



Evaluating Wearable Sensors for Combined Physical and Mental Fatigue Assessment

Ahmad Fauzi¹, Ohnishi Ayumi², Tsutomu Terada³, Masahiko Tsukamoto⁴

Graduate School of
Engineering
Kobe University, Japan¹
ahmad.fauzi@stu.kobe-u.ac.jp

Graduate School of
Engineering
Kobe University, Japan²
ohnishi@eedept.kobe-u.ac.jp

Graduate School of
Engineering
Kobe University, Japan³
tsutomu@eedept.kobe-u.ac.jp

Graduate School of
Engineering
Kobe University, Japan⁴
tuka@eedept.kobe-u.ac.jp

Abstract: Estimating mental fatigue levels during physical exertion is essential for ensuring safety and optimizing performance in sports and other high-demand environments. However, it remains challenging due to the overlapping physiological signals involved. This study examines the effectiveness of various wearable sensors, including ECG, PPG, and EEG, as well as eye trackers, in estimating mental fatigue levels under conditions of physical fatigue. Multimodal sensor data collected during controlled experiments that combined cognitive tasks with physical exercise were analyzed, with a focus on machine learning models for estimating mental fatigue levels on a 10-class scale. The results show that although EEG is the most accurate single sensor for estimating mental fatigue under physical fatigue, a combination of ECG and eye-tracking achieves higher accuracy (79.16%) with minimal error (MAE=0.52). This combination offers a balance between performance and wearability for practical applications.

Keywords: Fatigue Assessment, Mental Fatigue, Wearable Sensors.

I. INTRODUCTION

Mental fatigue, resulting from prolonged cognitive exertion, leads to measurable declines in cognitive performance. Studies demonstrate that intraindividual variability in reaction time, a marker of attentional instability, is a sensitive indicator of cognitive fatigability [1]. This variability correlates with subjective fatigue and reflects lapses in cognitive control, suggesting a breakdown in sustained attention under mental fatigue.

Mental fatigue poses serious risks in high-demand environments like air traffic control and maritime operations, where even minor performance degradation can have fatal consequences [2,3]. Research demonstrates that mental fatigue leads to measurable declines in operational performance, with EEG studies revealing significant neural changes after sustained task performance [3]. These changes correlate with impaired decisions and slower responses, which can make important activities unsafe. While individuals may temporarily compensate for fatigue through increased cognitive effort, evidenced by pupil dilation and subjective reports, this comes at a cost. Such

compensatory mechanisms are unstable under prolonged strain and can lead to sudden performance breakdowns, particularly in complex tasks [4].

Although mental fatigue has been vastly studied, detecting it in concurrence with physical fatigue remains a challenge. Wearable sensors, including those for electrocardiography (ECG), photoplethysmography (PPG), and electroencephalography (EEG), as well as eye-trackers, offer promising solutions for monitoring mental fatigue. However, their effectiveness in the presence of both physical and mental fatigue remains unclear. Previous research has shown the potential of wearable sensors for detecting isolated fatigue. For instance, mental fatigue was identified by analyzing heart rate variability (HRV) patterns through machine learning techniques [5]. This study indicates that wearable ECG devices can effectively monitor mental fatigue by tracking relevant physiological markers. In drivers, mental fatigue has been detected using SpO₂ (peripheral oxygen saturation) signals, which are indicative of autonomic nervous system activity associated with fatigue [6]. This approach highlights the sensitivity of SpO₂ to cardiovascular and respiratory changes during states of fatigue. Additionally, EEG

Corresponding Author: Tsutomu Terada. Graduate School of Engineering, Kobe University, 1-1 Rokkodai, Nada Ward, Hyogo, Japan, Tel/Fax: +81-3-5286-3759, E-mail: tsutomu@eedept.kobe-u.ac.jp

analysis has been utilized to detect mental fatigue by evaluating increases in theta and alpha power, using a machine learning classifier to recognize the development of fatigue [7]. Furthermore, fluctuations in pupil diameter have been monitored through infrared pupillography to assess mental fatigue [8].

However, adding physical activity can create complex signals in the body, such as exercise-induced heart rate variability (HRV) suppression, which complicates the direct application of existing methods [9]. This study evaluates the performance of wearable sensors in estimating mental fatigue levels in the context of physical fatigue, addressing two important gaps: a comparative analysis of various wearable sensors (including ECG, PPG, and EEG, as well as eye tracker) under conditions of physical fatigue, and an investigation into the potential of their combinations.

This work makes three main contributions. First, it compares single wearable sensors (ECG, PPG, EEG, and eye tracker) for estimating mental fatigue during physical fatigue. Second, it tests sensor combinations to identify the optimal pairs and groups in terms of accuracy and reliability. Third, it includes an evaluation of comfort and wearability. The rest of the paper is organized as follows: Section II reviews related work; Section III describes the experimental design and methods; Section IV presents the results; Section V discusses implications and limitations; and Section VI concludes with a summary of key findings and future directions.

II. RELATED RESEARCH

A. Wearable Sensors for Mental Fatigue Level Estimation

The emerging field of research into objective indicators of mental fatigue highlights the potential of various physiological measures. ECG-derived heart rate variability (HRV) has been recognized as a potential biomarker for mental fatigue. Studies have shown significant alterations in HRV following cognitive tasks [10]. However, the complexity introduced by exercise-induced physiological changes complicates the interpretation of these HRV measurements. Recent findings suggest that during simultaneous mental and physical tasks, physical fatigue often dominates HRV patterns, potentially masking the signatures of cognitive fatigue [11]. This indicates that conventional HRV metrics may lack the specificity necessary to detect mental fatigue in real-world settings accurately.

In addition, monitoring SpO₂ can help estimate mental fatigue levels. Research has shown that

extended work shifts are associated with consistent declines in SpO₂ levels, correlating with self-reported mental fatigue and physiological stress indicators, such as increased heart rate [12, 13]. However, SpO₂ measurements cannot adequately differentiate between the two, given that both forms of fatigue can significantly affect oxygenation levels [14].

