Classification of VR-Gaming Difficulty Induced Stress Levels using Physiological (EEG & ECG) Signals and Machine Learning

Sawon Pratiher [1], Ananth Radhakrishnan [1], Karuna P. Sahoo [1], SAZEDUL ALAM [2], Scott E. Kerick [1], nilanjan Banerjee [1], Nirmalya Ghosh [1], and amit patra [1]

[1]Affiliation not available
[2]University of Maryland

October 30, 2023

Abstract

"This work has been submitted to the IEEE Transactions on Affective Computing for possible publication. Copyright may be transferred without notice, after which this version may no longer be accessible."

Physiological sensing has long been an indispensable fixture for virtual reality (VR) gaming studies. Moreover, VR induced stressors are increasingly being used to assess the impact of stress on an individual's health and well-being. This study discusses the results of experimental research comprising multimodal physiological signal acquisition from 31 participants during a Go/No-Go VR-based shooting exercise where participants had to shoot the enemy and spare the friendly targets. The study encompasses multiple sessions, including orientation, thresholding, and shooting. The shooting sessions consist of tasks under low & high difficulty induced stress conditions with in-between baseline segments. Machine learning (ML) performance with heart rate variability (HRV) from electrocardiogram (ECG) and electroencephalogram (EEG) features outperform the prevalent methods for four different VR gaming difficulty-induced stress (GDIS) classification problems (CPs). Further, the significance of the HRV predictors and different brain region activations from EEG is deciphered using statistical hypothesis testing (SHT). The ablation study shows the efficacy of multimodal physiological sensing for different gaming difficulty-induced stress classification problems (GDISCPs) in a VR shooting task.

Classification of VR-Gaming Difficulty Induced Stress Levels using Physiological (EEG & ECG) Signals and Machine Learning

Sawon Pratiher, Ananth Radhakrishnan, Karuna P. Sahoo, Sazedul Alam
Scott E. Kerick, Nilanjan Banerjee, Nirmalya Ghosh and Amit Patra, *Senior Member, IEEE*

Abstract—Physiological sensing has long been an indispensable fixture for virtual reality (VR) gaming studies. Moreover, VR-induced stressors are increasingly being used to assess the impact of stress on an individual's health and well-being. This study discusses the results of experimental research comprising multi-modal physiological signal acquisition from 31 participants during a Go/No-Go VR-based shooting exercise where participants had to shoot the enemy and spare the friendly targets. The study encompasses multiple sessions, including orientation, threshold-ing, and shooting. The shooting sessions consist of tasks under low & high difficulty induced stress conditions with in-between baseline segments. Machine learning (ML) performance with heart rate variability (HRV) from electrocardiogram (ECG) and electroencephalogram (EEG) features outperform the prevalent methods for four different VR gaming difficulty-induced stress (GDIS) classification problems (CPs). Further, the significance of the HRV predictors and different brain region activations from EEG is deciphered using statistical hypothesis testing (SHT). The ablation study shows the efficacy of multimodal physiological sensing for different gaming difficulty-induced stress classification problems (GDISCPs) in a VR shooting task.

Index Terms—ECG; heart rate variability; EEG; brain waves; affective computing; VR shooting task; machine learning.

I. INTRODUCTION

A. Background

Since the 1970s, digital gaming's popularity as a leisure activity led game developers to explore new technologies and push the frontiers of game design [1]. Generally, digital gaming envisages two types of enhancements: game interface innovations for improved player experience and quantitative assessment of player's proficiency [2], [3].

Burgeoning research on physiological interfaces revolutionizes novel methods for digital play; player's mode of input [4], adaptability [5], player's affective state evolution [6], gaming experiences [7], and live streaming for inter-player communications [8]. The advent of self-tracking wearables for physiological data acquisition and feedback to the user in real-time via off-the-shelf gadgets such as the Fitbit[1] & Apple Watch[2] has mirrored the rising popularity and converged the ever-expanding affective computing research and incorporated physiological self-tracking into VR games [9].

S. Pratiher, A. Radhakrishnan, K.P. Sahoo, N. Ghosh, and A. Patra are with the Department of Electrical Engineering, Indian Institute of Technology Kharagpur, West Bengal 721302, India. (E-mail: sawon@ieee.org; ananth.8299@gmail.com; karuna.prasad@gmail.com; nirmalya@ee.iitkgp.ernet.in; amit.patra@ieee.org)

S. Alam and N. Banerjee are with the Computer science and Electrical engineering, University of Maryland-Baltimore County, 14701 Baltimore, Maryland, United States. (E-mail: sazedul1@umbc.edu; nilanb@umbc.edu)

S. E. Kerick is with the U.S. Combat Capabilities Development Command Army Research Laboratory, Aberdeen Proving Ground, Maryland, USA. (E-mail: scott.e.kerick.civ@mail.mil)

[1]https://www.fitbit.com/global/us/about-us
[2]https://www.apple.com/in/watch/

Affective gaming consolidates the psychological arousal and valence of a player's emotional state dynamics during the gameplay [10]. In general, VR games bring forth motivation, angst, frustration, and rage in the players [11], which make them inherently stimulating to implicitly impulse players to manage and relieve stress while pursuing their objectives [12].

In essence, VR games reward players for successfully down-regulating their negative emotional states by providing a secure environment to develop healthy emotion regulation [13]. Consequently, several studies are explored in [14], [15], concerning the linkages between the duration for which people are engaged in VR gaming, and affect down regulation in real-world reactions of emotional arousal.

B. Research Motivation and Contribution

One's interior physiological conditions are encoded as interoceptive awareness [16]. This interoceptive awareness level is directly proportional to one's internal state comprehensibility [17]. An enhanced physiological awareness manifests in better physiological management [12]. Individuals can be trained to view physiologically evocative stimuli and be more conscious of their internal temperaments (compared to control participants), and exhibit enhanced consistency between self-reported and physiological indicators of affect [18].

Typically, stress alters the brain's spontaneous electrical activity and manifests in cardiac repolarization heterogeneity and T-wave alterations, resulting in arrhythmia in the long run. As such, robust stress markers are essential for early-stage detection. EEG from the human head encodes the macroscopic neurological (brain) activities, and an ECG captures the electrical activity resulting from cardiac muscle depolarization and repolarization during each heartbeat. Therefore these physiological signals can capture the response of central and peripheral nervous system (CNS & PNS) to stress [19].

The interoceptive awareness is linked to adverse regulation effects and alexithymia (inability to recognize subjective events) [20]. Cannon-Bard's hypothesis provides empirical evidence of stress effect on the CNS and PNS. Hence, experimental research pushing the boundaries of interoceptive awareness realizes the significance of affective computing to be able to model mental state dynamics, and cognitive functionality [21], [22].

This paper examines the interoceptive awareness during a VR shooting mission by leveraging multimodal physiological signals. The current study expands on previous research on interoceptive awareness in physiological control by investigating an ML framework to infer GDIS levels in a VR shooting game. The workflow is depicted in Fig. 1. In addition, the following research contributions are derived from the design, application, and assessment of our system.
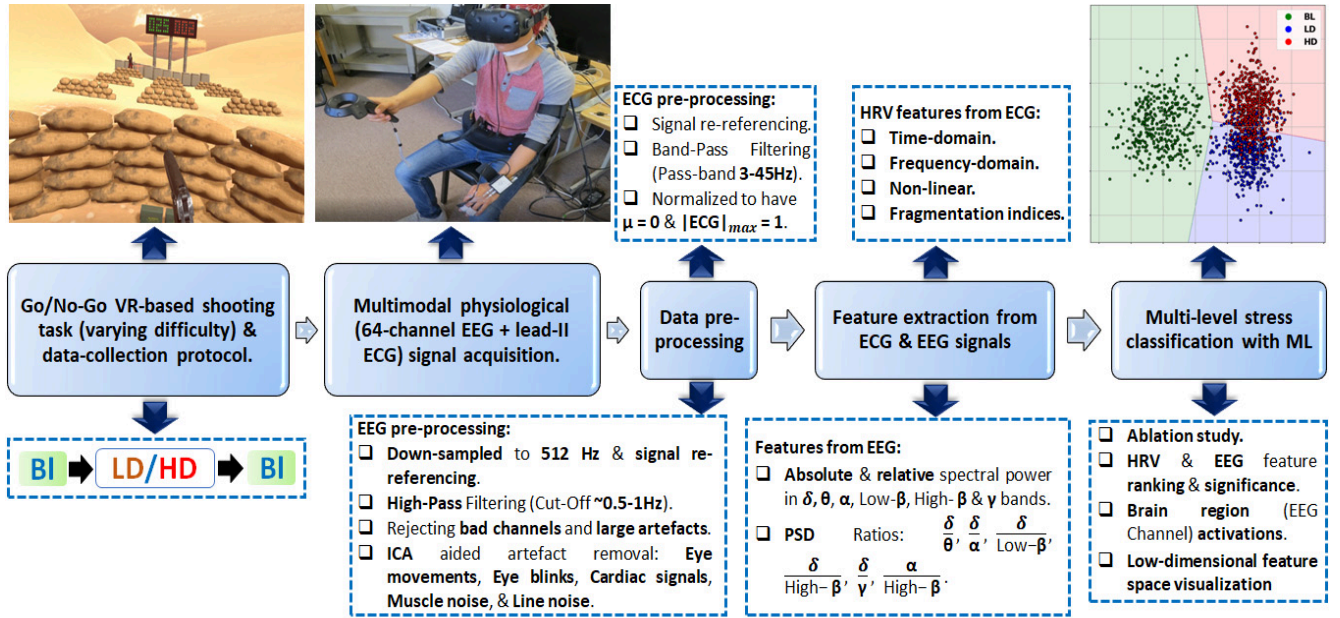
Figure 1: Block diagram for multimodal physiological signal-based GDISCPs using ML.

- To the best of our knowledge, the current study is the first to envisage an exhaustive multimodal GDIS classification using inter-class shooting task characterization, which was not reported in a consolidated way before.
- The framework was validated on physiological data acquired from 31 participants subjected to time-induced stress in a VR shooting task. The GDISCP scores outperform the prior art on stress delineation.
- SHT-based significance of HRV features and visualization of the brain region activations decode the ECG and EEG rhythms pertinent to various GDIS levels. The multimodal ablation study highlights the effect of individual (EEG/ECG) modalities for various GDISCPs.

Ongoing, the remaining paper is structured as follows: Section II reviews the prior art. Section III describes the data acquisition methodology. Section IV outlines the pre-processing and feature extraction. Experimental results on SHT and ML are discussed in Section V. Section VI concludes the paper.

II. RELATED WORK

The prior art on affective gaming dates back to CalmPrix, a racing game employing electrodermal activity (EDA) to gauge psychological arousal with a modified Apple II mouse [2]. The utility of motivational robotic assistance employing heart rate (HR) signals is explored in [23], and its alterations while playing Japanese Tetris 64 (by Nintendo 64) is studied in [24].

The competency of physiological gaming cues can be traced from the AffQuake [25], and Relax-to-Win [26]. These proliferating improvements in hardware and software led to the integration of various physiological signals into the digital gaming space as in Unity 3D[3], Source SDK[4], and Unreal Engine[5]. Physiological sensing using Empatica E4 wristband[6], Affdex Software Development Kit (SDK) by Affectiva [27],

and differentiation of positive and negative events using Microsoft Xbox's facial EMG [28] are prominent outcomes.

Investigating the role of physiological compliance and behavioral analytics in social gaming is an ongoing popular area of research [29], [30]. The VR literature spans several studies focused on stress detection using physiological signals [31], [32]. Compendious investigation on stress mitigation using ubiquitous bio-feedback methods in a mobile environment can be traced from [13].

A systematic literature review of existing contributions, research gaps, and recommended future thrust areas in the field of physiological sensing-based games and emotional gaming can be found in [24], [33]. A comprehensive comparison of prior art on stress delineation with commonly available modalities are summarized in Table IX.

The standards set by the European Society of Cardiology (ESC) and the North American Society of Pacing and Electrophysiology (NASPE) in 1996 provide guidelines for physiological and clinical interpretation of HRV measurements [34]. Standard HRV parameters include time-domain (refer Tables I & II), frequency-domain (refer Table III), and geometric indices (refer Tables IV & V).

Time-domain HRV features captures the temporal alterations in regular cardiac cycles. The standard deviation of the NN (normal-normal) intervals (SDNN) increases with increasing HRV fluctuations and measures the physiological resilience against stress [34]. The root mean square of successive NN interval differences (RMSSD) & percentage of successive NN intervals that differ by more than 50ms (pNN50) reflect the beat-to-beat variations [35], [36].

Frequency-domain HRV (spectral) features are associated with the sympathetic and parasympathetic branches of the autonomic nervous system [37]. The power in the high-frequency (HF) band is linked with the PNS activity and represents the vagus nerve activity, whereas the power in the low-frequency (LF) band indicates the sympathetic nervous system (SNS) activity [38]. An inverse relationship between the RR interval

and psychological stress has been reported, with a substantial rise in the LF/HF ratio, implying greater SNS activity during stressful times of the day [39]. However, HRV fluctuations has been shown to have a subjective nature with psychologically stressful circumstances and vary significantly between research [40].

Similarly, low parasympathetic activity is characterized by decreased HF power and increased LF power. During mental stress conditions, HR behavior has been shown to become more consistent and predictable with increased RR interval regularity and reduced complexity [41]. Claude Bernarde's pioneering research on associations between the brain and the heart has significantly motivated the present work [42].

