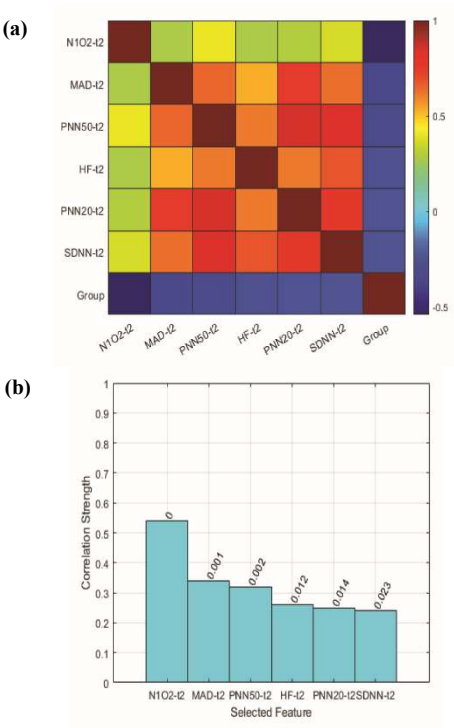


Fig. 4. (a) Correlation matrix of T2. (b) Correlation strength of T2.

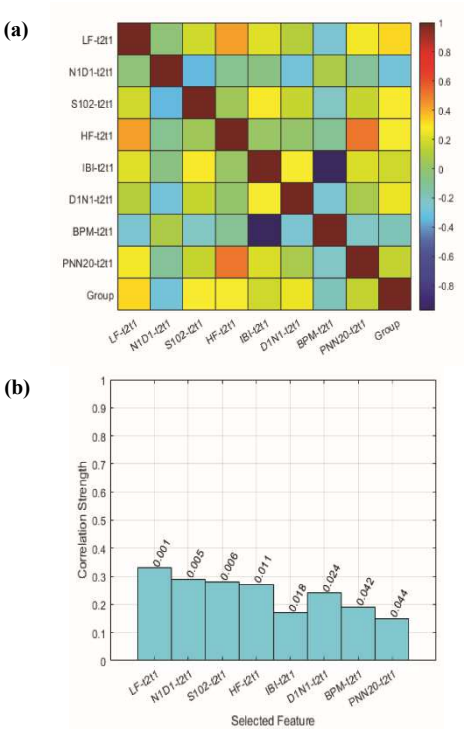


Fig. 5. (a) Correlation matrix of T3. (b) Correlation strength of T3.

Eight best features from T1 and four best features from T2 were used as mutual features in determining whether classification performance can be improved. Correlation matrix for the mutual features known as T4 is shown in Fig. 6.

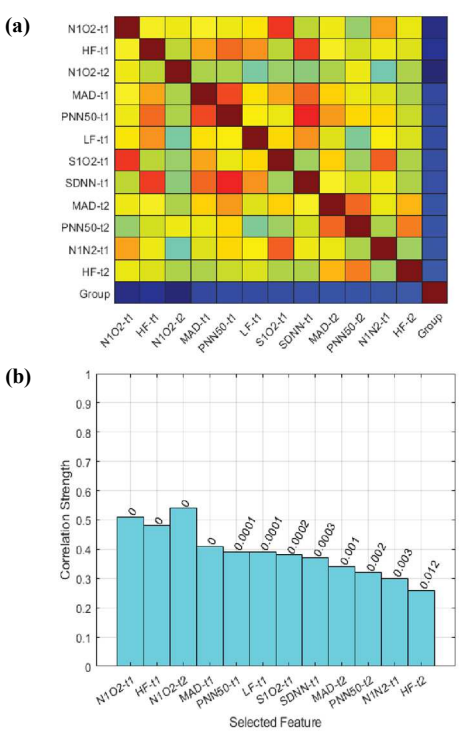


Fig. 6. (a) Correlation matrix of T4. (b) Correlation strength of T4.

No feature was correlated with a correlation strength of more than 0.8. Thus, all features are applied to classification. The average correlation strength of all 12 features was the highest (0.3925 ± 0.0817). The final list of the best features for T1, T2, and T4 is provided in Table 4. These features were used for classification using five ML techniques.

