



Published in final edited form as:

IEEE Sens J. 2019 October 1; 19(19): 8522–8531. doi:10.1109/jsen.2018.2872651.

Fusing Near-Infrared Spectroscopy with Wearable Hemodynamic Measurements Improves Classification of Mental Stress

Nil Z. Gurel* [Student Member, IEEE], Hewon Jung*, Sinan Hersek, Omer T. Inan [Senior Member, IEEE]

School of Electrical and Computer Engineering, Georgia Institute of Technology, Atlanta, GA, 30322.

Abstract

Human-computer interaction (HCI) technology, and the automatic classification of a person's mental state, are of interest to multiple industries. In this work, the fusion of sensing modalities that monitor the oxygenation of the human prefrontal cortex (PFC) and cardiovascular physiology was evaluated to differentiate between rest, mental arithmetic and N-back memory tasks. A flexible headband to measure near-infrared spectroscopy (NIRS) for quantifying PFC oxygenation, and forehead photoplethysmography (PPG) for assessing peripheral cardiovascular activity was designed. Physiological signals such as the electrocardiogram (ECG) and seismocardiogram (SCG) were collected, along with the measurements obtained using the headband. The setup was tested and validated with a total of 16 human subjects performing a series of arithmetic and N-back memory tasks. Features extracted were related to cardiac and peripheral sympathetic activity, vasomotor tone, pulse wave propagation, and oxygenation. Machine learning techniques were utilized to classify rest, arithmetic, and N-back tasks, using leave-one-subject-out cross validation. Macro-averaged accuracy of 85%, precision of 84%, recall rate of 83%, and F1 score of 80% were obtained from the classification of the three states. Statistical analyses on the subject-based results demonstrate that the fusion of NIRS and peripheral cardiovascular sensing significantly improves the accuracy, precision, recall, and F1 scores, compared to using NIRS sensing alone. Moreover, the fusion significantly improves the precision compared to peripheral cardiovascular sensing alone. The results of this work can be used in the future to design a multi-modal wearable sensing system for classifying mental state for applications such as acute stress detection.

Keywords

Sensor fusion; mental stress classification; near-infrared spectroscopy; wearable sensing

I. Introduction

HUMAN-COMPUTER interaction (HCI) is a growing field dedicated to the improvement of human performance. Recently, non-medical applications of HCI have gained considerable interest with a specific focus on improving the user's performance while executing a

nil@gatech.edu.

* Authors equally contributed to this work

mentally stressful task [1]. The design of proactive systems that measure a user's physiological parameters to decode mental state could provide feedback for HCI technologies to close the loop for performance improvement.

Current approaches to mental stress assessment are mainly based on brain computer interfaces (BCIs), and primarily focused on electroencephalogram (EEG) signals [2, 3]. Despite wide usage, EEG signals are known to have a highly noisy and variable nature with poor spatial resolution, making robust information extraction a very challenging task [4]. Near-infrared spectroscopy (NIRS) is an optical and non-invasive method for monitoring the changes in tissue oxygen dynamics, with high spatial resolution, as a complement or alternative to EEG [5]. NIRS provides information regarding blood flow to the pre-frontal cortex (PFC), and it was shown to be useful for assessing the effects of some mental stressors [6-10]. Nevertheless, while the performance of NIRS for mental stress assessment holds promise, improvements are still needed in its accuracy and sensitivity for the method to be viable for closed-loop HCI systems.

Complementing NIRS with other physiological signals may allow for such improvement to be achieved. Specifically, peripherally-measured non-invasive cardiovascular signals provide useful information related to mental stress that is complementary to NIRS. While the combined use of such cardiovascular signals and NIRS has not been explored for mental state monitoring, it has been investigated for improving the monitoring of exercise performance and oxygen transport [11-13]. Moreover, cardiovascular signals themselves have been studied in the context of mental stress, with some features changing in repeatable and predictable ways with stress [14]. The use of cardiovascular signals alone for mental stress assessment would likely result in a lack of specificity due to the fact that physiological perturbations can affect the signals in a similar manner to mental stress, thus confounding the results [15, 16]. Thus, we hypothesize that the *combination* of NIRS and peripherally-measured cardiovascular signals could potentially advance the state of the art for non-invasive mental stress assessment.

We anticipate that the fusion of these sensing modalities can be effective in decoding two types of mental stressors: first, mental arithmetic tasks, which are known to induce the strongest cardiovascular responses [14]; second, N-back memory tasks, which have been shown to cause activations in the PFC, as measured by NIRS [9, 17, 18]. In this work, we focus on the classification of rest state, mental arithmetic, and N-back memory tasks by fusing NIRS and peripherally-measured cardiovascular signals. We designed a setup containing NIRS, head photo-plethysmography (PPG), chest electrocardiogram (ECG), and seismocardiogram (SCG) signals. Furthermore, algorithms were designed to extract features and classify mental stress using these physiological signals. Specifically, algorithms using only NIRS signals, only peripherally-measured cardiovascular signals (ECG, PPG, SCG), and the fusion of both sensing modalities were implemented. We tested our system and algorithms on a total of 16 human subjects performing a series of arithmetic and N-back tasks. Our findings suggest that the fusion of NIRS and peripherally-measured cardiovascular signals significantly improves the classification performance of rest, arithmetic, and N-back tasks, as opposed to using each sensing modality alone. The outcomes could improve real-time decoding of mental state for HCI.

II. Methods

A. Experimental Protocol

The human subject study was performed under a protocol reviewed and approved by the Georgia Institute of Technology Institutional Review Board (IRB). All subjects read and signed a consent form before the data collection. Data were collected from 16 healthy subjects without cardiovascular disorders (six females, ten males, ages 26.7 ± 3.2 mean \pm std).

Each participant was seated on a comfortable chair, located 70cm away from a 21cm by 31cm monitor. Fig. 1 illustrates the experimental protocol and Fig. 2a shows the test setup for each subject. The experiment was divided into six parts that include mental tasks (mental arithmetic and N-back), with three-minute eyes closed rest breaks in between each task. The arithmetic task was chosen due to the high cardiovascular response observed in many clinical studies relative to other mental stress tests [14]. The N-back task was chosen because of its relevance with PFC activity that could be captured with NIRS [9, 17, 18]. All tasks were carried out on a laptop and the subjects used a keyboard to interact with custom graphical user interfaces (GUIs). The subjects were asked to remain silent, and to minimize posture changes during the protocol. Before the start of the experiment, the protocol was explained in detail to each subject, and the subjects practiced sample questions from each task. At the beginning of the experiment, each subject was instructed to sit comfortably with eyes closed for three minutes to obtain baseline rest signals. Then, the subjects underwent a series of arithmetic and N-back tasks, with difficulty levels ordered randomly. The questions in these tasks were different from the ones in the practice session. The three arithmetic tasks included 1-digit, 2-digit, or 3-digit algebraic calculations from a custom GUI. A series of arithmetic questions appeared on the screen for each difficulty level, for one-minute each, as illustrated in Fig. 1. The subjects entered the answers using the keyboard and pressed a key to progress onto the next question. The subjects were not allowed to progress to the next question if they did not enter the right answer. After completing the first three tasks that include arithmetic questions, subjects progressed to the N-back task.

The N-back task is a continuous performance task to measure working memory and working memory capacity. In this task, a three-by-three grid was presented to the subject on the computer screen via a GUI [19]. A sequence of squares at different spatial locations were highlighted consecutively, and the subject pressed a key when the location of the current highlighted square matched the one from N-steps earlier in the sequence (Fig. 1). For example, for $N = 2$, the subjects were required to remember the position of the square that was highlighted two turns ago. Our protocol included $N = 1, 2, 3$ -back tasks shuffled randomly, each lasting for a minute. Each trial (square appearance) was adjusted to last for a maximum of three seconds, therefore each N-back task includes approximately 20 trials within a minute.

To assign the subjective difficulty levels for the classification of each task, the subjects filled out the NASA Task Load Index (NASA-TLX) questionnaire at the end of the protocol [20]. NASA-TLX is a multidimensional assessment tool that rates the perceived workload for a task. For each of the six tasks (three arithmetic and three N-back), the total workload was

divided into six subcategories: mental demand, physical demand, temporal demand, performance, effort, frustration. The subjects rated each subscale with a score between 0-100, for each task. The ratings from the six subcategories were averaged for each task, referred to as RTLX scores. The average score indicates an estimate of overall workload for the corresponding task [20]. Using the RTLX scores, each task was assigned to a difficulty level as follows: Within each task type (arithmetic or N-back), the lowest RTLX score corresponded to easy, the medium score corresponded to medium, and the highest score to hard as difficulty level. The use of RTLX objectifies the difficulty assignment for each task type. For instance, after a subject completed the overall protocol that includes six mental stress tasks (three arithmetic and three N-back), she was asked to fill out the NASA-TLX questionnaire for each of the six tasks. Within each task type (i.e. arithmetic or N-back), RTLX scores (0-100) were calculated and ranked from minimum to maximum. The minimum RTLX score for a type of task resulted in a 'easy' label for that task (i.e. minimum RTLX among RTLX scores for arithmetic was labeled as 'easy'). Similarly, the medium RTLX score was labeled as 'medium' and the maximum RTLX score was labeled as hard. The labeled scores were used in statistical analyses to understand which difficulty level created the most drastic differences in mental workload, compared to the rest state. The highest difficulty levels for each task type from the statistical analyses were used for the classification.