The prior art on EEG-based stress detection depicts the alterations in spectral activity in all the EEG frequency bands during stress [43], [44], [45]. The α-wave (8-12 Hz) power is usually suppressed (negatively correlated) in the frontal lobe of the brain during active attention tasks [46] and increased during physically and mentally relaxed conditions [47]. Further, the alterations in left and right frontal activity of α-waves result in frontal asymmetry during stressful scenarios [21], [48]. On the contrary, the β-wave (12-30 Hz) power in the frontal lobe increases with higher demands on attention, cognitive load, and performance [45], [49].

The θ-wave (4-8 Hz) power increases in response to higher attention-demanding tasks and error-related processing [50]. The relative power in the γ band (30-100 Hz) also increases in the pre-frontal area during mental arithmetic tasks combined with negative social feedback [17], [51]. The δ-wave (0.5-4 Hz) power increases during cognitive tasks, which involve attention to internal processing with simultaneous inhibition of distractors [52], [53].

Apart from ECG & EEG, physiological modalities like electromyogram (EMG) in [31], [54], electrodermal activity (EDA) in [55], [56], [57], [31], [58], [59], galvanic skin response (GSR) in [49], [54], [60], [61], [62], [63], respiration rate in [31], [54], [64], [65], [63], blood pressure (BP) & blood volume pulse (BVP) in [57], [61], [31], [58], [59], [66], body temperature (BT) in [31] and photoplethysmogram (PPG) in [60] are reported for stress analysis. The present study exclusively analyzes the ECG-based HRV and EEG-based spectral features for GDISCPs in VR.

III. DATASET DESCRIPTION

A. Experimental Setup

An Institutional Review Board (IRB) approved human subject study was designed to evaluate the proposed multimodal GDISCPs framework as follows. The study was performed on (n=31) subjects (age: 24.99 +/- 3.21 years; male: 58%, female: 42%) at the University of Maryland Baltimore County (UMBC). There were no specific exclusion criteria for this study to retain the real-world noise inherent to actual VR-based training. The experimental setup is illustrated in Fig. 1.

The study comprised of multiple sessions, including an orientation session, a thresholding session, a shooting session, and a neuro-feedback session. The neuro-feedback session is outside the scope of this paper. The central task was a Go/No-Go VR-based shooting exercise where friendly and enemy targets pop up at random times in the VR environment. The participant had to shoot the enemy targets and not shoot the friendly targets (see Fig. 1). During the orientation session, the participants were familiarized with the experimental setup and allowed to perform practice trials, after which they participated in a thresholding session to adjust Target Exposure Time (TET) for similar stress for all subjects.

B. Data Acquisition Protocol

During the thresholding session, a participant played the shooting game in VR. Based on one's performance in that session, an individualized TET were determined, for enemy and friendly targets, from a Gaussian distribution using a target generation script. An individualized TET for each participant was established to normalize the individual differences in pre-experiment task performance for low and high-stress conditions. The psychophysical methods of limits or constant stimuli were followed [67].

Each shooting session consisted of low and high difficulty (LD/HD) tasks corresponding to low and high-stress conditions. The stress level was modulated by lengthening or shortening the TET. During the high stress (or HD) condition, a participant's TET was set according to their performance in the thresholding session, hitting approximately 50% of the enemy targets. Similarly, the TET was set for the low-stress (or LD) condition such that the player hits 90% of the enemy targets. A shorter TET was assumed to create higher stress on the participant as shooting the target would be difficult within that duration. TET ranged between 0.3-1.36 sec, with an average TET of 0.56 sec for HD and 0.86 sec for LD tasks. Since the low/high-stress condition was GDIS level, the term LD/HD task has been used synonymously with low/high-stress throughout this manuscript.

There were a total of six experimental sessions performed by each participant. Each session consisted of one LD task and one HD task interspersed with a one-minute resting baseline (Bl). Each LD/HD task had 4 blocks, and 90 targets per block (81 enemy, 9 friendly) appeared on the screen in sequence with delay governed by personalized TET. If the participant failed to hit the enemy, it shot back. The order of the LD and HD tasks was randomly counterbalanced across all participants and sessions. Each LD/HD task had an initial resting Bl, when the participants were asked to remain quiet, still, and stare at a fixed dot at the screen's center. The session concluded with another Bl. After artifact removal and signal pre-processing, a total of 160 sessions across 31 subjects are collected for analysis. A total of 521, 672, and 671 records for the Bl, LD, and HD classes, respectively, were used to validate the proposed method. A detailed description of the VR study and data collection protocol has been published in [68].

IV. METHOD OF ANALYSIS

A. ECG Record Statistics & Pre-processing

The current study involves lead II ECG signals acquisition using disposable spot electrodes on the thorax at a sampling frequency of 2048 Hz. ECG signals are first normalized (mean = 0, absolute maximum = 1) and denoised (see Fig. 1) by finite impulse response (FIR) bandpass filter (3-45 Hz). Then a multi-level Teager Energy Operator is used to delineate the P, Q, R, S, and T fiducial points on an ECG signal (refer Fig. 2). The positive maxima of the typical QRS complex constitute the R peaks and are used for RR interval abstraction [69].

The time between consecutive R-peaks gives a corresponding RR interval, and successive RR intervals constitute the RR interval time-series [69]. After removing the outlier points related to non-sinus-node depolarization (see Fig. 2), the NN interval time-series is obtained [70].
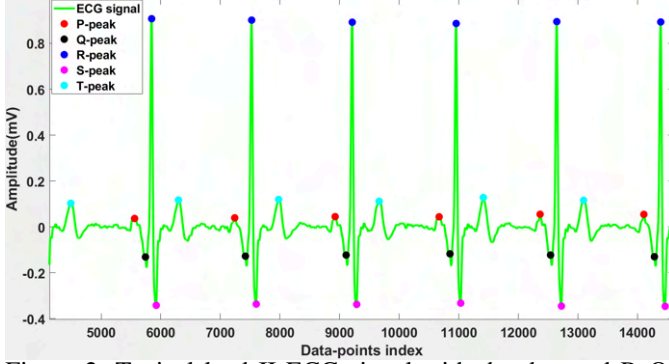

Figure 2: Typical lead II ECG signal with the detected P, Q, R, S, T fiducial points for subsequent feature extraction.

The NN interval time-series is generated from RR-interval time-series[7] using three outlier-removal filters in sequence: (a) range-filter with RR(n) within [0.32, 1.5] sec; (b) quotient-filter with ratio RR(n)/RR(n+1) within [0.75, 1.25]; and (c) moving-average filter with RR(n) within +/- 20% band around mean of RR [n-10, n+10] (see Fig. 3). The filtering parameters are selected from physiological limits [71].
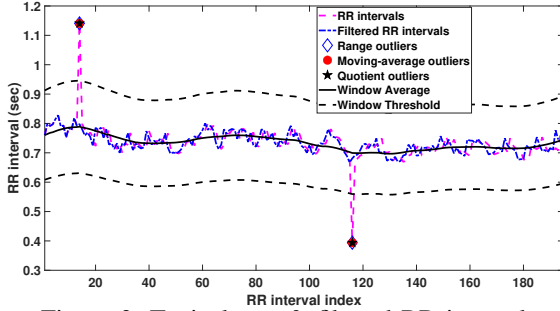

Figure 3: Typical raw & filtered RR intervals.

B. HRV Features Extracted from ECG Signal

The NN interval time series is used for HRV analysis [72], [73]. The HRV attributes measure the NN interval variability [74] and encapsulate the temporal fluctuation in a heartbeat sequence [75]. In this work, the following family of HRV descriptors are extracted for analysis: moment statistics (Table I; [76]), time-domain (Table II) and frequency domain features (Table III), non-linear NN interval spectral features (Table IV; [77], [78]), and time-domain fragmentation indices (Table V; [79]). For comprehensive HRV details, refer to [80], [81].

Similarly, time-domain features of NN-intervals (refer Table II) quantifies the degree of temporal variability in interbeat intervals (IBI) as illustrated in Fig. 4.

The power spectral density (PSD)-based frequency-domain HRV features comprehends the strength of the power distribution in different spectral components [37], [38]. This work explores three different spectrum estimation techniques: Lomb-scargle periodogram (Lomb), Yule-Walker autoregressive (AR) model, and Welch's method with overlapping windows.

---

[7]https://www.graduate.technion.ac.il/Theses/Abstracts.asp?Id=30274

Table I: Moment statistics (geometry) of HR distribution [76].

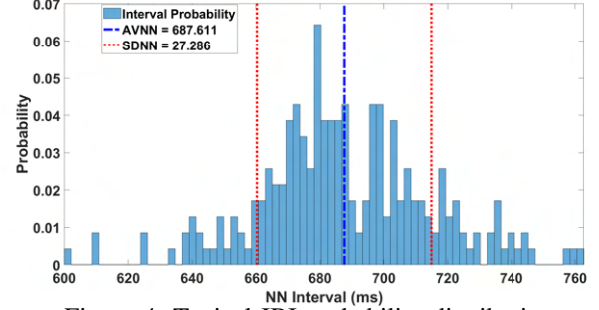| Feature	Description
μ_{HR}	Mean (expected value) of HR.
σ_{HR}	Standard deviation (dispersion/spread) of HR.
$Kurt_{HR}$	Kurtosis (tailedness).
$Skew_{HR}$	Skewness (asymmetry).

Figure 4: Typical IBI probability distribution.

Table II: Description of Time-domain HRV features.

Feature	Description
AVNN	Average duration of the NN intervals.
SDNN	Standard deviation of the NN intervals.
RMSSD	Root mean square of successive NN interval differences.
pNN50	Percentage of successive NN intervals that differ by more than 50ms.
SEM	Standard error of the mean NN intervals.

Following the physiological limits [71], a Hamming window with 50% overlap has been used for the Welch method. For the Yule-Walker AR method, linear detrending with an AR model order of 24 and no windowing is employed. The least square based Lomb-Scargle periodogram (LSP) method was found more robust compared to above-mentioned methods (Fig. 5) due to the presence of long-periodic noise in the long gapped ECG records [82]. Hence LSP method has been used in this study [39], [41], [42].
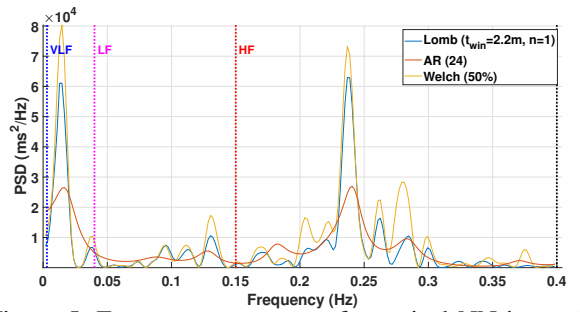

Figure 5: Frequency spectrum of a typical NN interval.

Different HRV spectral features (see Table III) are extracted in high-frequency band (HF; in [0.15, 0.4] Hz), low-frequency band (LF; in [0.04, 0.15] Hz) and very low-frequency band (VLF; in [0.003, 0.04] Hz). Fig. 6 depicts the slope of the log-log spectrum in the VLF range for typical NN intervals using different spectral estimation methods.

Non-linear HRV features measure the statistical randomness and periodicity in the NN intervals as well as the underlying complexity and non-linearity of the mechanisms governing

Table III: Description of Frequency-domain HRV features.

| Feature	Description
VLF_{Power}	Absolute power in the VLF band.
LF_{Power}	Absolute power in the LF band.
HF_{Power}	Absolute power in the HF band.
$Total_{Power}$	Combined total power in all the three bands .
$LF/HF Ratio$	$\frac{LF_{Power}}{HF_{Power}}$.
LF_{Peak}	Peak frequency in the LF band.
HF_{Peak}	Peak frequency in the HF band.
VLF_{Norm}	$100 * \frac{VLF_{Power}}{Total_{Power}}$.
LF_{Norm}	$100 * \frac{LF_{Power}}{Total_{Power}}$.
HF_{Norm}	$100 * \frac{HF_{Power}}{Total_{Power}}$.
β_{VLF}	Slope of log-log frequency plot in the VLF band.
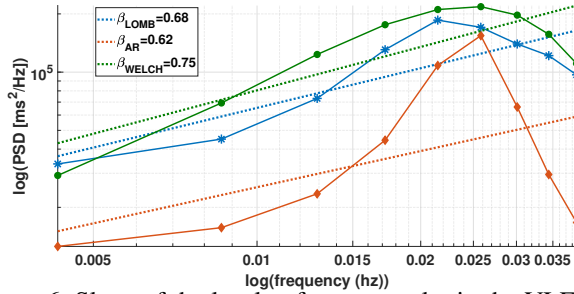
Figure 6: Slope of the log-log frequency plot in the VLF band.

the HRV regulation process. The non-linear HRV descriptors (refer Table IV) are abstracted from different non-linear signal processing techniques like detrended fluctuation analysis (DFA), Poincare plot, and multiscale sample entropy (MSE). Before employing the above-mentioned non-linear methods, the NN intervals are detrended to remove (any) systematic shift, resulting from electrodes (sensor) drift[8].

Table IV: Non-linear HRV features from NN intervals.

