Advanced Learner Assistance System's (ALAS) Recent Results

Aguilar-Herrera, Aime Judith
Tecnológico de Monterrey
Monterrey, México
aimeaguilar@tec.mx

Delgado-Jiménez, Esther Aimeé
Tecnológico de Monterrey
Monterrey, México
A00827948@itesm.mx

Candela-Leal, Milton Osiel
Tecnológico de Monterrey
Monterrey, México
milton.candela@tec.mx

Olivas-Martínez, Gustavo
Tecnológico de Monterrey
Monterrey, México
gustavo.olivas@tec.mx

Álvarez-Espinosa, Gabriela Jazmín
Tecnológico de Monterrey
Monterrey, México
A00825719@itesm.mx

Ramírez-Moreno, Mauricio Adolfo
Tecnológico de Monterrey
Monterrey, México
mauricio.ramirezm@tec.mx

Lozoya-Santos, Jorge de Jesús
Tecnológico de Monterrey
Monterrey, México
jorge.lozoya@tec.mx

Ramírez-Mendoza, Ricardo Ambrocio
Tecnológico de Monterrey
Monterrey, México
ricardo.ramirez@tec.mx

2021 Machine Learning-Driven Digital Technologies for Educational Innovation Workshop | 978-1-6654-2763-0/21/$31.00 ©2021 IEEE | DOI: 10.1109/IEEECONF53024.2021.9733770

***Abstract*—This work presents a real-time biofeedback tool that employs wearables and the Internet of Things with educational applications to improve students' learning and retention. We aimed to create a web platform using the Internet of Things (IoT) and Machine Learning (ML) architecture to predict students' performance, analyze mental fatigue, and provide real-time quantitative biofeedback to identify the best learning modality. Thus, the main goal was to develop a system that allows students to learn and improve their projects. We integrated the analysis of real-time biometric signals, machine learning algorithms, and web services as we observed their behavior under different learning modalities, seeking to improve cognitive performance. For this, 23 volunteers filled out the ten-question Fatigue Assessment Scale questionnaire about mental fatigue, validated with the P300 waves acquired during auditory-oddball (AO) tests. Synchronized data acquisition was achieved using Enophones and an E4 wristband. To develop predictive models, we collected the biometric data and incorporated it into an ML algorithm to visualize students' performance in real time. The system can accommodate other wearable systems with new features in further experiments. Thus, we believe this current development has the potential to further revolutionize traditional teaching with this methodology and future enhancements.**

***Index Terms*—EEG, cognitive Performance, education, neuroengineering, machine learning, higher education, educational innovation.**

I. Introduction

Education is a critical component of modern society, but traditional educational services have been lacked certain aspects, such as the development of systematized and non-personalized study plans [1]. Also, it has been one of the most altered services due to the COVID-19 pandemic [2]. Education needs to adapt constantly, to deliver a modern and fulfilling learning experience. A variety of trustworthy evaluation and teaching methods need to be employed [3], [4], [5].

Currently, educational institutions are transforming to adapt to actual social demands through innovative technologies so that learning processes are more efficient and attractive for students, teachers, and society [6].

Obtaining biometric data from students provides the opportunity to address learning-related problems, through stress and anxiety monitoring, optimizing the educational environment and allowing data-based decision making [6].

Much research has been conducted using single biometric devices to gather relevant cognitive information and validate its effectiveness in predicting stress, mental fatigue, and attention to a given task [7]. Their features might be used in an interactive IoT environment to support constant monitoring, and improve students' learning experience.

Research on biometric wearables show their reliability: For example, the Empatica E4 wristband detects convulsive seizures [3]. Regarding cognition, electroencephalography (EEG) signals are used to evaluate the student's level of meditation [5] and attention [8], further combined with an audio signal to increase the level of attention [9]. Thermal infrared imaging can also obtain students' levels of attention within a smart classroom [10], and electrocardiography (ECG) measures such as Heart Rate Variability (HRV) can identify stress conditions [7], which perform best when combined with electromyography (EMG) features [11]. These biometrics can then be used to manage learning-related issues such as assessment loads, which affect students' mental health, as negative emotions related to stress and anxiety deteriorate learning [6]. Monitoring biometrics in the classroom can create new learning approaches, as direct feedback to students with appropriate teaching methods improves learning and retention [1].

The validation of biometrics and cognitive features has been well established, with some studies even exploring their relationship to students' performance. Although there is scarce research about using IoT with multiple interconnected biometric devices in real-time, it does allow creating interactive environments in education to foster new learning methods, thus, leveraging biometrics to improve education [6].

