

Stress Detection during Motor Activity: Comparing Neurophysiological Indices in Older Adults

Rohith Karthikeyan, *Student Member, IEEE*, Anthony D. McDonald, and Ranjana K. Mehta, *Member, IEEE*

Abstract—The effects of cognitive stress are complex and multi-dimensional with nuanced neural and physiological representations across our lifespan. Chronic and instantaneous stressors are known to alter both executive function and motor performance — a particularly challenging prospect for older adults. Age, sex, and motor activity are critical yet under-represented dimensions in the domain of stress detection. Through the present work, we explore a subset of these variables and the relevance of brain hemodynamics and heart rate variability (HR/V) as biomarkers of stress in an aging population. We rely on a multimodal, sex-balanced, motor-stress data set (N = 59) and an exhaustive machine learning workflow to operationalize the unique neurophysiological states that form the human stress response. We found that a quadratic discriminant was sufficient to separate the two states across feature, demographic, and activity variables. We report a stress detection accuracy between 78 – 98% when using models trained independently on each feature-set. However, these models were highly sensitive to sex, and activity differences — with distinct regions, and features implicated in stress recognition. Both HR/V and fNIRS based features were excellent indices of cognitive stress, however neither generalized to a degree beneficial toward operational use. Our observations underscore the importance of task-context, age, and sex as factors in modeling stress detection tools for older adults.

Index Terms—affective inference, cognitive stress, sex, fNIRS, heart rate variability, aging, machine learning



1 INTRODUCTION

MENTAL stress is a complex, multi-dimensional phenomenon that triggers a cascade of physiological, behavioral, and cognitive adaptations [1]. These adaptations, which may be positive or negative, change across our lifespan, and influence our ability to manage situations and task demands. Despite substantial research on the interactions between stress and the human brain, there remains a limited understanding of its influence across demographic and physiological variables, and on motor coordination [2]. This is partly due to the nature of human stress-response, where genetic [3], experiential, and developmental [4] attributes moderate our lifelong adaptations and behaviors.

In older adults, perceived stress is known to trigger negative affect and play a significant role in mediating the influence of stressful life-events and psychological distress [5]. Epidemiological data point to similar concerns, with both chronic and instantaneous stressors shown to aggravate hypertension [6] and the incidence of acute car-

diovascular ailments [7]. These challenges are particularly relevant to older adults, who when cognitively burdened may require additional neural resources to perform what were normally automated tasks. For example, frail, senior patients often “stop walking, when they start talking” [8], presumably due to the attention demands of the conversation — a behavior that was found to be an early indicator of fall incidence. Furthermore, geriatric patients may also exhibit a decrease in voluntary muscle activation due to task workload [9], therefore, this motor-cognitive dual-task burden is a serious challenge in elderly care and one that demands further consideration. Above all, the repetitive performance of motor activity, such as knee-extension or hand-grip is central to musculoskeletal rehabilitation [10], stroke recovery [11], and the management of Alzheimer’s disease [12] in those populations. Hence, stress detection in task contexts that include motor activity can enable better mapping between task demands and the patient’s cognitive availability, and advance pertinent rehabilitation paradigms for the target demographic.

In our view, a critical bottleneck towards this goal is the lack of valid data sets that explore stress manifestation in populations of older adults during motor activity. Additionally, degeneracy in the representations of stress, e.g. those due to aging [13] or other underlying conditions [14]; activity-related artifacts, e.g. those due to movement

- R. Karthikeyan is with the Department of Mechanical Engineering, Texas A&M University, College Station, TX 77843. Email: rohithkarthikeyan@tamu.edu
- A.D. McDonald, and R. K. Mehta are with the Department of Industrial & Systems Engineering, Texas A&M University, College Station, TX 77843. Email: mcdonald, rmehta@tamu.edu

or breathing; and variability in experiment design and sample populations further obscure relevance and present as barriers to translation. In Section 2 we briefly expand on these issues and discuss the opportunities that motivate our current investigation.

In this work, we explore stress detection during motor exercises among older adults. We hypothesize that our observations will aid the development of task-agnostic stress detection tools for our study demographic. To this end, we rely on a multimodal, sex-balanced, motor-stress data set that includes functional near-infrared spectroscopy (fNIRS) as neural indices, and electrocardiography (ECG) features as physiological biomarkers. Salivary cortisol samples, a gold standard among stress indicators [15], provide ground truth for stress labels, along with the self-reported State-Trait and Anxiety inventory (STAI) responses [16]. We developed a supervised learning model and contrast the efficacy of neural and physiological biomarkers as indicators of stress across demographic and activity-related variables in a sample of older adults. Our observations build from the preliminary evidence introduced in the conference article [17] that focused on temporal and spectral components of HR/V as indicators of stress during hand-grip exercises.

Contributions

We summarize the main contributions of our work below:

- Evidence that cerebral hemodynamics, heart rate variability and derived measures are independently representative of the neurophysiological stress-response in older adults, beyond the effects of aging or motor activity-related perturbations.
- Insight on the influence of motor-activity and sex on stress representations, and the need for demographic-specific stress detection tools in operational contexts such as rehabilitation.
- A validated stress-induction protocol, an accepted standard in salivary cortisol as ground truth, a sex, physiology and age-balanced sample of older adults to support our observations.

The remainder of this paper is organized as follows: *Section 2* presents background related to stress detection and the current research gaps. *Section 3* details our experiment protocol, feature engineering, and machine learning workflow. *Section 4* presents our results and discussions. In *Section 5* we conclude with the limitations and future work.

2 BACKGROUND AND RESEARCH GAPS

The human stress response is mediated by two pathways, the Sympathetic Adrenal Medullary axis, and the Hypothalamic Pituitary Adrenal axis [2]. The former is responsible for fast adaptations under acute stress and instantiates our “flight or fight” behavior, while the latter moderates our slow adaptations to chronic stress. Both pathways drive

allostasis and are characterized by distinct neurophysiological signatures [18]. For example, the cardiovascular system adapts to stressors with an increase in heart rate and blood pressure [19], [20], the endocrine system mediates stress response by altering neurochemistry [21]. These changes are accompanied by neural mechanisms that underlie our perception and processing loops [22]. Researchers can reliably infer stress conditions using invasive methods that measure the release of neurotransmitters or stress hormones, e.g. blood samples for *catecholamines* [23] and salivary swabs for *corticosteroids* [24] to name a few, however, these methods remain cumbersome.

The need for fieldable stress indices has motivated researchers to look for other, non-invasive, approaches. Some have described statistically significant changes in heart rate variability, event related potentials, electroencephalography (EEG), photoplethysmograms, thermal imaging, and blood sugar (or) insulin levels as indicators of stress across multiple experiments [25], [26]. Indeed, a substantial amount of this research has focused on operationalizing biosignals for stress detection using existing, and novel machine learning algorithms. Groups have successfully used unobtrusive indicators such as heart rate variability (HR/V) [27], electrodermal activity (EDA) [28], and facial affect [29] with stress detection accuracy ranging between 80 – 95%. A subset have also explored the use of neural indices; Al-Shargie et al. report the use of a decision fusion architecture for fNIRS and EEG signals resulting in a classification accuracy of 96.45% [30]. For an overview of sensing modalities and algorithms, see [25]. Despite the widespread attention, these models are constrained by one or more of the following factors – the experiment conditions and their operational relevance, the participant demographic, the stress-induction paradigm, or the ground-truth that relates stress labels.

2.1 Experiment Conditions and Operational Relevance

The *experiment setting* is typically non-ambulatory, where the stressor, task, and environment are carefully controlled, e.g. workplaces, laboratories, or other supervised environments [31], [32]. While models developed under these constraints can be highly accurate, generalization is limited due to their sensitivity to external perturbations, and changes to the environment [33]. Some researchers have advanced model development toward ecologically valid, and ambulatory settings that are consistent with real-world conditions (e.g. [34], [35], [36]). These efforts aid in the expansion of existing benchmarks, however concerns related to (1) the role of contextual information such as motor activity; (2) *sensor choice*, given the fieldability and fidelity demands of operational use; and (3) *sensitivity* to subjective variations and demographic factors such as age remain.

Therefore, there is a need for studies that incrementally and methodically assess the effect of *controlled* perturbations, e.g. motor activity like those during an exercise or rehabilitation routine, on stress representations. Such

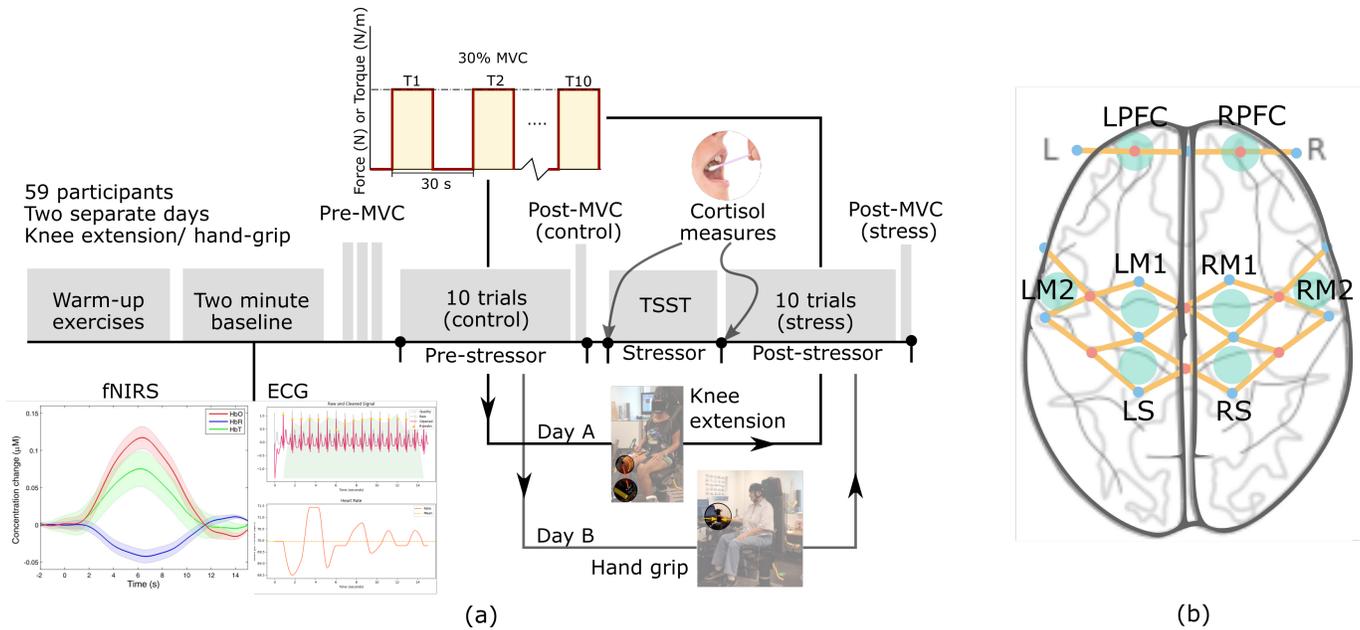


Fig. 1. (a) Schematic representation of the experiment timeline. Participants began each day with warm-up exercises and baseline measurements. Maximum voluntary contraction was measured before and after each block condition. Salivary cortisol samples were taken before and after the TSST block to verify stress induction. Motor activity pre- and post-stressor was identical in each session with hand grip or knee-extension exercises administered to the participant contingent on the experiment protocol. (b) The probemap used for capturing brain hemodynamics during the experiment. Brain regions of interest are as highlighted, with 26 channels on the motor cortex, and 4 channels on the prefrontal cortex.

considerations are especially relevant to older adults who suffer from the dual-task burden. We hypothesize that affective computing tools that incorporate motor activity will improve rehabilitation outcomes by advancing state-aware intervention methods for older adults, given that motor activity can aid in neurocognitive performance improvements over time [37], while stress is known to impede relevant faculties [38].