Feature	Description
α_1	Exponential scaling factor from DFA in the low-scale region (short-term correlations; see Fig. 7).
α_2	Exponential scaling factor from DFA in the high-scale region (long-term correlations; see Fig. 7).
SampEn	Sample entropy, which measures the statistical regularity and complexity.
SD_1	σ along the line perpendicular to the line of identity in the Poincare plot (see Fig. 8).
SD_2	σ along the line of identity in the Poincare plot (see Fig. 8).

Following the human physiological limits [71], the minimal and maximal DFA block-size is taken as 4 & 64, respectively, with incremental step = 2. The range for the exponential scaling factors α_1 & α_2 are taken as [4, 15] & [16, 64], respectively, for the low & high-scale regions (Fig. 7).

Following [83] a multiplicative factor of 2 and 3 are selected for the perpendicular line (SD_1) and the line of identity (SD_2) (see Fig. 8) for Poincare plots. Similarly, the MSE computation parameters are set to a normalized standard deviation of 1, a matching template threshold value of 2 for the distance between two points, and a maximal scale of 20 [71]. Fig. 9 exemplify the typical MSE variation across different scales with the sample entropy value.
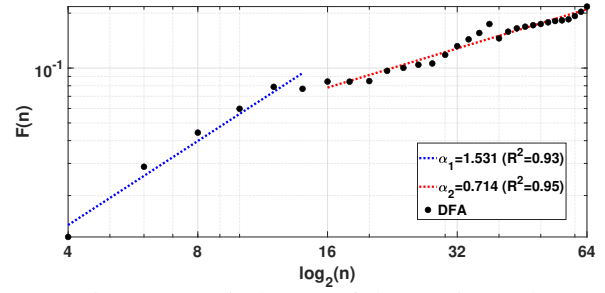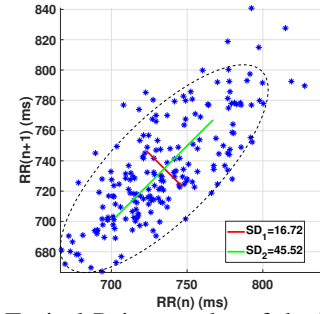

Figure 7: Typical DFA of the NN intervals.

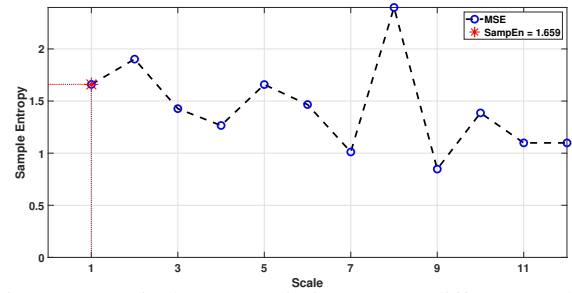Figure 8: Typical Poincare plot of the NN intervals.

Figure 9: Typical MSE variation across different scales.

Generally, short-term HRV manifests in physiologic vagal tone modulation [84]. However, the appearance of irregular short-term HRV with age and cardiovascular illness, which is consistent with the breakdown of the neuroautonomic-electrophysiologic regulatory mechanism, is recently identified in [79], as HR fragmentation indices (refer Table V).

Table V: Description of HRV fragmentation indices [79].

Feature	Description
PIP	Percentage of inflection points, i.e., % of zero-crossing points in the incremental NN interval time series.
IALS	Inverse of the average length of the acceleration/deceleration segments.
PSS	Percentage of short segments, i.e., % of NN intervals in acceleration/deceleration segments in less than 3 NN intervals.
PAS	Percentage of NN intervals in alternation segments, i.e., % of NN intervals in alternation (HR's acceleration changes sign) segments of at least 4 intervals.

C. EEG Signal Pre-processing and Artifact Rejection

The EEG signals are acquired using a 64 channel EEG cap from the Biosemi ActiveTwo System (BATS)[9] with scalp electrodes arranged according to the standard 10-10 layout

[8]https://physionet.org/content/pcst/1.0.0/

[9]https://www.biosemi.com/products.htm

system. The EEGLAB Toolbox[10] (ver. 14.1.2b) is used for signal pre-processing and artifact removal [85]. The raw EEG signals are passed through a high pass zero-phase Hamming-windowed sinc finite impulse response (FIR) filter with a cut-off frequency of 1Hz. The data per channel is downsampled from 2048 Hz to 512 Hz, after which they are subjected to the standard early-stage EEG Processing Pipeline (PREP), which includes removing line noise and robust re-referencing [86].

Then, the continuous data is epoched into non-overlapping 1 sec windows, and automatic artifact-rejection algorithms from EEGLAB are applied. The artifact-tagged epochs are visually inspected before being rejected. Next, the ocular artifacts due to eye movements and eye blinks are removed by using the Independent Component Analysis (ICA) technique for blind source separation [87] (see illustrative example in Fig. 10).
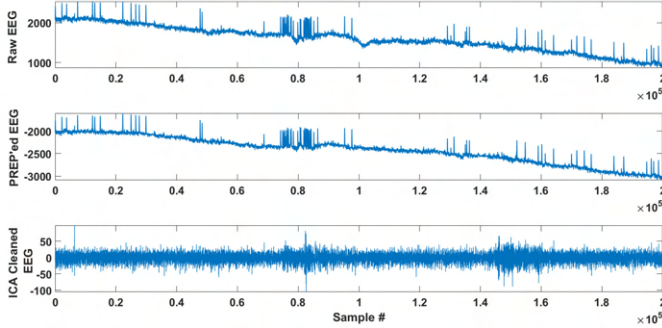

Figure 10: Typical raw EEG & its pre-processed versions.

D. Spectral Feature Extraction from Brain Waves (Table VI)

The EEG signals from each electrode are decomposed into different spectral bands having known distinct functional characteristics: δ **(1-4 Hz)**, θ **(4-8 Hz)**, α **(8-12 Hz)**, **Low-β (12-16 Hz)**, **High-β (16-30 Hz)**, and γ **(30-70 Hz)**. The PSD is estimated in each of these spectral bands using Welch's periodogram method with a Hanning window of size 4 sec and 50% overlap. For each band, absolute band power (ABP) is area-under-the PSD curve for that band, and relative band power $RBP = {}^{ABP}/_{Total\ Power}$. Three different feature

Table VI: Spectral features extracted per channel from EEG.

Feature	Description
ABP_k	Absolute power in the kth spectral band.
RBP_k	$ABP_k / Total\ Power$.
δ-to-θ	ABP_δ / ABP_θ.
δ-to-α	ABP_δ / ABP_α.
δ-to-Low-β	$ABP_\delta / ABP_{Low-\beta}$.
δ-to-High-β	$ABP_\delta / ABP_{High-\beta}$.
δ-to-γ	ABP_δ / ABP_γ.
α-to-High-β	$ABP_\alpha / ABP_{High-\beta}$.

Here k = δ, θ, α, Low-β, High-β, and γ.

sets have been used to study the BL-to-task variations. The first two sets consist of the ABPs and RBPs, computed on a per-channel basis for each frequency band and for each segment i.e Bl, LD task and HD task, in each session. The distribution of the ABPs and RBPs exhibited skewness [88]. Hence, they are transformed to decibel (dB) scale [89], to conform approximately to normality for statistical analysis

[90]. Comprehensive technical details on ABP and RBP-based various GDIS analysis is detailed in [88]. In this work, EEG feature set has been expanded to include the spectral ratios, i.e., δ/α, δ/θ, $\delta/Low-\beta$, $\delta/High-\beta$, δ/γ, and α/β (Table VI).

V. EXPERIMENTAL RESULTS AND DISCUSSION

A. Statistical Significance of Different HRV Features

The BoxPlots in Fig. 11, 12, 13, and 14 illustrate the variation in different HRV metrics corresponding to BL, LD, and HD classes. Further, the SHT scores depicted in Fig. 11, 12, 13, and 14 highlight the efficacy of a feature in inter-class segregation. Mathematically the SHT scores are computed as $SHT = -10 * log_{10}(pValue)$, where $pValues$ are obtained from Chi-square tests [91]. Chi-square tests are widely used in SHT to examine the statistical independence between a predictor variable and the response variable. The test statistic having a smaller $pValue$ translates to a larger SHT score and signifies a significant mutual dependence between the response variable and the corresponding predictor variable [92]. The mean feature values are computed by removing 4% of outliers at both the tail ends of the feature distribution.
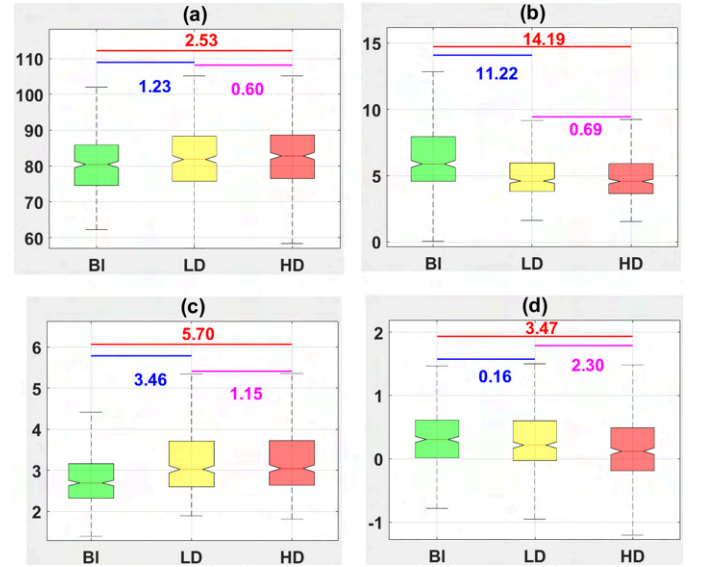

Figure 11: HRV distribution geometric features with differential significance $[-10 * log_{10}(pValue)]$ marked: (a) μ_{HR}, (b) σ_{HR}, (c) Kurt$_{HR}$ and (d) Skew$_{HR}$.

1) Moment statistics (Fig. 11(a)-(d)): Variations in HR moment statistics (Fig. 11) shows increasing trends in μ_{HR} (mean blood circulation; Fig. 11(a)) and Kurt$_{HR}$ (sharpness in activity around μ_{HR}) as well as decreasing trends in σ_{HR} (less variations, higher uniformity; Fig. 11(b)) and Skew$_{HR}$ (fewer outliers, always active; Fig. 11(c)) - especially for BL vs. (HD/LD) (see SHT), and less evident for LD vs. HD. The significant difference in SHT scores signifies variation in stress across BL vs. task, but less distinction between LD vs. HD. Kurt$_{HR}$ shows that, at LD/HD task, HR approaches Gaussian distribution (platykurtic; Kurt$_{HR}$ = 3; Fig. 11(d)) supported by σ_{HR} as well, while a sharp decrease of Skew$_{HR}$ in HD signifies all-time cardiac activity with increased GDIS level.

2) Non-linear HRV features (Fig. 12(a)-(e)): In general, SD_1 encapsulates the short-term HRV and is correlated with
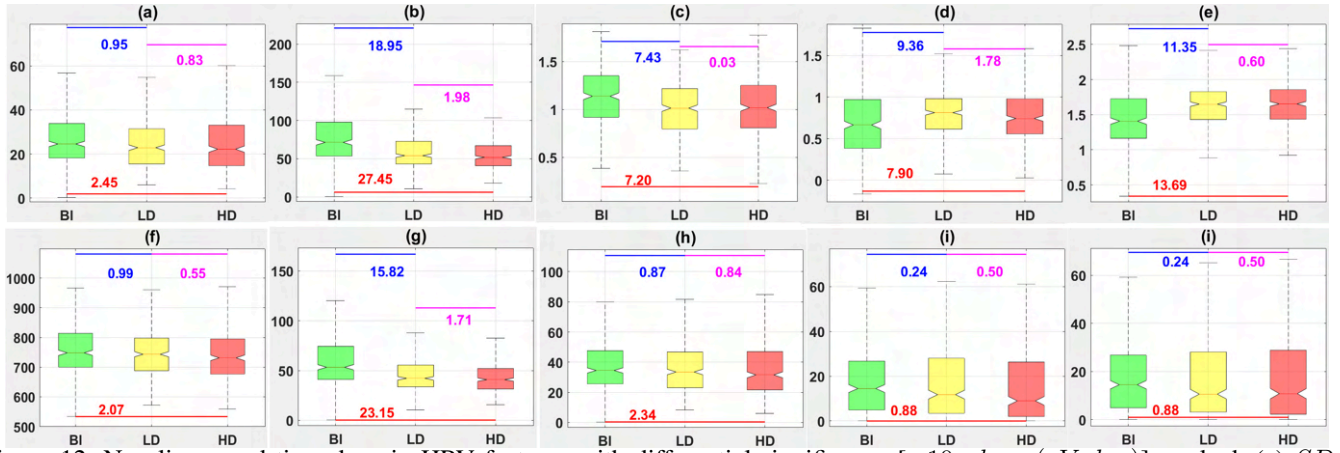
Figure 12: Non-linear and time-domain HRV features with differential significance $[-10*log_{10}(pValue)]$ marked: (a) SD_1, (b) SD_2, (c) α_1, (d) α_2, (e) SampEn, (f) AVNN, (g) SDNN, (h) RMSSD, (i) pNN50, (j) SEM.

the BRS and finds application in estimation of vagally mediated variations [93]. Similarly, SD_2 connects with the LF power and perceives both the short- and long-term HRV [94]. In the presented study, a decreasing SD_2 (signifying, LF power and BRS adaptation [95]), and non-discriminative variation in SD_1 is observed (Fig. 12(a)-(b)).