B. Instrumentation

Fig. 2a and b show the parts of headband and instrumentation blocks, namely NIRS and cardiovascular-peripheral blocks. The NIRS circuit consisted of a multi-chip near-infrared (NIR) light emitting diode (LED) (MTMD7885T38, Marktech Optoelectronics, Latham, NY), 10 photodiode-transimpedance amplifier (PD-TIA) chips (OPT101, Texas Instruments, Dallas, TX), 16-channel LED driver (TLC5940, Texas Instruments, Dallas, TX), 16-channel multiplexer (MUX, CD74HC4067, Texas Instruments, Dallas, TX), and a microcontroller (μ C, ATmega 2560, Arduino, NY) as shown in Fig. 2b. The multi-chip LED included six LEDs, with peak emission wavelengths (λ) of 770nm ($\times 2$), 810nm ($\times 2$), and 850nm ($\times 2$). The package had a diameter of 9mm, and all LEDs were separated by 45°. To maximize the level of reflections received by the detector, the LED driver was programmed such that the LEDs would operate with the maximum forward currents recommended by the manufacturer: 50mA for 770nm and 100mA for 810nm and 850nm. The MUX output was programmed to sequentially select which detector output would be read by the built-in analog-to-digital converter (ADC, 10-bit) of the microcontroller (time division multiplexing). A flexible headband was designed in SolidWorks (2016, Waltham, MA), as shown in Fig. 2a, with dimensions 140.5mm by 58.5mm by 4.5mm, using thermoplastic polyurethane (TPU) filament (NinjaFlex, Manheim, PA). The NIRS LED was placed to the center of the headband, and 10 of PD-TIA chips were distributed spatially around the NIRS LED. The distances between each source (LED)-detector (PD-TIA) combination were set at 1.5cm or 3cm. The distances were chosen due to the known optimal sensitivity of NIRS to intracranial brain tissues at these distances, and the NIRS implementations in literature [21].

In terms of firmware, upon the detection of a trigger signal from a data acquisition system (DAQ, MP150, Biopac Systems, Goleta, CA), the NIRS LEDs were programmed to turn on

sequentially. Once an LED was turned on, the MUX sequentially selected detectors to read their output signals. The NIRS signals were transmitted with 2Hz sampling rate.

For the peripherally-measured cardiovascular signals, both custom-built circuits and commercially available components were used to acquire a set of signals that would capture at least the following physiological features of relevance for mental stress assessment: heart rate (HR), the pre-ejection period (PEP), pulse transit time (PTT), pulse arrival time (PAT), blood volume (PPG amplitude). Moreover, the goal was to use sensors and electronics for measuring these signals that could ultimately be encapsulated in a wearable device, possibly even a headband with combined NIRS and cardiovascular signal sensing capability.

The headband used here contains the head PPG sensors, which include a multi-chip LED and PD combination (SFH 7070, OSRAM Opto semiconductors, Regensburg, Germany). The package includes two green emitters ($\lambda = 530\text{nm}$) and a matched PD, with an overall dimension of 7.5mm by 4mm by 0.9mm. The forward current through the LEDs were set to 20mA, per the datasheet suggestion, using a voltage divider and buffer combination. A TIA was designed to read the PD output using an operational amplifier (LT1885, Linear Technology, Milpitas, CA) with feedback components ($R_F = 350\text{k}\Omega$, $C_F = 10\text{nF}$), followed by a first-order passive low-pass filter ($f_c = 16\text{Hz}$). The head PPG signal was acquired using the DAQ. To measure SCG signals, a very low-noise 3-axis accelerometer evaluation board was used (ADXL354CZ, Analog Devices, Norwood, MA). The accelerometer was placed in a 2.8cm by 3cm by 1cm custom-printed rigid acrylonitrile butadiene styrene (ABS) plastic case. It was placed on the mid-sternum of each subject, but in future work could potentially obtain aortic valve opening information from the head as implemented previously by other groups [22]. For ECG data collection, a commercially available wireless 3-lead ECG amplifier was used (RSPEC-R, Biopac Systems, Goleta, CA). All peripherally-measured cardiovascular signals were transmitted through the DAQ with 2kHz sampling rate.

All custom circuits were powered using a benchtop power supply with $\pm 9\text{V}$ rails. For the components that require specific power levels (i.e. 3.3V for accelerometer and 5V for LED driver and PD-TIA), low drop-out regulators were used (LT1763, Linear Technology, Milpitas, CA).

C. Signal Processing and Feature Extraction

Pre-Processing and Feature Extraction: Data were processed in MATLAB (R2017b, MathWorks, Natick, MA). Fig. 2c gives an overview of the signal processing and feature extraction pipeline. The peripherally-measured cardiovascular parameters extracted are cardiac timing intervals and a signal amplitude, namely: the HR, R-Ao time interval (i.e., PEP), PAT, PTT, and head PPG amplitude. NIRS parameters are changes in concentrations of oxy-hemoglobin (HbO), deoxy-hemoglobin (HbR), and total hemoglobin ($Total Hb$).

Fig. 2d shows the parameters computed from the peripherally-measured cardiovascular signals. ECG, SCG and head PPG signals were filtered with finite impulse response (FIR) band-pass filters with cutoff frequencies 0.6-40 Hz, 0.8-20 Hz, and 0.8-10 Hz, respectively, to preserve the waveform shape and cancel the noise outside their bandwidths [23, 24]. The R-peaks of the ECG signals were detected using thresholding, and were used to calculate

HR. SCG and head PPG signals were ensemble averaged according to the R-peaks, using beat lengths of 300ms for SCG and 550ms for the head PPG. These lengths were sufficient to detect the fiducial points of each SCG and PPG beats. To reduce the effects of motion artifacts on the individually segmented beats, exponentially weighted moving ensemble averaging of successive beats was implemented [23]. Exponentially decreasing weighting gives more emphasis to the more recent beats, while still providing noise reduction based on the averaging. We determined 3-beat and 10-beat time constants for SCG and head PPG were sufficiently long to reduce the artifacts while short enough to still preserve the transient changes in the signals. Note that for all the parameters described below, the approximate first order derivatives (differences between the adjacent elements) were also computed to generate additional parameters.

Pre-Ejection Period: The PEP, measured by the time delay between the onset of electrical depolarization of the heart to the opening of the aortic valve, is a non-invasive measure of cardiac contractility and sympathetic activity. A decrease in PEP indicates increased contractility and cardiac sympathetic activity [25]. The gold-standard to measure PEP is the impedance cardiogram (ICG), which requires 4-8 electrodes on body. Instead of the ICG signal, the time interval between the R-peak of the ECG to the second peak of the SCG (known as aortic opening point, AO) is known to be highly correlated with PEP [26]. Therefore, we used R-Ao as a measure of sympathetic activity.

Pulse Arrival and Transit Times: PAT was measured as the time delay between the ECG R-peak to the foot of the head PPG signal. It represents the time delay from the electrical depolarization of the ventricles to the arrival of the pulse to forehead region, where the PPG signal is collected [27]. We also calculated PTT, which is the time taken for the pressure wave to travel between two arterial sites, measured by two blood pressure waveforms [27]. We calculated PTT as the time interval between the AO point of the SCG signal to the foot of the head PPG signal [27]. As a measure of peripheral sympathetic and vasomotor activity, the amplitude of PPG signal was extracted [24].

PFC Oxygenation Markers: The changes in oxyhemoglobin, deoxyhemoglobin, and total hemoglobin concentrations (*HbO*, *HbR*, *Total Hb*) were calculated from NIRS signals according to modified Beer-Lambert law (MBLL) [28]. The NIRS channel to process with MBLL was chosen manually due to interference from hair on forehead and incomplete contact of a few detectors for some subjects.