TABLE IV. SELECTED FEATURES AFTER CORRELATION ANALYSIS

Task	Selected Features
T1	N_iO_{i+1}_T1, HF_T1, MAD_T1, PNN50_T1, LF_T1, S_iO_{i+1}_T1, SDNN_T1, N_iN_{i+1}_T1
T2	N_iO_{i+1}_T2, MAD_T2, PNN50_T2, HF_T2
T4	N_iO_{i+1}_T1, HF_T1, N_iO_{i+1}_T2, MAD_T1, PNN50_T1, LF_T1, S_iO_{i+1}_T1, SDNN_T1, MAD_T2, PNN50_T2, N_iN_{i+1}_T1, HF_T2

E. Classification

Five different ML models are developed to classify case and control subjects using the selected features of T1, T2, and T4. Each model was trained with increasing input numbers, started from the least *p*-value, and randomly divided into trained and test data. The same random values were selected for each task. The accuracy of the best-trained models for each ML with different numbers of features is shown in Fig. 7. In general, ANN and KNN showed the best performance in T1 classification. KNN produced the highest accuracy (96.43%) in four features: N_iO_{i+1}_T1, HF_T1, MAD_T1, and PNN50_T1.

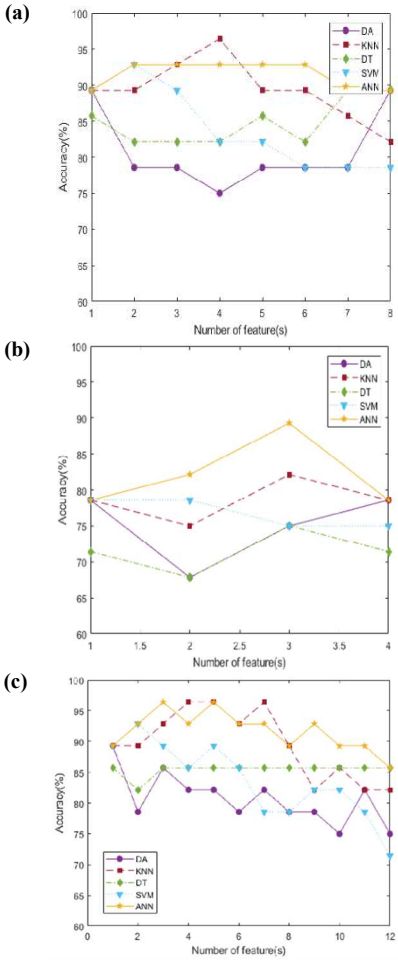
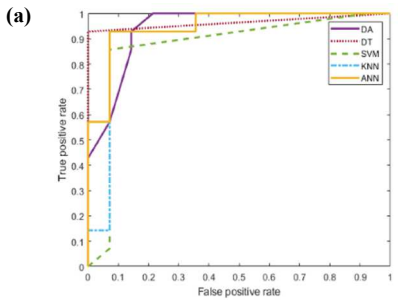


Fig. 7. Classification performance with an increasing number of features for (a) T1, (b) T2, and (c) T4, for each ML.

In T2 classification, the highest accuracy of KNN with seven neighbors and linear SVM and ANN with hidden nodes of nine and five was up to 89.29% when N_iO_{i+1}_T2, MAD_T2, and PNN50_T2 were used. Meanwhile, ML of T4 has able to achieve the same accuracy as T1 using ANN (hidden nodes = 11 at both hidden layers) with three features of N_iO_{i+1}_T1, HF_T1 and N_iO_{i+1}_T2. The sensitivity (92.86%) of the ANN model and its AUC (0.9949) are the highest as compared to KNN in T1 classification. The ROC curves and performance of the best trained model for each ML in T1, T2 and T4 classification are shown in Fig. 8 and Table 5, respectively.



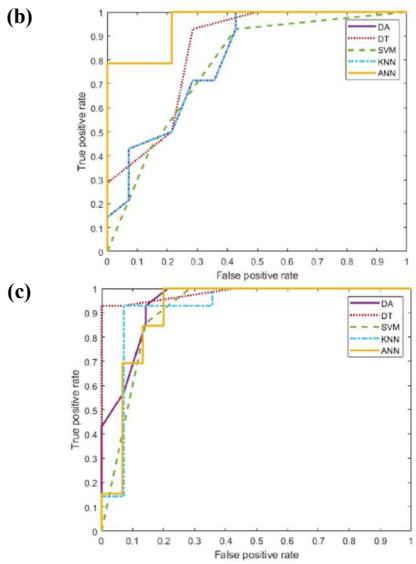


Fig. 8. ROC curves of the best trained model for each ML classification of a) T1, b) T2, and c) T4.