Likewise, DFA decodes the multi-scale associations in the NN interval time series [77]. α_1 is correlated with the baroreflex sensitivity (BRS), and α_2 reflects the regulatory mechanisms limiting the HR fluctuations [95]. The BRS is defined as the change in IBI (milliseconds) per unit change in BP. Here, a decreasing α_1 (BRS alterations; Fig. 12(c)), with an increasing α_2 (regulatory mechanism limiting HR fluctuation [84]; Fig. 12(d)), and a rising SampEn (i.e., regularity in the complexity of HR; Fig. 12(e)) are contemplated. In essence, all these trends point towards less HRV during LD/HD tasks (compared to BL) and consistent cardiac effort to handle higher physiological demands during shooting. However, the above-mentioned non-linear domain features perform poorly for the inter-class separation between LD and HD classes.

3) Time-domain HRV features (Fig. 12(f)-(j)): As expected, like SD_1, RMSSD (as well as AVNN, pNN50, SEM) did not show any changing trend as they characterize the very sudden transient change in HR (not expected), while the decreasing trend of SDNN matches with σ_{HR}.

4) HR fragmentation indices (Fig. 13(a)-(d)): The HR fragmentation indices are atypical short-term HRV signatures confounded with the disparities in HR acceleration sign dynamics. Physiological demands (and hence HR) were expected to rise/fall with appearing/disappearing targets and exhibit short-time acceleration and deceleration phases during shooting tasks. An increasing trend in all HRV fragmentation indices signifies an increase in short-term bursts in HR acceleration/deceleration phases during VR shooting exercise (Bl vs. LD/HD). However, the LD/HD inter-class separation is still fuzzy owing to the subjective nature of GDIS levels in VR.

5) Frequency-domain HRV features (Fig. 14(a)-(k)): A high differential significance $[-10*log_{10}(pValue)]$ with a decreasing β_{VLF} as well as reducing VLF_{Power} for BL vs. LD/HD signify reduced VLF activity of HR and increasing PNS activity [96]. In contrast, LF_{Norm} and HF_{norm} (signifying SNS activity) appear to be non-informative. Further, an in-

creasing HF_{Peak} and a decreasing trend in LF_{Peak}, LF_{Power}, HF_{Power} and $Total_{Power}$ (from BL to LD/HD) is observed. These observations signify decreasing biofeedback and HRV with shooting task (same as reducing σ_{HR} in Fig. 11(b)) and matches with prior published work [97]. A slightly decreasing trend in LF/HF over BL vs. LD/HD tasks signifies reducing LF (SNS activity) and an increasing HF (PNS activity) and hence increasing parasympathetic dominance during shooting mission, again echoing the current understanding [38], [98].
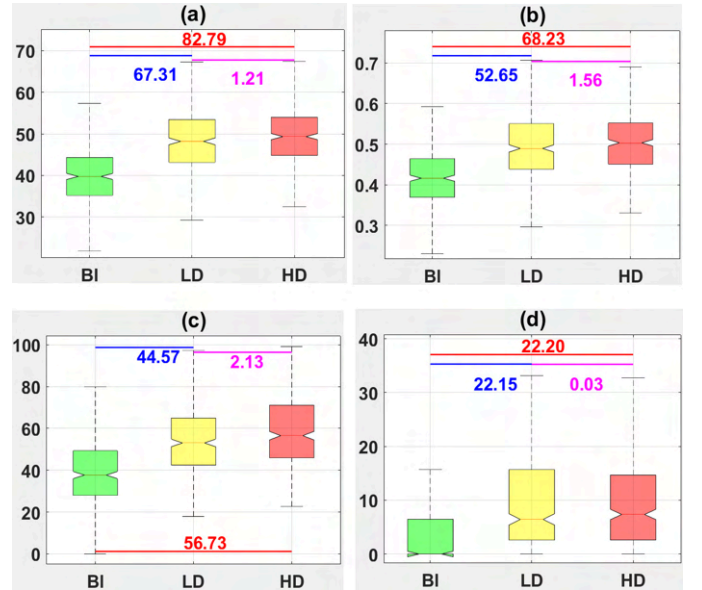
Figure 13: HR Fragmentation indices with differential significance $[-10*log_{10}(pValue)]$ marked: (a) PIP, (b) IALS, (c) PSS, and (d) PAS.

B. HRV Sensitivity Towards Different GDISCPs

The competence of σ_{HR} from moment statistics, PIP, IALS, PSS, & PAS from fragmentation indices, HF_{Peak}, LF_{Power}, & $Total_{Power}$ from frequency-domain HRV measures, SD_2 & SampEn from non-linear HRV metrics and SDNN & SEM from time-domain HRV attributes are apparent from the SHT scores (refer Fig. 11, 12, 13, & 14) for various GDISCPs. The SEM from time-domain HRV and the HR fragmentation

Figure 14: Frequency-domain HRV features with differential significance $[-10 * log_{10}(pValue)]$ marked: (a) β_{VLF}, (b) HF_{Norm}, (c) HF_{Peak}, (d) HF_{Power}, (e) LF_{Norm}, (f) LF_{Peak}, (g) LF_{Power}, (h) LF/HF, (i) $Total_{Power}$, (j) VLF_{Norm} and (k) VLF_{Power}.

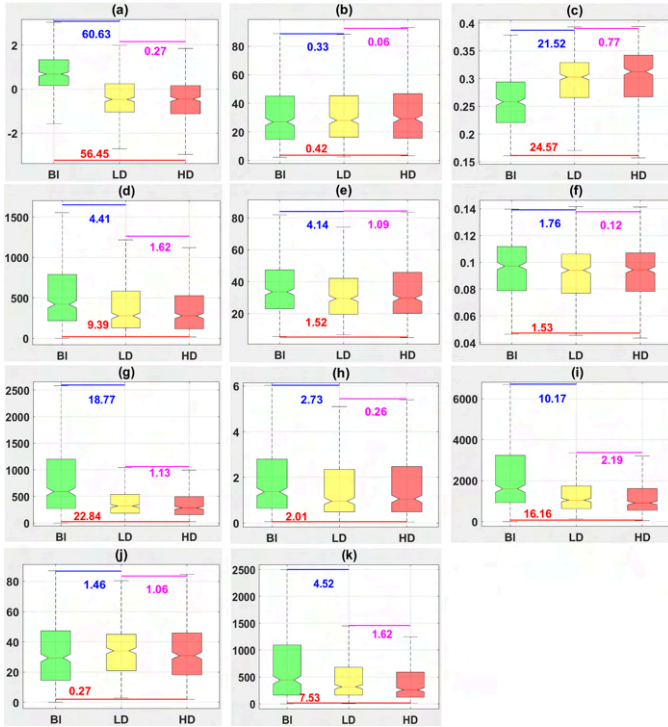indices are the most discriminative VR-stress metrics for different GDISCPs. These observations are consistent with the prior art in [37], [38], [39], [40], [41], & [42]. Most of the significant HRV descriptors show the maximum latent class separability for BL vs. HD and BL vs. LD classification. However, the sensitivity towards LD vs. HD classification is poor, as evidenced by the SHT scores, and may be attributed to the subjective nature of the GDIS levels with psychologically stressful circumstances [40].
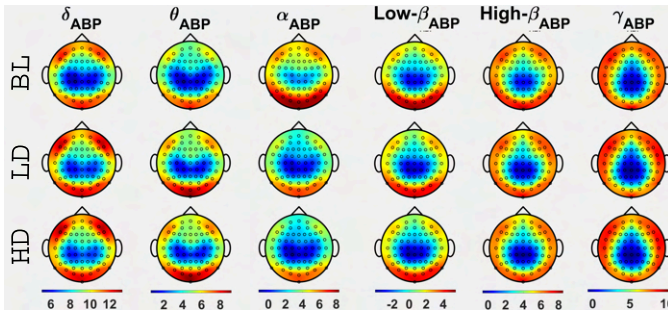

Figure 15: Variation in ABP values (dB) extracted per channel from 64-channel EEG signals.

C. Statistical Analysis of Spectral Features from EEG Signals (Fig. 15, 16, and 17)

Among EEG-ABP features (Fig. 15), except θ_{ABP} and α_{ABP}, no other frequency bands are discriminatory between BL vs. LD/HD tasks. θ_{ABP} shows increasing occipital-lobe activity (visual perception) and decreasing α_{ABP} indicates increasing activity in occipital lobe, in frontal lobe (intellectual

task) and in the central area (motor-control) during the LD and HD tasks compared to Bl.

For EEG-RBP features (Fig. 16), except for the γ_{RBP}, all the RBPs evince discriminatory brain region activations for the BL to tasks (LD/HD) case and lower disparities for in-between the tasks i.e. LD vs. HD case. δ_{RBP} exhibits the least relative spectral sensitivity in the parietal lobes for the BL class compared to other brain regions. However, during GDIS levels, the frontal lobe and the central cortex regions portray higher relative spectral activity than the rest.

Likewise, the θ_{RBP} manifests at least in the frontal and parietal lobes with a relatively enhanced spectral sensitivity in the central cortex region during the Bl. Conversely, during LD and HD tasks, the relative θ_{RBP} distribution is boosted throughout the brain. The frontal and parietal lobes show a slightly increased relative spectral power. There is a significant increase in θ_{RBP} in the occipital lobes, parieto-occipital, central pre-motor, motor cortex regions, and along the midline. The θ_{RBP} is slightly more prominent for the HD-based GDIS level compared to the LD-based GDIS level.
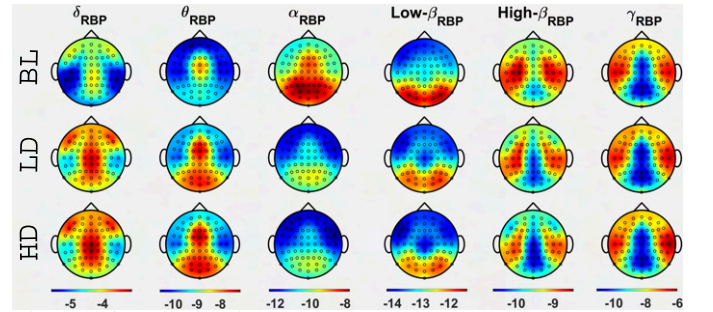

Figure 16: Variation in RBP values (dB) extracted per channel from 64-channel EEG signals.

The α_{RBP} exhibit relatively higher spectral power content in the occipital lobes, parieto-occipital, central pre-motor, and motor cortex regions compared to the frontal lobes during Bl. During the LD and HD tasks, the α_{RBP} values are reduced with the minimum relative power in the frontal lobes and a slightly sensitive central-parietal region. For the LD vs. HD case, HD-based GDIS evinces the least central-parietal region sensitivity compared to the LD-based GDIS level.

The Low-β_{RBP} exemplifies a comparatively higher relative power content during BL compared to LD and HD tasks. The Low-β_{RBP} values are minimum in the frontal lobes, and it increases as one traverses through the central cortex and parietal lobes and is maximum at the occipital lobe. From Bl-to-LD-to-HD classes, the apparent gradient of this trend varies as evident from the 4th **col** of Fig. 16. The Low-β_{RBP} values in the occipital lobes diminishes for LD and HD tasks and is the least for HD-based GDIS level. Similarly, the Low-β_{RBP} values in the frontal, central cortex, and parietal lobes are reduced for LD and HD tasks and are lowest for HD-based GDIS level.

The high β_{RBP} demonstrates a comparatively increased relative spectral power in the parietal lobe compared to the other brain regions across BL, LD and HD tasks. The spectral content in the parietal lobes reduces with an apparent disparity along the midline for LD and HD tasks. This apparent midline disparity is more pronounced for the HD-based GDIS level.

The γ_{RBP} shows the least discriminative sensitivity towards different GDIS levels (refer 6th **col** of Fig 16).
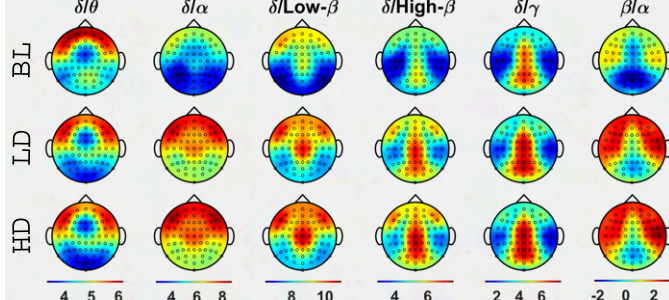

Figure 17: Variation in Spectral Ratios extracted per channel from 64-channel EEG signals.

Fig. 17 shows the variation in the different spectral ratios for Bl, LD, and HD classes. Here, δ/α, $\delta/Low-\beta$, $\delta/High-\beta$, and α/β exhibit the most discriminating features for GDIS level characterization. Across BL and LD/HD task levels, the δ/θ demonstrates higher values in the frontal lobe and lower values in the central cortex region. However, for LD and HD tasks, the δ/θ ratio reduces in the occipital lobe and is the lowest for the HD-based GDIS level. Similarly, the δ/α value in the frontal and occipital region increases from BL to tasks and is maximum during HD shooting task. Likewise, the $\delta/Low-\beta$ value in the pre-frontal, lateral-frontal, temporal, and occipital regions increases from BL to tasks. However, the distinction between LD and HD tasks is not substantial. For the $\delta/High-\beta$ and δ/γ, an apparent disparity along the midline for LD and HD tasks is observed. The parietal lobes activations for the $\delta/High-\beta$ increases during LD and HD tasks compared to Bl. However, the parietal lobes activations for the δ/γ are similar across BL and across different task levels. The α/β ratio evince relatively lower values in the occipital lobe and in the peripheral regions of the parietal, temporal and frontal lobes during Bl, and it increases with GDIS levels.

