2022 17th Annual System of Systems Engineering Conference (SOSE) | 978-1-6654-9623-0/22/$31.00 ©2022 IEEE | DOI: 10.1109/SOSE55472.2022.9812694

Diagnostic Human Fatigue Classification using Wearable Sensors for Intelligent Systems

Likhitha Nagahanumaiah
Rochester Institute of Technology
Rochester, New York
ln2047@rit.edu

Saurav Singh
Rochester Institute of Technology
Rochester, New York
ss3337@rit.edu

Jamison Heard
Rochester Institute of Technology
Rochester, New York
jrheee@rit.edu

***Abstract*—Adaptive intelligent systems seek to optimize team performance by adapting to a human teammate. Typically, these adaptations are based on factors known to impact human performance, such as adapting autonomy levels based on human workload. These systems have yet to focus on adapting based on human fatigue; however, fatigue can be just as detrimental to performance as workload. This paper presents a multi-modal approach to fatigue classification in order to provide an intelligent system with the necessary information to optimize team performance over long time frames. The results indicate that mental and physical fatigue can be classified accurately for two fatigue levels, but cross-fatigue classification is a more difficult problem.**

***Index Terms*—intelligent systems, fatigue, human-state assessment**

I. Introduction

Intelligent systems, such as adaptive automation [1] need to interact with humans for long periods of time, while adapting their interaction strategies. Human performance may be negatively impacted during this prolonged interaction period, due to fatigue effects [2], [3]. An intelligent system capable of understanding a human's fatigue level may perform more impactful adaptations, such as appropriating more task load when a person is fatigued or allocating a mental task to a physically fatigued human. This paper seeks to detect mental and physical fatigue using multi-modal sensor data in order to permit these impactful adaptation strategies.

Human fatigue can be categorized as physical or mental. Physical fatigue occurs during intense physical activity and negatively impacts muscle functions. Similarly, cognitive fatigue is the result of prolonged cognitive activity and can lead to attention problems. Both fatigue types can occur when humans are interacting with intelligent systems; however, most related work focuses on one or the other.

Fatigue has typically been measured using subjective approaches, where a questionnaire is completed after a task. However, these approaches are intrusive and have poor time resolution. Thus, there has been a shift in the community to focus on physiological metrics (i.e, respiration rate, electromyography (EMG), and heart rate). These metrics correlate to mental and/or physical fatigue and can be collected using wearable sensors. Such a collection scheme allows continuous monitoring of an individual's fatigue state while minimally impacting the primary task.

Physiological metrics are often fed into machine-learning pipelines which produce a fatigue classification (e.g., high or low fatigue). However, typical approaches focus solely on physical [4] or mental fatigue [5], [6]. This study enhances the current research works by building a machine-learning model pipeline that classifies both mental and physical fatigue. Data from a human-subjects experiment where both fatigue types are induced is used to train and validate the machine-learning pipeline.

This paper is organized as follows. Section II reviews the literature, while Section III details the human subjects experiment. Section IV presents the machine-learning approach where the results are presented in Section V. The results are discussed and concluded in Section VI.

II. Related Work

Prolonged work or sleep loss can lead to a gradual or sudden onset of fatigue [10]. These onsets impact a human's physiological state. However, most research on fatigue from a physiological perspective has focused on muscles [11]. The spectral modification (change in magnitude and phase) of the electromyography (EMG) signals shifts towards a lower frequency [12]. Heart-rate and respiration rate may also increase with physical fatigue [13], [14]. Heart-rate variability decreases with mental and physical fatigue [13]. Mental fatigue has typically been assessed using electroencephalography (EEG) measures. Specifically, the relative wavelet packet energy in the frequency band decreases, and wavelet packet entropy decreases [15].

Machine learning is typically used to map physiological signals to a fatigue level. Table I presents an overview of various algorithms, fatigue type, levels, metrics, and overall classification accuracy. Most work focuses on classifying physical or mental fatigue using classical machine-learning approaches, such as random forest or support-vector machines. The physical fatigue approaches rely on muscle or acceleration information, while mental fatigue primarily relies on brain activity measures. There tends to be a large range of classification accuracy, from 62% to 91%.