Normalization and Dataset for Classification: After the difficulty level assignment for each task using RTLX, the tasks that resulted in the maximum perceived workload were selected to be used in the classification. Specifically, the parameters used for classification were extracted from rest (one minute), hard arithmetic (one minute), and hard N-back (one minute). Then, peripherally-measured cardiovascular parameters were normalized using a baseline reference interval. The one-minute reference interval for peripherally-measured cardiovascular parameters was collected before the protocol started. This interval is different than the rest class interval. NIRS parameters were used as is. To equalize the length of each parameter within an interval, extracted parameters were resampled to the length of the

parameter that has the maximum length. Then instances were created by using 10-sample sliding windows with 50% overlap. The features used in the classification consisted of the mean, standard deviation (std), maximum (max), minimum (min), area under curve (auc) and slope of the extracted parameters in each window. There was a total of 88 features (33 NIRS features, 55 peripherally-measured cardiovascular features), and the total number of instances were 842 (333 rest instances, 228 arithmetic instances, 281 N-back instances).

D. Feature Selection and Classification

For the classification of mental tasks and rest state, a feature matrix was constructed from all extracted features and the corresponding labels as classes (rest, arithmetic, N-back classes). This matrix included instances as rows and features as columns, and it was used to build the classification model. To eliminate irrelevant features that could decrease the accuracy of the classification model, we performed feature selection. Univariate feature selection was performed by calculating p-values for each feature using analysis of variance (ANOVA), and applying the Benjamini-Hochberg procedure ($\alpha = 0.005$) for multiple comparisons [29]. A univariate statistics-based feature selection method rather than manual selection automates the feature selection process, making it possible to treat the data blindly without assumptions. To visualize the dataset, we performed the aforementioned feature selection on the whole dataset and applied t-distributed stochastic neighbor embedding (t-SNE), reducing the dimensionality of the dataset to two [30].

To classify each instance to one of three mental tasks using the selected features, a random forest classifier was used. A random forest classifier is an ensemble learning method that trains multiple decision trees and determines the classification result through a majority vote amongst all individual trees. Each tree is trained on a bootstrap sample drawn from the dataset, and at each node of the tree, a random subset of the features is considered for a split [31]. In our algorithm, we trained 50 trees as a part of the random forest classifier. A single hyperparameter of the trees, maximum depth, was tuned using a leave-one-subject-out cross validation (LOSO-CV) grid-search scheme. In this scheme, we first defined the parameter grid values between three to ten. For each value on the grid, we performed LOSO-CV and found the parameter that maximizes the LOSO-CV accuracy, to use that parameter in the final model. The maximum depth parameter controls the complexity of each tree in the forest where increased depth corresponds to more complicated models.

Random forests are ensemble learning models that are often hard to interpret, especially when they consist of many trees. To get more insight on what the model learned, we performed feature importance ranking using a random forest classifier that was trained on the whole dataset. This was done by evaluating the improvement in the gini-index metric at each node of each tree within the forest. These improvements were accumulated across all nodes of all trees within the forest to rank the feature importance, with the most important features resulting in the largest improvements in the gini-index [31]. Feature selection, classification and t-SNE dimensionality reduction were all implemented using the scikit-learn library for Python [32].

E. Model Evaluation

We evaluated our algorithm using LOSO-CV, where one subject is left out of the training then used for testing. Without this subject's data, we first performed feature selection followed by hyperparameter tuning via grid-search within an inner LOSO-CV loop. With the optimal hyperparameters and the selected features, we trained our random forest classifier which was then used to calculate the performance metrics: accuracy, precision, recall and F1 score for each subject. The final scores were calculated by averaging the scores from each CV fold.

It should be noted that each CV fold implements an inner CV loop on the subjects that are not left out, which results in a nested CV protocol [33]. This procedure was performed by using only NIRS features, only peripherally-measured cardiovascular features and the fusion of both sensing modalities. The results were compared using statistical analyses to identify which sensing modalities perform better in differentiating among the rest, arithmetic, N-back tasks.

F. Statistical Analysis

We performed statistical analyses on the classification results to compare each sensing modality alone and their fusion. Specifically, each LOSO-CV fold results in one data point (accuracy, precision, recall, or F1 score metrics) per subject. We obtained 16 data points for 16 subjects per metric. This scheme was repeated for NIRS alone, cardiovascular (cardio) alone, and the fusion of both. These samples were used for statistical testing to assess the performance of the sensing modalities. Friedman Test was chosen to detect if any difference exists between the performance of the sensing modalities from the outcomes for each subject, using the same classifier model and validation method [34]. A follow-up multiple comparison based on the Nemenyi Test was performed using the ranks generated by the Friedman Test [34]. A similar statistical analysis was performed on the RTLX scores as well to understand if the mental tasks induce significantly different workloads between the difficulty levels. For all analyses, p-values lower than 0.05 were considered statistically significant.

III. Results

Fig. 3 shows the extracted parameters for a subject transitioning from baseline rest to task. As the task starts, HR increases, R-Ao and PPG amplitude decrease, indicating increased cardiac and peripheral sympathetic activity. HbO, HbR, and Total Hb also show change during this transition, due to the change in oxygenation levels in the PFC. It should be noted that the directions in PTT, PAT, or concentration changes were not necessarily identical for each subject.

A. NASA-TLX Scores

Fig. 4 shows the RTLX scores for each task. There are significant differences between the following intervals: easy-medium arithmetic ($p < 0.05$), easy-hard arithmetic ($p < 0.001$), easy-medium N-back ($p < 0.05$), medium-hard N-back ($p < 0.05$), easy-hard N-back ($p < 0.001$), according to Friedman followed by Nemenyi test. There is no statistically significant

difference for task-wise comparisons of the same difficulty level (i.e. hard arithmetic versus hard N-back).

B. Dimensionality Reduction From Selected Features

Fig. 5 shows the t-SNE plots for NIRS features alone, cardiovascular features alone, and the fusion features to gain intuitive understanding for each sensing modality's ability to separate between classes (clusters). The plots were constructed from the features selected by applying the univariate feature selection method described in section II.D to the whole dataset. The total number of selected features was 53, which was used to construct the t-SNE plot corresponding to the fusion features (Fig. 5c). There were 23 NIRS features and 30 cardiovascular features, each were used to construct the corresponding t-SNE plots in Fig. 5a and 5b, respectively. When only NIRS features are used (Fig. 5a), there is overlap between the "rest" and "MAT" clusters, unlike the cardiovascular (Fig. 5b) or the fusion (Fig. 5c) results; accordingly, using NIRS only might result in misclassification. For only cardiovascular features, there is separation between the "MAT" and "rest" clusters, compared to NIRS features alone. As both fusion and cardiovascular sensing have clear separation between the tasks and rest, the statistical comparison of the performance of machine learning algorithms to classify the data will reveal whether the fusion will outperform the cardiovascular sensing [35].

C. Classification Results

Macro-averaged accuracy, precision, recall, and F1 scores for each class and sensing modality are shown in Fig. 6. The fusion results in accuracy scores of $85\% \pm 9\%$, recall rate of $84\% \pm 14\%$, precision of $83\% \pm 10\%$, and F1 score of $80\% \pm 13\%$ (mean \pm std). According to the Friedman and follow-up Nemenyi tests, accuracy ($p < 0.0001$), precision ($p < 0.001$), recall rate ($p < 0.001$), and F1 score ($p < 0.001$) significantly improve when fusion is used, as opposed to only NIRS. Moreover, there is significant improvement in precision from using only cardiovascular sensing to the fusion ($p < 0.05$).

D. Feature Importance Ranking

Fig. 7a shows the importance ranking for the top 10 features. There are four cardiovascular-related features (R-Ao, PAT, HR, PTT), and five NIRS-related features (HbO and Total Hb). Fig. 7b shows the boxplots of these features. The medians of the three cardiovascular features show notable difference between rest and other tasks (AUC R-Ao, AUC PAT, AUC HR). The medians of these features for arithmetic and N-back are close to each other. For other cardiovascular features (AUC PTT, Min PTT), medians of the rest class are not as separable, unlike the case with R-Ao, PAT, HR. Notably, medians for MAT and N-back lie in opposite directions. For NIRS-related features (Mean, Max, Min, AUC HbO and Min Total Hb), medians of two mental stressors (MAT and N-back) look more differentiable. They do not show as good separation for rest class, however.

IV. Discussion and Conclusion

In this work, enhancement in separating rest, arithmetic, and N-back tasks through the fusion of NIRS and peripherally-measured cardiovascular sensing was investigated. Our

hypothesis was supported through a custom designed NIRS-PPG headband along with cardiovascular measurements. Our results indicate that the fusion of sensors results in significant improvements in the classification of rest, arithmetic, and N-back tasks. The fusion of these sensing modalities seem compatible to merge the advantages of both worlds. Peripherally-measured cardiovascular sensing represents a more central mechanism (due to the inclusion of the electrical activity and the mechanical motion of the heart), while NIRS sensing mainly target the PFC, which has an important role in working memory [36]. The improvements for accuracy, precision, recall rate, and F1 score from either of the sensing modalities to the fusion result from the ability to capture the PFC activity together with the central changes.