Electroencephalography (EEG) has also emerged as a valid instrument for estimating mental fatigue levels. Research indicates that EEG parameters can effectively capture the progression of mental fatigue, with physical exertion intensifying these neural markers [15]. Even though advanced analytical techniques can achieve high levels of detection accuracy in controlled environments, real-world applications face limitations due to the interference caused by physical activity [16]. The effects of physical activity can alter EEG patterns in ways that may mimic or amplify indicators of mental fatigue.

Lastly, recent advancements in pupil diameter dynamics have revealed their potential as indicators of mental fatigue. Research indicates that while baseline pupil size remains steady, task-evoked pupillary responses (TERPs) decrease progressively with increasing mental fatigue, signaling cognitive resource depletion. However, it is crucial to acknowledge the confounding influence of physical fatigue, which can independently activate sympathetic responses and further alter pupil dynamics, thereby obscuring accurate indicators of cognitive fatigue [17].

While promising advancements have been made in identifying objective measures of mental fatigue, significant confounding factors, particularly those arising from physical exertion, remain a crucial concern in accurately assessing and interpreting these indicators. The overlapping of physiological responses to cognitive and physical stressors also complicates the interpretation of individual signals. Therefore, combining data from multiple sensors may provide a more robust alternative, potentially improving the accuracy and reliability of assessments in real-world scenarios by providing a more comprehensive physiological profile.

B. Methods for Inducing Mental and Physical Fatigue

There are various established protocols for inducing mental and physical fatigue. Physical fatigue is commonly induced through sustained or repetitive physical exertion, such as treadmill exercise, cycle ergometry, or resistance exercises [18-25]. Mental

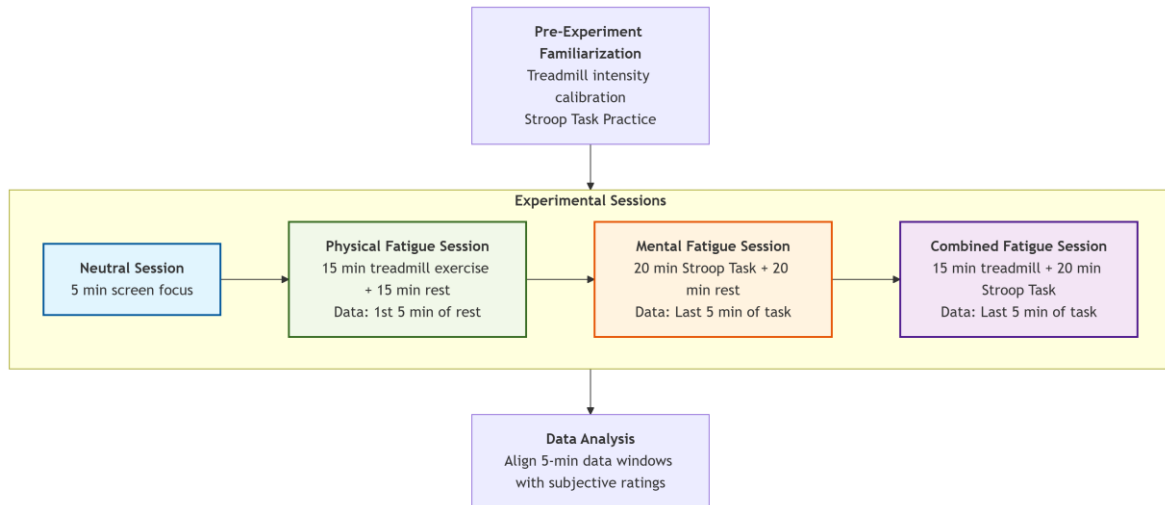


Fig. 1 Experimental Procedure

fatigue, by contrast, is typically induced via prolonged cognitive tasks that demand sustained attention, working memory, or executive control. Common paradigms include the Psychomotor Vigilance Task (PVT), n-back tasks, and the Stroop Task [26-32]. The Stroop Task is a well-validated method for imposing a high cognitive load by creating cognitive conflict. In this task, participants must name the ink color of a color word while inhibiting the automatic tendency to read the word itself (e.g., the word "RED" printed in blue ink). This interference requires greater cognitive control and attention, leading to mental fatigue, tiredness, and poorer performance, including slower responses and fewer correct answers [31, 32]. Compared with other tasks, the Stroop Task imposes greater demands on cognitive control and response inhibition, resulting in more pronounced and persistent fatigue effects and making it suitable for studies requiring robust mental fatigue induction.

In this study, the combination of treadmill running and the Stroop Task was selected to induce physical and mental fatigue. Treadmill exercise provides a controllable and reproducible means of inducing physical fatigue while allowing concurrent cognitive task performance. The Stroop Task was chosen for its well-established ability to deplete cognitive resources efficiently and for its compatibility with a physically active setting. This dual-induction approach enables the isolation and examination of physiological signals under distinct fatigue states, physical, mental, and combined, which is the aim of this study.

III. PROPOSED METHOD

A. Experiment Design

The experiment was arranged into four different sessions to investigate the effects of physical fatigue, mental fatigue, and their combined impact on physiological signals within the context of mental fatigue, as illustrated in Fig. 1. In the neutral session, participants focused on a screen while remaining physically relaxed for five minutes, establishing a baseline for low levels of both physical and mental fatigue.

During the physical fatigue session, participants engaged in 15 minutes of treadmill exercise at a predetermined intensity, followed by a 15-minute rest period. Physiological data from the first five minutes of this rest period were analyzed, capturing what is assumed to be the peak of physical fatigue while minimizing movement artifacts, thereby representing a state of high physical fatigue alongside low mental fatigue.

The mental fatigue session involved a twenty-minute digital Stroop Task designed to induce cognitive load, followed by an additional twenty minutes of rest. For analysis, data from the last five minutes of the Stroop Task were used to represent high mental fatigue with minimal physical fatigue, as this timeframe was considered to reflect maximal cognitive strain.