D. EEG Feature Significance for different GDISCPs

Fig. 18 & 19 depict the distribution of spectral feature importance scores across various channels for different GDIS-CPs. In essence, both these figures exhibit the sensitivity of the brain regions corresponding to different EEG predictors for different GDISCPs using the Chi-square test [91], [92]. Except for the LD vs. HD case, all the other GDISCPs shows discriminative brain region sensitivity in Fig. 18 & 19. The central motor cortex activations are observed for δ_{RBP}, α_{RBP}, Low-β_{RBP}, High-β_{RBP}, γ_{RBP}, δ_{ABP}, δ-to-α, δ-to-Low-β, δ-to-High-β, and δ-to-γ features for different GDISCPs with relatively higher values for BL vs. HD and BL vs. LD vs. HD cases. The α_{RBP} and the δ-to-α features show activations in the pre-motor cortex areas as well. These activations may be due to the voluntary actions and sequence of movements planned, coordinated, and executed by the subject during the shooting task.

Significant activations are seen for occipital and parieto-occipital regions for θ_{RBP}, θ_{ABP}, and δ-to-θ, for all the GDISCPs. Occipital theta activation could be due to higher cognitive demands for identifying and processing visual stimuli during the shooting task [99]. Similar significant activations are also visible in α_{RBP} and δ-to-α in occipital and parietal

Figure 18: Distribution of spectral feature importance scores (via Chi-square test) across channels for different GDISCPs.

regions for all the GDISCPs. They could be attributed to cognitive processes, and increased attention demands during the execution of the shooting task [99]. Some discriminative activations are also observed in the frontal region for α_{RBP}, δ-to-α and β-to-α features for all the GDISCPs and have been associated with internally directed attention and higher load on the working memory [99]. Increased β-to-α power ratio in the frontal and parietal lobes could also indicate higher cognitive load combined with a higher state of arousal during the shooting tasks [100].

E. Feature selection and ML-based GDISCPs

*1) **SHT-based Predictor Selection**:* As discussed in section V-A, the efficacy of a particular HRV or EEG feature in inter-class separation for various GDISCPs is comprehended from the SHT scores [91]. All the HRV and EEG descriptors are ranked based on the SHT values. The top $k-$ranked features are used for ML-based GDISCPs. Here $'k'$ is varied between 1 and ($\leq \#predictors$) in steps of 5, and an optimal value of $'k'$
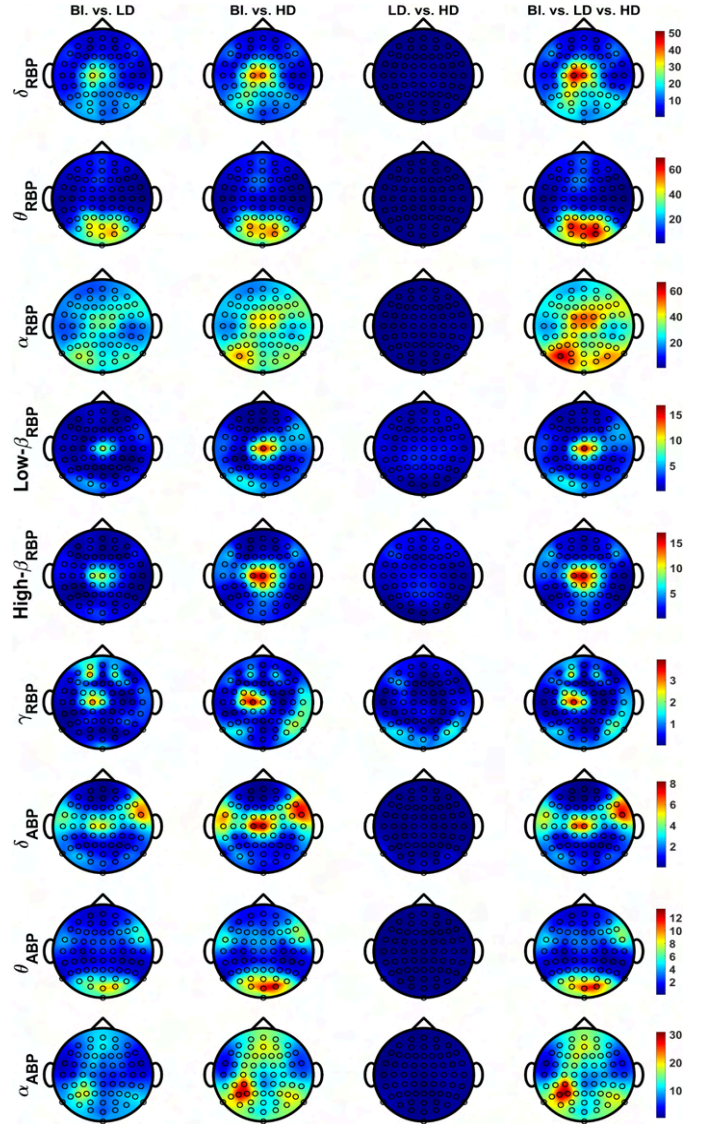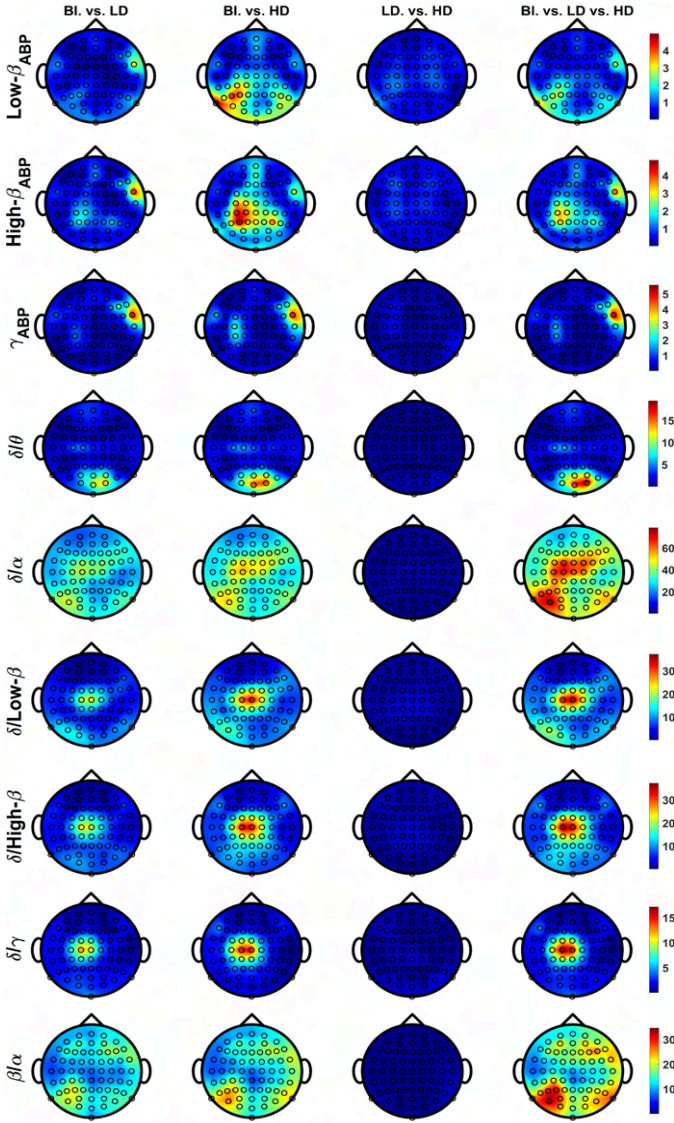
Figure 19: Distribution of spectral feature importance scores (via Chi-square test) across channels for different GDISCPs.

maximum splits. The parameter tweaking in KNN involves nearest-neighborhood size, which is varied between 1 to 20, and distance functions, L_1 & L_2 penalty. Likewise, the linear SVC hyperparameter tuning involves convergence tolerance and regularization constraint-based optimization.

The optimal RF hyperparameters are 195 trees with a maximum depth of 3. A minimum of 2 samples is required for an internal node to split with Gini impurity criterion and entropy-based information gain to measure the quality of a split. Random bootstrapping is envisaged for building trees with a maximum of $\sqrt[2]{\#features}$ for the best split. The fine-tuned GB hyperparameters embrace a learning rate of 0.09 with 110 trees, a minimum leaf size of 1, a maximum depth of 3, and a maximum of 2 splits with Friedman mean squared error for measuring the split quality. The optimal KNN parameters vary between 6 and 10 nearest neighbors with uniform neighborhood weights and L_2 distance for different GDISCPs. A L_2 penalty function with squared-hinge loss and the unity regularization parameter is used for linear SVC, and the tolerance criterion for convergence is set to 10^{-5}.

F. Training/Testing Protocol and Performance Metrics

Labeled multimodal features were randomly shuffled and underwent standard 10-fold cross-validation with $training : validation = 7 : 3$ for training using performance evaluating metrics (PEMs). The ML classifiers predicted labels can be categorized into TP (true-positive), TN (true-negative), FP (false-positive), and FN (false-negative). Based on TP, TN, FP, and FN values, the following PEMs are defined for assessment and consistency testing of the various GDISCPs: $Accuracy\ (Acc.) = \frac{TP+TN}{TP+TN+FP+FN}$, $Precision\ (Pr.) = \frac{TP}{TP+FP}$, $Recall\ (Re.) = \frac{TP}{TP+FN}$, $Macro\ Precision = \frac{\sum_{i=1}^{k} Precision_i}{k}$, and $Macro\ Recall = \frac{\sum_{i=1}^{k} Recall_i}{k}$. Here, $k = 3$ corresponds to the three classes (Bl, LD, & HD) [106].

G. Classification Results and Modality-based Ablation Study

The PEM scores ($\mu \pm \sigma$) are summarized in Table VII for different ML classifiers and compared across different GDISCPs with different modality combination. For the three-class CP (Bl vs. LD vs. HD), the macro-averaged PEM scores are reported. GB and linear SVC perform better than RF and KNN, with the latter manifesting better for BL vs. tasks (LD/HD). GB is more robust towards intra-task LD vs. HD and BL vs. LD vs. HD classification. From Table VII, it may be noted that for BL to tasks (LD/HD) CPs (i.e., LD vs. BL and HD vs. BL), the explored descriptors segregate the LD/HD-based GDIS levels vs. BL with a very high PEM score. However, for the LD vs. HD case, the PEM scores fall drastically due to the subjective nature of GDIS levels with psychologically stressful circumstances [40].

Using only ECG-based HRV features, an average classification accuracy of 96.4%, 95.5%, 71.0%, and 76.8% is obtained for BL vs. LD, BL vs. HD, LD vs. HD, and BL vs. LD vs. HD GDISCPs, respectively. Correspondingly, the EEG modality-based classification accuracies are 92.4%, 93.7%, 69.7%, and 72.9%, respectively is achieved. This performance comparison shows the efficacy of ECG compared to EEG in decoding the various GDISCPs. A similar trend is also observed for precision and recall scores. However, with combined features

is used for different GDISCPs and across various modalities. The classification metrics are tabulated in Table VII.

2) ML Classifiers for GDISCPs: This study uses four commonly used ML classifiers for different GDISCPs (i.e., BL vs. LD, BL vs. HD, LD vs. HD, and BL vs. LD vs. HD); K-nearest neighbor (KNN), random forest (RF), gradient boosting (GB), and linear support vector classifier/machine (SVC/SVM). Comprehensive details can be found in [101] for KNN, [102] for RF, [103] for GB and [104] for linear SVC. For different GDISCPs, the optimal hyperparameters of the ML classifiers are obtained using Bayesian minimization of the cross-validation loss function [105].

To abstract the prior/posterior distribution over the objective function, the hyperparameters are sampled from a mixture of the Gaussian Processes model. The hyperparameter in RF tuning comprises the number of trees ($\leq \#predictors$) in the ensemble and the forest's depth/complexity (≤ 20). Similarly, the GB hyperparameter fine-tuning encompasses learning rate, minimum leaf size, tree-complexity level, and

Table VII: Classification results with different MLCs & various modality combination for different GDISCPs.