2.2 Subject Diversity

Stress-responses and manifestation are individualistic and influenced by demographic factors [39]. A critical limitation in prior stress detection research is subject diversity. Typical populations recruited for these experiments are restricted to the college-going age group (e.g. [30], [31]) and group balance considerations across demographic attributes such as sex, age or physiology are not widely reported. This perpetuates the deficiencies of representation seen in human subjects literature [40].

Cognitive stress is known to affect males and females to different extents, and along distinct dimensions. Studies have reported differences in strategies adopted by males and females in coping with these stressors (e.g. [41]). A meta-analysis from Liu et al. [42] found that the Trier Social Stress Test (TSST), an established stress induction paradigm in behavioral research, elicits different salivary cortisol levels across the sexes and that the peak levels, and recovery patterns were different across task modalities. In addition, we know that most neurophysiological indices

are sex-specific, e.g., resting state functional connectivity differences in the brain [43] and the activation and recovery patterns of HR/V [44], etc.

Age is a relevant demographic because pathophysiological conditions and oxidative stress can mask the typical neurophysiological indicators of stress [44], yet, few stress detection models have been developed using data from older adults. In one recent study Delmastro et al. [45] developed a stress detection model with nine participants under mild cognitive impairment, however the work relied on implicit stress labels derived from the experiment protocol which resulted in an imbalanced data set. Nevertheless, the preliminary results were encouraging with a state estimation accuracy of 88.2% when using HR/V and EDA-based features. Furthermore underlying conditions in older adults may obscure the representations of stress, which demands considerations toward factors like levels of physical activity or an aging population specific comorbidities such as obesity in our samples. Needless to say, older adults remain under-represented in this domain, and ensuring balance along other dimensions such as sex or physiology present as additional barriers.

2.3 Stress Labels and Induction Methods

Another limitation in existing research relates to the ground truth reference for stress. Some studies rely on the experiment protocol as "implicit" ground truth (e.g. [45]), i.e. states prior to a stressor are regarded as "no-stress" states while states during or after the stressor are labeled

as “stress” states. This approach places emphasis on the type of stressor used, where uncertainty due to individual differences moderate our overall confidence [46]. Other studies rely on subjective responses as the index for ground truth stress. A battery of surveys and questionnaires have been adopted to reflect on these state changes (e.g. the STAI or NASA Task Load Index), but such methods suffer from recall bias, add to the participant’s cognitive burden, and are not directly relevant to their experience of the stressor. Methods that use neurochemical indices, e.g. salivary cortisol or salivary-alpha amylase, both widely used measures in the clinical diagnosis of emotional stress events [47], [48] are optimal. However, few studies rely on these mechanisms given the associated challenges.

The choice of ground truth stress is also impacted by the choice of stressor, not all stress-induction methods are equivalent. Techniques that rely on the use of emotional imagery, simulated stress events, driving tasks or other tasks may be inherently limited by a lack of standardization and subject variability. There is a growing scientific consensus behind validated techniques such as the Triers Social Stress Test (TSST), which relies on a combination of public interview and mental arithmetic [49], and is known to increase activation of the hypothalamic-pituitary adrenal axis, consistent with an acute and instantaneous stress response [50].

In their review on psychological stress detection with bio-signals, Giannakakis et al. identified 41 studies on stress detection with machine learning tools (see Table 3 in [25]), however only six studies among them (14.63 %) relied on a validated stress-induction protocol (e.g. TSST or the Montreal Imaging Stress Task). Therefore, the choice of stressor remains a major impediment to the broader relevance of earlier findings. Incidentally, among those six studies, not one study achieved a sex balance, in fact, only two among the six included a female population, and only one in all 41 reported an overall sex balance. Further, in those six studies, the participant age spanned 20 to 30 years with only two out of 41 relying on a sample that included senior adults (> 65 years), highlighting the importance of representation and use-case relevant subject demographics. Finally, none among those six studies relied on a neurochemical stress index for labeling their observations, and only two out of the 41 reported the use of a related neurochemical index for their ground truth. These observations highlight the deficiencies and opportunities that motivate our current work.

2.4 Summary

There are applied and methodological challenges in protocol design, experimentation, data collection, feature identification, and the development of machine learning models robust to the challenges discussed in the prior sections. Through our current work we wish to improve on some of the existing limitations – (1) we leverage a motor-intensive

exercise and explore stress detection during concurrent motor activity; (2) we extend subject diversity by considering an under-represented sub-group, i.e. older adults; (3) we ensure a sex and body-type balance in our sample; (4) we adopt a validated stress induction protocol; and (5) we rely on salivary cortisol measures as ground truth to label stress states. This way we hope to advance the collective understanding on the role of sex and motor activity on the neurophysiological representations of stress in older adults.

3 METHODS

3.1 Data set

The data operationalized in this investigation derives from a comprehensive resource introduced in [51]. Below we detail elements of a stress protocol and the corresponding participant pool relevant to our present work.

3.1.1 Experiment protocol

On informed consent, all participants completed two experiment sessions. Each session entailed an identical motor *exercise*, either hand-grip or knee-extension before and after a stress-induction period. Participants came in on separate days for each session and the order of the motor exercise was counterbalanced to account for any familiarization bias. The sessions began with participants wearing relevant instruments, warm-up exercises, baseline capture of bio-signals, and strength testing using maximum voluntary contraction (MVC) protocols (see Fig. 1). After the MVC, participants completed a control task where they were instructed to maintain a force level at 30% MVC for 15s followed by 15s at rest for ten trials i.e. the *pre-stressor* phase for a total time of 300s. Participants then engaged in the stress induction activity, a Triers Social Stress Test (TSST) [49] for the *stressor* phase. In the TSST, participants were given 10 minutes to prepare and deliver a five minute speech on an unknown topic to an independent review panel. Following the speech, participants were engaged by the panelists in a series of challenging mental arithmetic tests for another five minutes. During the stressor activity, the panelists maintained a neutral affect and took notes on participant behavior, while periodically monitoring a video camera. Salivary cortisol samples were taken before and after the TSST block to serve as ground truth for cognitive stress. After the *stressor* phase, participants repeated the motor task for 10 trials resulting in the *post-stressor* experiment block. During both *pre-* and *post-stressor* conditions participants received visual feedback about the current force/ torque levels to help them maintain their hand-grip force and knee-extension torque at 30% MVC. Importantly, the force/ torque threshold was limited to 30% their MVC so that participants could reliably maintain these levels without getting fatigued for the five minute duration. Participants also reported their State-Trait and Anxiety Inventory (STAI) responses after both the *pre-stressor* and *post-stressor* activities.

3.1.2 Participants

The study included a total of 60 older adults, with 30 females and 30 males. In each sex, 15 individuals were obese with a body mass index (BMI) > 30.0. One male, obese participant dropped out during the experiment. All participants were right hand dominant; and reported no musculoskeletal injuries for at least an year prior to the experiments. The median age among them was 71 years ($\mu = 72.23$, $\sigma = 5.69$ years), the median BMI was 26 ($\mu = 30.18$, $\sigma = 7.61$), the median height was 1.68m ($\mu = 1.69$, $\sigma = 0.11m$), and the median weight was 85kg ($\mu = 86.89$, $\sigma = 23.82kg$), these numbers together form a representative sample of older adults (> 65 years) in the United States [52]. All procedures were approved by University's Institutional Review Board (IRB Number: IRB2015-0647F), and proceeded in accordance with the *Ethics Code* of the *American Psychological Association*.

3.1.3 Bio-instrumentation

Participants were instrumented with a continuous wave fNIRS device (TechEn CW6 system TechEn inc., Milford, MA, USA) with a probemap focused on the motor cortex (supplementary motor area (SMA), primary motor cortex (PMC)) along with four channels on the prefrontal cortex (PFC) that captured brain hemodynamic activity, these regions were chosen given the motor-intensive nature of the task. The participant's heart rate and its variability (HR/V) were monitored using a three lead ECG, and amplifier interface (BIOPAC ECG100C, BIOPAC Systems Inc., Aero Camino Goleta, CA, USA). For the hand-grip protocol, participants interacted with a hand dynamometer (SSL25B, BIOPAC Systems Inc., Aero Camino Goleta, CA, USA), while an isokinetic dynamometer (HUMAC NORM, Computer Sports Medicine Inc., Stoughton, MA, USA) was used during the knee-extension exercises.

3.2 Preprocessing of biosignals

3.2.1 Brain hemodynamics

Cerebral hemodynamics was captured using the fNIRS device at 50Hz. Near-infrared spectra transmitted at two wavelengths ($\lambda_{1,2} = 690, 830nm$), from 8 emitters, passing through neural tissue and blood vessels, were detected by 13 separate detectors. The transmitted signal was used to characterize vascular hemodynamics across a network of 30 channels. Firstly, the light intensity registered by the instrument ($\mathbf{I}_o(\lambda)$) was converted into optical density ($\mathbf{OD}(\lambda)$) using a log transform on the ratio of received (output) to transmitted (input) intensities [54],

$$\mathbf{OD}(\lambda) = \log_{10} \left(\frac{\mathbf{I}_o}{\mathbf{I}} \right)_{\lambda_{1,2}} \quad (1)$$

The optical density signal was low-pass filtered at 3Hz to reduce high frequency noise that could mask the underlying signal. Motion artifacts that showed abrupt change

TABLE 1
Features derived across each window from the time series brain hemodynamics data.