In the combined fatigue session, participants simultaneously performed treadmill exercise and Stroop Task for 15 and 20 minutes, respectively. The treadmill exercise stopped after fifteen minutes, while the cognitive task continued. Physiological data from the

Congruent	Incongruent
RED	RED
GREEN	GREEN
BLUE	BLUE
YELLOW	YELLOW

Fig 2. Stroop Task

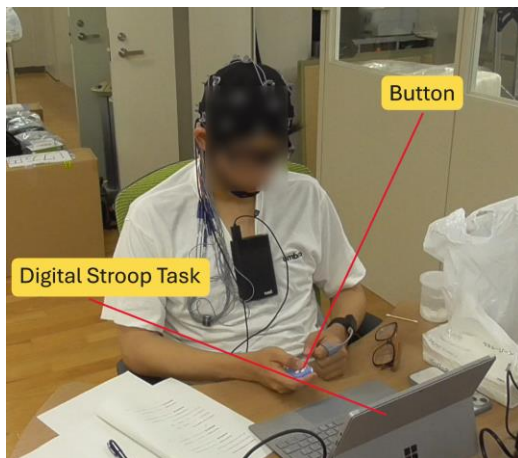


Fig 3. Mental Fatigue Session

final five minutes of the Stroop Task were analyzed to assess the combined effects of high physical and mental fatigue, thereby simulating real-world situations where both types of fatigue occur concurrently.

Before the main experiment, participants completed a brief treadmill trial to determine individualized exercise intensities. They also underwent a practice session for the Stroop Task to become familiar with the response mappings. This methodology facilitated the controlled isolation of various fatigue states while minimizing confounding factors such as learning effects or uneven levels of exertion.

B. Fatigue Induction

In this study, physical fatigue was induced through two sessions of treadmill exercise, each lasting 15 minutes (referred to as the Physical Fatigue session and the Combined Fatigue Session). The speed for each participant was customized according to their limits. Before the experiment, participants completed a brief treadmill trial to establish their optimal exercise speed. The lowest default speed set for all participants was 6



Fig 4. Combined Fatigue Session

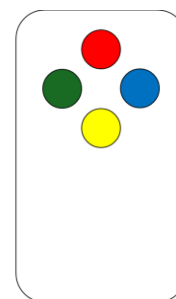


Fig 5. Button Mapping

km/h, which was gradually increased until participants indicated they could no longer sustain the exercise for the full 15 minutes. This ensures maximum physical exertion while allowing participants to fully engage in the experimental sessions.

To induce mental fatigue, participants completed a digital Stroop Task, shown in Fig. 2. In this task, participants viewed color words (e.g., "RED," "GREEN," "BLUE," "YELLOW") printed in incongruent ink colors. For instance, the word "RED" may be displayed in blue ink. The primary instruction is to name the ink color of the word while disregarding the word itself. This creates a cognitive conflict, as reading is an automatic process, making it challenging to suppress the meaning of the word and concentrate solely on the color. The task consists of two types of trials: congruent trials, in which the word and ink color match (e.g., "BLUE" in blue), and incongruent trials, where they do not (e.g., "BLUE" in red).

Participants were required to complete the Stroop Task during both the Mental Fatigue Session and the

Table 1. Questionnaire Items used for Sensor Usability Evaluation

It was comfortable to wear during physical activity (Comfort)						
Strongly Disagree	1	2	3	4	5	Strongly Agree
It interfered with my movements (Interfere)						
Strongly Agree	1	2	3	4	5	Strongly Disagree
The sensor material (e.g., strap, cap) felt intrusive (Intrusive)						
Strongly Agree	1	2	3	4	5	Strongly Disagree
The weight of the sensor was unnoticeable (Weight)						
Strongly Disagree	1	2	3	4	5	Strongly Agree
I would be willing to wear this sensor for extended periods (e.g., > 4 hours) (Extended)						
Strongly Disagree	1	2	3	4	5	Strongly Agree
The sensor was easy to wear without assistance (Easy)						
Strongly Disagree	1	2	3	4	5	Strongly Agree

Combined Fatigue Session. The screen was positioned in front of the participant, as illustrated in Figs. 3 and 4. Participants responded by pressing a button, as shown in Fig. 5.

C. Subjective Fatigue Ratings

To ensure accurate alignment between subjective reports and physiological measures, the 5-minute segments of physiological data from each experimental session (i.e., neutral, physical fatigue, mental fatigue, combined fatigue) were matched with their corresponding subjective fatigue ratings. For the physical fatigue condition, we aligned the first 5 minutes of post-exercise recovery data with the immediately following subjective rating. In the mental fatigue condition, physiological recordings from the last 5 minutes of the Stroop Task were paired with post-task ratings. The combined fatigue session used the same alignment method as the mental condition.

D. Comfort and Usability Questionnaire

Following the experiment, each participant was asked to complete a questionnaire assessing the comfort

and usability of each sensor. This post-experiment assessment was intended to capture the participants' subjective experiences and perceptions with the sensor after extended use. For each sensor, the following six practical considerations were assessed: comfort during physical activity, interference with movement, intrusiveness of the sensor material, perceived weight, willingness to wear for extended periods (more than four hours), and ease of wearing without assistance. These criteria were chosen to reflect the physical factors that may influence a user's willingness to use the wearable sensors in everyday settings. Participant responses were recorded using a 5-point Likert scale (1 = strongly disagree, 5 = strongly agree), as shown in Table 1.

E. Participants

A total of ten adult participants (all male, aged 22 to 27 years) were recruited for the experiment. Each participant provided written informed consent after receiving thorough explanations of the experiment procedures and associated risks. To minimize confounding physiological effects, participants were instructed to refrain from consuming caffeine, alcohol, and engaging in any activities that could impact their performance before the testing sessions.

F. Sensor Setup

Multimodal wearable sensors were used to capture physiological signals:

ECG: Cardiac activity was monitored using the MyBeat WHS-3 single-lead ECG sensor, a lightweight, chest-worn device that recorded the heart's electrical signals. This sensor was designed to accurately detect R-peaks even during physical movement, facilitating reliable heart rate variability (HRV) analysis across various experimental conditions. Data was transmitted wirelessly, allowing participants full mobility during both stationary cognitive tasks and treadmill exercises. The system provided a wide range of HRV metrics, including time-domain parameters (R-R intervals) and frequency-domain components (LF, HF, and LF/HF ratio), which served as indicators of autonomic nervous system modulation in various states of fatigue. With the capability of continuous operation for an extended period, the sensor ensured uninterrupted monitoring throughout long testing sessions without the need for battery changes. The ECG data were synchronized with other sensors using timestamp alignment, ensuring temporal precision in the assessment of fatigue.