MLC	PEM	Features from ECG signal				Features from EEG signal				Features from both ECG and EEG signal			
		Case-1	Case-2	Case-3	Case-4	Case-1	Case-2	Case-3	Case-4	Case-1	Case-2	Case-3	Case-4
KNN	Acc.	83.2±2.2	86.7±2.3	53.8±3.1	54.1±2.9	81.0±2.6	84.8±3.1	49.8±3.4	52.9±2.7	83.7±2.7	86.7±3.0	49.7±3.3	60.0±2.5
	Re.	93.3±2.0	91.6±2.8	49.7±5.1	52.9±2.8	92.9±2.6	91.5±2.0	48.8±5.8	54.1±2.8	93.7±2.4	91.8±1.7	48.5±5.7	58.0±2.7
	Pr.	79.9±3.1	81.4±3.3	55.3±4.9	60.3±2.5	77.7±3.5	81.0±4.0	49.7±4.4	60.6±2.4	80.0±3.5	81.7±4.0	49.6±4.5	60.6±2.7
RF	Acc.	86.8±2.2	87.2±2.1	56.6±3.5	60.1±2.4	84.8±2.2	86.1±2.0	57.0±3.0	57.6±2.3	87.1±2.4	88.4±2.0	57.3±3.0	61.4±2.3
	Re.	92.9±2.7	94.0±2.1	54.8±8.2	57.7±2.7	92.7±2.2	93.1±1.5	58.6±8.2	55.6±2.4	93.5±2.0	94.2±1.5	58.5±8.2	59.2±2.5
	Pr.	84.7±3.0	85.4±3.0	57.1±5.0	62.4±2.6	82.0±3.3	84.2±3.1	57.0±4.7	60.2±2.5	85.8±3.5	85.7±3.1	57.5±4.9	62.3±2.6
GB	Acc.	95.3±1.3	94.2±1.4	**71.0±3.1**	**76.8±2.6**	90.5±1.8	**93.7±1.5**	69.7±2.9	72.9±2.3	95.5±1.0	95.01±1.1	69.0±3.1	**78.7±2.3**
	Re.	94.7±1.4	**95.5±1.4**	**72.4±5.8**	76.3±2.8	94.4±1.7	94.8±1.4	**68.3±5.1**	73.6±2.2	95.6±0.9	96.4±1.3	68.5±5.0	77.0±2.5
	Pr.	95.6±1.5	94.1±1.8	**71.1±4.5**	78.2±2.5	89.6±2.6	91.9±2.3	**69.2±4.4**	75.7±2.4	96.1±1.4	94.4±1.7	69.2±4.5	**79.2±2.3**
Linear SVC	Acc.	**96.4±0.5**	95.5±0.3	70.4±3.1	74.9±2.5	**92.4±1.5**	92.9±1.8	66.0±3.0	67.9±2.7	**97.8±0.9**	96.3±1.4	**72.2±3.0**	76.4±2.6
	Re.	96.2±0.8	95.4±0.5	71.4±4.2	**76.9±2.2**	95.5±2.0	95.1±2.2	65.6±4.4	69.2±2.6	**98.7±0.9**	97.1±1.6	71.4±4.6	**77.9±2.4**
	Pr.	**97.8±0.3**	95.9±0.3	70.8±4.2	74.8±2.2	91.8±2.3	92.1±2.2	66.1±4.6	70.1±2.6	97.6±1.2	96.3±1.6	**72.5±4.5**	78.4±2.3

Case-1: BL vs. LD, **Case-2:** BL vs. HD, **Case-3:** LD vs. HD, **Case-4:** BL vs. LD vs. HD. **BOLD** signifies the best PEM score for a particular GDISCP. MLC = machine learning classifier, KNN = K-nearest neighbor, RF = random forest, GB = gradient boosting, linear SVC = support vector classifier, Acc. = Accuracy, Pr. = Precision, Re. = Recall.
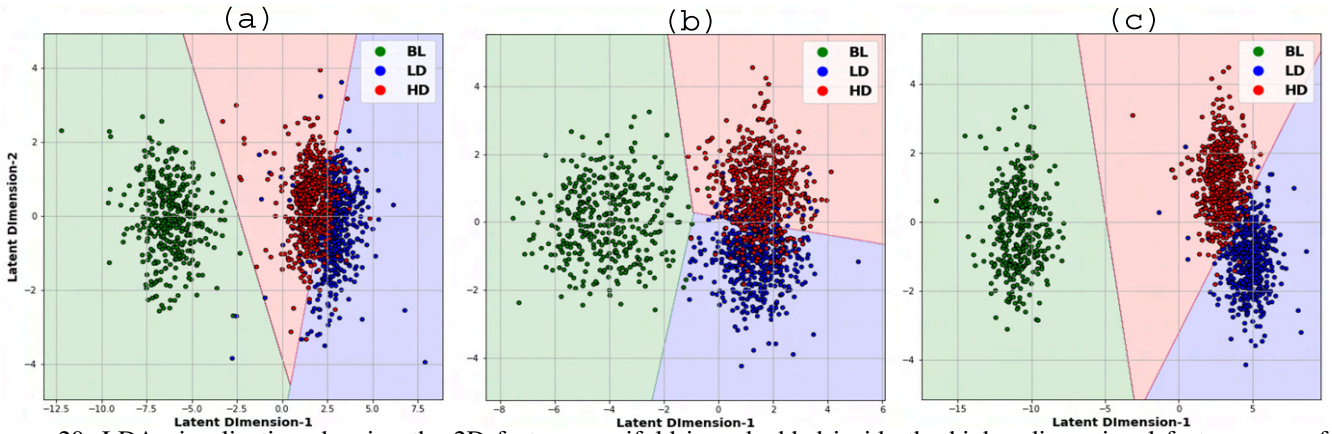

Figure 20: LDA visualization showing the 2D feature manifold is embedded inside the higher-dimensional feature space for (a) ECG; (b) EEG; and (c) combined EEG + ECG, respectively. The x-axis and y-axis are the two latent dimensions.

from both EEG and ECG modalities, the accuracies obtained are 97.8%, 96.3%, 72.2%, and 78.7%, respectively for BL vs. LD, BL vs. HD, LD vs. HD, and BL vs. LD vs. HD cases.

The PEM scores in columns 2–14 (see Table VII) and the ablation study evince the competency of the multimodal approach for physiological sensing-based stress delineation. Combining features from both EEG and ECG boosts the classification accuracy scores by 1.4%, 0.8%, 1.2%, and 1.9%, for different GDISP-CPs corresponding to BL vs. LD, BL vs. HD, LD vs. HD, and BL vs. LD vs. HD, respectively. Very high precision and recall score of more than 96% for BL vs. tasks (LD/HD), 71% for LD vs. HD, and 77% for BL vs. LD vs. HD entails an impeccable agreement between the predicted and actual (Bl/LD/HD) categories.

H. Visual Explanations of the Physiological Feature Space

The utility of multimodal physiological feature space in deciphering the underlying manifold for various GDIS levels is visually depicted in Fig. 20 and quantified in Table VIII. Fig. 20 shows the linear discriminant analysis (LDA) visualization of the 2D feature manifold, which is embedded inside the higher-dimensional feature space. LDA envisages dimensionality reduction by projection feature vectors onto new axes by maximizing the distance between the distinct cluster means and minimizing inter-class variances in the LDA feature space. 2D feature manifold embedding has been used for understanding the class separability with different feature modalities. Table VIII shows ECG is better for BL vs. task separation, while EEG for LD vs. HD classification. The compactness of the different GDIS levels and BL clusters are evaluated by its variance around each cluster mean. LD is compact with EEG features, while HD with ECG, combined with EEG and ECG features.

Table VIII: Exploratory LDA feature space cluster analysis.

Modality	$d_{Bl,LD}$	$d_{Bl,HD}$	$d_{LD,HD}$	σ_{Bl}^2	σ_{LD}^2	σ_{HD}^2
EEG	5.53	5.71	1.94	2.85	**1.61**	1.77
ECG	9.05	7.82	1.40	2.37	2.04	**1.68**
EEG+ECG	**15.68**	**13.85**	**2.98**	2.62	1.78	1.76

$d_{i,j}$ = Distance between the centroids of the ith and j^{th} clusters.
σ_i^2 = Total variance in the ith cluster.
$\sigma_i^2 = \sigma_{i,x}^2 + \sigma_{i,y}^2$, and x & y are the 2 orthogonal axis.
BOLD signifies the best inter and intra class cluster compactness measure.

I. Comparative Evaluation with the State-of-the-art

The literature on physiological sensing-based stress analysis includes stressors from both acute and perceived types

in an indoor or outdoor environment. The commonly used stress stimuli are the Montreal imaging stress task (MIST), Stroop Color and Word Test (SCWT), mental arithmetic test, public speaking, VR-induced stressors including roller coaster stimulation, shooting games, etc. VR systems rapidly integrate into various disciplines like education, digital marketing, entertainment, virtual communities, clinical therapies, occupational safety, social science, etc., and psychology.

However, gaming difficulty-induced stress in VR systems is sparsely studied. To the best of our knowledge, an exhaustive multimodal GDIS classification using inter-class shooting task representation was not reported in a consolidated way before. Generally, stress representation learning comprises both hand-crafted physiological features with ML models and deep neural networks (DNN). Comparison of classification methodology and accuracy of the current study with the prior art for physiological sensing-based GDIS classification in VR environment is summarized in Table IX.

VI. CONCLUSION

This paper discusses the experimental findings from a Go/No-Go VR shooting task exercise using ECG and EEG modalities. The research comprises multimodal physiological signal acquisition using commercially available wearable sensors, pre-processing, feature extraction, and classification. The individualized TET modulation for each participant induces multi-level shooting difficulty-based stress conditions. Accordingly, the acquired records are split and labeled into three (Bl, LD, & HD) classes.

This work provides a data-driven ML framework to characterize the alterations in physiological dynamics during a VR shooting task. Several HRV descriptors from ECG and spectral EEG power derived features from the δ, θ, α, low-β high-β, and γ bands are explored for four different GDISCPs in VR. Linear SVC exhibits the best ML classification results to identify the different stress levels associated with task difficulty from a resting BL segment, referred to as Bl-to-task variations (LD/HD). An average 10-fold cross-validation accuracy of more than 96% for the BL vs. LD/HD task, 72% for the inter-class GDIS segregation outperforms the prevalent art (refer Table IX). Also, our results reveal that multimodal study is more efficient than individual ECG or EEG modality-based classification, with ECG being more competent than EEG in most cases for decoding the various GDIS levels. This study will aid in modeling interoceptive awareness associated with VR environment stimuli, improve human training, and better manage gaming stress levels. Coupling between brain and heart operations at multi-granular levels might reveal further insight into VR-related stress classification. Investigations on neurological modulation of other physiological signals like GSR, BP, and the subject's motivation as a causal factor might also be explored for improved GDIS classification in the future.

Table IX: Comparative evaluation with the prior art.

Ref. (YoP)	Modality	Methodology	Accuracy (# class)	Stress or Stimuli
[107] (2013)	ECG	KNN	94.6% (2)	-
[55] (2014)	ECG, EDA, Respiration	SVM, KNN	73% (2)	-
[43] (2015)	EEG	θ, α, β-ABP, RBP, Coherence FD, SVM, NB, LR	85.2% (2) 67.1 (4) 94.6% (2) 83.4% (3)	Acute, MIST, Stroop Test
[44] (2015)	EEG	β PSD, SVM	71.4% (2)	Perceived
[56] (2015)	EDA	LDA	98.8 % (2)	-
[61] (2015)	BVP, ECG, EDA	SVM	80% (2)	-
[54] (2015)	EDA, EMG, HR, Respiration	SVM	98.4% (3)	-
[64] (2015)	Respiration, ECG	SVM	72% (2)	-
[45] (2016)	EEG	α/β Ratio, SVM	88% (2) 96% (2) 75% (3)	Stroop test, arithmetic test
[108] (2016)	ECG	KNN	94.6% (2)	-
[65] (2016)	ECG, EDA Respiration	DT	89.1% (2)	-
[109] (2016)	Activity	SVM	83.3% (2)	-
[58] (2016)	BVP, EDA, ECG	RF	76% (2)	-
[110] (2017)	EEG	α, β PSD, SVM	83.3% (2)	Acute
[111] (2017)	EEG	β PSD, NB	71.4% (2)	Perceived
[59] (2017)	EDA, BVP, Speech, Activity	AdaBoost	94% (2)	-
[57] (2017)	EDA,BVP	SVM	88.8% (2)	-
[66] (2017)	BVP, Activity	SVM	95% (2)	-
[63] (2017)	ECG, EDA, Respiration	HRV & SVM	99% (3)	-
[53] (2018)	EEG	low/high-β, low-γ PSD, SVM	78.5% (2)	Perceived
[31] (2018)	ECG, BT, BVP, EDA, EMG, Respiration	LDA, Adaboost	93% (2) 80% (3)	Speaking & mental arithmetic task
[17] (2018)	EEG	α power, SVM	94.8% (3)	Mental arithmetic
[49] (2018)	ECG, EDA, EEG	HRV & SVM	86% (2)	-
[48] (2018)	EEG	RBP, Entropy, Assymetry, ERD/ERS, RF	93% (2) 83% (8)	Outdoor Environment
[21] (2019)	EEG	MLP, SVM, NB, DA, RA, θ-PSD, ABP, Correlation	92.8% (2) 64.3% (3)	Perceived Mental Stress
[62] (2019)	EDA, Activity	DNN	98.3%(2)	-
[112] (2020)	ECG	CNN	85.4% (5)	VR roller coaster stimulation
[60] (2020)	EDA, PPG	RF	92% (2)	-
This Work	**ECG, EEG**	**HRV, ABP, RBP, Power Ratio, KNN, RF, GB, SVC**	**97.8% (2) 96.3% (2) 72.2% (2) 78.7% (3)**	**VR Shooting task**

Ref. = reference, YoP = year of publication, MIST = Montreal imaging stress task, Stroop = stroop color and word test, LDA = linear discriminant analysis, DA = differential asymmetry, RA = rotational asymmetry, FD = fractal dimension, MLP = multi-layer perceptron, NB = Naive Bayes, DT = decision tree, LR = Logistic Regression.