The significant enhancement for all performance metrics achieved via sensor fusion was particularly notable. Although the NIRS performance might arguably be improved by using multiple channels, using a single channel NIRS has its advantages for user convenience: multi-channel NIRS devices are often uncomfortable due to the size and high power requirements to feed multiple sources. A setup consisting of an accelerometer attached on the chest and a 3-lead ECG provide a simpler and practical means of data collection that would not block the forehead of the user.

Another remarkable result was the higher classification performance in all metrics between only cardiovascular and only NIRS sensing. Cardiovascular sensing outperformed NIRS in the macro-averaged scores, although not statistically significant. This difference suggests that peripherally-measured cardiovascular physiology might provide more consistent biomarkers of mental stress, compared to PFC activity biomarkers. A downside for using only cardiovascular sensing is that other types of stressors (i.e. temperature change or physical exercise) elicit cardiovascular responses similar to mental stress, thus, cardiovascular reactivity is not reliable across different stressors [15, 16, 37]. For instance, physical exercise is shown to increase HR and decrease PEP (R-Ao) by multiple studies, resulting in the same directional changes as mental stress [15, 38]. The addition of NIRS sensing by the fusion might rule out changes due to factors other than mental stress. Additionally, our results indicated improved precision from using fusion as compared to only cardiovascular sensing, highlighting that the fusion significantly improves the classification performance compared to using either of the sensing modalities alone. It should be noted that the other performance metrics, accuracy, recall rate, and F1 scores, were higher with the fusion, compared to using only cardiovascular sensing, although not statistically significant.

The interpretation of the t-SNE plots and boxplots of top 10 features seem consistent with clinical research on cardiovascular mental stress testing and the neuroscience literature: arithmetic tasks induce high cardiovascular reactivity [14], therefore cardiovascular sensing differentiates this type of stressor well, which appears as a separate cluster than the rest interval for cardiovascular t-SNE (Fig. 5b) and different median (Fig. 7b), compared to NIRS t-SNE or median of rest class for NIRS (Fig. 5a and Fig. 7b). Similarly, there is better separation between the clusters or medians of arithmetic and N-back NIRS features, when compared to cardiovascular features. This might be due to the high oxygenation activity in the PFC during N-back, as verified with comparison of multiple stressors in prior work [39].

The difference between the sensitivity of the sensing modalities to each task class might be due to the different regions activated in brain during these tasks. Moreover, it could be anticipated that the fusion would enhance the classification of each mental task from the rest state (i.e. task versus rest classification). Due to the clear separation in t-SNE plots, our model showed high classification scores for rest class. Specifically, we obtained macro-averaged scores of 98%, 99%, 98.5% for precision, recall, and f1 respectively. This indicates that our model could achieve highly accurate detection of mental stress (arithmetic or N-back) from the resting state.

According to the feature importance rankings, peripherally-measured cardiovascular features have the highest importance for the classification (R-Ao, PAT, HR). These are followed with NIRS features and PTT. R-Ao (i.e., PEP) ranks the highest, consistent with its definition to quantify cardiac sympathetic activity. PEP is a non-invasive measure of cardiac contractility, which reflects cardiac sympathetic activity. Particularly, decreased PEP reflects sympathetic (beta-adrenergic) receptor stimulation to the left ventricle of the heart, hence increased cardiac sympathetic activity [25]. On the other hand, HR, a parameter always assumed to be vital in mental stress studies, does not have the highest ranking. This result is also consistent with the literature, as HR is controlled by both branches of the autonomic nervous system: sympathetic (fight-or-flight) and parasympathetic (rest-and-digest) nervous system activity. The effect of the autonomic nervous system on the heart rate is the net balance between these two opposing branches [40]. The interplay between the two might mask the true sympathetic activation during mental stress, compared to the effects seen in PEP [41].

PAT and PTT were also selected in the top features, both of which were obtained by the SCG and head PPG signals. PTT is widely studied in the literature: it is inversely related to blood pressure, such that a decrease in PTT reflects increase in blood pressure [27]. Blood pressure is also known to be modulated by mental stress, largely due to the changes in vasomotor tone [42]. PAT, on the other hand, contains influences from both PEP and PTT. Accordingly, it does not have a precise relationship with blood pressure, unlike PTT. Since it contains both vasomotor-related (elevated blood pressure, decreased PTT) and contractility-related (elevated sympathetic activity, decreased PEP) influences, the contribution from both in the same direction might be the reason for the higher feature importance for PAT, compared to HR or PTT. Additionally, feature importance scores drop to below half for the NIRS features and PTT, highlighting the relative importance of the top three features. It should be noted that features related to PPG amplitude and HbR were also among the selected features in each LOSO-CV loop, although they are not among the ten most important features. PPG amplitude reflects the variations in blood volume at the region of measurement (head). Decrease in the head PPG amplitude indicates local vasoconstriction at the vicinity of the sensor, and an increase in PPG amplitude reflects vasodilation [24]. In addition to vasomotor function, multiple studies noted that PPG amplitude variations are linked to sympathetic activity due to mental stress [43]. Mental stress affects the cardiac, vascular, and autonomic nervous system activity, hence the PPG signal stands as a rich source of physiological information as it is influenced by all these activities. Lastly, the appearance of NIRS features (HbO , HbR , $Total Hb$) in selected features is not surprising, as multiple studies pointed out significant changes in the PFC oxygenation levels for different types of mental stressors [44].

An important strength of our methodology is the use of a statistical feature selection method (based on Benjamini-Hochberg p-value correction), for each LOSO-CV loop. Unlike manual (ad hoc) feature selection, this method automatically selects the most useful features in each iteration. Another strength is in our validation scheme based on LOSO-CV, which is ideal for assessing future performance with naive users. There is no need to collect calibration signals for a new user based on the methods described here.

One limitation of our current study is the size and homogeneity in demographics of the study population. In future studies, our methods will be validated with larger populations of subjects, and also will include persons with cardiovascular and neurological disorders. Additionally, we intend to classify minimal mental stress changes in future work, perhaps the perceived workload levels that may not be apparent from RTLX scores. Another limitation is that the instrumentation requires placement of the sensors on multiple areas of the body, which is not convenient for the users. Future work will investigate the integration of the different sensing modalities – including both NIRS and cardiovascular signal acquisition – onto a single, head-worn device such as a headband.

Ease-of-use, accuracy, and length of training period are key criteria for HCI research and development. The fusion of NIRS and peripherally-measured cardiovascular sensing significantly increases the classification performance of rest, arithmetic, and N-back tasks. The parameter that reflects cardiac sympathetic activity has the highest feature importance, followed by the parameters that are influenced by sympathetic and vasomotor tone, and PFC activity. NIRS and cardiovascular sensing complement each other strongly for this purpose. The addition of wearable hemodynamic measurements to NIRS sensing provides easy-to-use, higher performance HCIs that require no training for these types of mental stressors. HCIs could translate these physiological signals into a control signal for an external aid during the presence of such mental stressors, thus resulting in improved performance and successful augmentation of the human for challenging tasks.

Acknowledgments

This work is based on material supported in part by the National Heart, Lung and Blood Institute under R01HL130619. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

Biography



Nil Z. Gurel (S' 16) received the B.S. degree in electrical and electronics engineering from Bogaziçi University, Istanbul, Turkey, in 2014, and the M.S. degree in electrical and computer engineering from University of Maryland, College Park, MD, in 2016. Her M.S. work focused on bio-inspired sensing for micro-aerial vehicles. Since 2016, she is pursuing her Ph.D. in electrical and computer engineering at the Georgia Institute of Technology,

under the guidance of Dr. Omer T. Inan. Her research interests include non-invasive physiologic modulation, monitoring, active sensing, biomedical instrumentation, signal processing, and machine learning.



Hewon Jung received the B.S. degree in electrical engineering from the Korea Advanced Institute of Technology (KAIST), Daejeon, Korea, in 2016, and the M.S. degree in electrical and computer engineering in 2018 from the Georgia Institute of Technology, Atlanta, GA, where she is currently pursuing the Ph.D. She is currently a Research Assistant with Inan Research Laboratory, Georgia Institute of Technology, under the guidance of Dr. Omer.T. Inan. Her research interests include non-invasive physiologic monitoring, signal processing, and machine learning.