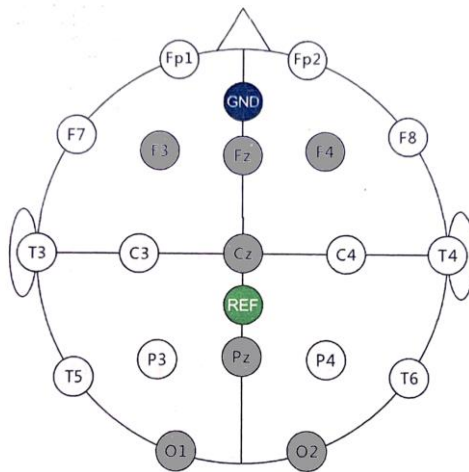


Fig 6. EEG Electrode Placement

PPG: The Ring O₂ PPG sensor was used to monitor peripheral oxygen saturation (SpO₂) levels. Participants wore this ring-shaped sensor on the index finger of their non-dominant hand to enhance comfort and minimize interference with task performance. The sensor continuously recorded SpO₂ levels and provided averaged readings every two seconds.

The ring design offered distinct advantages over traditional finger-clip PPG sensors, particularly in reducing motion artifacts associated with hand movements during physical activities. With an extended battery life, the sensor enabled uninterrupted monitoring throughout all experimental sessions without the need for charging breaks.

EEG: EEG signals were recorded using an OpenBCI system equipped with saltwater electrodes and a Cyton 8-channel amplifier, operating at a sampling rate of 250 Hz. The system targeted several electrode positions, including frontal (F3, Fz, F4), central (Cz), parietal (Pz), and occipital (O1, O2) regions, as shown in Fig. 6, to capture neural activity patterns related to mental fatigue. Before the experiment, the quality of electrode contact was verified using OpenBCI GUI v6.0.0, ensuring that impedance remained below 300 kΩ to facilitate optimal signal acquisition. The design of the saltwater electrodes offered several advantages for studying fatigue, notably enhancing participant comfort during prolonged wear. This system was capable of detecting subtle neural changes relevant to mental fatigue assessment, especially within the theta (4-7 Hz), alpha (8-12 Hz), and beta (13-30 Hz) frequency bands. Under physical activity conditions, the saltwater solution helped maintain electrode-scalp contact despite



Fig. 7 Sensor Setup

participant movement, all while preserving comfort. Moreover, the wireless configuration allowed for natural head movements throughout the experiment, making this system particularly well-suited for the study.

Eye Tracker: The Tobii Pro Glasses 3 eye tracker was used to record binocular pupillometry and gaze behavior at a frequency of 100 Hz, providing accurate measurements of pupil diameter dynamics and visual attention patterns. To ensure the validity of pupillometry data, all experiments were conducted in a controlled environment with constant ambient lighting. External light sources were minimized, and the brightness of the stimulus screen was standardized across all participants and sessions to avoid variations in pupil diameter that were not related to mental fatigue. This device features an integrated scene camera that captures the participant's field of view, enabling accurate tracking of gaze behavior in relation to experimental stimuli, such as the Stroop Task display. The lightweight frame of the glasses minimizes additional fatigue during extended wear, while a small carrying bag for the central hub enhances comfort and positioning.

Automatic calibration ensures precise gaze tracking without the need for extensive manual setup. Synchronization with other sensors is achieved by aligning timestamps across all recording devices. The eye tracker provides continuous measurements of pupil

Table 2. Performance Metrics for Individual Sensors

Sensor	Accuracy (%)	MAE
ECG	56.29	0.99
PPG	69.52	0.91
Eye Tracker	71.54	0.95
EEG	74.27	0.79

diameter for both the left and right eyes, which is particularly beneficial for estimating mental fatigue levels.

Although the system delivers excellent data quality, two practical challenges need to be addressed, including limited battery life and significant heat generation during prolonged use. To mitigate these issues, scheduled maintenance breaks were implemented during rest periods. During these intervals, the device was powered down to allow for cooling, and spare batteries were swapped in to ensure uninterrupted data collection throughout the complete session duration. The overall sensor setup and placement are illustrated in Fig. 7. For clarity, the ECG sensor is shown outside the shirt. The EEG circuit board was placed behind the participants' heads.

G. Data Processing and Feature Extraction

ECG and PPG: MyBeat WHS-3 ECG and Ring O₂ PPG provided precomputed features (e.g., R-R intervals, SpO₂), omitting raw signal preprocessing. Statistical features (i.e., mean, median, variance, skewness, kurtosis) were derived directly from these outputs for each experimental session.

EEG: The raw EEG data was processed through a three-stage pipeline: (1) Detrending: Trends were removed in segments of 250 samples (1 second each); (2) Filtering: High-pass (4 Hz), band-stop (59–61 Hz), and band-pass (4–30 Hz) filters were applied to isolate the theta, alpha, and beta frequency bands; (3) Quality Control: Sessions or channels that showed artifacts (e.g., disconnected electrodes) were excluded from analysis. The features extracted included channel-wise relative band power (theta: 4–7 Hz, alpha: 8–12 Hz, beta: 13–30 Hz) utilizing sliding windows with 50% overlap, global average power, and various statistical features (i.e., mean, median, variance, skewness, and kurtosis).

Eye-Tracker: The pupillometry data collected during the experiment underwent preprocessing to enhance data quality and reliability. This process

included the identification and removal of blink artifacts, which can introduce significant noise into pupil diameter measurements. After artifact removal, statistical features were extracted from the cleaned dataset.

H. Post Processing

All features were first balanced using the Synthetic Minority Oversampling Technique (SMOTE). Features were then normalized using the z-score method for each sensor. For the EEG data, the relative power, average power, and statistical features were normalized separately. The dataset was then divided into 80% for training and 20% for testing in a 5-fold cross-validation approach while maintaining the order of the data. Top features were selected based on their importance in scoring for evaluation.

I. Evaluation

The performance of the sensors was assessed using a Random Forest classifier, focusing on two key metrics (i.e., accuracy (overall prediction correctness) and mean absolute error (MAE)). The model was trained to predict the ground truth of subjective fatigue ratings, which consisted of a 10-class ordinal scale ranging from 1 ("none") to 10 ("very exhausted"). The dataset was divided into five sequential segments, with each segment used as the test set (20%) in turn, while the remaining data (80%) served as the training set for the following iterations. The evaluation process was carried out in two phases: (1) assessing each sensor (EEG, ECG, PPG, and eye-tracking), and (2) evaluating combinations of sensors.