The reviewed works focus on physical or mental fatigue classification, but do not consider classifying both fatigue types. Additionally, the mental fatigue classification uses EEG data, which has low operator acceptability after hours of

978-1-6654-9623-0/22/$31.00 ©2022 IEEE

TABLE I
LITERATURE SUMMARY

Algorithm	Fatigue type	Fatigue Levels	Metrics	Accuracy
Random Forest	Physical fatigue [7]	2	IMU,EEG, EMG	0.88
Support Vector machine	Mental fatigue [8]	2	EEG	0.91
	Physical fatigue [7]	2	EMG, EEG, IMU	0.78
K-Nearest Neighbor(K-NN)	Mental fatigue [9]	3	EEG, PPG	0.62

use [16]. The presented algorithmic approach focuses on using non-intrusive wearable sensors (such as a chest-strap) to classify both mental and physical fatigue.

III. Human Subjects Experiment

The goal of this study is to create a fatigue detection model that detects both mental and physical fatigue by capturing physiological signals from wearable sensors. Two tasks were used to collect data corresponding to mental and physical fatigue: (i) Jigsaw puzzle-solving task, (ii) Pick and Place task. Each participant completed the tasks in a random order in order to counterbalance the fatigue conditions.

The experiment started with collecting demographic information and a baseline session, where physiological signals are collected from the participants for 5 minutes while sitting still. A five-minute break then occurred before completing the physical or mental fatigue task. Each physical and mental task is conducted for an hour, and followed by a baseline collection and a 5-minute break. Each task was designed in such a way that each round consists of the same task demand and one round is approximately 60 seconds. A consistent task demand level allows for fatigue effects to arise without being confounded by varying task demand levels. Participants rated their fatigue levels and workload on a Likert-scale from 1 (little to no) to 5 (high) after each task round.

Participants were fitted with a Zephyr BioHarness and Myo armband in order to collect the physiological signals during the experiment. The BioHarness is a chest-strap device that collects preprocessed heart rate, respiration rate, heartbeat interval, and posture measures at 1 Hz. The Myo armband was fitted on the participant's forearms in order to collect 8-channel EMG and IMU data at a rate of 200 Hz and 50 Hz, respectively.

A. Pick and Place task

Two 7.5 lbs adjustable dumbbells (shown in Figure 1) must be carried around a 10-meter u-shaped indoor track. The subject performs a curl with each arm at the beginning of the lap and walks to the end of the u-shape track and back. The weights are then placed on a standard height table. The subject then walks the track again without the weights, after which the lap is restarted with picking up the weights. One lap with weights and one lap without weights is considered to be one complete round, and at the end of each round, the time taken to walk with weights and without weights are recorded using timestamps from the ultrasonic sensors (teensy board) mounted on the wooden box. The task is completed for an entire hour in order to elicit physical fatigue.

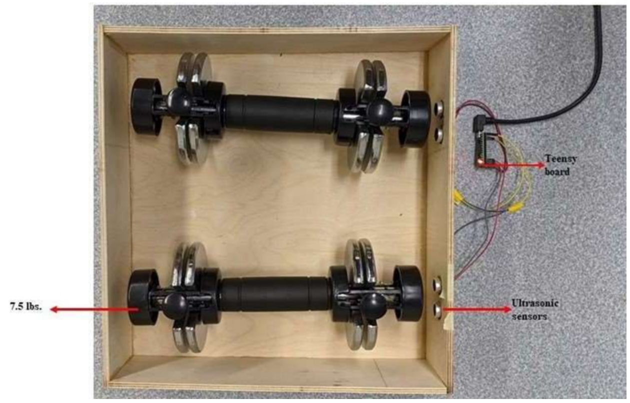

Fig. 1. Weights and Teensy board layout.

B. Puzzle-solving task

The subjects are given a jigsaw mind game puzzle (provided in Figure 2) which can be designed with varying difficulty levels, but for this experiment, we use 16 pieces(4x4) of a jigsaw puzzle to match the physical task demand of 60 seconds. The game is designed using the Unity game engine, where each movement of each puzzle piece is timestamped and is recorded as a correct and wrong move, game start, restart, game duration, and game exit. The subject completes the same puzzle repetitively for an hour, which elicits the desired mental fatigue levels.