REFERENCES

[1] D. R. Flatla, C. Gutwin, L. E. Nacke, S. Bateman, and R. L. Mandryk, "Calibration games: making calibration tasks enjoyable by adding motivating game elements," in *Proceedings of the 24th annual ACM symposium on User interface software and technology*, 2011, pp. 403–412.

[2] L. E. Nacke, M. Kalyn, C. Lough, and R. L. Mandryk, "Biofeedback game design: using direct and indirect physiological control to enhance game interaction," in *Proceedings of the SIGCHI conference on human factors in computing systems*, 2011, pp. 103–112.

[3] J. Cusveller, C. Gerritsen, and J. de Man, "Evoking and measuring arousal in game settings," in *International Conference on Serious Games*. Springer, 2014, pp. 165–174.

[4] J. Arroyo-Palacios and D. M. Romano, "Exploring the use of a respiratory-computer interface for game interaction," in *2009 International IEEE Consumer Electronics Society's Games Innovations Conference*. IEEE, 2009, pp. 154–159.

[5] P. Jemioło, B. Giżycka, and G. J. Nalepa, "Prototypes of arcade games enabling affective interaction," in *International Conference on Artificial Intelligence and Soft Computing*. Springer, 2019, pp. 553–563.

[6] S. Kaczmarek and S. Petroviča, "Promotion of learning motivation through individualization of learner-game interaction," in *2018 IEEE Conference on Computational Intelligence and Games (CIG)*. IEEE, 2018, pp. 1–8.

[7] F. Balducci, C. Grana, and R. Cucchiara, "Classification of affective data to evaluate the level design in a role-playing videogame," in *2015 7th International Conference on Games and Virtual Worlds for Serious Applications (VS-Games)*. IEEE, 2015, pp. 1–8.

[8] R. Robinson, Z. Rubin, E. M. Segura, and K. Isbister, "All the feels: designing a tool that reveals streamers' biometrics to spectators," in *Proceedings of the 12th International Conference on the Foundations of Digital Games*, 2017, pp. 1–6.

[9] R. L. Mandryk and L. E. Nacke, "Biometrics in gaming and entertainment technologies," in *Biometrics in a Data Driven World*. Chapman and Hall/CRC, pp. 215–248.

[10] K. M. Gilleade and A. Dix, "Using frustration in the design of adaptive videogames," in *Proceedings of the 2004 ACM SIGCHI International Conference on Advances in computer entertainment technology*, 2004, pp. 228–232.

[11] A. Al Mahmud, O. Mubin, J. R. Octavia, S. Shahid, L. Yeo, P. Markopoulos, and J.-B. Martens, "amazed: designing an affective social game for children," in *Proceedings of the 6th international conference on Interaction design and children*, 2007, pp. 53–56.

[12] A. Dekker and E. Champion, "Please biofeed the zombies: enhancing the gameplay and display of a horror game using biofeedback," in *DiGRA'07-Proceedings of the 2007 DiGRA International Conference: Situated Play*, 2007, pp. 550–558.

[13] H. Al Osman, H. Dong, and A. El Saddik, "Ubiquitous biofeedback serious game for stress management," *IEEE Access*, vol. 4, pp. 1274–1286, 2016.

[14] K. Gilleade, A. Dix, and J. Allanson, "Affective videogames and modes of affective gaming: assist me, challenge me, emote me," *DiGRA 2005: Changing Views–Worlds in Play.*, 2005.

[15] K. Kuikkaniemi, T. Laitinen, M. Turpeinen, T. Saari, I. Kosunen, and N. Ravaja, "The influence of implicit and explicit biofeedback in first-person shooter games," in *Proceedings of the SIGCHI conference on human factors in computing systems*, 2010, pp. 859–868.

[16] K. L. Amon and A. Campbell, "Can children with ad/hd learn relaxation and breathing techniques through biofeedback video games?." *Australian Journal of Educational & Developmental Psychology*, vol. 8, pp. 72–84, 2008.

[17] F. Al-Shargie, T. B. Tang, N. Badruddin, and M. Kiguchi, "Towards multilevel mental stress assessment using svm with ecoc: an EEG approach," *Medical & biological engineering & computing*, vol. 56, no. 1, pp. 125–136, 2018.

[18] J. L. Andreassi, *Psychophysiology: Human behavior and physiological response*. Psychology Press, 2010.

[19] W. B. Cannon, "The wisdom of the body," 1939.

[20] F. Arai, T. Fukuda, Y. Yamamoto, T. Naito, and T. Matsui, "Interactive adaptive interface using recursive fuzzy reasoning," in *Proceedings of IEEE Virtual Reality Annual International Symposium*. IEEE, 1993, pp. 104–110.

[21] A. Arsalan, M. Majid, A. R. Butt, and S. M. Anwar, "Classification of perceived mental stress using a commercially available EEG headband," *IEEE journal of biomedical and health informatics*, vol. 23, no. 6, pp. 2257–2264, 2019.

[22] R. Sioni and L. Chittaro, "Stress detection using physiological sensors," *Computer*, vol. 48, no. 10, pp. 26–33, 2015.

[23] A. Tapus *et al.*, "Stress game: the role of motivational robotic assistance in reducing user's task stress," *International Journal of Social Robotics*, vol. 7, no. 2, pp. 227–240, 2015.

[24] R. Robinson, K. Wiley, A. Rezaeivahdati, M. Klarkowski, and R. L. Mandryk, ""let's get physiological, physiological!" a systematic review of affective gaming," in *Proceedings of the Annual Symposium on Computer-Human Interaction in Play*, 2020, pp. 132–147.

[25] R. Picard, "Affetive computing. cambridge," 1997.

[26] J. Sharry, M. McDermott, and J. Condron, "Relax to win treating children with anxiety problems with a biofeedback video game," *Eisteach*, vol. 2, pp. 22–26, 2003.

[27] D. McDuff, A. Mahmoud, M. Mavadati, M. Amr, J. Turcot, and R. e. Kaliouby, "Affdex sdk: a cross-platform real-time multi-face expression recognition toolkit," in *Proceedings of the 2016 CHI conference*

extended abstracts on human factors in computing systems, 2016, pp. 3723–3726.

[28] T. Tijs, D. Brokken, and W. IJsselsteijn, "Creating an emotionally adaptive game," in *International Conference on Entertainment Computing*. Springer, 2008, pp. 122–133.

[29] D. G. Arellano, L. Tokarchuk, and H. Gunes, "Measuring affective, physiological and behavioural differences in solo, competitive and collaborative games," in *International conference on intelligent technologies for interactive entertainment*. Springer, 2016, pp. 184–193.

[30] G. Chanel, J. M. Kivikangas, and N. Ravaja, "Physiological compliance for social gaming analysis: Cooperative versus competitive play," *Interacting with Computers*, vol. 24, no. 4, pp. 306–316, 2012.

[31] P. Schmidt, A. Reiss, R. Duerichen, C. Marberger, and K. Van Laerhoven, "Introducing wesad, a multimodal dataset for wearable stress and affect detection," in *Proceedings of the 20th ACM international conference on multimodal interaction*, 2018, pp. 400–408.

[32] L.-W. Ko, P.-W. Lai, B.-J. Yang, and C.-T. Lin, "Mobile EEG & ecg integration system for monitoring physiological states in peforming simulated war game training," in *2015 IEEE Conference on Computational Intelligence and Games (CIG)*. IEEE, 2015, pp. 542–543.

[33] M. Garbarino, M. Lai, D. Bender, R. W. Picard, and S. Tognetti, "Empatica e3—a wearable wireless multi-sensor device for real-time computerized biofeedback and data acquisition," in *2014 4th International Conference on Wireless Mobile Communication and Healthcare-Transforming Healthcare Through Innovations in Mobile and Wireless Technologies (MOBIHEALTH)*. IEEE, 2014, pp. 39–42.

[34] T. F. o. t. E. S. o. C. t. N. A. S. o. P. Electrophysiology, "Heart rate variability: standards of measurement, physiological interpretation, and clinical use," *Circulation*, vol. 93, no. 5, pp. 1043–1065, 1996.

[35] R. E. Kleiger, P. K. Stein, M. S. Bosner, and J. N. Rottman, "Time domain measurements of heart rate variability," *Cardiology clinics*, vol. 10, no. 3, pp. 487–498, 1992.

[36] K. Hnatkova, X. Copie, A. Staunton, and M. Malik, "Numeric processing of lorenz plots of rr intervals from long-term ecgs: comparison with time-domain measures of heart rate variability for risk stratification after myocardial infarction," *Journal of electrocardiology*, vol. 28, pp. 74–80, 1995.

[37] Z. Ŏri, G. Monir, J. Weiss, X. Sayhouni, and D. H. Singer, "Heart rate variability: frequency domain analysis," *Cardiology clinics*, vol. 10, no. 3, pp. 499–533, 1992.

[38] A. Malliani, M. Pagani, F. Lombardi, and S. Cerutti, "Cardiovascular neural regulation explored in the frequency domain." *Circulation*, vol. 84, no. 2, pp. 482–492, 1991.

[39] R. Sloan, P. Shapiro, E. Bagiella, S. Boni, M. Paik, J. Bigger Jr, R. Steinman, and J. Gorman, "Effect of mental stress throughout the day on cardiac autonomic control," *Biological psychology*, vol. 37, no. 2, pp. 89–99, 1994.

[40] T. Kageyama, N. Nishikido, T. Kobayashi, Y. Kurokawa, T. Kaneko, and M. Kabuto, "Self-reported sleep quality, job stress, and daytime autonomic activities assessed in terms of short-term heart rate variability among male white-collar workers," *Industrial health*, vol. 36, no. 3, pp. 263–272, 1998.

[41] D. Dimitriev and E. Saperova, "Heart rate variability and blood pressure during mental stress," *Rossiiskii fiziologicheskii zhurnal imeni IM Sechenova*, vol. 101, no. 1, pp. 98–107, 2015.

[42] J. F. Thayer and R. D. Lane, "Claude bernard and the heart–brain connection: Further elaboration of a model of neurovisceral integration," *Neuroscience & Biobehavioral Reviews*, vol. 33, no. 2, pp. 81–88, 2009.

[43] X. Hou, Y. Liu, O. Sourina, Y. R. E. Tan, L. Wang, and W. Mueller-Wittig, "EEG based stress monitoring," in *2015 IEEE International Conference on Systems, Man, and Cybernetics*. IEEE, 2015, pp. 3110–3115.

[44] S. M. U. Saeed, S. M. Anwar, M. Majid, and A. M. Bhatti, "Psychological stress measurement using low cost single channel EEG headset," in *2015 IEEE International Symposium on Signal Processing and Information Technology (ISSPIT)*. IEEE, 2015, pp. 581–585.

[45] G. Jun and K. G. Smitha, "EEG based stress level identification," in *2016 IEEE international conference on systems, man, and cybernetics (SMC)*. IEEE, 2016, pp. 003 270–003 274.

[46] W. Klimesch, "EEG alpha and theta oscillations reflect cognitive and memory performance: a review and analysis," *Brain research reviews*, vol. 29, no. 2-3, pp. 169–195, 1999.

[47] G. Giannakakis, D. Grigoriadis, K. Giannakaki, O. Simantiraki, A. Roniotis, and M. Tsiknakis, "Review on psychological stress detection using biosignals," *IEEE Transactions on Affective Computing*, 2019.

[48] C. Saitis and K. Kalimeri, "Multimodal classification of stressful environments in visually impaired mobility using EEG and peripheral biosignals," *IEEE Transactions on Affective Computing*, 2018.

[49] S. Betti, R. M. Lova, E. Rovini, G. Acerbi, L. Santarelli, M. Cabiati, S. Del Ry, and F. Cavallo, "Evaluation of an integrated system

of wearable physiological sensors for stress monitoring in working environments by using biological markers," *IEEE Transactions on Biomedical Engineering*, vol. 65, no. 8, pp. 1748–1758, 2017.

[50] A. Gevins, M. E. Smith, L. McEvoy, and D. Yu, "High-resolution EEG mapping of cortical activation related to working memory: effects of task difficulty, type of processing, and practice." *Cerebral cortex (New York, NY: 1991)*, vol. 7, no. 4, pp. 374–385, 1997.

[51] J. Minguillon, M. A. Lopez-Gordo, and F. Pelayo, "Stress assessment by prefrontal relative gamma," *Frontiers in computational neuroscience*, vol. 10, p. 101, 2016.

[52] T. Harmony, "The functional significance of delta oscillations in cognitive processing," *Frontiers in integrative neuroscience*, vol. 7, p. 83, 2013.

[53] S. M. Umar Saeed, S. M. Anwar, M. Majid, M. Awais, and M. Al-nowami, "Selection of neural oscillatory features for human stress classification with single channel EEG headset," *BioMed research international*, vol. 2018, 2018.

[54] A. Ghaderi, J. Frounchi, and A. Farnam, "Machine learning-based signal processing using physiological signals for stress detection," in *2015 22nd Iranian Conference on Biomedical Engineering (ICBME)*. IEEE, 2015, pp. 93–98.

[55] A. Muaremi, A. Bexheti, F. Gravenhorst, B. Arnrich, and G. Tröster, "Monitoring the impact of stress on the sleep patterns of pilgrims using wearable sensors," in *IEEE-EMBS international conference on biomedical and health informatics (BHI)*. IEEE, 2014, pp. 185–188.