IV. RESULTS

In this study, the effectiveness of several wearable sensors, including ECG, PPG, EEG, and an eye-tracker, was evaluated for their performance to estimate mental fatigue levels under physical fatigue. The experimental procedures were designed to capture data across different fatigue conditions. The results are presented in three subsections: (A) Individual Sensor Performance, (B) Sensor Combination Performance, and (C) Sensor Comfort and Usability Evaluation, which considers the practical implications for long-term or real-world deployment. Model performance was assessed using a Random Forest classifier with 5-fold cross-validation, maintaining the temporal sequence of the data to ensure

Table 3. Performance Metrics for Sensor Combinations

Sensor	Accuracy (%)	MAE
ECG-PPG	72.09	0.77
PPG-EEG	72.87	0.80
ECG-PPG-EEG	74.11	0.68
ECG-EEG	74.18	0.66
EEG-Eye Tracker	75.85	0.73
PPG-Eye Tracker	76.23	0.74
ECG-EEG-Eye Tracker	76.48	0.61
PPG-EEG-Eye Tracker	77.40	0.68
All Sensors	78.74	0.61
ECG-Eye Tracker	79.16	0.52
ECG-PPG-Eye Tracker	79.67	0.62

a robust, generalizable evaluation of the model's predictive capabilities.

A. Individual Sensor Performance

The accuracy and Mean Absolute Error (MAE) for each sensor are outlined in Table 2. The findings indicate an order in the predictive capabilities of each sensor. While a Kruskal-Wallis test revealed no statistically significant main effect of sensor configuration on accuracy ($p = 0.991$) or MAE ($p = 0.813$), analysis of effect sizes (Cohen's d) revealed some differences. The EEG sensor appeared as the most effective single sensor, achieving the highest accuracy of 74.27% and the lowest mean absolute error (MAE) of 0.79. Effect size analysis confirmed its advantage, showing a large improvement over ECG (Accuracy $d = 1.07$; MAE $d = 0.40$) and a small-to-medium improvement over PPG (Accuracy $d = 0.24$; MAE $d = 0.18$). This suggests that the physiological data obtained from the EEG suit as direct and robust biomarkers of

mental fatigue. Additionally, these data showed greater resilience against interference from the physiological noise associated with physical fatigue compared to other sensors.

The Eye Tracker also showed excellent performance, achieving an accuracy of 71.54% and an MAE of 0.95. This indicates that pupil diameter variability can serve as a sensitive secondary indicator of cognitive load and fatigue. However, it is important to note that while the Eye Tracker's performance was robust, it did not exceed EEG in accuracy or MAE.

In contrast, the cardiovascular sensors (ECG and PPG) showed lower effectiveness, especially ECG. The relatively poor accuracy (56.29%) and higher MAE value (0.99) highlights a limitation where autonomic nervous system (ANS) activity, as measured by heart rate variability (HRV), is overwhelmed by the body's response to physical fatigue. This confound was explicitly pronounced in this study, as the physical fatigue induction method directly engaged the cardiovascular system. This strong cardiovascular noise might have masked the more subtle ANS variations that are related to purely mental fatigue. This result shows that cardiovascular data are insufficient as standalone measures for mental fatigue level estimation in some specific cases. The data is indistinguishably correlated to the physical fatigue, making it difficult to separate the two fatigue sources based on cardiovascular data alone.

B. Sensor Combination Performance

The combination of multiple sensors has resulted in an improvement of model performance, as shown in Table 3. Although the Kruskal-Wallis test was not significant, the effect sizes demonstrated a benefit of sensor combination. Although EEG independently proved to be the most powerful predictor, the model's performance was improved with sensor combination. Additionally, more effective outcomes can be achieved through combinations that exclude EEG, providing a practical alternative for real-world situations where wearing an EEG may be difficult.

The best performance achieved was with the combination of ECG, PPG, and Eye Tracker, achieving an accuracy of 79.67% and an MAE of 0.62. This was closely followed by the pairing of ECG and Eye Tracker, which achieved an accuracy of 79.16% and an MAE of 0.52. Notably, the effect size for the difference between these two combinations was negligible for both accuracy ($d = 0.03$) and MAE ($d = 0.16$). These results show that the model is capable of learning complex and non-linear relationships between the

Table 4. Average Comfort and Usability Ratings for Each Sensor

Sensor	Comfort	Interfere*	Intrusive*	Weight	Extended	Easy	Overall (mean \pm SD)
ECG	3.8	4.0	3.7	3.4	2.9	3.3	3.5 \pm 1.0
PPG	3.7	4.0	4.0	3.0	3.7	4.0	3.7 \pm 0.7
EEG	2.3	2.5	2.8	2.6	1.8	1.4	2.2 \pm 0.7
Eye Tracker	2.5	2.6	2.7	3.2	1.9	3.7	2.8 \pm 0.8

body's autonomic responses as measured by ECG and PPG, and the behavioral outputs captured by the eye tracker. This combined system produces a more robust and complete model of mental fatigue levels, providing better insights than just looking at brain activity alone.

While the combination of ECG, PPG, and eye tracking showed the best accuracy compared to other combinations, the results support the pairing of ECG and eye tracking as the most advantageous combination for practical implementation and long-term settings. This suggestion is based on the balance between how well they work and how easy they are to wear. The data shows that the performance difference between the two combinations is slight, with a difference in accuracy of 0.51%. This negligible difference, confirmed by effect size analysis, may not be meaningful in many contexts, but it may influence user convenience. The PPG sensor can be uncomfortable and may affect the user experience. It can interfere with small hand tasks and cause discomfort when used for long periods. Therefore, for any scenario that prioritizes long-term uses, convenience, and minimal disturbance to daily activities, the combination of ECG and Eye Tracker appears as the optimal suggestion. This pairing not only provides a robust estimation of mental fatigue with an accuracy of 79.16%, but it also greatly enhances the possibility that users will frequently use the system. The small compromise in absolute accuracy is a beneficial trade-off for considerably enhanced practicality and user experience, placing it as a tolerable solution for real-world applications.