Temporal (\mathbf{X}_{fnirs})	
Mean (μ)	$\frac{1}{n} \sum_{i=1}^n x_i$
Variance (σ^2)	$\mathbb{E}[(\mathbf{X} - \mu)^2]$
Max.	Maximum value in window
Min.	Minimum value in window
Kurtosis	$\mathbb{E} \left[\left(\frac{\mathbf{X} - \mu}{\sigma} \right)^4 \right]$
Skewness	$\mathbb{E} \left[\left(\frac{\mathbf{X} - \mu}{\sigma} \right)^3 \right]$
AUC	$\approx \frac{1}{2} \frac{x_0 - x_n}{n} \left[x_0 + 2 \cdot \sum_{i=1}^{n-1} x_i + x_n \right]$
Functional connectivity (\mathbf{X}_{FC})	
Corr.	$\frac{\mathbb{E}[(\mathbf{X}_1 - \mu_1)(\mathbf{X}_2 - \mu_2)]}{\sigma_1 \sigma_2}$

* \mathbb{E} = the expected value; AUC = area under the curve, Corr. = linear correlation coefficient.

(peaks) were detected and corrected using a spline interpolation algorithm [55], and further smoothed using wavelet transforms [56]. The rectified signals were band-pass filtered at 0.5 – 0.016Hz to reduce the effect of physiological noise and slow wave drifts. Lastly, change in oxygenated, deoxygenated, and total ($\Delta\text{HbO/R/T}$) hemoglobin was calculated across all 30 channels using the modified Beer-Lambert principle [57]:

$$\Delta\text{OD}(\lambda) = \epsilon(\lambda) \cdot \Delta\mathbf{c} \cdot \mathbf{d} \cdot \text{DPF}(\lambda) + \mathbf{g}(\lambda) \quad (2)$$

Where $\Delta\mathbf{c}$ is the attenuation due to the hemoglobin (Hb) protein, $\Delta\text{OD}(\lambda)$ is the change in optical density, $\epsilon(\lambda)$ is the attenuation coefficient of HbO/R/T at a particular wavelength (λ), $\text{DPF}(\lambda)$ is the differential path length factor, \mathbf{d} is the source-detector separation distance, and $\mathbf{g}(\lambda)$ is the scattering coefficient for a given wavelength. The change in concentration time series data was further smoothed using an exponential weighted moving average function such that,

$$\kappa_t = \begin{cases} \Delta c_t, & \text{at } t = 0 \\ \alpha \cdot \Delta + (1 - \alpha) \cdot \kappa_{t-1}, & \text{otherwise} \end{cases} \quad (3)$$

Where κ_t is the exponential weighted moving average of Δc at time t , and α ($= 0.85$) is the weighting coefficient. Cortical positions for the 21 optodes (i.e., 8 emitters and 13 detectors; see Fig. 1 (a)) is determined based on the 10/20 international EEG system using AtlasViewer [58]; see Fig. 2 (a). Further, in the present study, the time-series data across 30 channels were grouped into eight regions of interest (ROIs) that were defined based on the functions of a sensorimotor network from both hemispheres [59], namely left/right PFC, left/right medial motor area (LM1 and RM1, respectively), left/right lateral motor area (LM2

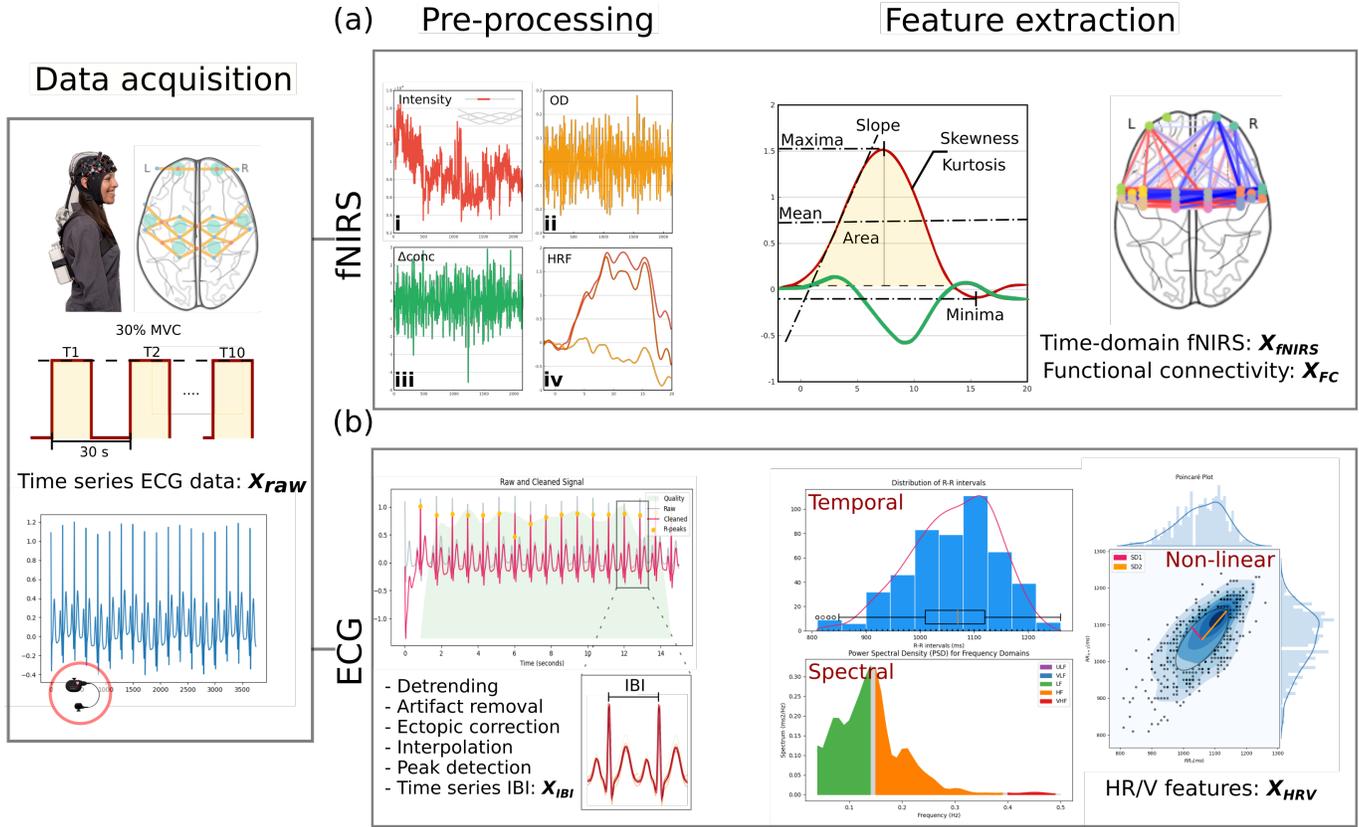


Fig. 2. Feature extraction workflow employed in this study. (a) Cerebral oxygenation from fNIRS signal resulted in time-domain fNIRS, and functional connectivity features. (b) The ECG data resulted in temporal, spectral, and nonlinear HR/V features using the Neurokit2 library [53]. The workflows feed into a training, and validation pipeline.

and RM2, respectively), and left/right sensory area (LS and RS, respectively).

3.2.2 Heart rate variability

Cardiac electrical activity was obtained from the ECG probe and amplifier interface at $1000Hz$. The electrodes were positioned at the base of the sternum and over the left pectoralis minor muscle. The raw ECG signal was filtered for motion-related artifacts [60], and corrected for ectopics with polynomial interpolation [61]. Subsequently, a peak detection algorithm was used to isolate the R peaks from the ECG signal [62]. The time between successive R - R peaks, i.e. the inter-beat-interval (X_{IBI}) or normal-to-normal (NN) interval was then derived from the processed peak signals for subsequent feature engineering.

3.3 Feature extraction

3.3.1 Time-domain fNIRS and functional connectivity

The pre-processed fNIRS signal was subject to participant-level feature scaling (min-max normalization) to account for individual differences and consolidated into sliding windows of size 150, and a step size of 75 which resulted in 50% overlap between windows. The windows were used

to extract relevant time-domain fNIRS features [63]. These features were derived for all 30 channels, and signal types (HbO/R/T), resulting in 12,980 labeled observations (59 participants \times 20 trials \times 11 windows) on the knee data, and the hand data respectively, including *Stress* and *No Stress* conditions. There were 630 unique temporal features (30 channels \times HbO/R/T measures \times 7 features) in all (see *Table 1*; Fig. 2 (a)). Further, for each window, channel, and signal type we derived pairwise Pearson correlation statistics to measure functional connectivity [64] which resulted in 4,005 additional connectivity features ($= (30 \times 3)^2 / 2 - 90$). The fNIRS feature matrix had 12,980 observations, with $X_{fNIRS} \in \mathbb{R}^2$ having 630 features, and $X_{FC} \in \mathbb{R}^2$ having 4,005 features.

3.3.2 Temporal, spectral, and nonlinear HR/V

The corrected inter-beat-interval (X_{IBI}) obtained from the continuous ECG signal was subject to participant-level feature-scaling (min-max normalization) and partitioned into time windows of 60s with a step size of 30s, resulting in 50% overlap between windows. Temporal, spectral, and nonlinear features were then extracted from each window resulting in 50 features for heart rate and its variability (HR/V) using the Neurokit2 library [53]. The non-linear

TABLE 2
Short-term heart rate variability (HR/V) features for each window (60s) of the filtered ECG signal.