This was the purpose of the current project: to develop a web platform for biometric data obtained through wearable devices and integrate the data in a real university setting better than other studies [12]. We developed a real-time performance report that can be used by the teacher or the students to assess the current teaching materials and methods. By receiving constant biofeedback from each student, the teacher would improve the quality of classroom instruction.

Only EEG and ECG-related signals were used in the current study, as these are related to mental fatigue measurements. However, the system can add other biometric systems with new features of interest to the platform.

II. Methods

A. Data

Twenty-three healthy undergraduate students (mean age: 19.8 ± 2 years; gender: M = 14, F = 10) volunteered for this work. One student took painkillers for muscle pain and was removed from the study. The data integrity from the resulting 22 students was verified using EEG data as the primary data type. Fifteen students had their complete EEG data recorded and were not removed from subsequent analyses.

Before starting the experiments, we provided each participant with a consent form. The form contained detailed information about the experimental procedures and the tasks to be conducted. The volunteers were asked to sign it if they agreed to continue the experiments and gave permission to be recorded in photos or videos.

Before taking each recording, the students' mental fatigue was assessed using the Fatigue Assessment Scale (FAS), with a total score ranging from 10 to 50 [1]. This score was obtained via a ten-question self-answered questionnaire (Spanish version). This numerical discrete target variable was encoded into one of the two ordinal categorical variables: 2-Class (No Fatigue, Substantial Fatigue) or 3-Class (No Fatigue, Moderate Fatigue, Extreme Fatigue).

B. Experimental Protocol

A rapid mental fatigue assessment in a resting state was used (reported in "Evaluation of a Fast Test Based on Biometric Signals to Assess Mental Fatigue at the Workplace - a Pilot Study," in which correlations of mental fatigue with spectral parameters of EEG and P300 were found) [7].It is essential to mention that although the P300 parameters were not used in this study, the same proposed experimental protocol was used.

The test was designed to collect biometric signals during an auditory-oddball (AO) task that elicits the P300 waveform, a positive deflection in the human event-related potential. It is most commonly elicited in an" oddball" paradigm when a subject detects an occasional" target" stimulus in a regular train of standard stimuli. The P300 wave only occurs if the subject is actively engaged in detecting the targets. It is important to realize that the amplitude varies with the improbability of the targets, and its latency varies with the difficulty of discriminating the target stimulus from the standard stimuli [13].

The test is composed of four phases: 30 seconds to stabilize the EEG sensors (as recommended by Eno) [14]), followed by 30 seconds measuring Eyes Closed (EC) and 30 seconds of Eyes Open (EO). Afterwards, there is a four-minute AO task. The task consists of two types of stimuli with different tones. Both tones were presented with different frequencies (baseline and oddball). The baseline tone was repeated more frequently than the oddball, in an 80:20 ratio. The system then evoked the P300 waveform when the oddball sound was presented. On the other hand, the EO and EC are used as calibration steps to normalize signals, taking data from each subject's signals.

C. Infrastructure

The created infrastructure is an integrative system that uses multi-processing in Python to gather data from the biometric devices used: the Empatica E4 wristband (Empatica) and Enophones (EEG). These data were plotted in real-time while gathering FAS score predictions every 10 seconds from data in those time windows. An example of the complete system in use is shown in Fig. 1.

Fig. 1: Example of a subject using the biometric devices: Empatica E4 wristband on her left hand and a pair of Enophones on her head, while real-time signals are gathered and plotted in a computer running the integrative system.

1) Enophones: Noise-cancelling headphones connected to a laptop with Windows via Bluetooth. The Media Access Control (MAC) address of the device was used to acquire, process, and clean its signal, based on *Brainflow*'s Python library. Thus, EEG band powers from 4-channel gold-plated dry electrodes could be extracted, following the 10-20 International System of Electrode Placement: $A1$, $A2$, $C3$, $C4$ [14].

After removing atypical data on pre-processing, the power spectral density (PSD) of each signal was calculated using the following band power ranges: Delta δ (1–4 Hz); Theta θ (4–7

Hz); Alpha α (8–12 Hz); Beta β (13–29 Hz); and Gamma γ (30–50 Hz).

2) Empatica E4: A smart band manufactured by the company Empatica. It provides a variety of signals for an accelerometer, Blood Volume Pulse (BVP), ElectroDermal Activity (EDA) and skin temperature from which ECG-related features can identify if a subject is under stress, fatigue, or in a relaxation period. The EDA and BVP could be cross-checked to assess stress and fatigue [15].