The PPG sensor can also be uncomfortable and may affect the user experience. It can interfere with small hand tasks and cause discomfort when used for long periods. Therefore, for any scenario that prioritizes long-term uses, convenience, and minimal disturbance to daily activities, the combination of ECG and Eye Tracker appears as the optimal suggestion. This pairing not only provides a robust estimation of mental fatigue with an accuracy of 79.16%, but it also greatly

enhances the possibility that users will frequently use the system. The small compromise in absolute accuracy is a beneficial trade-off for considerably enhanced practicality and user experience, placing it as a tolerable solution for real-world applications.

C. Sensor Comfort and Usability Evaluation

As described in the earlier section, a post-experiment questionnaire was administered to evaluate the comfort and usability of each sensor. The average scores for each point and the overall comfort score (mean of all six points) are presented in Table 4. Higher scores indicate better comfort/usability. See Table 1 for a detailed overview of the questionnaire items and definitions.

The overall comfort scores were highest for PPG (3.7) and ECG (3.5), followed by Eye Tracker (2.8) and EEG (2.2). Participants noted that the ECG chest strap was generally comfortable, though some found the weight noticeable. The PPG ring was considered comfortable and easy to wear, but it interfered with hand movements during tasks. The EEG headset was rated least comfortable, with frequent comments about its intrusiveness, weight, and reluctance to wear for extended periods. The Eye Tracker, while relatively easy to wear (3.7), received low scores for comfort (2.5) and extended wear (1.9).

V. DISCUSSION

A. Sensor Performance and Combination Benefits

This study evaluated the performance of various sensors both in isolation and in combination. While the Kruskal-Wallis test did not yield significant p-values, the patterns observed in the effect sizes (Cohen's d) for both accuracy and MAE provide evidence for the reported performance and the benefits of sensor combination.

EEG has proved itself as the gold standard, showing the highest performance among single sensors. Its superiority derives from its ability to directly capture neural oscillatory changes, which serve as some of the most specific biomarkers for mental fatigue [20]. Our findings show that these neural signatures remain detectable, even when masked by the physiological noise associated with physical fatigue. This was evidenced by its superior mean accuracy (74.27%) and MAE (0.79), supported by medium to large effect sizes over other single sensors.

The robust standalone performance of eye-tracking is quite noteworthy. While physical fatigue may influence pupil response, pupillometry has demonstrated itself to be a reliable method for assessing mental fatigue. This indicates that pupillometry maintains its effectiveness, even in complex situations where fatigue states overlap.

In contrast, cardiovascular measures (i.e., ECG and PPG) were less effective on their own, mostly because their signals were greatly overwhelmed by the body's response to physical fatigue. This confounding factor greatly limits their usefulness as standalone indicators for purely mental fatigue in active scenarios. This was reflected in the very large effect sizes when comparing ECG to most other sensors.

However, there was a considerable improvement in accuracy achieved through the combination of sensors. The effect size analysis confirms this, showing, for example, a massive improvement ($d > 1.3$ for accuracy, $d < 0.6$ for MAE) when moving from ECG alone to the top combinations. This suggests that, despite its noise, cardiovascular data offers valuable complementary information. This allows the combined model to overcome the limitations of each sensor, creating a more robust and accurate assessment system than a single sensor can achieve.

When considering practical implementation, our results support the pairing of ECG and eye tracking as the most advantageous combination. Although the combination of ECG, PPG, and eye tracking achieved the highest accuracy (79.67%), the marginal improvement over the ECG-eye tracker pair (79.16%) is negligible (a difference of 0.51%, Cohen's $d = 0.03$ for accuracy, $d = 0.16$ for MAE) when weighed against the practical drawbacks of the PPG sensor, which can cause discomfort during extended wear and may hinder manual tasks. Therefore, in scenarios that prioritize long-term use, convenience, and minimal disturbance to daily activities, the combination of ECG and Eye Tracker emerges as the optimal choice. This pairing offers a reliable estimation of mental fatigue levels

while increasing the likelihood of user adherence. The small compromise in absolute accuracy is a beneficial trade-off for considerably improved practicality and user experience, making this combination an effective solution for real-world applications.

B. MAE Value

The notably low mean absolute error (MAE) value of 0.52 obtained from the pairing of ECG and eye tracker is arguably as important as the high accuracy achieved by this pairing. In practical and real-world applications, especially those in high-demanding environments, it is not only important to achieve high accuracy, but also to reduce the degree of error in estimating mental fatigue levels. A system that consistently misclassifies a "highly fatigued" state (e.g., level 8) as "moderately fatigued" (e.g., level 6) is far more helpful and safer than one that sometimes misclassifies it as "low fatigued" (e.g., level 2). The MAE shows that our model's estimations are, on average, very close to the actual subjective ratings. This makes it notably well-suited for reliable fatigue assessment in critical settings.

C. Sensor Comfort

The post-experiment comfort and usability evaluation revealed practical considerations for real-world deployment. Although the EEG sensor was the most accurate, its low comfort score (2.2) means people may not want to use it for long periods. Conversely, the combination of electrocardiogram (ECG) and eye-tracking shows a reasonable compromise in terms of wearability. The ECG sensor attained a relatively high comfort score (3.5), and the eye tracker, despite its lower overall comfort score (2.8), was rated as easy to wear (3.7). However, the low willingness to wear the eye tracker for extended periods (1.9) indicates that other combinations can be considered for long-term use. These findings reinforce the recommendation to pair the ECG with the eye tracker, as this configuration offers an optimal balance between technical performance and user acceptability.

It should be noted that the comfort and usability ratings presented in this paper are specific to the sensor models and wearing configurations used in this study (MyBeat WHS-3 ECG, Ring O₂ PPG, OpenBCI EEG with saltwater electrodes, and Tobii Pro Glasses 3 eye tracker). Different models or wearing positions may yield different comfort profiles, though the relative trends observed here are expected to inform future wearable systems.