TABLE V. EACH ML HAD THE BEST TRAINED MODEL CLASSIFICATION PERFORMANCE IN T1, T2, AND T4

ML	DA	KNN	DT	SVM	ANN
Task	**T1 (*N*=8)**				
Trained features	1	4	7	2	2
ML Setting	Type= Linear	k=1	Split=1	Kernel= 3rd order Polynomial	HN1=5 HN2=8
Val MSE	0.250	0.308	0.231	0.154	0.154
Val ACC	75.00	69.23	76.92	84.62	84.62
Test MSE	0.107	0.036	0.108	0.071	0.071
%Test SN	92.86	92.86	85.71	92.86	92.86
%Test SP	85.71	100.0	92.86	92.86	92.86
%Test ACC	89.29	96.43	89.29	92.86	92.86
AUC	0.941	0.964	0.870	0.918	0.949
Task	**T2 (*N*=4)**				
Trained features	1	3	3	1	3
ML Setting	Type= Linear	k=7	Split=1	Kernel= Linear	HN1=9 HN2=5
Val MSE	0.417	0.154	0.500	0.385	0.308
Val ACC	58.33	84.62	50.00	61.54	69.23
Test MSE	0.214	0.179	0.250	0.214	0.107
%Test SN	100.0	92.86	92.86	100.0	78.57
%Test SP	57.14	71.43	57.14	57.14	100.0
%Test ACC	78.57	82.14	75.00	78.57	89.29
AUC	0.804	0.842	0.776	0.804	0.954
Task	**T4(*N*=12)**				
Trained features	1	4	1	2	3
ML Setting	Type= Linear	k=10	Split=1	Kernel= 3rd order Polynomial	HN1=11 HN2=11
Val MSE	0.250	0.154	0.077	0.154	0.154
Val ACC	75.00	84.62	92.31	84.62	84.62
Test MSE	0.107	0.036	0.143	0.071	0.036
%Test SN	92.86	92.86	85.71	92.86	92.86
%Test SP	85.71	100.0	85.71	92.86	100.0
%Test ACC	89.29	96.43	85.71	92.86	96.43
AUC	0.941	0.982	0.908	0.918	0.995

IV. DISCUSSIONS

The current study found that PPG features measured during T1 and T2 distinguished patients with PD from healthy controls, suggesting that PPG features are biomarkers for PD. We used an experimental protocol that included T1 and T2 to extract PPG features that responded to ANS activity and determined whether data from PPG signals can predict current PD during the stimulus period. In the physical assessments, the participants had similar mean values for gender, age, BMI, and cognitive level. The Stroop test confirmed that measures of attention, skill capacity, and processing speed ability did not differ between the case and control groups. All subjects in this study had almost the same cognitive level.

For the selection of features from a PPG signal, the correlation matrix showed that four out of the 27 features had high significance. N_iO_{i+1}, MAD, and PNN50 from the time domain and HF from the frequency domain, which had the lowest *p*-value and the highest correlation strength for the classification group. Additionally, all these features had lower mean values in the case group than in the control group. These results were consistent with previous studies [35], in which the PRV feature of depressed patients were lower than those of healthy people. N_iO_{i+1} is a potential feature for assessing stress influenced by changes in cardiovascular properties [36]. Moreover, PRV features (MAD, PNN50, and HF) are associated with ANS and respiration and may serve as the markers of pathological conditions in mental health [16], [37–39]. It has been used in previous studies related to anxiety [40], depression [41] and bipolar disorder [12]. Thus, by consolidating all the important features into T4, it can increase ability to predict PD.

The measure for feature changes (T3) from T1 to T2 was insignificant in terms of the strength of correlation to a group, suggesting no significant difference between the relaxed and stimulated states in both groups. These results were not consistent with previous studies in which significant differences in the stress task with respect to the relaxation state from healthy people only were observed [42]. The possible reason was the low level of stimulation, which may have induced participants to stress at a lower level. Other factors, such as the complex human response system to stress, should be considered as they vary between individuals. Therefore, the relationship of PPG features with various stress stimuli can be further studied.