[56] A. Liapis, C. Katsanos, D. Sotiropoulos, M. Xenos, and N. Karousos, "Stress recognition in human-computer interaction using physiological and self-reported data: a study of gender differences," in *Proceedings of the 19th Panhellenic Conference on Informatics*, 2015, pp. 323–328.

[57] B. Egilmez, E. Poyraz, W. Zhou, G. Memik, P. Dinda, and N. Alshurafa, "Ustress: Understanding college student subjective stress using wrist-based passive sensing," in *2017 IEEE International Conference on Pervasive Computing and Communications Workshops (PerCom Workshops)*. IEEE, 2017, pp. 673–678.

[58] M. Gjoreski, H. Gjoreski, M. Luštrek, and M. Gams, "Continuous stress detection using a wrist device: in laboratory and real life," in *proceedings of the 2016 ACM international joint conference on pervasive and ubiquitous computing: Adjunct*, 2016, pp. 1185–1193.

[59] O. M. Mozos, V. Sandulescu, S. Andrews, D. Ellis, N. Bellotto, R. Dobrescu, and J. M. Ferrandez, "Stress detection using wearable physiological and sociometric sensors," *International journal of neural systems*, vol. 27, no. 02, p. 1650041, 2017.

[60] R. K. Nath, H. Thapliyal, and A. Caban-Holt, "Validating physiological stress detection model using cortisol as stress bio marker," in *2020 IEEE International Conference on Consumer Electronics (ICCE)*. IEEE, 2020, pp. 1–5.

[61] V. Sandulescu, S. Andrews, D. Ellis, N. Bellotto, and O. M. Mozos, "Stress detection using wearable physiological sensors," in *International work-conference on the interplay between natural and artificial computation*. Springer, 2015, pp. 526–532.

[62] L. Rachakonda, S. P. Mohanty, E. Kougianos, and P. Sundaravadivel, "Stress-lysis: A dnn-integrated edge device for stress level detection in the iomt," *IEEE Transactions on Consumer Electronics*, vol. 65, no. 4, pp. 474–483, 2019.

[63] L.-l. Chen, Y. Zhao, P.-f. Ye, J. Zhang, and J.-z. Zou, "Detecting driving stress in physiological signals based on multimodal feature analysis and kernel classifiers," *Expert Systems with Applications*, vol. 85, pp. 279–291, 2017.

[64] K. Hovsepian, M. Al'Absi, E. Ertin, T. Kamarck, M. Nakajima, and S. Kumar, "cstress: towards a gold standard for continuous stress assessment in the mobile environment," in *Proceedings of the 2015 ACM international joint conference on pervasive and ubiquitous computing*, 2015, pp. 493–504.

[65] M. Abouelenien, M. Burzo, and R. Mihalcea, "Human acute stress detection via integration of physiological signals and thermal imaging," in *Proceedings of the 9th ACM international conference on pervasive technologies related to assistive environments*, 2016, pp. 1–8.

[66] D. S. Lee, T. W. Chong, and B. G. Lee, "Stress events detection of driver by wearable glove system," *IEEE Sensors Journal*, vol. 17, no. 1, pp. 194–204, 2016.

[67] W. H. Ehrenstein and A. Ehrenstein, "Psychophysical methods," in *Modern techniques in neuroscience research*. Springer, 1999, pp. 1211–1241.

[68] D. P. Spangler, S. Alam, S. Rahman, J. Crone, R. Robucci, N. Banerjee, S. E. Kerick, and J. R. Brooks, "Multilevel longitudinal analysis of shooting performance as a function of stress and cardiovascular responses," *IEEE Transactions on Affective Computing*, vol. 12, no. 3, pp. 648–665, 2020.

[69] H. Sedghamiz, "Biosigkit: a matlab toolbox and interface for analysis of biosignals," *Journal of Open Source Software*, vol. 3, no. 30, p. 671, 2018.

[70] H. Sedghamiz and D. Santonocito, "Unsupervised detection and classification of motor unit action potentials in intramuscular electromyography signals," in *2015 e-health and bioengineering conference (ehb)*. IEEE, 2015, pp. 1–6.

[71] J. A. Behar, A. A. Rosenberg, I. Weiser-Bitoun, O. Shemla, A. Alexandrovich, E. Konyukhov, and Y. Yaniv, "Physiozoo: a novel open access platform for heart rate variability analysis of mammalian electrocardiographic data," *Frontiers in physiology*, vol. 9, p. 1390, 2018.

[72] C. M. van Ravenswaaij-Arts, L. A. Kollee, J. C. Hopman, G. B. Stoelinga, and H. P. van Geijn, "Heart rate variability," *Annals of internal medicine*, vol. 118, no. 6, pp. 436–447, 1993.

[73] M. Malik, "Heart rate variability." *Current opinion in cardiology*, vol. 13, no. 1, pp. 36–44, 1998.

[74] F. Shaffer and J. P. Ginsberg, "An overview of heart rate variability metrics and norms," *Frontiers in public health*, vol. 5, p. 258, 2017.

[75] B. M. Saykrs, "Analysis of heart rate variability," *Ergonomics*, vol. 16, no. 1, pp. 17–32, 1973.

[76] C. Schubert, M. Lambertz, R. Nelesen, W. Bardwell, J.-B. Choi, and J. Dimsdale, "Effects of stress on heart rate complexity—a comparison between short-term and chronic stress," *Biological psychology*, vol. 80, no. 3, pp. 325–332, 2009.

[77] C.-K. Peng, J. M. Hausdorff, and A. Goldberger, "Fractal mechanisms in neuronal control: human heartbeat and gait dynamics in health and disease," *Self-organized biological dynamics and nonlinear control*, pp. 66–96, 2000.

[78] M. Costa, A. L. Goldberger, and C.-K. Peng, "Multiscale entropy analysis of biological signals," *Physical review E*, vol. 71, no. 2, p. 021906, 2005.

[79] M. D. Costa, R. B. Davis, and A. L. Goldberger, "Heart rate fragmentation: a new approach to the analysis of cardiac interbeat interval dynamics," *Frontiers in physiology*, vol. 8, p. 255, 2017.

[80] M. Malik and A. J. Camm, "Heart rate variability," *Clinical cardiology*, vol. 13, no. 8, pp. 570–576, 1990.

[81] H. M. Stauss, "Heart rate variability," *American Journal of Physiology-Regulatory, Integrative and Comparative Physiology*, vol. 285, no. 5, pp. R927–R931, 2003.

[82] N. R. Lomb, "Least-squares frequency analysis of unequally spaced data," *Astrophysics and space science*, vol. 39, no. 2, pp. 447–462, 1976.

[83] U. R. Acharya, K. P. Joseph, N. Kannathal, C. M. Lim, and J. S. Suri, "Heart rate variability: a review," *Medical and biological engineering and computing*, vol. 44, no. 12, pp. 1031–1051, 2006.

[84] S. W. Porges, "Cardiac vagal tone: a physiological index of stress," *Neuroscience & Biobehavioral Reviews*, vol. 19, no. 2, pp. 225–233, 1995.

[85] A. Delorme and S. Makeig, "EEGLAB: an open source toolbox for analysis of single-trial EEG dynamics including independent component analysis," *Journal of neuroscience methods*, vol. 134, no. 1, pp. 9–21, 2004.

[86] N. Bigdely-Shamlo, T. Mullen, C. Kothe, K.-M. Su, and K. A. Robbins, "The prep pipeline: standardized preprocessing for large-scale EEG analysis," *Frontiers in neuroinformatics*, vol. 9, p. 16, 2015.

[87] T.-P. Jung, S. Makeig, C. Humphries, T.-W. Lee, M. J. Mckeown, V. Iragui, and T. J. Sejnowski, "Removing electroencephalographic artifacts by blind source separation," *Psychophysiology*, vol. 37, no. 2, pp. 163–178, 2000.

[88] P. Sahoo, Karuna, A. Radhakrishnan, S. Pratiher, S. Alam, S. Kerick, N. Ghosh, D. Chhan, N. Banerjee, and A. Patra, "Alterations in multichannel EEG dynamics during a stressful shooting task in virtual reality systems," in *2021 43rd Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, p. (**To Appear**).

[89] G. Dumermuth and L. Molinari, "Spectral analysis of the eeg," *Neuropsychobiology*, vol. 17, no. 1-2, pp. 85–99, 1987.

[90] F. Changyong, W. Hongyue, L. Naiji, C. Tian, H. Hua, L. Ying *et al.*, "Log-transformation and its implications for data analysis," *Shanghai archives of psychiatry*, vol. 26, no. 2, p. 105, 2014.

[91] D. Sharpe, "Chi-square test is statistically significant: Now what?" *Practical Assessment, Research, and Evaluation*, vol. 20, no. 1, p. 8, 2015.

[92] N. Mantel, "Chi-square tests with one degree of freedom; extensions of the mantel-haenszel procedure," *Journal of the American Statistical Association*, vol. 58, no. 303, pp. 690–700, 1963.

[93] M. P. Tulppo, T. H. Makikallio, T. Seppanen, R. T. Laukkanen, and H. V. Huikuri, "Vagal modulation of heart rate during exercise: effects of age and physical fitness," *American Journal of Physiology-Heart and Circulatory Physiology*, vol. 274, no. 2, pp. H424–H429, 1998.

[94] M. Brennan, M. Palaniswami, and P. Kamen, "Do existing measures of poincare plot geometry reflect nonlinear features of heart rate variability?" *IEEE transactions on biomedical engineering*, vol. 48, no. 11, pp. 1342–1347, 2001.

[95] T. Kuusela, "-methodological aspects of heart rate variability analysis," in *Heart rate variability (HRV) signal analysis*. CRC Press, 2016, pp. 28–61.

[96] S. Akselrod, D. Gordon, F. A. Ubel, D. C. Shannon, A. Berger, and R. J. Cohen, "Power spectrum analysis of heart rate fluctuation: a quantitative probe of beat-to-beat cardiovascular control," *science*, vol. 213, no. 4504, pp. 220–222, 1981.

[97] P. M. Lehrer and R. Gevirtz, "Heart rate variability biofeedback: how and why does it work?" *Frontiers in psychology*, vol. 5, p. 756, 2014.

[98] F. Shaffer, R. McCraty, and C. L. Zerr, "A healthy heart is not a metronome: an integrative review of the heart's anatomy and heart rate variability," *Frontiers in psychology*, vol. 5, p. 1040, 2014.

[99] Y.-K. Wang, T.-P. Jung, and C.-T. Lin, "Theta and alpha oscillations in attentional interaction during distracted driving," *Frontiers in behavioral neuroscience*, vol. 12, p. 3, 2018.

[100] D. O. Bos *et al.*, "Eeg-based emotion recognition," *The influence of visual and auditory stimuli*, vol. 56, no. 3, pp. 1–17, 2006.

[101] T. Cover and P. Hart, "Nearest neighbor pattern classification," *IEEE transactions on information theory*, vol. 13, no. 1, pp. 21–27, 1967.

[102] L. Breiman, "Random forests," *Machine learning*, vol. 45, no. 1, pp. 5–32, 2001.

[103] J. H. Friedman, "Greedy function approximation: a gradient boosting machine," *Annals of statistics*, pp. 1189–1232, 2001.

[104] C. Cortes and V. Vapnik, "Support-vector networks," *Machine learning*, vol. 20, no. 3, pp. 273–297, 1995.

[105] J. Snoek, H. Larochelle, and R. P. Adams, "Practical bayesian optimization of machine learning algorithms," *Advances in neural information processing systems*, vol. 25, 2012.

[106] L. Chen, N. Qin, X. Dai, and D. Huang, "Fault diagnosis of high-speed train bogie based on capsule network," *IEEE Transactions on Instrumentation and Measurement*, vol. 69, no. 9, pp. 6203–6211, 2020.

[107] P. Karthikeyan, M. Murugappan, and S. Yaacob, "Analysis of stroop color word test-based human stress detection using electrocardiography and heart rate variability signals," *Arabian Journal for Science and Engineering*, vol. 39, no. 3, pp. 1835–1847, 2014.

[108] R. Castaldo, W. Xu, P. Melillo, L. Pecchia, L. Santamaria, and C. James, "Detection of mental stress due to oral academic examination via ultra-short-term hrv analysis," in *2016 38th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*. IEEE, 2016, pp. 3805–3808.

[109] B.-G. Lee and W.-Y. Chung, "Wearable glove-type driver stress detection using a motion sensor," *IEEE Transactions on Intelligent Transportation Systems*, vol. 18, no. 7, pp. 1835–1844, 2016.

[110] A. Secerbegovic, S. Ibric, J. Nisic, N. Suljanovic, and A. Mujcic, "Mental workload vs. stress differentiation using single-channel EEG," in *CMBEBIH 2017*. Springer, 2017, pp. 511–515.

[111] S. M. U. Saeed, S. M. Anwar, and M. Majid, "Quantification of human stress using commercially available single channel EEG headset," *IEICE Transactions on Information and Systems*, vol. 100, no. 9, pp. 2241–2244, 2017.

[112] Z. Ahmad and N. M. Khan, "Multi-level stress assessment using multi-domain fusion of ecg signal," in *2020 42nd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC)*. IEEE, 2020, pp. 4518–4521.