D. Limitation

Several limitations need to be acknowledged in this study. First, although the experimental setup was intended to replicate real-world conditions, it may not fully capture their complexity. In addition, the homogeneity of the participant sample (only young adult males) limits the generalizability of the findings to broader populations, including females and individuals from other age groups. A further limitation concerns participants with visual impairments. The eye tracker utilized in this study did not address these issues, which may have influenced the progression of mental fatigue throughout the experimental procedure. Moreover, while we have outlined the subjective fatigue ratings, there may still be subjectivity in the ratings used as the ground truth in this study. Finally, although environmental factors such as lighting and screen brightness were strictly controlled to minimize confounding effects on pupillometry, the specific cognitive task used in this study may influence eye-tracking metrics. This task involves visual conflict, response inhibition, and sustained gaze fixation, and is likely to elicit pupil and gaze responses that are particularly sensitive to mental fatigue. If a different cognitive task were employed, the eye tracker's performance in estimating mental fatigue level might differ. Consequently, the reported eye-tracking performance should be interpreted in the context of the specific experimental task.

VI. CONCLUSION AND FUTURE WORKS

A. Conclusion

In conclusion, this study shows that although wearables can moderately estimate mental fatigue levels in the context of physical fatigue using a single sensor, a multi-sensor approach is necessary to achieve higher accuracy and reliability. By combining ECG data with eye tracking, a model can be developed that achieves an accuracy of 79.16% and a mean absolute error of only 0.52 on a 10-point scale. This study shows that the combination of ECG and eye-tracker offers a practical and accurate method for assessing mental fatigue during physical fatigue. This leads to the development of reliable fatigue monitoring systems in critical domains, such as occupational safety and athletic training. This study provides an empirical basis for designing usable fatigue monitoring systems capable of handling the complex interactions between mental and physical fatigue.

B. Future Works

Future work should test the sensor combinations on other cognitive tasks (e.g., working memory, sustained attention). This would indicate whether the findings from this study are applicable not just to the Stroop task, and help determine whether the sensors, especially the eye tracker, perform well across different situations.

ACKNOWLEDGEMENTS

This work was supported in part by JST Moonshot R&D Program Grant Number JPMJMS239F.

REFERENCES

- [1] Wang, M. Ding, and B. M. Kluger: Change in Intraindividual Variability Over Time as a Key Metric for Defining Performance-based Cognitive Fatigability, *Brain and Cognition*, Vol. 85, pp. 251-258 (Mar. 2014).
- [2] T.G. Monteiro, G. Li, C. Skourup, and H. Zhang: Investigating an Integrated Sensor Fusion System for Mental Fatigue Assessment for Demanding Maritime Operations, *Sensors*, Vol 20, No. 9, pp. 1-18 (May 2020).
- [3] D. Dasari, G. Shou, and L. Ding: Investigation of Independent Components based EEG Metrics for Mental Fatigue in Simulated ATC Task, *Proc. of the 6th International IEEE/EMBS Conference on Neural Engineering*, pp. 1331-1334 (Nov. 2013).
- [4] E. Muñoz-de-Escalona, J. J. Cañas, and P. Noriega: Inconsistencies between Mental Fatigue Measures under Compensatory Control Theories, *Psicológica*, Vol. 41, No. 2, pp. 103-126 (July 2020).
- [5] S. Huang, J. Li, P. Zhang, and W. Zhang: Detection of Mental Fatigue State with Wearable ECG Devices, *Medical Informatics*, Vol. 119, pp. 39-46 (Nov. 2018).
- [6] M.M. Bundeale and R. Banerjee: Detection of Fatigue of Vehicular Driver using Skin Conductance and Oximetry Pulse: a Neural Network Approach, *Proc. of the 11th International Conference on Information Integration and Web-based Applications & Services*, pp. 739-744 (Dec. 2009).
- [7] L.J. Trejo, K. Kubitz, R. Rosipal, R.L. Kochavi, and L.D. Montgomery: EEG-Based Estimation and Classification of Mental Fatigue, *Psychology*, Vol. 6, No. 5, pp. 572-589 (April 2015).