Temporal		
Measures of central tendency	RMSSD	RMS of successive differences (SD) of NN intervals
	SDNN	Standard deviation of NN
	SDSD	Standard deviation of SD
	MeanNN	Mean of NN
	pNN50	Proportion of intervals > 50ms
	MadNN	Median absolute value of SD
	CVNN	SDNN/MeanNN
	CVSD	RMSSD/MeanNN
	TINN	Base width of NN histogram
	HTI	HRV triangular index
IQRNN	Inter-quartile range of NN	
Spectral		
Power density	ULF	Ultra low (0 – 0.0033Hz)
	VLF	Very low (0.0033 - 0.04Hz)
	LF	Low (0.04 – 0.15Hz)
	HF	High (0.15 – 0.4 Hz)
	VHF	Very high (0.4 – 0.5 Hz)
Nonlinear		
Poincaré geometry	SD1,2	Transverse and longitudinal variability of poincaré plot
	CVI	Cardiac vagal index
	CSI	Cardiac sympathetic index
Asymmetry	SD1d, a	Short-term deceleration (d) and acceleration (a)
	C1d/C2d	Contribution of HR <i>d</i> , <i>a</i> on short-term, and long-term HR/V
	C1a/C2a	Total variance from <i>d</i> or <i>a</i> changes
	SDNNd/a	
Fragmentation	PIP	Percentage of inflection points
	IALS	Inverse of average length of <i>a</i> , <i>d</i> segments
	PSS	Percentage of short segments
	PAS	Percentage of NN intervals in alternating segments
Complexity	ApEn	Approximate entropy
	SampEn	Sample entropy

* *NN* = Normal to Normal interval (*ms*); *F* = frequency (*Hz*).

features were further categorized as measures of asymmetry [65], fragmentation [66], complexity and geometry [67] (see Table 2). The composite HR/V feature matrix $\mathbf{X}_{HRV} \in \mathbb{R}^2$ had 1,300 labeled observations with 50 features each; see Fig. 2 (b).

3.4 Machine learning

3.4.1 Ground truth definition

Ground truth for each observation was derived from salivary cortisol samples and self-reports (STAI) taken during the experiments. Pair-wise t-tests on pre- and post-TSST cortisol measures confirmed a significant increase in cortisol levels after the stress-induction period ($p = 0.049$). This finding also concurred with STAI statistics that point to elevated state-anxiety after the TSST exercise ($p = 0.026$). Hence, we used two class labels — *Stress* or *No Stress*.

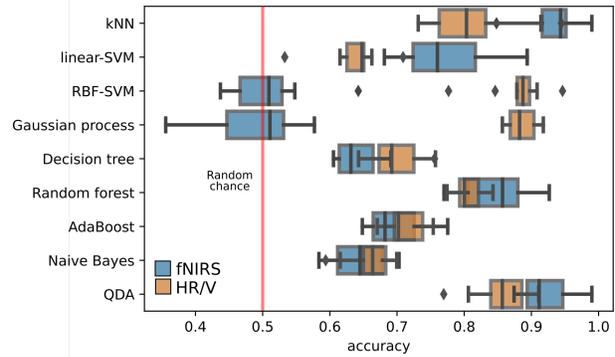


Fig. 3. Stress detection accuracy of each model on fNIRS, and HR/V based features. The parametric Quadratic Discriminant (QDA), and the non-parametric Random Forest (RF), and k-Nearest Neighbors (kNN) models performed well across all data.

Observations from the *pre-stressor* phase were labeled as *No Stress*, while observations from the *post-stressor* phase were labeled as *Stress* (see Section 3.1.1). This labeling strategy followed the scientific consensus around the TSST protocol, where the effects of stress are instantaneous and known to persist for at least 10 minutes after the induction period [68]. Secondly, although the experiment protocol comprised of rest and contraction events, post-hoc sensitivity testing showed that these states do not substantially influence model outcomes.

3.4.2 Model training and optimization

Using the labeled data set, we built a machine learning pipeline to train, optimize parameters, and evaluate nine machine learning algorithms on fNIRS and HR/V data independently. We used 10-fold, stratified, cross validation on the data, with 75% of participants in training, and the remaining as a “hold out” test set in each fold to avoid the bias of highly correlated samples coming from the same participants [69]. We optimized for accuracy by tuning a subset of hyper-parameters in each model using a grid-search process. Here accuracy was defined as,

$$accuracy = \mathbb{E}[\varepsilon(\mathbf{y}^{pred}, \mathbf{y}^{ground\ truth})] \quad (4)$$

where, \mathbb{E} is the expected value (mean) of the error (ε) when $\mathbf{y}^{pred} \neq \mathbf{y}^{ground\ truth}$. The observed accuracy, and optimization targets for each model are provided in Table 3. All models were developed using *scikit-learn* 0.23.2 on *Python* 3.7.8. In subsequent discussion and tables we report the mean cross-validated testing accuracy and the standard error of the mean (SEM) unless otherwise indicated.

3.4.3 Model evaluation and selection

We evaluated the models with accuracy as our metric and contrasted the top performing model using the 5×2 CV paired t-test method proposed in [70]. Overall, most

TABLE 3

Mean 10-fold CV accuracy (\pm std) across selected algorithms and their optimized hyper parameters with independent training on fNIRS and HR/V data set from hand-grip exercises

Algorithm	Params.	Accuracy (%)	
		fNIRS	HR/V
kNN	k = 3	0.941 \pm 0.005	0.813 \pm 0.027
L-SVM	c=10	0.773 \pm 0.013	0.638 \pm 0.042
RBF-SVM	c=10, $\gamma = 0.1$	0.551 \pm 0.000	0.882 \pm 0.012
GP	kernel=1*RBF(0.5)	0.448 \pm 0.000	0.885 \pm 0.021
DT	criterion = gini, max_depth = 50, min_samples_split = 2, min_samples_leaf = 2, max_features = 'auto'	0.647 \pm 0.015	0.0697 \pm 0.036
RF	n_estimators = 300, criterion = gini, max_depth = 50, min_samples_split = 2, min_samples_leaf = 5, max_features = 'auto', bootstrap = False	0.845 \pm 0.005	0.806 \pm 0.021
AB	$\alpha = 0.75$	0.695 \pm 0.012	0.712 \pm 0.028
NB	—	0.645 \pm 0.013	0.659 \pm 0.032
QDA	reg_param = 0.4	0.922 \pm 0.079	0.855 \pm 0.031

*kNN = k-nearest neighbors; L-SVM = linear support vector machine; RBF = radial basis function; GP = gaussian process; DT = decision tree; RF = random forest; AB = adaboost; NB = naive bayes; QDA = quadratic discriminant analysis; k = No. of neighbors; c = penalty; γ = similarity radius; α = learning rate.

algorithms had a classification accuracy substantially better than chance across the two data sets (see Table 3 and Fig. 3). Among them, the k-Nearest Neighbors (kNN), Random Forest (RF), and the Quadratic Discriminant (QDA) performed the best across both fNIRS and HR/V data sets with a mean accuracy of 0.903 ± 0.006 and 0.825 ± 0.027 respectively. We found that QDA outperformed RF and kNN models on HR/V data with p -values of 0.026, and 0.038 (pre-determined threshold of 0.05). On fNIRS data both QDA and kNN methods outperformed the RF model. For the remainder of the analysis we focus on the QDA results given its observed balance of accuracy, interpretability, complexity, and classification efficiency.

3.4.4 Model comparisons and feature importance analysis

On selecting QDA as our algorithm of choice, we trained separate models to assess classification performance across – i. feature (\mathbf{X}_{fnirs} , \mathbf{X}_{FC} , $\mathbf{X}_{HR/V}$); ii. sex (male, female); and iii. exercise-related (hand or knee data) variables. The mean classification accuracy for each model and training combination is reported in Table 4. In our comparisons, we explicitly train and evaluate models within sex or exercise-specific data subsets and in turn, evaluate model generalizability by testing across those subsets (see Table 4). Furthermore, we quantified the variables of importance for each model using a permutation importance index, where

we iteratively replaced each variable in the feature set (\mathcal{D}) with random noise, and tracked changes in model accuracy.

Consider the model m a product of cross-validated training on $\mathcal{D} = \{(x_i, y_i) \mid i = 1 : N\}$ where $\mathcal{D} \ni (\mathbf{X}_{fnirs}, \mathbf{X}_{FC}, \mathbf{X}_{HR/V})$, let s be the accuracy of the classifier m when trained on \mathcal{D} . For each feature $j \in \mathcal{D}$, we randomly shuffle the j^{th} column in \mathcal{D} to generate an altered data set \mathcal{D}_j . We repeat this process K times to generate a randomized combination $\mathcal{D}_{j,k}$ for each iteration. We then determine the accuracy of model m on the altered data set to obtain an updated accuracy score $s_{j,k}$ for feature j on the k^{th} shuffle. We define an importance measure i_j for the chosen variable as follows:

$$i_j = s - \frac{1}{K} \sum_{k=1}^K s_{j,k} \quad (5)$$

This score, effectively the mean classification error on permutation, is used to rank feature importance across models.

4 RESULTS AND DISCUSSION

4.1 State estimation with fNIRS

We found that the \mathbf{X}_{fnirs} features enabled a stress classification accuracy of $98.06 \pm 0.4\%$ with a specificity of 99.04%, and sensitivity of 96.23% across combinations of the input data set when trained, and tested within subsets. Table 4 shows the pairwise train-test accuracy across feature, sex, and exercise-type distinctions, where the rows indicate the training data, and the columns indicate the test data, therefore, each cell presents the classification accuracy for a model trained on the row-header, and tested on the column-header. We rely on accuracy as our metric for discussion given the balance in our data set and the stratified evaluation process. Furthermore, the accuracy observed here compared favorably with prior work that involved the use of brain hemodynamics (e.g. [32], [71]) — this is notable especially given the relative simplicity of the QDA algorithm and our probemap which was not focal to brain regions typically associated with human stress response.

The PFC is responsible for top-down executive control and our highest order of cognitive functions, and is the region most sensitive to detrimental effects of stress with both functional and structural changes over time [72]. In our study we employed a probe configuration (see Fig. 1 (b)) primed toward the motor cortex, with only four channels on the PFC. We found that this did not hinder our classification accuracy. Interestingly, channels from the PFC region were not significant to outcomes altogether, we retained a cross-validated classification accuracy of $97.42 \pm 0.004\%$ on explicitly removing them from our training pipeline. Fig. 4 (a), and (b) provide an overview of the observed accuracy when relying on features from the ROIs introduced in Fig. 1 (c), here, sex and exercise-related differences are apparent.

TABLE 4

Accuracy across temporal fNIRS features, functional connectivity (FC) features, and HR/V indices respectively. The rows indicate the *training* data, and columns indicate the *test* data, therefore each cell indicates the testing accuracy for a model trained on the row-header, and tested on the column-header. For diagonal elements we report a mean 10-fold, cross-validated accuracy when using a 75-25 train-test split, for off-diagonal elements we report the mean accuracy from testing on the relevant subset using the 10 trained models with $SEM \in [0.0001, 0.07]$.