Sinan Hersek received the B.S. degree in electrical and electronics engineering from Bilkent University, Ankara, Turkey, in 2013, and the M.S. and Ph.D. degrees in electrical and computer engineering from the Georgia Institute of Technology, Atlanta, GA, in 2015 and 2017, respectively. During his senior undergraduate year, he focused on developing digitally controlled, efficient, on-coil power amplifiers for MRI systems with the National Magnetic Resonance Research Center, Bilkent University. The same year he was a Part-time Engineer at ASELSAN (Turkish Military Electronic Industries), Ankara, in the radar and electronic warfare systems business sector. He is currently a Post-Doctoral Researcher with the Department of Electrical and Computer Engineering, Georgia Institute of Technology. His research interests include analog electronics, embedded systems, wearable device design, signal processing, and machine learning.



Omer T. Inan (S'06, M'09, SM'15) received the B.S., M.S., and Ph.D. degrees in electrical engineering from Stanford University, Stanford, CA, in 2004, 2005, and 2009, respectively. He joined ALZA Corporation (A Johnson and Johnson Company) in 2006, where he designed micropower circuits for iontophoretic drug delivery. In 2007, he joined Countryman Associates, Inc., Menlo Park, CA where he was Chief Engineer, involved in designing and developing high-end professional audio circuits and systems. From

2009-2013, he was also a Visiting Scholar in the Department of Electrical Engineering, Stanford University. From 2013-2018, Dr. Inan was Assistant Professor of Electrical and Computer Engineering at the Georgia Institute of Technology, where he is currently an Associate Professor. He is also an Adjunct Associate Professor in the Wallace H. Coulter Department of Biomedical Engineering. His research focuses on non-invasive physiologic sensing and modulation for human health and performance, including for chronic disease management, acute musculoskeletal injury recovery, and pediatric care.

Dr. Inan is an Associate Editor of the IEEE Journal of Biomedical and Health Informatics, Associate Editor for the IEEE Engineering in Medicine and Biology Conference and the IEEE Biomedical and Health Informatics Conference, Invited Member of the IEEE Technical Committee on Translational Engineering for Healthcare Innovation and the IEEE Technical Committee on Cardiopulmonary Systems, and Technical Program Committee Member or Track Chair for several other major international biomedical engineering conferences. He has published more than 125 technical articles in peer-reviewed international journals and conferences, and has six issued patents. Dr. Inan received the Gerald J. Lieberman Fellowship in 2009, the Lockheed Dean's Excellence in Teaching Award in 2016, the Sigma Xi Young Faculty Award in 2017, the IEEE Sensors Early Career Award in 2018, the Office of Naval Research Young Investigator Award in 2018, and the National Science Foundation CAREER Award in 2018. He was a National Collegiate Athletic Association (NCAA) All-American in the discus throw for three consecutive years (2001-2003).

References

- [1]. Blankertz B et al., "The Berlin Brain-Computer Interface: Non-Medical Uses of BCI Technology," *Front Neurosci*, vol. 4, p. 198, 2010. [PubMed: 21165175]
- [2]. Curran EA and Stokes MJ, "Learning to control brain activity: a review of the production and control of EEG components for driving brain-computer interface (BCI) systems," *Brain Cogn*, vol. 51, no. 3, pp. 326–36, 4 2003. [PubMed: 12727187]
- [3]. Nicolas-Alonso LF and Gomez-Gil J, "Brain computer interfaces, a review," *Sensors (Basel)*, vol. 12, no. 2, pp. 1211–79, 2012. [PubMed: 22438708]
- [4]. Jackson AF and Bolger DJ, "The neurophysiological bases of EEG and EEG measurement: a review for the rest of us," *Psychophysiology*, vol. 51, no. 11, pp. 1061–71, 11 2014. [PubMed: 25039563]
- [5]. Scholkmann Fet al., "A review on continuous wave functional near-infrared spectroscopy and imaging instrumentation and methodology," *Neuroimage*, vol. 85 Pt 1, pp. 6–27, 1 15 2014. [PubMed: 23684868]
- [6]. Miller EK and Cohen JD, "An integrative theory of prefrontal cortex function," *Annu Rev Neurosci*, vol. 24, pp. 167–202, 2001. [PubMed: 11283309]
- [7]. Ayaz H, Onaral B, Izzetoglu K, Shewokis PA, McKendrick R, and Parasuraman R, "Continuous monitoring of brain dynamics with functional near infrared spectroscopy as a tool for neuroergonomic research: empirical examples and a technological development," *Front Hum Neurosci*, vol. 7, p. 871, 12 18 2013. [PubMed: 24385959]
- [8]. Tomita N, Imai S, Kanayama Y, Kawashima I, and Kumano H, "Use of Multichannel Near Infrared Spectroscopy to Study Relationships Between Brain Regions and Neurocognitive Tasks of Selective/Divided Attention and 2-Back Working Memory," *Percept Mot Skills*, vol. 124, no. 3, pp. 703–720, 6 2017. [PubMed: 28347211]

- [9]. Herff C, Heger D, Fortmann O, Hennrich J, Putze F, and Schultz T, "Mental workload during n-back task-quantified in the prefrontal cortex using fNIRS," *Front Hum Neurosci*, vol. 7, p. 935, 2013. [PubMed: 24474913]
- [10]. Schudlo LC and Chau T, "Development of a Ternary Near-Infrared Spectroscopy Brain-Computer Interface: Online Classification of Verbal Fluency Task, Stroop Task and Rest," *Int J Neural Syst*, vol. 28, no. 4, p. 1750052, 5 2018. [PubMed: 29281922]
- [11]. von Luhmann A, Wabnitz H, Sander T, and Muller KR, "M3BA: A Mobile, Modular, Multimodal Biosignal Acquisition Architecture for Miniaturized EEG-NIRS-Based Hybrid BCI and Monitoring," *IEEE Trans Biomed Eng*, vol. 64, no. 6, pp. 1199–1210, 6 2017. [PubMed: 28113241]
- [12]. Safaie J, Grebe R, Abrishami Moghaddam H, and Wallois F, "Toward a fully integrated wireless wearable EEG-NIRS bimodal acquisition system," *J Neural Eng*, vol. 10, no. 5, p. 056001, 10 2013. [PubMed: 23893764]
- [13]. Pollonini L, Re R, Simpson RJ, and Dacso CC, "Integrated device for the measurement of systemic and local oxygen transport during physical exercise," *Conf Proc IEEE Eng Med Biol Soc*, vol. 2012, pp. 3760–3, 2012.
- [14]. Liao LM and Carey MG, "Laboratory-induced Mental Stress, Cardiovascular Response, and Psychological Characteristics," *Rev Cardiovasc Med*, vol. 16, no. 1, pp. 28–35, 2015. [PubMed: 25813794]
- [15]. Atterhog JH, Eliasson K, and Hjemdahl P, "Sympathoadrenal and cardiovascular responses to mental stress, isometric handgrip, and cold pressor test in asymptomatic young men with primary T wave abnormalities in the electrocardiogram," *Br Heart J*, vol. 46, no. 3, pp. 311–9, 9 1981. [PubMed: 7295424]
- [16]. Kelsey RM, Ornduff SR, and Alpert BS, "Reliability of cardiovascular reactivity to stress: internal consistency," *Psychophysiology*, vol. 44, no. 2, pp. 216–25, 3 2007. [PubMed: 17343705]
- [17]. Owen AM, McMillan KM, Laird AR, and Bullmore E, "N-back working memory paradigm: a meta-analysis of normative functional neuroimaging studies," *Hum Brain Mapp*, vol. 25, no. 1, pp. 46–59, 5 2005. [PubMed: 15846822]
- [18]. Ayaz H, Shewokis PA, Bunce S, Izzetoglu K, Willems B, and Onaral B, "Optical brain monitoring for operator training and mental workload assessment," *Neuroimage*, vol. 59, no. 1, pp. 36–47, 1 2 2012. [PubMed: 21722738]
- [19]. Hoskinson P. (04.05.2018). Brain Workshop. Available: <http://brainworkshop.sourceforge.net/>
- [20]. Hart SG, "Nasa-Task Load Index (NASA-TLX); 20 Years Later," *Proceedings of the human factors and ergonomics society annual meeting*, vol. 50, no. 9, 2006.
- [21]. Strangman GE, Li Z, and Zhang Q, "Depth sensitivity and source-detector separations for near infrared spectroscopy based on the Colin27 brain template," *PLoS One*, vol. 8, no. 8, p. e66319, 2013. [PubMed: 23936292]
- [22]. He DD, Winokur ES, and Sodini CG, "An Ear-Worn Vital Signs Monitor," *IEEE Trans Biomed Eng*, vol. 62, no. 11, pp. 2547–52, 11 2015. [PubMed: 26208264]
- [23]. Sornmo L, *Bioelectrical Signal Processing in Cardiac and Neurological Applications*, 1st ed. Burlington, MA: Elsevier, 2005.
- [24]. Allen J, "Photoplethysmography and its application in clinical physiological measurement," *Physiol Meas*, vol. 28, no. 3, pp. R1–39, 3 2007. [PubMed: 17322588]
- [25]. Sherwood A, Allen MT, Fahrenberg J, Kelsey RM, Lovallo WR, and van Doornen LJ, "Methodological guidelines for impedance cardiography," *Psychophysiology*, vol. 27, no. 1, pp. 1–23, 1 1990. [PubMed: 2187214]
- [26]. Inanetal OT., "Ballistocardiography and seismocardiography: a review of recent advances," *IEEE J Biomed Health Inform*, vol. 19, no. 4, pp. 1414–27, 7 2015. [PubMed: 25312966]
- [27]. Mukkamala Ret al., "Toward Ubiquitous Blood Pressure Monitoring via Pulse Transit Time: Theory and Practice," *IEEE Trans Biomed Eng*, vol. 62, no. 8, pp. 1879–901, 8 2015. [PubMed: 26057530]
- [28]. Kocsis L, Herman P, and Eke A, "The modified Beer-Lambert law revisited," *Phys Med Biol*, vol. 51, no. 5, pp. N91–8, 3 7 2006. [PubMed: 16481677]