PD was better predicted in T1 than in T2. The results were consistent with previous studies, where more significant differences in relaxation and recovery tasks than in a stimulated state, T2 [43]. The task may help differentiate responses in ANS activity between case and control groups.

Table 6 shows a comparison of performance between MLs that have been used in studies related to mental health. The best performance of the ML model in this study was observed in ANN with SN of 92.86%, SP of 100%, and ACC of 96.43%. The model used three combined features of N_iO_{i+1}_T1, HF_T1 and N_iO_{i+1}_T2. The performance of this study was higher than that reported in other studies.

TABLE VI. COMPARISONS WITH RELATED WORKS IN TERMS OF SN, SP AND ACC

Study	Method	Focus	Signal	Performance
[44]	SVM	Major Depressive Disorder	EDA	SN = 70% SP = 71% ACC = 70%
[45]	SVM-RFE	Major Depressive Disorder	EDA	SN = 74% SP = 71% ACC = 74%
[46]	KNN	Depression	EEG	SN (NA) SP (NA) ACC = 79.27%
[12]	SVM	Mental Disorder	ECG, PPG	SN = 71.40% SP = 93.80% ACC = 87%
Proposed Method	**ANN**	**Psychiatric Disorder**	**PPG**	**SN = 92.86%** **SP = 100%** **ACC = 96.43%**

V. CONCLUSION

Direct comparisons with PD classifications are relatively difficult because the development of predictive models mostly uses different bio signals, questionnaires, and social lifestyle information as inputs. PD classification using PPG signals is rarely reported. KNN produced comparable results to ANN because of its capability to handle noisy instances. An SVM algorithm was not fully optimized to minimize the classification error, as the c hyperparameter did not vary. DT performed better on mutually exclusive classes, but the data did not. DA required a normal distribution assumption on features, but about one-third of the total features were not normally distributed.

The limitation to this study was that it did not further classify PPG signals by PD type. This study included only a few categories of PD, which may not be optimal and reduces the variance of the classification measurements. Future work is needed to emphasize the collection of PPG signals to provide reliable SN, SP, and ACC classifications. This study shows that ANN provides 92.86% sensitivity, 100.00% specificity, and 96.43% accuracy compared with other ML models. Therefore, a PD prediction model was developed using ML techniques from PPG morphological extraction.

ACKNOWLEDGMENT

This research was supported by the Ministry of Education Malaysia's Fundamental Research Grant Scheme FRGS/1/2019/TK04/UKM/02/4.

REFERENCES

[1] A. S. B. Moni et al., "Psychological distress, fear and coping among Malaysians during the COVID-19 pandemic," PLoS One, vol. 16, no. 9, pp. 1–21, September 2021.

[2] Z. Khan, M. Makhbul, and Z. A. Rawshdeh, "Mental stress post-COVID-19," Int. J. Public Heal. Sci., vol. 10, no. 1, pp. 194–201, March 2021.

[3] H. Yaribeygi, Y. Panahi, H. Sahraei, T. P. Johnston, and A. Sahebkar, "The impact of stress on body function: A review," EXCLI Journal, vol. 16, pp. 1057–1072, July 2017.

[4] Z. Miao, Y. Wang, and Z. Sun, "The relationships between stress, mental disorders, and epigenetic regulation of bdnf," Int. J. Mol. Sci., vol. 21, no. 4, pp. 1375, February 2020.

[5] J. Aníbal González-Rivera, O. M. Pagán-Torres, and E. M. Pérez-Torres, "Depression, Anxiety and Stress Scales (DASS-21): Construct Validity Problem in Hispanics," Eur. J. Investig. Heal. Psychol., vol. 10, no. 1, pp. 375–389, January 2020.

[6] P. Kalra and V. Sharma, "Mental Stress Assessment Using PPG Signal a Deep Neural Network Approach," IETE J. Res., vol. 4, pp. 1-7, November 2020.

[7] M. Elgendi et al., "The use of photoplethysmography for assessing hypertension," npj Digit. Med., vol. 2, no. 1, pp. 1–11, June 2019.

[8] M. H. Chowdhury et al., "Estimating blood pressure from the photoplethysmogram signal and demographic features using machine learning techniques," Sensors, vol. 20, no. 11, pp. 3127, June 2020.