		Hand			Knee			
		All	Female	Male	All	Female	Male	
fNIRS	Hand	All	0.978±0.005	0.991	0.987	0.456	0.398	0.523
		Female	0.764	0.993±0.004	0.521	0.414	0.328	0.512
		Male	0.766	0.543	0.994±0.004	0.406	0.342	0.478
	Knee	All	0.530	0.510	0.552	0.945±0.005	0.975	0.946
		Female	0.526	0.518	0.533	0.797	0.988±0.005	0.574
		Male	0.538	0.515	0.562	0.661	0.369	0.985±0.003
FC	Hand	All	0.790±0.004	0.668	0.644	0.462	0.463	0.461
		Female	0.644	0.775±0.003	0.500	0.500	0.488	0.516
		Male	0.590	0.490	0.758±0.005	0.468	0.456	0.481
	Knee	All	0.485	0.483	0.488	0.810±0.002	0.824	0.735
		Female	0.487	0.483	0.481	0.764	0.824±0.004	0.725
		Male	0.486	0.483	0.488	0.735	0.823	0.797±0.003
HR/V	Hand	All	0.751±0.05	0.800	0.748	0.625	0.615	0.636
		Female	0.760	0.819±0.04	0.656	0.604	0.574	0.637
		Male	0.759	0.719	0.802±0.05	0.643	0.635	0.652
	Knee	All	0.684	0.700	0.662	0.689±0.06	0.716	0.721
		Female	0.684	0.703	0.659	0.725	0.753±0.04	0.661
		Male	0.679	0.691	0.665	0.686	0.621	0.733±0.05

Between sexes, we found that \mathbf{X}_{fnirs} enabled consistent state estimation for models trained and evaluated *within* each subset with a mean accuracy of $99 \pm 0.31\%$. When evaluating the model across sexes the mean accuracy reduced to $53 \pm 0.10\%$, with poor classification in both directions i.e. for models trained on males and evaluated on females or vice-versa. In addition, the models were found to be highly sensitive to the the exercise-type involved — accuracy when trained and evaluated *within* the hand-grip data subset was $97.81 \pm 0.56\%$, while that *within* the knee-extension subset was $94.56 \pm 0.54\%$. When evaluating across exercises, i.e. training on the hand subset and testing on the knee subset or vice-versa, the classification accuracy was poor bidirectionally with a mean value of $41 \pm 0.16\%$.

4.1.1 Relevant brain regions

We used the permutations of importance measure introduced in Section 3.4.4 on each model to identify ranked regions of importance as shown in Fig. 4 (c). The top fifty model features were selected and mapped to the 10/10 EEG space contingent on the channel they were sourced from. Each coordinate received a measure which was the mean value of the importance index (i_j) of all features derived from that region. With this approach we observed that features from bilateral medial regions of the motor cortex appear significant to outcomes on hand data, and left-lateral ventral regions appear significant to outcomes with knee data, with a mean error increase of 15.23% and 17.84% on

permutation respectively. Further, the four channels located on the frontal lobe, a region more consistently associated with stress and its cognitive underpinnings, were not coherently relevant to classification outcomes. We observed that in males, the medial frontal region resulted in a mean error increase of 9.46% on permutation, while in females it resulted in a marginal mean error increase of 0.15%. Conversely in females we found a greater contribution from the medial PFC when using knee data with a mean error increase of 6.52% on permutation, while remaining negligible in males with a mean error increase of 0.0027%.

4.1.2 Functional correlations

In this section, we present the results of our classification attempts when relying on multi-channel, time-series correlation metrics Section 3.3.1. We obtained a mean state detection accuracy of $79.27 \pm 2.39\%$ for models trained *within* data subsets using \mathbf{X}_{FC} . Across exercises we observed a classification accuracy of $79.03 \pm 0.42\%$ with hand data, and an accuracy of $81.06 \pm 0.21\%$ with knee data. On validation across the sexes, we obtained a mean accuracy of $49.32 \pm 0.17\%$ in hand data and a mean accuracy of $77.42 \pm 0.56\%$ for knee data under consistent input conditions.

Overall, \mathbf{X}_{FC} appeared to be less sensitive to the stressor than \mathbf{X}_{fnirs} , we found a decrease in model accuracy of $\approx 20\%$. Unlike with \mathbf{X}_{fNIRS} , we observed that on knee data, models trained with connectivity features generalized

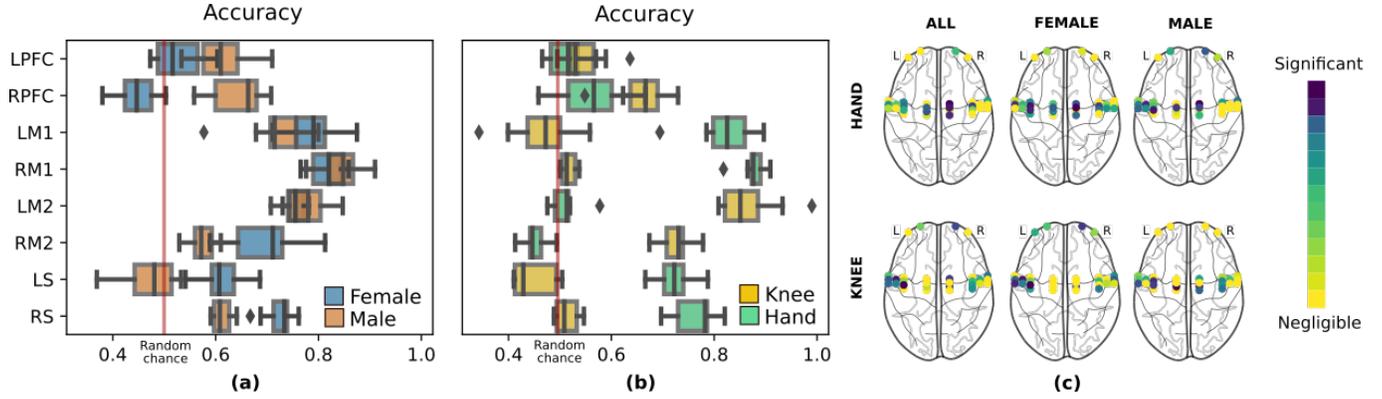


Fig. 4. Stress detection accuracy across (a) the sexes, and (b) limb types, when relying on \mathbf{X}_{fnirs} features sourced from the regions of interest described in Fig. 1. (c) The permutations of importance score was mapped to 10/10 EEG coordinate-space to determine regions of importance from the fNIRS probemap. Regions from the medial aspect were significant to outcomes from hand data, while those from the ventral aspect of the motor cortex were relevant to outcomes on the knee data. The frontal cortex was not coherently important to the classifiers output across the data.

better than with other fNIRS features with a mean accuracy of $77.42 \pm 0.56\%$ across the sexes. We hypothesized that this is due to the nature of each activity; the hand-grip exercises are known to exhibit a relatively distributed cortical footprint, while the knee-extension task recruited a focused muscle group and therefore a limited cortical footprint, perhaps resulting in differences that transcend sex-specific neural activity under cognitive stress. This hypothesis is visually supported by the connectivity feature importance maps shown in Fig. 5, where the network density is focal with knee data while diffuse and distributed for hand data. Prior research supports this idea, where central motor and cognitive structures are known to contribute differentially to isolated knee and elbow movements (e.g. [73]). However, future work is needed to assess this hypothesis with more rigor.

4.1.3 Discussion on stress detection with fNIRS

The results introduced thus far suggest that our algorithm, workflow, and the temporal fNIRS features are sensitive to the cortical effects of stress beyond hemodynamic activity associated with a motor task. Further, although there are known cortical differences in the representations of motor activity within the fNIRS signal [74], the findings show that this does not hinder the classifier’s sensitivity to stress states. In some ways, the cognitive effects of the stressor exhibit a form of motor-task *independence* — where motor actions do not interfere with cortical representations of stress identified by the trained models.

However, we also note that models trained on \mathbf{X}_{fnirs} features do not generalize well across sex or exercise-related variables; we believe that this is tied to the sensitivity of temporal hemodynamics to both motor-activity and demographic differences [43]. The regions of importance (see Fig. 4 (c)) further reinforce this hypothesis; we identified distinct ROIs across the sexes, and a sensitivity to the limb type engaged during the experiment. However,

with connectivity metrics (\mathbf{X}_{FC}), although we observed a decrease in overall classification accuracy relative to \mathbf{X}_{fnirs} , we also note improved generalizability on knee data. This we believe is explained by the focality of the knee-extension exercise that was shown to recruit a limited set of muscle groups, unlike the distributed neuromuscular footprint of the hand-grip exercises. In some ways, we find that the nature of the task can alter the sensitivity of neural indicators to demographic variables, and that this influence appears to be stronger on connectivity than on other attributes of the fNIRS signal.

These observations reinforce the view that traditional region-specific denominations of cognitive constructs (e.g. stress [72]) and their influence need reconsideration; perhaps one that transitions into a view of our cortical resources as a distributed connectome, with stress having broader implications on neural activation than previously understood. In addition, the lack of generalizability across sex or exercise-type differences underscores the sensitivity of model development to those variables, and the task-specificity of the human neurophysiological stress response. These factors are especially relevant to rehabilitation programs that target older adults, where motor-cognitive synergies are known to play a central role across exercises (e.g. [10], [11]).

4.2 State estimation with HR/V

With temporal, spectral, and non-linear HR/V features (\mathbf{X}_{HRV}), when relying on a quadratic discriminant we achieved a mean cross-validated classification accuracy of $76.37 \pm 5.58\%$ *within subsets*. Between sexes we see a mean accuracy of $67.16 \pm 4.28\%$ across exercise-types. Between exercise types we see a mean accuracy of $65.72 \pm 4.6\%$ across the sexes. Overall, models built on \mathbf{X}_{HRV} features generalized better than those developed on fNIRS data across both sex and exercise-related distinctions.

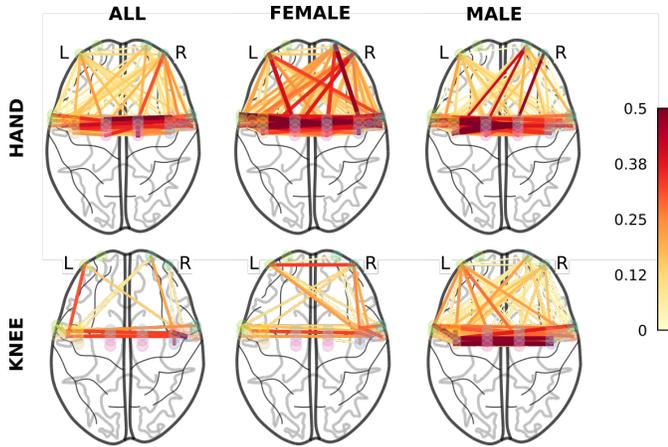


Fig. 5. Feature importance for functional connectivity measures. Functional connectivity appears to be diffuse, and collectively relevant to the classifier’s output when using hand data, but focal on knee data.