- [29]. Benjamini Yoav and Hochberg Y, "Controlling the false discovery rate: a practical and powerful approach to multiple testing," *Journal of the royal statistical society. Series B (Methodological)*, vol. 57, no. 1, pp. 289–300, 1995.
- [30]. der Maaten Laurens van and Hinton G, "Visualizing data using t-SNE," *Journal of machine learning research*, vol. 9, pp. 2579–2605, 2008.
- [31]. Hastie Trevor, Tibshirani Robert, and Friedman J, *The Elements of Statistical Learning*, 2 ed. New York: Springer, 2001.
- [32]. Pedregosa F, "Scikit-learn: Machine learning in Python," *Journal of machine learning research*, vol. 12, pp. 2825–2830, 2011.
- [33]. Krstaji Damjan, Buturovic Ljubomir J., Leahy David E., and Thomas S, "Cross-validation pitfalls when selecting and assessing regression and classification models," *Journal of Cheminformatics*, vol. 6, no. 10, 2014.
- [34]. Demsar J, "Statistical Comparisons of Classifiers over Multiple Data Sets," *Journal of Machine Learning Research*, vol. 7, pp. 1–30, 2006.
- [35]. Wattenberg M, Viégas F, and Johnson I, "How to use t-sne effectively," *Distill*, vol. 1, no. 10, p. e2, 2016.
- [36]. Barbey AK, Koenigs M, and Grafman J, "Dorsolateral prefrontal contributions to human working memory," *Cortex*, vol. 49, no. 5, pp. 1195–205, 5 2013. [PubMed: 22789779]
- [37]. Houtveen JH, Rietveld S, and de Geus EJ, "Contribution of tonic vagal modulation of heart rate, central respiratory drive, respiratory depth, and respiratory frequency to respiratory sinus arrhythmia during mental stress and physical exercise," *Psychophysiology*, vol. 39, no. 4, pp. 427–36, 7 2002. [PubMed: 12212635]
- [38]. Willemsen G, Ring C, Carroll D, Evans P, Clow A, and Hucklebridge F, "Secretory immunoglobulin A and cardiovascular reactions to mental arithmetic and cold pressor," *Psychophysiology*, vol. 35, no. 3, pp. 252–9, 5 1998. [PubMed: 9564745]
- [39]. Cui X, Bray S, Bryant DM, Glover GH, and Reiss AL, "A quantitative comparison of NIRS and fMRI across multiple cognitive tasks," *Neuroimage*, vol. 54, no. 4, pp. 2808–21, 2 14 2011. [PubMed: 21047559]
- [40]. Gordan R, Gwathmey JK, and Xie LH, "Autonomic and endocrine control of cardiovascular function," *World J Cardiol*, vol. 7, no. 4, pp. 204–14, 4 26 2015. [PubMed: 25914789]
- [41]. Goedhart AD, Willemsen G, Houtveen JH, Boomsma DI, and De Geus EJ, "Comparing low frequency heart rate variability and preejection period: two sides of a different coin," *Psychophysiology*, vol. 45, no. 6, pp. 1086–90, 11 2008. [PubMed: 18823417]
- [42]. Gasperin D, Netuveli G, Dias-da-Costa JS, and Pattussi MP, "Effect of psychological stress on blood pressure increase: a meta-analysis of cohort studies," *Cad Saude Publica*, vol. 25, no. 4, pp. 715–26, 4 2009. [PubMed: 19347197]
- [43]. Charlton PH, Celka P, Farukh B, Chowienczyk P, and Alastruey J, "Assessing mental stress from the photoplethysmogram: a numerical study," *Physiol Meas*, vol. 39, no. 5, p. 054001, 5 15 2018. [PubMed: 29658894]
- [44]. Naseer N and Hong KS, "fNIRS-based brain-computer interfaces: a review," *Front Hum Neurosci*, vol. 9, p. 3, 2015. [PubMed: 25674060]

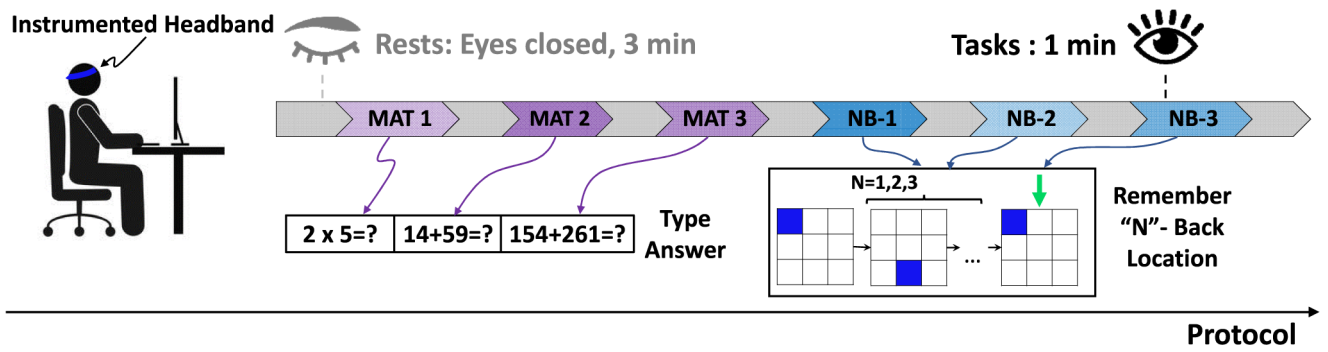


Fig. 1.

Protocol description: Each subject completed a series of arithmetic and N-back tasks with difficulty levels shuffled, each lasting a minute. Before the start of the tasks and between each task, subjects rested for three minutes with their eyes-closed. MAT: Mental arithmetic, NB: N-back task.