[9] P. Madhan Mohan, V. Nagarajan, and S. R. Das, "Stress measurement from wearable photoplethysmographic sensor using heart rate variability data," International Conference Communication Signal Process Processing (ICCSP), 6-8 April 2016, Melmaruvathur, India. pp. 1141–1144.

[10] D. Perpetuini et al., "Prediction of state anxiety by machine learning applied to photoplethysmography data," PeerJ, vol. 9, pp. e10448, January 2021.

[11] M. Rinkevičius et al., "Photoplethysmogram Signal Morphology-Based Stress Assessment," 2019 Computing in Cardiology Conference (CinC), 8-11 September 2019, Matrix, Singapore. vol. 45, pp. 116-119.

[12] M. Kobayashi, G. Sun, T. Shinba, T. Matsui, and T. Kirimoto, "Development of a Mental Disorder Screening System Using Support Vector Machine for Classification of Heart Rate Variability Measured from Single-lead Electrocardiography," 2019 IEEE Sensors Applications Symposium (SAS), 11-13 March 2019, Sophia Antipolis, France. pp. 1–6.

[13] A. Stautland et al., "Heart rate variability as biomarker for bipolar disorder," medRxiv, pp. 2022.02.14.22269413, February 2022.

[14] P. P. Filntisis et al., "Identifying differences in physical activity and autonomic function patterns between psychotic patients and controls over a long period of continuous monitoring using wearable sensors," Website, 10 2020. [Online]. Available: https://deepai.org/publication/identifying-differences-in-physical-activity-and-autonomic-function-patterns-between-psychotic-patients-and-controls-over-a-long-period-of-continuous-monitoring-using-wearable.

[15] S. A. Akar, S. Kara, F. Latifoğlu, and V. Bilgiç, "Analysis of heart rate variability during auditory stimulation periods in patients with schizophrenia," J. Clin. Monit. Comput., vol. 29, no. 1, pp. 153–162, February 2015.

[16] I. Liu, S. Ni, and K. Peng, "Happiness at Your Fingertips: Assessing Mental Health with Smartphone Photoplethysmogram-Based Heart Rate Variability Analysis," Telemed. e-Health, vol. 26, no. 12, pp. 1483–1491, December 2020.

[17] T. M. Ghazal et al., "IoT for Smart Cities: Machine Learning Approaches in Smart Healthcare—A Review," *Futur. Internet*, vol. 13, no. 8, pp. 218, August 2021.

[18] T. M. Ghazal et al., "Hep-pred: Hepatitis C staging prediction using fine gaussian SVM," Comput. Mater. Contin., vol. 69, no. 1, pp. 191–203, June 2021.

[19] L. S. C. Woon, H. Sidi, N. R. N. Jaafar, and M. F. I. L. Bin Abdullah, "Mental health status of university healthcare workers during the covid-19 pandemic: A post–movement lockdown assessment," Int. J. Environ. Res. Public Health, vol. 17, no. 24, pp. 1–20, December 2020.

[20] J. R. Stroop, "Studies of interference in serial verbal reactions," J. Exp. Psychol., vol. 18, no. 6, pp. 643–662, 1935.

[21] B. Zhang, Y. Morère, L. Sieler, C. Langlet, B. Bolmont, and G. Bourhis, "Reaction time and physiological signals for stress recognition," Biomed. Signal Process. Control, vol. 38, pp. 100–107, September 2017.

[22] Y. Liu *et al.*, "The salivary-α-amylase level after stroop test in anxious patients can predict the severity of anxiety," Neurosci. Lett., vol. 715, pp. 134613, January 2020.

[23] E. Pakdamanian, S. Sheng, and S. Baee, "Deeptake: Prediction of driver takeover behavior using multimodal data," Conf. Hum. Factors Comput. Syst. - Proc., 8-13 May 2021, Yokohama, Japan. pp. 1–14.

[24] Q. Li and G. D. Clifford, "Dynamic time warping and machine learning for signal quality assessment of pulsatile signals," Physiol. Meas., vol. 33, no. 9, pp. 1491–1501, September 2012.

[25] H. Ab Hamid and N. A. Nayan, "Methods of extracting feature from photoplethysmogram waveform for non-invasive diagnostic applications," Int. J. online Biomed. Eng., vol. 16, no. 9, pp. 39–62, August 2020.