Among the variables of importance, short term measures of heart rate asymmetry were found most important, Poincaré plot geometry measures were next followed by spectral and temporal attributes. Fig. 6 (a) presents the feature importance rank for the domains introduced in Section 3.3.2, where we found that the feature-groups exhibit identical levels of importance across each data set, with nonlinear attributes appearing more prominently in each subset.

4.2.1 Discussion on stress detection with HR/V

HR/V is a secondary indicator of cognitive state, in that, the dynamics of inter beat interval are mediated by the vagal nerve that provides both a functional and structural link between cognitive processes in the brain and the stochasticity of heart rate variability [75]. These interactions represent the balance between sympathetic and parasympathetic activity and in some ways this response is a few steps removed from its neural origins. Under the effects of stress, we expect an increased sympathetic footprint due to the cognitive burden associated with the *TSST* [76], and a decrease in parasympathetic tone, however this is not a zero-sum system and the effects are easily confounded by perturbations due to physical activity or breathing. Typically the effects of stress are associated with changes in the spectral power density over the HF and LF regimes [77], however in our observations the spectral components do not appear to strongly influence the classifier’s output. In fact, operating on just Poincaré plot geometry and heart rate asymmetry based features we observed a comparable stress detection accuracy of $78.76 \pm 0.04\%$ (see Fig. 6 (b)).

The Poincaré plot represents the nonlinear fluctuations of the RR interval over the time course of the experiment, and is a graphical indicator, where a scatter plot shaped like an ellipse describes the nonlinear dynamics of heartbeat. Here, the dispersion of points along the identity line (i.e.,

the major axis of the ellipse) indicates long-term variability while dispersion along the minor axis indicates short-term variability [78]. Prior studies suggest that the former is indicative of parasympathetic regulation [79], while the latter and the ratio of the two points to sympathetic cardiac regulation [80]. In our study, we found statistically significant differences in *SD1* and *SD2*, i.e. the dispersion along the major and minor axes respectively, between *Stress* and *No Stress* conditions ($p = 0.037$ for *SD1*, and $p = 0.048$ for *SD2*), with a decrease in *SD1* and an increase in *SD2* post-stressor, which agrees with expectations, and explains the ability of the models to separate the two classes.

Heart rate asymmetry variables are also tied to sympathovagal balance control, where increased sympathetic activity does not occur concurrently with vagal withdrawal resulting in a time lag or time asymmetry in heart rate. Prior studies (e.g. [81]) reported that acute mental stress results in elevated HRA which supports its relevance as an index of stress in our data set. There are some advantages to the use of nonlinear measures; the empirical evidence that connects spectral and temporal HR/V characteristics to sympathovagal balance often rely on longer time-windows (5 min. - 24 hrs), and are suspect as short-term measures [67]; nonlinear HR/V indices provide a richer template to explore beat-to-beat fluctuations under limiting conditions (E.g. window size), but require larger empirical studies toward wider adoption and relevance.

In our workflow, we make accommodations to account for activity related changes in the cardiac response, and our findings reflect the sensitivity of the underlying signal to the effects of stress beyond those influences. The improved generalizability of HR/V indices is not surprising given broader systemic change during the tasks, the nature of the response, and the non-focal characteristics of vagal nerve regulation, unlike the correlation and activation dependent attributes of the fNIRS signal. Another attractive property of HR/V based indices is the fieldability of the sensing instrument, between fNIRS and ECG devices, most research grade ECG sensors tend to be unobtrusive which make them an attractive prospect for the extension of dry lab models toward naturalistic settings.

5 CONCLUSIONS

We explored stress detection on older adults engaged in a motor task using brain hemodynamics (fNIRS) and heart rate variability (ECG). To this end, we built a machine learning workflow that enabled independent model development and validation across demographic (sex) and activity (limb type) variables. Further, we contrasted the efficacy of physiological and neural biomarkers along the same dimensions.

In general, we found that temporal characteristics of cerebral hemodynamics, connectivity measures, and heart rate variability are independently sensitive to the stressor, with a mean classification accuracy of $\approx 97\%$, 78% , and

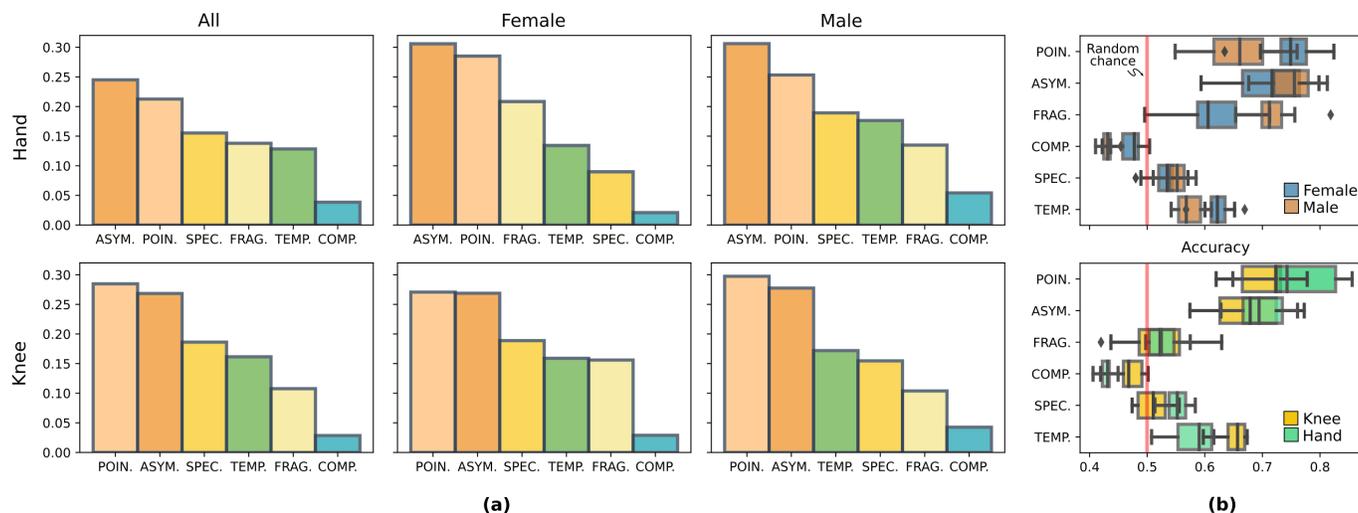


Fig. 6. (a) Variables of importance from the HR/V data set based on the permutations of importance index introduced in Section 3, where measures are grouped together based on their source – (i) Poincaré plot geometry, (ii) heart rate asymmetry, (iii) fragmentation, (iv) complexity, (v) spectral, and (vi) temporal characteristics. (b) Stress detection accuracy when relying on specific sections of the HR/V feature space. Nonlinear features appear to be consistently sensitive to the effects of stress above and beyond other attributes across the sexes (TOP), and limb types (BOTTOM).

72% respectively for models trained and tested *within* sexes or experiment conditions. Our findings suggest that the physiological and neural indices identified in this study are indeed representative of the neurophysiological changes that accompany stress-induction and its after-effects in older adults. Importantly, we found that neural effects of stress are not restricted to the frontal areas of the brain, and that probemaps focused on the motor cortex remain responsive to the effects of mental stress. Further, we found that a simple quadratic discriminant was sufficient to separate the two states.

However, we note that these models do not generalize well across demographic or activity variables. Notably, features that represent the cerebral hemodynamic response appear to be strongly influenced by sex, and limb-type distinctions, with accuracy no better than random chance when evaluating across *subsets*. Functional connectivity measures on the other hand had one exception to this trend, where, on knee data we found that the models performed equivalently across both sexes. We believe that focused motor activity can override sex-specific neural differences due to a cognitive stressor, which further underscores the challenge of task-specificity in the model development process. In contrast, HR/V based measures were seen to transfer well across both demographic and activity variables, however, the overall classification accuracy was lower than their neural counterparts. These observations are significant to sensorimotor rehabilitation programs for older adults, where robust stress-detection workflows can play an instrumental role in mapping task demands to patient cognitive states, and in enabling adaptive exercises that promote accelerated recovery.

In summary, we observed that, (i) demographic and activity variables do influence model sensitivity to stress,

but the extent of this effect is specific to the nature of the sensing modality, where temporal neural indices were more susceptible to these differences than connectivity or secondary physiological indices (e.g. HR/V); and (ii) on the whole, we observed a motor-task *independence*, where actions in the motor domain did not obscure the neural or physiological representations of stress. Although these observations indicate positively toward stress detection in older adults, there remain some key limitations which demand our attention to foster real-world solutions; they are discussed in the section to follow.

5.1 Limitations and future work

First, the fNIRS instruments in use today remain cumbersome for any real world stress detection, although they permit ambulatory use, they are obtrusive and some users may find the probes uncomfortable or painful when worn for long time periods. Optimized probemaps, with a minimal footprint may offset this issue in practice; our observations that found relevance in the motor cortex toward stress classification highlights this possibility.

Second, our inferences are limited by the experiment protocol. While we employ a motor-intensive task, the neurophysiological representations of these tasks are only partially representative of real world behaviors due to our instruction, the environment, and the overall participant experience; therefore the benefits of this workflow are primarily relevant to controlled rehabilitation programs. Wider testing with larger, ecologically valid, data sets could enable a path towards *true* task-agnosticity. Additionally, the post-stressor state is an aggregation of participant experience over the pre-stressor states and the TSST which in our demographic could add to the overall discriminability.

Third, while we explore the under-represented dimensions of sex and age, recent trends point to the need for personalization (e.g. [45]), nevertheless, our observations remain a valid extension of previous efforts on stress-detection in older adults under dual-task demands.

Fourth, we explore stress detection in an offline setting that involves task and sensor-specific processing, however the compatibility of these methods toward online stress detection requires further investigation.

Finally, our study was limited to data from a senior population, therefore the transferability of the models to a wider age group remains suspect. Aging is known to perturb our neurophysiological responses, however, our protocol explicitly considers stress and no stress conditions within the participant pool of older adults; therefore, providing confidence to the overall stress detection ability, above and beyond aging related factors within our demographic.

In conclusion, the synergistic interactions between motor task and cognitive stress have long remained unresolved and this research offers a potential window into those interactions. We believe that with improvements in fNIRS hardware, multimodal explorations that include temporal hemodynamics, functional connectivity, and HR/V based measures offer a robust solution to the stress detection problem in older adults, especially in the context of sensorimotor rehabilitation.