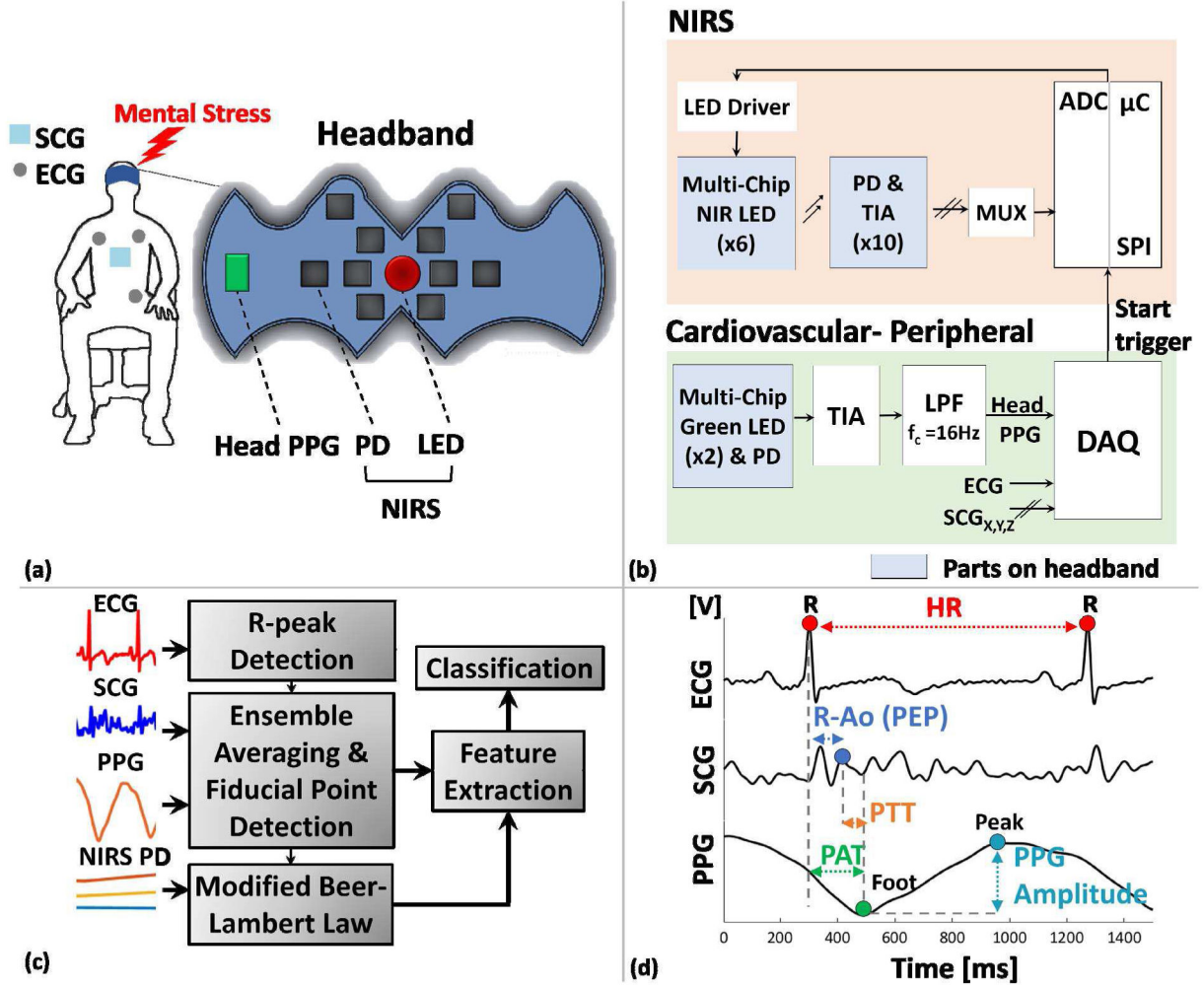


Fig. 2.

a) Illustration of the sensing modalities attached to each subject: headband PPG, NIRS, ECG, SCG signals were collected throughout the mental stress protocol. b) Block-diagrams of each sensing modality. The headband includes PPG and NIRS parts. c) Block diagram of signal processing and feature extraction. d) Summary of peripherally-measured cardiovascular parameters. HR, R-Ao (PEP), PTT, PAT, PPG amplitude were extracted from the peripherally-measured cardiovascular sensing part.

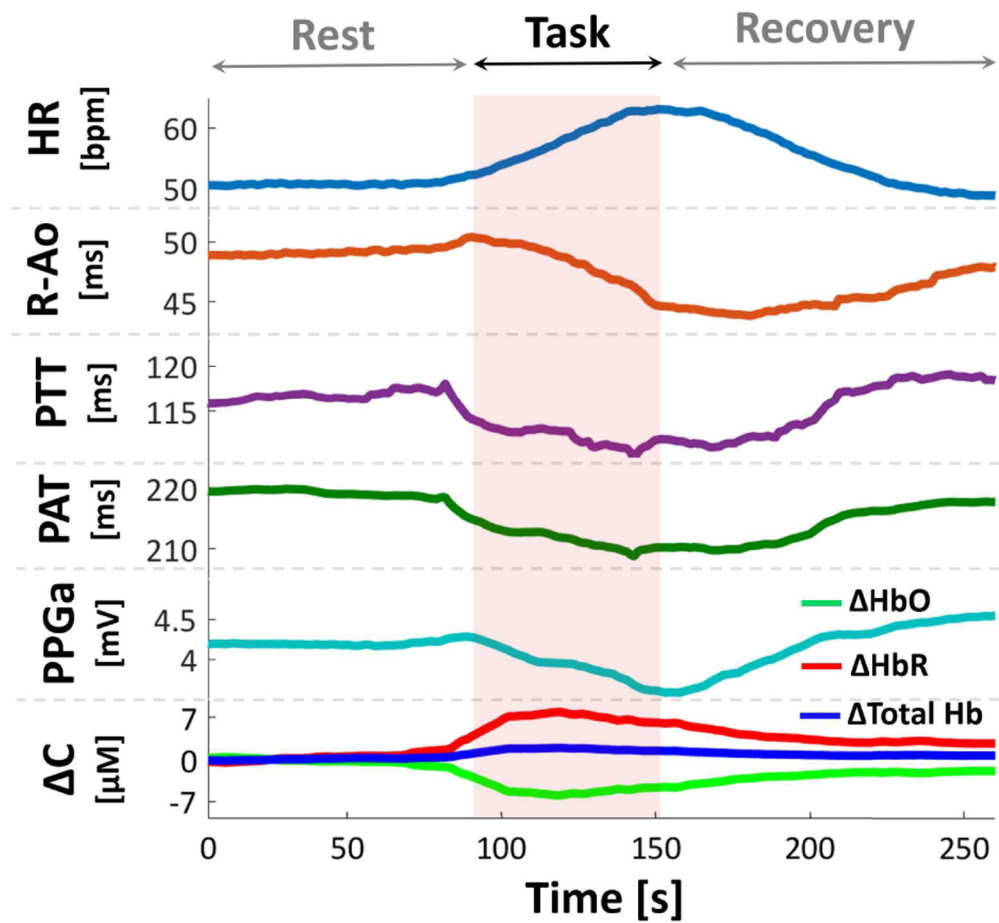


Fig. 3.

Extracted parameters for a subject transitioning from rest to mental stress. Changes in HR, R-Ao, PTT, PAT, PPG amplitude (PPGa), oxyhemoglobin (HbO), deoxyhemoglobin (HbR), total hemoglobin (Total Hb) were observed during stress. Data in figure were smoothed with moving average filter for better visualization. C: Change in concentration, [M]: Molar.

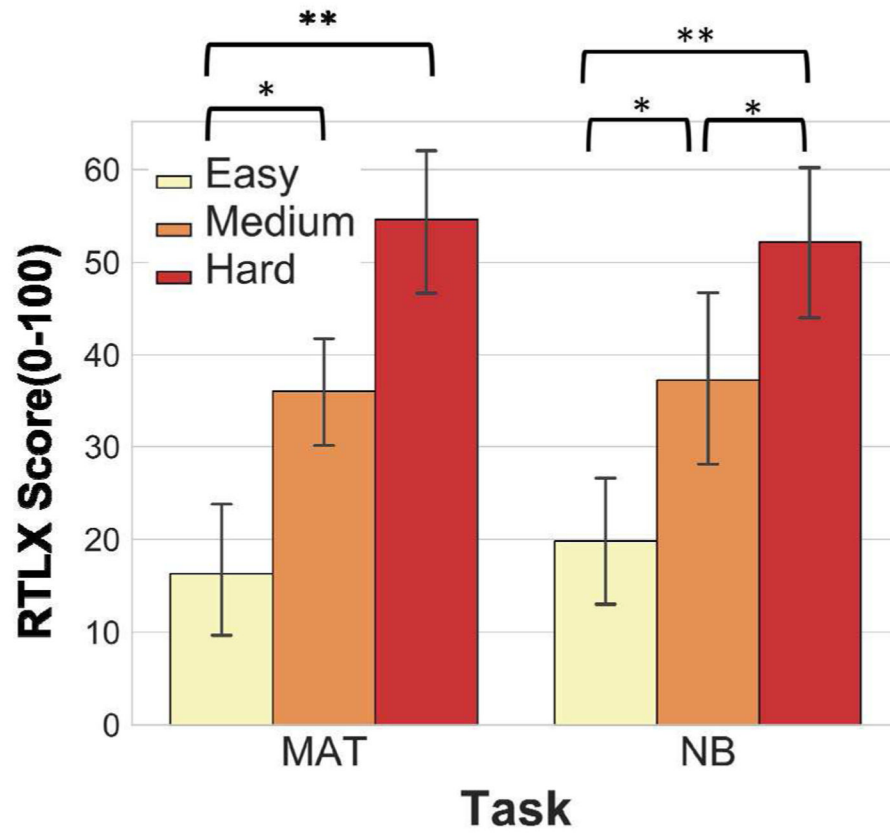


Fig. 4. RTLX scores for arithmetic and N-back tasks, for difficulty levels easy, medium, hard for each task. * indicates statistical significance, * $p < 0.05$, ** $p < 0.001$. Error bars: std.