Fig. 2. The Jigsaw Puzzle for the Mental Fatigue Task.

C. Participant Information

22 healthy subjects completed the IRB-approved study. The average age of the participants was 22.27 years. There were 11 undergraduates, 9 Masters students, and 2 PhD students. Participants were excluded from the study if they were unable

to lift the weights, walk for an hour, were pregnant, or had a nickle sensitivity.

IV. Methodology

The collected physiological information is windowed for 1-minute with a 10-second stride. This segmented data is then preprocessed to reduce noise and to support feature extraction. The extracted features are then used in a machine-learning algorithm that classifies mental or physical fatigue as low or high.

A. Data Preprocessing

The segmented physiological data is preprocessed in order to fill in missing information and reduce noise. The raw EMG signals obtained from the Myo armband are sampled at the rate of 200 Hz. These signal are passed through a notch-filter at 60 Hz in order to reduce power-line noise. No filter is applied to the BioHarness data, since this data is preprocessed onboard the device. Any missing samples from the Myo or BioHarness are imputed with linear regression techniques before features are extracted.

B. Feature Extraction and Reduction

The mean, standard deviation, variance, gradient, and slope is calculated for the processed heart-rate and respiration-rate. The mean is the arithmetic average of a set of given numbers.

$$Arithmeticmean = \frac{(x_1 + x_2 + ... + x_n)}{n} \tag{1}$$

The variance is the average of the squared differences from the mean.

$$\sigma^2 = \frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{n-1} \tag{2}$$

The term "gradient" refers to a graded difference in physiological activity along an axis. A line's slope is the ratio of how much y increases as x increases by some amount.

$$m = \frac{y_2 - y_1}{x_2 - x_1} \tag{3}$$

Heart-rate variability time, frequency, and non-linear domain features are extracted from the raw ECG signal that the BioHarness collects. However, only the mean HRV, standard deviation HRV, HRV very low frequency and very high frequency components, the ratio of the HRV low and high frequency, and the pNN 50 and 20 features were used in the classification process, due to their high correlations with the fatigue levels.

The median frequency of each Myo EMG channel is extracted. The median frequency f_m of a power spectrum $P(f)$ is defined as the frequency satisfying the following equation [17]:

$$\int_0^{f_m} P(f)df = \int_{f_m}^{\infty} P(f)df = \frac{1}{2}\int_0^{\infty} P(f)df \tag{4}$$

The chosen features are known to correlate with physical and/or mental fatigue, where a total of 42 features are extracted. This number is reduced using Principal component analysis (PCA), in order to develop less complex models and help prevent overfitting.

PCA reduces dimensionality by finding projections that maximize variance (information) along an axis. This is done by computing eigen vectors and values of a dataset. Larger eigen values represent more information in the corresponding principle component; thus, the proportion of variance can be computed as a ranked ratio: $PoV = \frac{\lambda_1 + \lambda_2 + ... + \lambda_i}{\lambda_1 + \lambda_2 + ... + \lambda_N}$, where λ_i is the ith largest eigen value. This work found the PoV corresponding to 0.9 or explains 90% of the variance. The corresponding eigen vectors are then used to project the 42 features into a smaller dimensional space. The PCA algorithm found 20 principle components for physical fatigue and 22 principle components for mental fatigue.

C. Fatigue classification Models

The machine learning models considered for this study are: Random Forest(RF), and Support Vector Machine (SVM). Each model was trained to classify either physical or mental fatigue as low (0) or high (1). The low labels were from the first 5-minute resting baseline collection and the high labels were from the last 5-minutes from the mental or physical fatigue conditions. The intuition is that a person will not be fatigued when they first come into the experiment, but will be fatigued at the end of each fatigue condition. This formulation also produces a balanced training/testing paradigm.

1) Random Forest: A random forest algorithm is composed of multiple decision trees, where each tree predicts a class using simple decision rules (e.g., if heart-rate is greater than 80, classify high fatigue) and no two trees are the same. A final prediction is the max vote of all of the decision trees in the random forest.

A grid search was performed to find the fewest number of trees with the least amount of depth on the training dataset for one LOSO CV fold. The number of trees varied from 10 to 500 in steps of 10, while max depth varied from 5 to 25 (greater than than the max number of components). It was found that 50 trees with a max depth of 25 produced the best results without diminishing returns (increasing complexity without increasing performance).