3) Python Programming language: We used Python programming language to acquire data, as it established a connection with the Lab Stream Layer (LSL) module. The LSL, in turn, can interact with the" E4 streaming server" app [15] available on Windows and Mac OS platforms. The E4 streaming server acquires the data from the smart band at a frequency of approximately 1 Hz; it is further processed using Python. As more data would be obtained from the Empatica E4 than the Enophones, we used various statistical functions for adequate granularity in a single data frame.

D. Feature Extraction

1) Data Pre-Processing: Min-max feature scaling was done using *MinMaxScaler* class from the *scikit-learn* package in Python, to normalize data depending on both subject and feature using the values from the Eyes Open (EO) phase of calibration ($30s$). Data normalization is helpful to transform each subject's data into a common space; moreover, the scaler transforms values into a domain $1 < x < 0$. We needed to equations to perform the feature scaling, two equations would be needed: First, an X_{std} value was calculated as shown in Eq. 1.

$$X_{std} = \frac{X - X_{min}}{X_{max} - X_{min}} \tag{1}$$

Second, the max and min for each subject and feature were also calculated, to take each X_{std} value and transform it into an X_{scaled} value, calculated as shown in Eq. 2.

$$X_{scaled} = X_{std} \cdot (max - min) + min \tag{2}$$

A quantile method used Q_1, Q_3 and InterQuartile Range (IQR) to remove outliers. This was done to set a Lower Fence (LF) and Higher Fence (HF) pair of values for each feature F and subject S. LF was calculated as shown in Eq. 3, while HF was calculated as shown in Eq. 4.

$$LF_{F,S} = Q_{1,F,S} - 1.5 \cdot IQR_{F,S} \tag{3}$$

$$HF_{F,S} = Q_{3,F,S} + 1.5 \cdot IQR_{F,S} \tag{4}$$

Thus, each statistically accepted sample must be between LF and HF in the form of: $LF_{F,S} < x_{i,F,S} < HF_{F,S}$, according to each feature. Each value (Q_1, Q_3, IQR) was acquired during most of the calibration phase: Eyes Closed (EC), EO ($60s$).

2) Feature Generation & Selection: A set of combined features were generated using the combinations as shown in Table I. For each ith feature F, four main combined functions were created: Inverse, natural logarithm, multiplication with jth feature, and division with jth feature. Encoded names were useful to identify which features performed best in Sect. III.

TABLE I: Combined features based on four combined functions, using base EEG and Empatica features.

Encoded name	Combined function
F_i–I	$\frac{1}{F_i}$
F_i–L	$\ln \lvert F_i + 1 \rvert$
F_i–M–F_j	$F_i \cdot F_j$
F_i–D–F_j	$\frac{F_i}{F_j + \varepsilon}$

It is worth noting that a small number in the form of an epsilon ($\varepsilon = 1 \times 10^-6$) was used in the denominator of the fourth combined feature, as division by 0 causes errors. Additionally, a $+1$ was added in the natural logarithm of i-th feature, to avoid $\ln(0)$, which cause errors. Moreover, a separate list of *combinations* was used to keep track of the combinations of i-th with j-th features, so as not to generate duplicate features when multiplying the i-th feature with the j-th feature and vice versa (as they produce the same number).

The feature selection method used was Random Forest Regressor, a class from the *scikit-learn* package in Python, highlighting the Mean Decrease Impurity (MDI) in variance, via the Mean Squared Error (MSE), which shows the difference between real and predicted values. It was calculated as shown in Eq. 5.

$$MSE = \frac{1}{n} \sum_{i=1}^{n} (y_i - \hat{y}_i)^2 \tag{5}$$

Where y_i is the ith value of sample y, n is the number of samples, and $\hat{y}_i$ is the predicted value. The reduction of this criterion means a decrease in MSE error, so this criterion's computation is helpful to determine which feature has the most significant decrease in impurity.

3) Performance Metrics: A set of performance metrics was used to evaluate the model's performance. The metrics included classification (for the categorical target variables) and regression (for the numerical target variables). The first set of performance metrics corresponded to classification problems: Accuracy, Precision, Recall and F1-Score. Confusion matrices (See Table II) helped explain these performance metrics. The samples were pre-labeled with their real value and predicted value, resulting in four types of possible outcomes; these are usually arranged in a confusion matrix in which one axis represents the true value, and the other represents the predicted value. There are four types of errors: true positives (TP), true negatives (TN), false positives (FP), which are also known as false discoveries or type-I errors, and false negatives (FN), which are also known as missed discoveries or type-II errors.