ACKNOWLEDGMENT

Research reported in this publication was partially supported by the *National Institute on Aging* of the *National Institutes of Health* under award number R15AG047553

REFERENCES

- [1] B. S. McEwen, "Physiology and neurobiology of stress and adaptation: central role of the brain," *Physiological reviews*, vol. 87, no. 3, pp. 873–904, 2007.
- [2] L. D. Godoy, M. T. Rossignoli, P. Delfino-Pereira, N. Garcia-Cairasco, and E. H. de Lima Umeoka, "A comprehensive overview on stress neurobiology: basic concepts and clinical implications," *Frontiers in behavioral neuroscience*, vol. 12, p. 127, 2018.
- [3] W. Gerin and T. G. Pickering, "Association between delayed recovery of blood pressure after acute mental stress and parental history of hypertension," *Journal of hypertension*, vol. 13, no. 6, pp. 603–610, 1995.
- [4] A. M. Stankiewicz, A. H. Swiergiel, and P. Lisowski, "Epigenetics of stress adaptations in the brain," *Brain Research Bulletin*, vol. 98, pp. 76–92, 2013.
- [5] K. H. Kwag, P. Martin, D. Russell, W. Franke, and M. Kohut, "The impact of perceived stress, social support, and home-based physical activity on mental health among older adults," *The International Journal of Aging and Human Development*, vol. 72, no. 2, pp. 137–154, 2011.
- [6] T. M. Spruill, "Chronic psychosocial stress and hypertension," *Current hypertension reports*, vol. 12, no. 1, pp. 10–16, 2010.
- [7] A. Steptoe and M. Kivimäki, "Stress and cardiovascular disease," *Nature Reviews Cardiology*, vol. 9, no. 6, pp. 360–370, 2012.
- [8] L. Lundin-Olsson, L. Nyberg, Y. Gustafson, et al., "Stops walking when talking as a predictor of falls in elderly people," *Lancet*, vol. 349, no. 9052, p. 617, 1997.
- [9] K. Z. Li, L. Bherer, A. Mirelman, I. Maidan, and J. M. Hausdorff, "Cognitive involvement in balance, gait and dual-tasking in aging: a focused review from a neuroscience of aging perspective," *Frontiers in neurology*, vol. 9, p. 913, 2018.
- [10] N. M. Fisher, D. R. Pendergast, G. E. Gresham, and E. Calkins, "Muscle rehabilitation: its effect on muscular and functional performance of patients with knee osteoarthritis," *Archives of physical medicine and rehabilitation*, vol. 72, no. 6, pp. 367–374, 1991.
- [11] J. Vinstrup, J. Calatayud, M. D. Jakobsen, E. Sundstrup, J. R. Jørgensen, J. Casaña, and L. L. Andersen, "Hand strengthening exercises in chronic stroke patients: Dose-response evaluation using electromyography," *Journal of Hand Therapy*, vol. 31, no. 1, pp. 111–121, 2018.
- [12] E. L. McGough, S.-Y. Lin, B. Belza, K. M. Becofsky, D. L. Jones, M. Liu, S. Wilcox, and R. G. Logsdon, "A scoping review of physical performance outcome measures used in exercise interventions for older adults with alzheimer disease and related dementias," *Journal of Geriatric Physical Therapy*, vol. 42, no. 1, pp. 28–47, 2019.
- [13] W. A. Pedersen, R. Wan, and M. P. Mattson, "Impact of aging on stress-responsive neuroendocrine systems," *Mechanisms of ageing and development*, vol. 122, no. 9, pp. 963–983, 2001.
- [14] C. N. Bairey Merz, O. Elboudwarej, and P. Mehta, "The autonomic nervous system and cardiovascular health and disease: a complex balancing act," 2015.
- [15] N. Smyth, F. Hucklebridge, L. Thorn, P. Evans, and A. Clow, "Salivary cortisol as a biomarker in social science research," *Social and Personality Psychology Compass*, vol. 7, no. 9, pp. 605–625, 2013.
- [16] C. D. Spielberger, "State-trait anxiety inventory for adults," 1983.
- [17] A. Kalatzis, L. Stanley, R. Karthikeyan, and R. K. Mehta, "Mental stress classification during a motor task in older adults using an artificial neural network," in *Adjunct Proceedings of the 2020 ACM International Joint Conference on Pervasive and Ubiquitous Computing and Proceedings of the 2020 ACM International Symposium on Wearable Computers*, pp. 244–248, 2020.
- [18] B. S. McEwen and J. C. Wingfield, "The concept of allostasis in biology and biomedicine," *Hormones and behavior*, vol. 43, no. 1, pp. 2–15, 2003.
- [19] J. F. Thayer, F. Åhs, M. Fredrikson, J. J. Sollers III, and T. D. Wager, "A meta-analysis of heart rate variability and neuroimaging studies: implications for heart rate variability as a marker of stress and health," *Neuroscience & Biobehavioral Reviews*, vol. 36, no. 2, pp. 747–756, 2012.
- [20] B. Falkner, G. Onesti, E. Angelakos, M. Fernandes, and C. Langman, "Cardiovascular response to mental stress in normal adolescents with hypertensive parents. hemodynamics and mental stress in adolescents," *Hypertension*, vol. 1, no. 1, pp. 23–30, 1979.
- [21] E. Charmandari, C. Tsigos, and G. Chrousos, "Endocrinology of the stress response," *Annu. Rev. Physiol.*, vol. 67, pp. 259–284, 2005.
- [22] M. G. Calvo and A. Gutierrez-Garcia, "Cognition and stress," in *Stress: Concepts, cognition, emotion, and behavior*, pp. 139–144, Elsevier, 2016.
- [23] J. E. Dimsdale and J. Moss, "Short-term catecholamine response to psychological stress," *Psychosomatic Medicine*, 1980.
- [24] E. R. de Kloet, H. Karst, and M. Joëls, "Corticosteroid hormones in the central stress response: quick-and-slow," *Frontiers in neuroendocrinology*, vol. 29, no. 2, pp. 268–272, 2008.
- [25] 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.
- [26] J. Vogel, A. Auinger, and R. Riedl, "Cardiovascular, neurophysiological, and biochemical stress indicators: a short review for information systems researchers," *Information Systems and Neuroscience*, pp. 259–273, 2019.
- [27] P. Melillo, M. Bracale, and L. Pecchia, "Nonlinear heart rate variability features for real-life stress detection. case study: students under stress due to university examination," *Biomedical engineering online*, vol. 10, no. 1, p. 96, 2011.
- [28] C. Setz, B. Arnrich, J. Schumm, R. La Marca, G. Tröster, and U. Ehlert, "Discriminating stress from cognitive load using a