2) Support Vector Machine: Support Vector Machines find a hyperplane that linearly separates the dataset with some slack (cost). This hyperplane is determined by support vectors that maximize the margin between the hyperplane and the edge of the examples. Similar to the random forest, a grid search was performed by changing the cost values, gamma (influence of a sample based on distance), and kernel (radial basis function or polynomial). It was found that a radial basis function with a cost value of 10 and gamma value of 0.001 produced the best results.

V. Results

Two validation paradigms are performed: Leave-One-Subject-Out Cross validation (LOSO CV) and Cross-Fatigue validation (CFV). LOSO CV trains a model on all but one participant and validates on the remaining participant. This

TABLE II
LOSO CV RESULTS BY MODEL TYPE AND FATIGUE TYPE.

Model	Fatigue	Precision	Recall	F-1 score	Support	Accuracy(%)
Random Forest	Mental	0.79	0.80	0.79	535	80.2
	Physical	0.94	0.94	0.95	347	94.5
Support Vector Machine	Mental	0.84	0.85	0.84	535	85.0
	Physical	0.96	0.96	0.96	347	96.5

TABLE III
CROSS-FATIGUE CLASSIFICATION RESULTS BY MODEL TYPE AND FATIGUE TYPE.

Model	Fatigue	Precision	Recall	F-1 score	Support	Accuracy(%)
Random Forest	Mental	1.00	0.186	0.31	403	18.6
	Physical	1.00	0.63	0.77	192	64.5
Support Vector Machine	Mental	1.00	0.10	0.18	403	10.1
	Physical	1.00	0.33	0.49	192	32.8

approach is repeated such that each participant is tested on once. The LOSO CV results provide insight into a model's population generalizability (how well it performs on an previously unseen human).

The CFV paradigm examines how a model performs classifying any fatigue type by training on data corresponding to a baseline and one fatigue condition and testing on the remaining condition. For example, a model may be trained on the baseline condition and the last 5-minutes of the mental fatigue condition. The model is then evaluated on the last 5-minutes of the physical fatigue condition.

Sensor failures required removing data that was either NaN or where below theoretical thresholds (i.e., Heart-rate was determined to be 0). This produced a smaller dataset than expected, primarily due to either ECG signal noise or one of the Myo (EMG) devices stopped sending samples. However, there are sufficient samples to determine the accuracy of the fatigue models.

A. *Leave-One-Subject-Out Cross-Validation*

LOSO CV permits understanding how a model may perform on a previously unseen human. The corresponding results are presented in Table II. Overall, the support vector machine achieved higher performance than the random forest for both fatigue types. However, each model achieved $> 80\%$ accuracy, which is in line with the state-of-the-art results in Table I. Physical fatigue was classified better than mental fatigue, which may be attributed to the inclusion of the EMG signals.

The LOSO CV was also performed for classifying overall fatigue, by combining the mental with the physical fatigue data. This paradigm creates a large class imbalance; thus, undersampling was performed to better balance the classes. The random forest model classified overall fatigue correctly 86% of the time, while the support vector machine classified overall fatigue correctly 90% of the time. These results indicate that the support vector machine achieves greater performance than the random forest classifier.

B. *Cross-Fatigue Validation*

Overall fatigue is comprised of mental and physical components. However, there may be an interdependence between the two; thus, it is useful to determine how well a model may perform on a different fatigue type. The corresponding results are presented in Table III, where the Fatigue column indicates the testing set (i.e., Mental means the model was trained on baseline/physical data and tested on the mental fatigue data). Overall. neither model achieved state-of-the-art results and had difficulty classifying a fatigue type it was not rained on. Classifying mental fatigue when trained on physical fatigue data produced the lowest accuracy, while classifying physical fatigue when trained on mental fatigue data produced higher accuracy. This result is attributed to physical fatigue impact a person's EMG and cardiac related signals more than mental fatigue impacts them.

Additional analysis was performed by removing the EMG signals and found that the random forest model decreased in accuracy when classifying Physical fatigue (51%) and that the support vector machine increased in accuracy (70%). A similar trend occurred two both models when trained on physical and tested on mental fatigue.