[26] P. van Gent, H. Farah, N. van Nes, and B. van Arem, "HeartPy: A novel heart rate algorithm for the analysis of noisy signals," Transp. Res. Part F Traffic Psychol. Behav., vol. 66, pp. 368–378, October 2019.

[27] B. Nan, M. Chui, and M. I. Vai, "Biomedical Signal Processing and Control On an automatic delineator for arterial blood pressure waveforms," Biomed. Signal Process. Control, vol. 5, vol. 1, pp. 76–81, January 2010.

[28] A. Laurin, "BP_annotate - File Exchange - MATLAB Central." https://www.mathworks.com/matlabcentral/fileexchange/60172-bp_annotate (accessed Apr. 15, 2022)

[29] D. George and P. Mallery, *SPSS for Windows step by step: A simple guide and reference. 11.0 update*. Boston, MA: Pearson Allyn and Bacon, 2003.

[30] M. Kuhn and K. Johnson, *Applied Predictive Modeling with Applications in R*, vol. 26. New York: Springer, 2013.

[31] S. Uddin, A. Khan, M. E. Hossain, and M. A. Moni, "Comparing different supervised machine learning algorithms for disease prediction," BMC Med. Inform. Decis. Mak., vol. 19, no. 1, pp. 1–16, December 2019.

[32] W. Zhu, N. Zeng, and N. Wang, "Sensitivity, specificity, accuracy, associated confidence interval and ROC analysis with practical SAS® implementations," NESUG proceedings: Heal. Care Life Sci, 14-17 November 2010, Baltimore, Maryland. pp. 1–9.

[33] M. R. Salleh, "The Burden of Mental Illness: An Emerging Global Disaster," J. Clin. Heal. Sci., vol. 3, no. 1, pp. 5–12, June 2018.

[34] C. Orphanidou, T. Bonnici, P. Charlton, D. Clifton, D. Vallance, and L. Tarassenko, "Signal-quality indices for the electrocardiogram and photoplethysmogram: Derivation and applications to wireless monitoring," IEEE J. Biomed. Heal. Informatics, vol. 19, no. 3, pp. 832–838, May 2015.

[35] R. Hartmann, F. M. Schmidt, C. Sander, and U. Hegerl, "Heart rate variability as indicator of clinical state in depression," Front. Psychiatry, vol. 9, pp. 1–8, January 2019.

[36] P. H. Charlton, "Assessing mental stress from the photoplethysmogram: a numerical study," Physiol. Meas., vol. 39, no. 5, pp. 1–16, May 2018.

[37] M. Nardelli, A. Lanata, G. Bertschy, E. P. Scilingo, and G. Valenza, "Heartbeat Complexity Modulation in Bipolar Disorder during Daytime and Nighttime," Sci. Rep., vol. 7, no. 1, pp. 1–11, December 2017.

[38] N. Dillen, M. Ilievski, E. Law, L. E. Nacke, K. Czarnecki, and O. Schneider, "Keep Calm and Ride Along: Passenger Comfort and Anxiety as Physiological Responses to Autonomous Driving Styles," Conf. Hum. Factors Comput. Syst. - Proc., 25-30 April 2020, Honolulu, Hawaii. pp. 1–13.

[39] J. Firth et al., "The efficacy of smartphone-based mental health interventions for depressive symptoms: a meta-analysis of randomized controlled trials," World Psychiatry, vol. 16, no. 3, pp. 287–298, October 2017.

[40] A. Arza et al., "Measuring acute stress response through physiological signals: towards a quantitative assessment of stress," Med. Biol. Eng. Comput., vol. 57, no. 1, pp. 1–24, January 2019.

[41] A. Y. Kim et al., "Skin conductance responses in Major Depressive Disorder (MDD) under mental arithmetic stress," PLoS One, vol. 14, no. 4, pp. 1–13, April 2019.

[42] A. Y. Kim et al., "Automatic detection of major depressive disorder using electrodermal activity," Sci. Rep., vol. 8, no. 1, pp. 1–9, November 2018.

[43] H. Cai et al., "A Pervasive Approach to EEG-Based Depression Detection," Complexity, vol. 3, pp. 1–13, February 2018.