- [8] L. A. Torres-Salomao, M. Mahfouf, and E. El-Samahy: Pupil Diameter Size Marker for Incremental Mental Stress Detection, *Proc. Of 17th International Conference on E-health Networking, Application & Services*, pp.286-291 (Oct. 2015).
- [9] D.A. Crawford, K. M. Heinrich, N.B. Drake, J. DeBiau, and M.J. Carper: Heart Rate Variability Mediates Motivation and Fatigue throughout a High-Intensity Exercise Program, *Applied Physiology, Nutrition, and Metabolism*, Vol. 45, No. 2, pp. 193-202 (Feb. 2020).
- [10] C. Zhao, M. Zhao, J. Liu, and C. Zheng: Electroencephalogram and Electrocardiograph Assessment of Mental Fatigue in Driving Simulator, *Accident Analysis & Prevention*, Vol. 45, pp. 83-90 (Mar. 2012).
- [11] H. Jin, M. Xiao, L. Liu, S. Kan, Y. Fu, and D. Zhang: Relationship between Physical Fatigue and Mental Fatigue based on Multimodal Measurement under Different Load Levels, *Ergonomics*, Vol. 67, No. 11, pp. 1748-1763 (Nov. 2024).
- [12] I. Wojtasz, S. Cofta, P. Czudaj, K. Jaracz, and R. Kaźmierski: Effect of Face Masks on Blood Saturation, Heart Rate, and Well-Being Indicators in Health Care Providers Working in Specialized COVID-19 Center, *Environmental Research and Public Health*, Vol. 19, No. 3, pp. 1-9 (Jan. 2022).
- [13] G. Moshtaghi-Kashanian, N. Moshtaghi-Kashanian, M.H. Nejad, N. Moshtaghi-Kashanian, and H.N. Oscuii: Reduction of Oxygen Saturation and Increase of Heart Rate in Hospital Workers Wearing Face Mask during Routine Shift. *Research in Applied and Basic Medical Sciences*, Vol. 7, No. 1, pp. 11-17 (Mar. 2021).
- [14] L.M. Barker and M.A. Nussbaum: Fatigue, Performance and the Work Environment: a Survey of Registered Nurses, *Advanced Nursing*, Vol. 67, No. 6, pp. 1370-1382 (Mar. 2011).
- [15] R. Xu, C. Zhang, F. He, X. Zhao, H. Qi, P. Zhou, L. Zhang, and D. Ming: How Physical Activities affect Mental Fatigue based on EEG Energy, Connectivity, and Complexity, *Frontiers in Neurology*, Vol. 9, No. 915, pp. 1-13 (Oct. 2018).
- [16] A. Setiawan, A.D. Wibawa, E.S. Pane, and M.H. Purnomo: EEG-based Mental Fatigue Detection using Cognitive Tests and RVM Classification, *Proc. of Artificial Intelligence and Information Technology*, pp. 180-185 (Mar. 2019).
- [17] J.F. Hopstaken, D.V.D. Linden, A.B. Bakker, and M.A.J. Kompier: The Window of My Eyes: Task Disengagement and Mental Fatigue Covary with Pupil Dynamics, *Biological Psychology*, Vol. 110, pp. 100-106 (Sep. 2015).
- [18] L. M. Stirling, V. V. Tscharner, J. R. Fletcher, and B. M. Nigg: Quantification of the Manifestations of Fatigue during Treadmill Running, *European Journal of Sport Science*, Vol. 12, No. 5, pp. 418-424 (Sep. 2012).
- [19] S. Ameli, F. Naghdy, D. Stirling, G. Naghdy, and M. Aghmesheh: Quantitative and Non-Invasive Measurement of Exercise-Induced Fatigue, *Sports Engineering and Technology*, Vol. 233, No. 1, pp. 34-45 (June 2018).
- [20] G. Bilgin, I. E. Hindistan, Y. G. Özkaya, E. Köklükaya, Ö. Polat, and Ö. H. Çolak: Determination of Fatigue Following Maximal Loaded Treadmill Exercise by Using Wavelet Packet Transform Analysis and MLPNN from MMG-EMG Data Combinations, *Medical Systems*, Vol. 39, No. 108, pp. 1-10 (Dec. 2015).
- [21] S. Mahtumovic, E. Y. Sprout, J. C. Fontaine, T. M. Buskirk, S. S. Galen, and M. H. Malek: Test Retest Reliability of the Electromyographic Fatigue Threshold for Cycle Ergometry, *Muscle & Nerve*, Vol. 53, No. 5, pp. 803-807 (Mar. 2016).
- [22] M. J. Briscoe, M. S. Forgach, E. Trifan, M. H. Malek: Validating the EMG_{FT} from a Single Incremental Cycling Test, *Sports Medicine*, Vol. 35, No. 7, pp. 566-570 (Nov. 2013).
- [23] A. C. Coelho, D. T. Cannon, R. Cao, J. Porszasz, R. Casaburi, M. M. Knorst, and H. B. Rossiter: Instantaneous Quantification of Skeletal Muscle Activation, Power Production, and Fatigue during Cycle Ergometry, *Applied Physiology*, Vol. 118, No. 5, pp. 646-654 (Mar. 2015).
- [24] A. B. H. Itagi, N. A. Patil, R. K. Kotian, S. K. Reddy, S. Abhyankar, and R. S. Parveen: Physical Exhaustion Induced Variations in Event-Related Potentials and Cognitive Task Performance in Young Adults, *Annals of Neurosciences*, Vol. 25, No. 4, pp. 299-304 (Nov. 2018).
- [25] D. Doyle-Baker, J. Temesi, M. E. Medysky, R. J. Holash, and G. Y. Millet: An Innovative Ergometer to Measure Neuromuscular Fatigue Immediately after Cycling, *Medicine & Science in Sports & Exercise*, Vol. 50, No. 2, pp. 375-387 (Feb. 2018).
- [26] G. N. Dimitrakopoulos, I. Kakkos, Z. Dai, H. Wang, K. Sgarbas, N. Thakor, A. Bezerianos, and Y. Sun: Functional Connectivity Analysis of Mental Fatigue Reveals Different Network Topological Alterations between Driving and

- Vigilance Tasks, *Transactions on Neural Systems and Rehabilitation Engineering*, Vol. 26, No. 4, pp. 740-749 (Apr. 2018).
- [27] K. L. Kowalski and A. D. Christie: Force Control and Motor Unit Firing Behaviour following Mental Fatigue in Young Female and Male Adults, *Frontiers in Integrative Neuroscience*, Vol. 14, No. 15, pp. 1-11 (Mar. 2020).
 - [28] J. F. Hopstaken, D. V. D. Linden, A. B. Bakker, and M. A. J. Kompier: A Multifaceted Investigation of the Link between Mental Fatigue and Task Disengagement, *Psychophysiology*, Vol. 52, No. 3, pp. 305-315 (Sept. 2014).
 - [29] A. Ishii, M. Tanaka, Y. Shigihara, E. Kanai, M. Funakura, and Y. Watanabe: Neural Effects of Prolonged Mental Fatigue: A Magnetoencephalography Study, *Brain Research*, Vol. 1529, pp. 105-112 (Sep. 2013).
 - [30] N. V. Janczewski, J. Wittmann, Arnd. Engln, M. Baumann, L. Krauß: A Meta-Analysis of the N-Back Task while Driving and Its Effects on Cognitive Workload, *Transportation Research Part F: Traffic Psychology and Behaviour*, Vol. 76, pp. 269-285 (Jan. 2021).
 - [31] M. Slimani, H. Znazen, N.L. Bragazzi, M.S. Zguira, and D. Tod: The Effect of Mental Fatigue on Cognitive and Aerobic Performance in Adolescent Active Endurance Athletes: Insights from a Randomized Counterbalanced, Cross-Over Trial, *Clinical Medicine*, Vol. 7, No. 12, pp. 1-10 (Dec. 2018).
 - [32] M.R. Smith, R. Chai, H.T. Nguyen, S.M. Marcora, and A.J. Coutts: Comparing the Effects of Three Cognitive Tasks on Indicators of Mental Fatigue, *Psychology*, Vol. 153, No. 8, pp. 759-783 (Jun. 2019).
 - [33] T.G. Monteiro, C. Skourup, and H. Zhang: Using EEG for Mental Fatigue Assessment: A Comprehensive Look Into the Current State of the Art, *Transactions on Human-Machine Systems*, Vol. 49, No. 6, pp. 599-610 (Dec. 2019).