VI. DISCUSSION AND CONCLUSION

Two machine learning models were developed to classify mental and physical fatigue as high or low using data collected during a human-subjects experiment. Overall, both models were able to classify a previously unseen person's fatigue level (validated using Leave-One-Subject-Out Cross-Validation) using cardiac, respiration, and electromyography information. This result indicates that both fatigue components can be incorporated into an intelligent system of systems in order to help monitor a human's performance level. The system may be able to allocate a physically demanding task to a human if the human has been classified as mental fatigued (or vice-versa). Similarly, the system may invoke autonomy for the current task in order to help mitigate the human's fatigued state.

Physical fatigue was classified better than mental fatigue, which is attributed to relying on physiological measures that are more sensitive to physical fatigue. Typical state-of-the-art mental fatigue classifiers rely on EEG information; however, these sensors can be intrusive and have low operator acceptability [18]. This is especially true during long-duration tasks where fatigue effects are more likely to manifest.

Both machine-learning models were unable to classify fatigue cross components (e.g., trained on physical and tested on mental fatigue). This result indicates that there is a disassociation between the fatigue components meaning that a model trained to classify one fatigue component may not achieve high performance on a different component. This disassociation is attributed to how the fatigue components impact a human's physiological signals. For example, a person's resting heart-rate may be 60 BPM, mental fatigue heart-rate may be 85 BPM, and physical fatigue heart-rate may be 110 BPM. A fatigue classification model trained to classify mental fatigue may perform well on the physical fatigue data due to a larger increase in heart-rate from the the baseline (similar to the results in Table III). However, going from physical fatigue classification to mental fatigue classification is more challenging due to the lower margin from the baseline. This means that the current state-of-the-art classification algorithms may have poor domain transferability and separate models are needed for physical and mental fatigue classification.

The majority of the paper has focused on classifying mental or physical fatigue, but the models can classify overall fatigue as well. The LOSO CV results indicate that the developed models achieved state-of-the-art performance to classifying overall fatigue (both mental and physical fatigue). Although the classification performance was high, the models have poor diagnosticity [19] and cannot discern if a person is fatigued due to mental or physical demands. Diagnosticity is an important factor if an intelligent system of systems will be deployed in task domains comprised of both mental and physical demanding conditions.

The developed approach to fatigue classification does have some limitations. First, collecting more data will provide a more robust analysis of the algorithm's capabilities. Especially if the data is collected from more realistic scenarios, such as industrial human-robot collaboration settings. Second, data collected from the last 5-minutes of each fatigue condition were considered to be high fatigue, which was verified by the subjective ratings. However, the participants experience a rise in fatigue levels throughout the conditions. A more granular approach (e.g., 5 fatigue classes) may permit intelligent systems to understand if a person is likely to become highly fatigued in the near future and adapt in order to prevent a high fatigue state from ever occurring.

The developed human fatigue classification approach can be characterized as a System of Systems architecture. Each individual sensor (BioHarness and Myo) are measurement systems that provide necessary preprocessed data to the fatigue classification system. The classification system extracts relevant features from the preprocessed data and uses a machine learning model to predict fatigue levels. This system of systems architecture can then be incorporated into a larger system that uses the fatigue classifications in its decision making processes.

Overall, this paper presented a diagnostic approach to fatigue classification by developing machine-learning models that classify either mental or physical fatigue. Data from a human-subjects study was used to validate the models using two validation schemes. The results indicate that the developed models are able to classify mental and physical fatigue for an unseen participant. This result expands on the current literature, as most works have either focused on mental or physical and have not analyzed their algorithm's population generalizability. The developed models are a necessary step towards to intelligent systems of systems that adapt to a human's internal state in order to optimize team performance.

REFERENCES

[1] D. B. Kaber and M. R. Endsley, "The effects of level of automation and adaptive automation on human performance, situation awareness and workload in a dynamic control task," *Theoretical Issues in Ergonomics Science*, vol. 5, no. 2, pp. 113–153, 2004.

[2] C. D. Wickens, J. D. Lee, Y. Liu, and S. E. G. Becker, *An Introduction to Human Factors Engineering*, 2nd ed. Pearson Education, Inc., 2004.