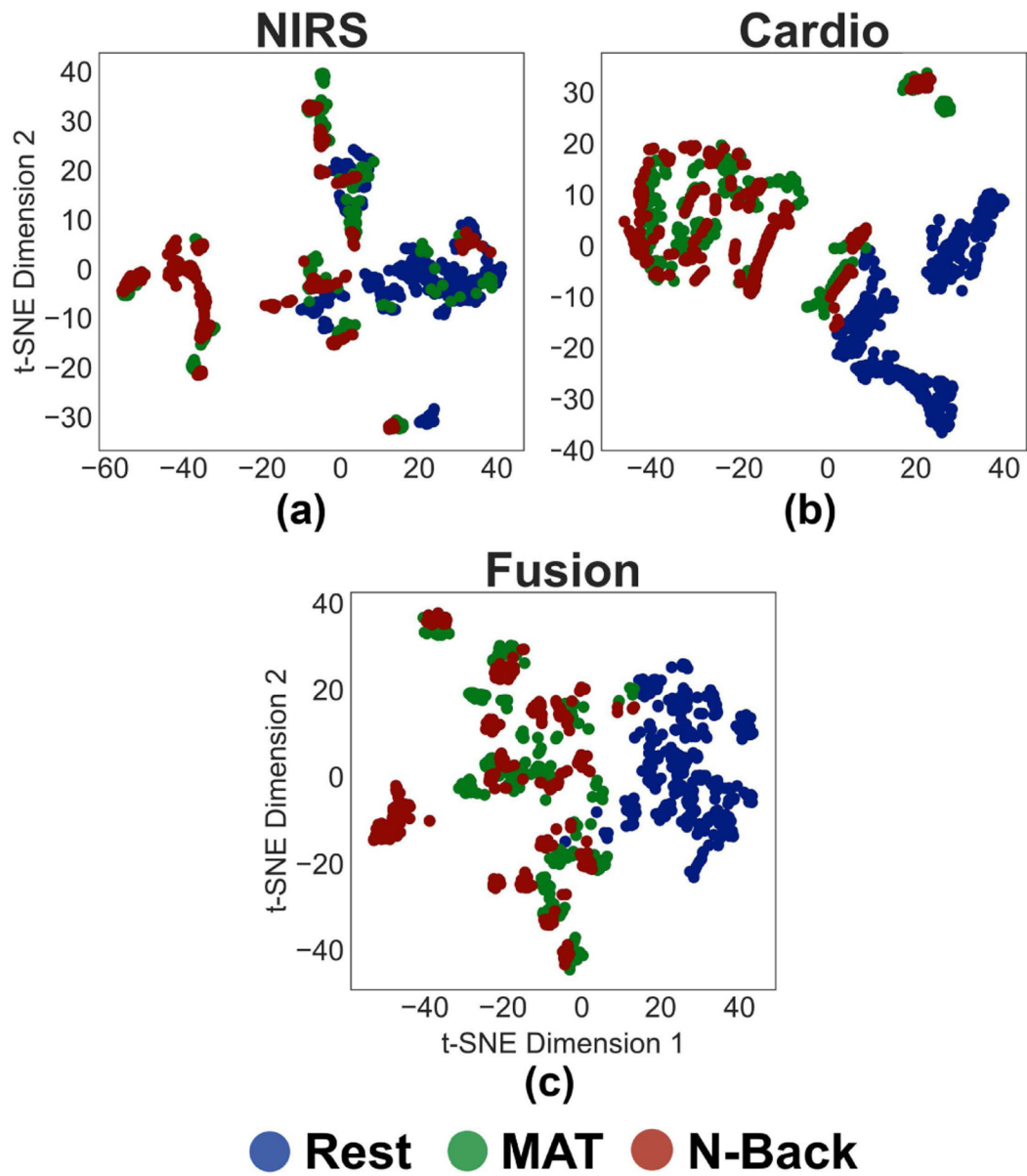


Fig. 5. t-SNE plots for features related to a) NIRS sensing, b) peripherally-measured cardiovascular sensing, and c) the fusion of both sensing modalities for rest, MAT, and N-back classes.

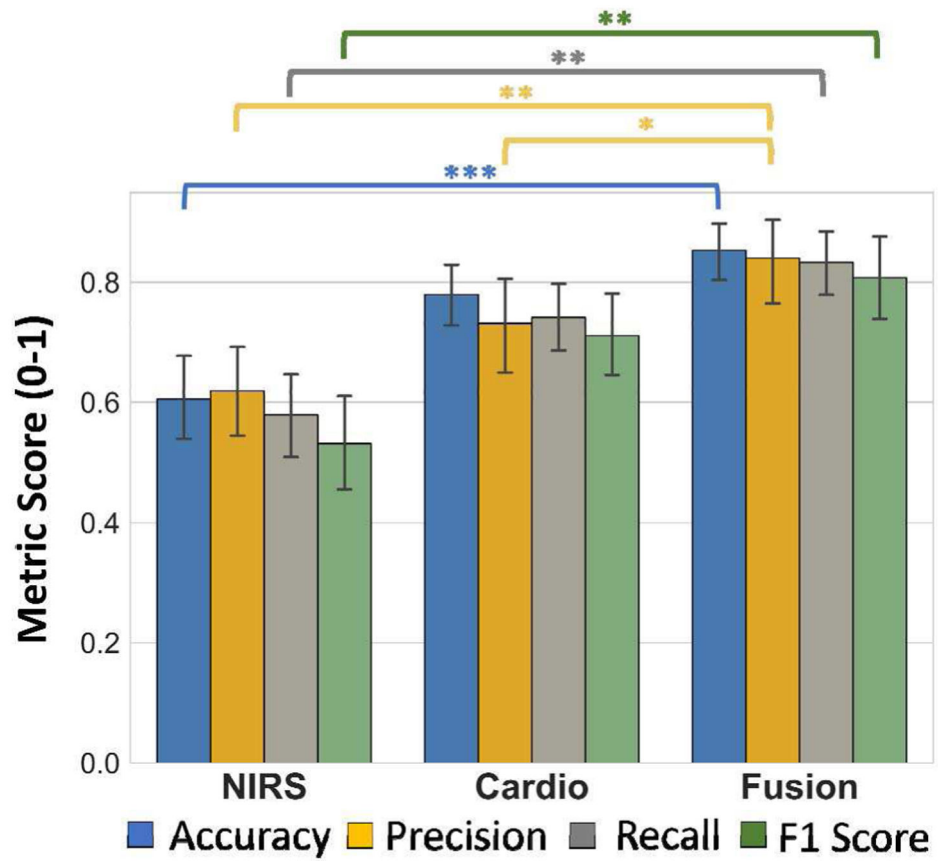


Fig. 6. Summary of the performance metrics for classifier outputs for each sensing modality. * indicates statistical significance, * $p < 0.05$, ** $p < 0.001$, *** $p < 0.0001$. Error bars: std.

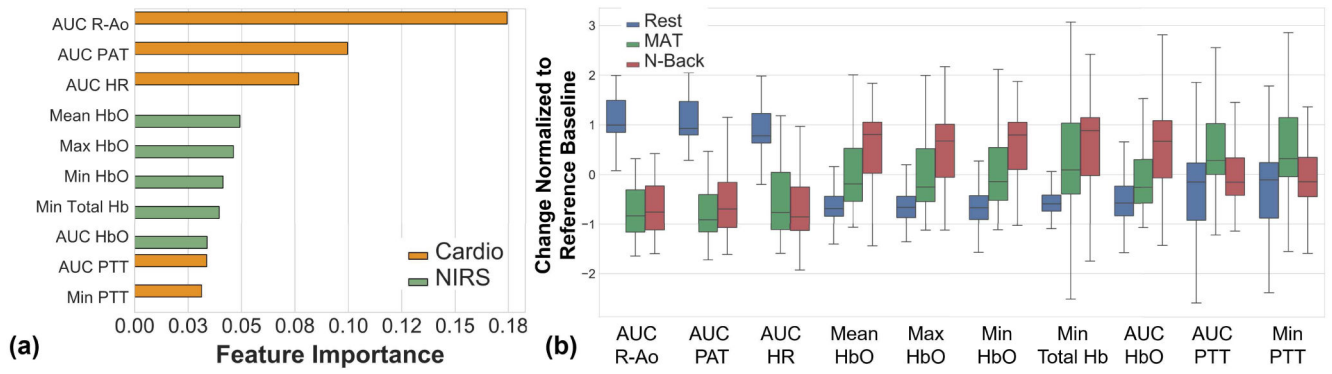


Fig. 7.
 a) Feature importance ranking for the top 10 features. Top 10 include both cardiovascular-peripheral and NIRS features, with cardiovascular-peripheral features having the highest importance. b) Boxplots of top 10 features. Error bars: std