- wearable eda device," *IEEE Transactions on information technology in biomedicine*, vol. 14, no. 2, pp. 410–417, 2009.
- [29] G. Giannakakis, M. Padiaditis, D. Manousos, E. Kazantzaki, F. Chiarugi, P. G. Simos, K. Marias, and M. Tsiknakis, "Stress and anxiety detection using facial cues from videos," *Biomedical Signal Processing and Control*, vol. 31, pp. 89–101, 2017.
- [30] F. Al-Shargie, T. B. Tang, and M. Kiguchi, "Stress assessment based on decision fusion of eeg and fnirs signals," *IEEE Access*, vol. 5, pp. 19889–19896, 2017.
- [31] S. Koldijk, M. A. Neerinx, and W. Kraaij, "Detecting work stress in offices by combining unobtrusive sensors," *IEEE Transactions on Affective Computing*, vol. 9, no. 2, pp. 227–239, 2016.
- [32] F. Al-Shargie, M. Kiguchi, N. Badruddin, S. C. Dass, A. F. M. Hani, and T. B. Tang, "Mental stress assessment using simultaneous measurement of eeg and fnirs," *Biomedical optics express*, vol. 7, no. 10, pp. 3882–3898, 2016.
- [33] V. Mishra, T. Hao, S. Sun, K. N. Walter, M. J. Ball, C.-H. Chen, and X. Zhu, "Investigating the role of context in perceived stress detection in the wild," in *Proceedings of the 2018 ACM International Joint Conference and 2018 International Symposium on Pervasive and Ubiquitous Computing and Wearable Computers*, pp. 1708–1716, 2018.
- [34] J. A. Healey and R. W. Picard, "Detecting stress during real-world driving tasks using physiological sensors," *IEEE Transactions on intelligent transportation systems*, vol. 6, no. 2, pp. 156–166, 2005.
- [35] P. Zontone, A. Affanni, R. Bernardini, A. Piras, and R. Rinaldo, "Stress detection through electrodermal activity (eda) and electrocardiogram (ecg) analysis in car drivers," in *2019 27th European Signal Processing Conference (EUSIPCO)*, pp. 1–5, IEEE, 2019.
- [36] M. Gjoreski, M. Luštrek, M. Gams, and H. Gjoreski, "Monitoring stress with a wrist device using context," *Journal of biomedical informatics*, vol. 73, pp. 159–170, 2017.
- [37] P. J. Smith, J. A. Blumenthal, B. M. Hoffman, H. Cooper, T. A. Strauman, K. Welsh-Bohmer, J. N. Brownlyke, and A. Sherwood, "Aerobic exercise and neurocognitive performance: a meta-analytic review of randomized controlled trials," *Psychosomatic medicine*, vol. 72, no. 3, p. 239, 2010.
- [38] R. Y. Akinlosotu, N. Alissa, S. R. Waldstein, R. A. Creath, G. F. Wittenberg, and K. P. Westlake, "Examining the influence of mental stress on balance perturbation responses in older adults," *Experimental Gerontology*, vol. 153, p. 111495, 2021.
- [39] B. S. McEwen and P. J. Gianaros, "Central role of the brain in stress and adaptation: links to socioeconomic status, health, and disease," *Annals of the New York Academy of Sciences*, vol. 1186, p. 190, 2010.
- [40] H. M. Khalid, "Embracing diversity in user needs for affective design," *Applied ergonomics*, vol. 37, no. 4, pp. 409–418, 2006.
- [41] J. K. Nuamah, W. Mantooh, R. Karthikeyan, R. K. Mehta, and S. C. Ryu, "Neural efficiency of human-robotic feedback modalities under stress differs with gender," *Frontiers in Human Neuroscience*, vol. 13, p. 287, 2019.
- [42] J. J. Liu, N. Ein, K. Peck, V. Huang, J. C. Pruessner, and K. Vickers, "Sex differences in salivary cortisol reactivity to the trier social stress test (tsst): a meta-analysis," *Psychoneuroendocrinology*, vol. 82, pp. 26–37, 2017.
- [43] C. Zhang, C. C. Dougherty, S. A. Baum, T. White, and A. M. Michael, "Functional connectivity predicts gender: Evidence for gender differences in resting brain connectivity," *Human brain mapping*, vol. 39, no. 4, pp. 1765–1776, 2018.
- [44] I. Antelmi, R. S. De Paula, A. R. Shinzato, C. A. Peres, A. J. Mansur, and C. J. Grupi, "Influence of age, gender, body mass index, and functional capacity on heart rate variability in a cohort of subjects without heart disease," *The American journal of cardiology*, vol. 93, no. 3, pp. 381–385, 2004.
- [45] F. Delmastro, F. Di Martino, and C. Dolciotti, "Cognitive training and stress detection in mci frail older people through wearable sensors and machine learning," *IEEE Access*, vol. 8, pp. 65573–65590, 2020.
- [46] R. M. Sapolsky, "Stress and the brain: individual variability and the inverted-u," *Nature neuroscience*, vol. 18, no. 10, pp. 1344–1346, 2015.
- [47] V. A. Nejtek, "High and low emotion events influence emotional stress perceptions and are associated with salivary cortisol response changes in a consecutive stress paradigm," *Psychoneuroendocrinology*, vol. 27, no. 3, pp. 337–352, 2002.
- [48] K. Obayashi, "Salivary mental stress proteins," *Clinica chimica acta*, vol. 425, pp. 196–201, 2013.
- [49] C. Kirschbaum, K.-M. Pirke, and D. H. Hellhammer, "The 'trier social stress test'—a tool for investigating psychobiological stress responses in a laboratory setting," *Neuropsychobiology*, vol. 28, no. 1-2, pp. 76–81, 1993.
- [50] B. M. Kudielka, D. H. Hellhammer, and C. Kirschbaum, "Ten years of research with the trier social stress test—revisited," 2007.
- [51] J. Rhee and R. K. Mehta, "Functional connectivity during hand-grip motor fatigue in older adults is obesity and sex-specific," *Frontiers in human neuroscience*, vol. 12, p. 455, 2018.
- [52] L. G. Martin, R. F. Schoeni, and P. M. Andreski, "Trends in health of older adults in the united states: Past, present, future," *Demography*, vol. 47, no. 1, pp. S17–S40, 2010.
- [53] D. Makowski, T. Pham, Z. J. Lau, J. C. Brammer, F. Lespinasse, H. Pham, C. Schölzel, and A. S. H. Chen, "Neurokit2: A python toolbox for neurophysiological signal processing," 2020.
- [54] A. D. McNaught, A. Wilkinson, et al., *Compendium of chemical terminology*, vol. 1669. Blackwell Science Oxford, 1997.
- [55] F. Scholkmann, S. Spichtig, T. Muehlemann, and M. Wolf, "How to detect and reduce movement artifacts in near-infrared imaging using moving standard deviation and spline interpolation," *Physiological measurement*, vol. 31, no. 5, p. 649, 2010.
- [56] A. M. Chiarelli, E. L. Maclin, M. Fabiani, and G. Gratton, "A kurtosis-based wavelet algorithm for motion artifact correction of fnirs data," *Neuroimage*, vol. 112, pp. 128–137, 2015.
- [57] D. T. Delpy, M. Cope, P. van der Zee, S. Arridge, S. Wray, and J. Wyatt, "Estimation of optical pathlength through tissue from direct time of flight measurement," *Physics in Medicine & Biology*, vol. 33, no. 12, p. 1433, 1988.
- [58] C. M. Aasted, M. A. Yücel, R. J. Cooper, J. Dubb, D. Tsuzuki, L. Becerra, M. P. Petkov, D. Borsook, I. Dan, and D. A. Boas, "Anatomical guidance for functional near-infrared spectroscopy: Atlasviewer tutorial," *Neurophotonics*, vol. 2, no. 2, p. 020801, 2015.
- [59] M. A. Mayka, D. M. Corcos, S. E. Leurgans, and D. E. Vaillancourt, "Three-dimensional locations and boundaries of motor and premotor cortices as defined by functional brain imaging: a meta-analysis," *Neuroimage*, vol. 31, no. 4, pp. 1453–1474, 2006.
- [60] F. Strasser, M. Muma, and A. M. Zoubir, "Motion artifact removal in eeg signals using multi-resolution thresholding," in *2012 Proceedings of the 20th European Signal Processing Conference (EUSIPCO)*, pp. 899–903, IEEE, 2012.
- [61] V. Marked, "Correction of the heart rate variability signal for ectopics and missing beats," *Heart rate variability*, 1995.
- [62] C. Li, C. Zheng, and C. Tai, "Detection of eeg characteristic points using wavelet transforms," *IEEE Transactions on biomedical Engineering*, vol. 42, no. 1, pp. 21–28, 1995.
- [63] Y. Zhu, J. K. Jayagopal, R. K. Mehta, M. Erraguntla, J. Nuamah, A. D. McDonald, H. Taylor, and S.-H. Chang, "Classifying major depressive disorder using fnirs during motor rehabilitation," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 28, no. 4, pp. 961–969, 2020.
- [64] W. R. Shirer, S. Ryali, E. Rykhlevskaia, V. Menon, and M. D. Greicius, "Decoding subject-driven cognitive states with whole-brain connectivity patterns," *Cerebral cortex*, vol. 22, no. 1, pp. 158–165, 2012.
- [65] J. Piskorski and P. Guzik, "Asymmetric properties of long-term and total heart rate variability," *Medical & biological engineering & computing*, vol. 49, no. 11, pp. 1289–1297, 2011.
- [66] 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.
- [67] F. Shaffer and J. Ginsberg, "An overview of heart rate variability metrics and norms," *Frontiers in public health*, vol. 5, p. 258, 2017.
- [68] A. P. Allen, P. J. Kennedy, S. Dockray, J. F. Cryan, T. G. Dinan, and G. Clarke, "The trier social stress test: principles and practice," *Neurobiology of stress*, vol. 6, pp. 113–126, 2017.

- [69] G. Varoquaux, P. R. Raamana, D. A. Engemann, A. Hoyos-Iidrobo, Y. Schwartz, and B. Thirion, "Assessing and tuning brain decoders: cross-validation, caveats, and guidelines," *NeuroImage*, vol. 145, pp. 166–179, 2017.
- [70] E. Alpaydm, "Combined 5×2 cv f test for comparing supervised classification learning algorithms," *Neural computation*, vol. 11, no. 8, pp. 1885–1892, 1999.
- [71] M. Mirbagheri, N. Hakimi, S. K. Setarehdan, A. Jodeiri, and V. Zakeri, "Accurate stress assessment based on functional near infrared spectroscopy using deep learning approach," in *2019 26th National and 4th International Iranian Conference on Biomedical Engineering (ICBME)*, pp. 4–10, IEEE, 2019.
- [72] A. F. Arnsten, "Stress signalling pathways that impair prefrontal cortex structure and function," *Nature reviews neuroscience*, vol. 10, no. 6, pp. 410–422, 2009.
- [73] A. R. Luft, G. V. Smith, L. Forrester, J. Whittall, R. F. Macko, T.-K. Hauser, A. P. Goldberg, and D. F. Hanley, "Comparing brain activation associated with isolated upper and lower limb movement across corresponding joints," *Human brain mapping*, vol. 17, no. 2, pp. 131–140, 2002.
- [74] D. R. Leff, F. Orihuela-Espina, C. E. Elwell, T. Athanasiou, D. T. Delpy, A. W. Darzi, and G.-Z. Yang, "Assessment of the cerebral cortex during motor task behaviours in adults: a systematic review of functional near infrared spectroscopy (fnirs) studies," *NeuroImage*, vol. 54, no. 4, pp. 2922–2936, 2011.
- [75] 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.
- [76] J. Taelman, S. Vandeput, A. Spaepen, and S. Van Huffel, "Influence of mental stress on heart rate and heart rate variability," in *4th European conference of the international federation for medical and biological engineering*, pp. 1366–1369, Springer, 2009.
- [77] H.-G. Kim, E.-J. Cheon, D.-S. Bai, Y. H. Lee, and B.-H. Koo, "Stress and heart rate variability: a meta-analysis and review of the literature," *Psychiatry investigation*, vol. 15, no. 3, p. 235, 2018.
- [78] M. Brennan, M. Palaniswami, and P. Kamen, "Poincare plot interpretation using a physiological model of hrv based on a network of oscillators," *American Journal of Physiology-Heart and Circulatory Physiology*, vol. 283, no. 5, pp. H1873–H1886, 2002.
- [79] P. W. Kamen, H. Krum, and A. M. Tonkin, "Poincare plot of heart rate variability allows quantitative display of parasympathetic nervous activity in humans," *Clinical science*, vol. 91, no. 2, pp. 201–208, 1996.
- [80] S. Rahman, M. Habel, and R. J. Contrada, "Poincaré plot indices as measures of sympathetic cardiac regulation: Responses to psychological stress and associations with pre-ejection period," *International Journal of Psychophysiology*, vol. 133, pp. 79–90, 2018.
- [81] Z. Visnovcova, M. Mestanik, M. Javorka, D. Mokra, M. Gala, A. Jurko, A. Calkovska, and I. Tonhajzerova, "Complexity and time asymmetry of heart rate variability are altered in acute mental stress," *Physiological measurement*, vol. 35, no. 7, p. 1319, 2014.



driver behavioral modeling.

Anthony D. McDonald received his Ph.D. in Industrial Engineering from the University of Wisconsin-Madison. He is currently an Assistant Professor and Corrie and Jim Furber '64 Faculty Fellow with the Wm Michael Barnes '64 Department of Industrial and Systems Engineering at Texas A&M University. Prior to his current appointment, he worked for 3 years at the Oracle Corporation in Software Development. His research interests include integrating Human Factors and Machine Learning and



and in under-served populations.

Ranjana K. Mehta, received her BE degree in production engineering from University of Mumbai, India, the MS degree in industrial engineering from University at Buffalo, and a PhD degree in industrial engineering from Virginia Tech. She is currently the director of the NeuroErgonomics Lab at Texas A&M University, which examines mind-motor-machine interactions to understand, quantify, and predict human performance, particularly under stress, when interacting with emerging and wearable technologies,



Rohith Karthikeyan received his bachelor's degree in Mechanical Engineering at the BMS College of Engineering, Bengaluru, India and his Master's in Mechanical Engineering at Texas A&M University, College Station, TX where he's currently pursuing a PhD in Mechanical Engineering as a graduate research assistant in the NeuroErgonomics laboratory. His research interests include human factors, robotics, computational neuroscience, and human-robot interaction.