[3] K. S. Pasupathy and L. M. Barker, "Impact of fatigue on performance in registered nurses: Data mining and implications for practice," *Journal for Healthcare Quality*, vol. 34, no. 5, pp. 22–30, 2012.

[4] L. S. Caldwell and J. M. Lyddan, "Serial isometric fatigue functions with variable intertrial intervals," *Journal of Motor Behavior*, vol. 3, no. 1, pp. 17–30, 1971, pMID: 23941345. [Online]. Available: https://doi.org/10.1080/00222895.1971.10734888

[5] V. J. Gawron, J. French, and D. Funke, "An overview of fatigue," *Stress, workload, and fatigue*, 2001.

[6] J. A. Stern, D. Boyer, and D. Schroeder, "Blink rate: a possible measure of fatigue," *Human factors*, vol. 36, no. 2, pp. 285–297, 1994.

[7] Z. S. Maman, Y.-J. Chen, A. Baghdadi, S. Lombardo, L. A. Cavuoto, and F. M. Megahed, "A data analytic framework for physical fatigue management using wearable sensors," *Expert Systems with Applications*, vol. 155, p. 113405, 2020.

[8] C. Zhang, C. Zheng, X. Yu, and Y. Ou, "Mental fatigue estimation based on multichannel linear descriptors and support vector machine with optimizing parameters," in *Third International Conference on Natural Computation (ICNC 2007)*, vol. 2, 2007, pp. 309–313.

[9] Y. Zheng, T. C. Wong, B. H. Leung, and C. C. Poon, "Unobtrusive and multimodal wearable sensing to quantify anxiety," *IEEE Sensors Journal*, vol. 16, no. 10, pp. 3689–3696, 2016.

[10] H. Van Dongen, G. Maislin, J. M. Mullington, and D. F. Dinges, "The cumulative cost of additional wakefulness: dose-response effects on neurobehavioral functions and sleep physiology from chronic sleep restriction and total sleep deprivation," *Sleep*, vol. 26, no. 2, pp. 117–126, 2003.

[11] J.-J. Wan, Z. Qin, P.-Y. Wang, Y. Sun, and X. Liu, "Muscle fatigue: general understanding and treatment," *Experimental & molecular medicine*, vol. 49, no. 10, pp. e384–e384, Oct 2017, 28983090[pmid]. [Online]. Available: https://pubmed.ncbi.nlm.nih.gov/28983090

[12] T. Sadoyama and H. Miyano, "Frequency analysis of surface emg to evaluation of muscle fatigue," *European Journal of Applied Physiology and Occupational Physiology*, vol. 47, no. 3, pp. 239–246, 1981.

[13] Y. Yue, D. Liu, S. Fu, and X. Zhou, "Heart rate and heart rate variability as classification features for mental fatigue using short-term ppg signals via smartphones instead of ecg recordings," in *2021 13th International Conference on Communication Software and Networks (ICCSN)*, 2021, pp. 370–376.

[14] M. Grassmann, E. Vlemincx, A. Von Leupoldt, J. M. Mittelstädt, and O. Van den Bergh, "Respiratory changes in response to cognitive load: a systematic review." *Neural plasticity*, 2016.

[15] C. Zhang and X. Yu, "Estimating mental fatigue based on electroencephalogram and heart rate variability," *Polish Journal of Medical Physics And Engineering*, vol. 16, no. 2, p. 67, 2010.

[16] B. Cain, "A review of mental workoad literature," Defence Research and Development Toronto, Technical Report RTO-TR-HFM-121-Part-II, 2007.

[17] A. Georgakis, L. K. Stergioulas, and G. Giakas, “Fatigue analysis of the surface emg signal in isometric constant force contractions using the averaged instantaneous frequency,” *IEEE transactions on biomedical engineering*, vol. 50, no. 2, pp. 262–265, 2003.
[18] M. Castor, *GARTEUR Handbook of Mental Workload Measurement*, ser. GARTEUR technical publications. Group for Aeronautical Research and Technology in Europe, 2003.
[19] J. Heard, C. E. Harriott, and J. A. Adams, “A survey of workload assessment algorithms,” *IEEE Transactions on Human-Machine Systems*, vol. 48, no. 5, pp. 434–451, 2018.