TABLE II: Confusion Matrix that describes components depending on combinations of predicted and true labels.

		Predicted label: Positive	Predicted label: Negative
True label	Positive	True Positive (TP)	False Negative (FN)
	Negative	False Positive (FP)	True Negative (TN)

Our first classification metric is accuracy, which is the average correct classification percentage. This metric could be represented as the number of predicted labels that match the correct classification divided by the total number of samples. In this case, based on the confusion matrix, this would be represented as the sum of TP and TN divided by the total possibilities N, which is the sum of all classifications (see Eq. 6).

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{6}$$

Precision is a classification metric that measures the number of TP classifications per FP; it determines which of the positive predicted labels are actually positive by the true label. This metric was calculated as shown in Eq. 7.

$$Precision = \frac{TP}{TP + FP} \tag{7}$$

On the other hand, recall is quite similar to precision, although this metric measures the number of TP classifications with respect to FN, penalizing when the models miss predicting a positive label on a true labeled positive sample.

This metric is widely used in the medical field, as FNs are dangerous when diagnosing a deadly disease. It is usually preferred to have an FP label and do more testing to determine if it is TP or FP, rather than not diagnosing it at all. The metric was calculated as shown in Eq. 8.

$$Recall = \frac{TP}{TP + FN} \tag{8}$$

This last classification metric combines precision and recall into a single metric called the F1-Score, which takes the harmonic mean of these metrics. Thus, it is a more balanced approach when using precision and recall separately. The metric was calculated as shown in Eq. 9.

$$F1 = 2 \cdot \frac{precision \cdot recall}{precision + recall} \tag{9}$$

The coefficient of determination (R^2) is a widely used performance metric in regression metrics to measure how well the model predictions fit into the actual data. The metric's domain is $0 \leq R^2 \leq 1$, where $R^2 = 1$ means a perfect prediction. The calculation is shown in Eq. 10.

$$R^2 = \frac{SSR}{SST} \tag{10}$$

Above, SSR refers to the Sum of Squared Regression (SSR), which is also the variation explained by the model. SSR was calculated as shown in Eq. 11.

$$SSR = \sum_{i=1}^{n} (\hat{y}_i - \bar{y})^2 \tag{11}$$

On the other hand, SST refers to the Sum of Squared Total (SST), also known as the total variation in the data. SST was calculated as shown in Eq. 12.

$$SST = \sum_{i=1}^{n} (y_i - \bar{y})^2 \tag{12}$$

4) Predictive models: Multiple Linear Regression (MLR): This statistical technique uses source variables to predict a target variable. It models a linear relationship using coefficient β for each feature to a maximum p features, which would be applied for each sample i. MLR was modeled as shown in Eq. 13.

$$\hat{y}_i = \beta_0 + \beta_1 x_{i,p} + \beta_2 x_{i,p} + \ldots + \beta_p x_{i,p} \tag{13}$$

In the equation, $\hat{y}_i$ is the predicted target variable, which would be compared to each real observation y_i. Additionally, the β_0 coefficient is a constant term representing the y-intercept, or the base values that the predicted score would have if each feature were equal to 0.

5) Validation: To validate the model, we employed Leave-One-Out (LOO) validation, which uses all but one subject's data to train a Machine Learning (ML) model, and the removed subject's data is used for testing. So, considering 10 subjects, ten different 90:10 training-testing data splits were applied to obtain a predicted score for each subject. Also, a subject could have multiple samples, so the trained model would predict scores for each of the subject's samples and then obtain the mean of predictions. When outliers are present, the predicted score is not within the target variable's score; in this case, we used the median instead of the mean.

III. RESULTS & DISCUSSION

Based on the data from 15 students, we created two models: a) using EEG features (model A) and b) using both EEG and Empatica features (model B). The reason behind not creating a model based on Empatica features alone was due to instability and poor correlation with the target variable with respect to the EEG's features correlation. The data was first pre-processed, removing outliers via the quantile method described in Sect II-D1, using Eq. 3 and Eq. 4 with each subject's data in their calibration phases (EC and EO). After this outlier removal, the number of samples decreased significantly. Some subjects even had 0 samples and so were removed. This was due to a significant difference in EEG bandpowers or Empatica features during the test compared to the calibration phase. From model A, four subjects were removed, which resulted in 11 total subjects. From model B, five subjects were removed, which resulted in 10 total subjects. The remaining subjects' fatigue levels are shown in Table III. It can be seen that there is a balance among the various levels of fatigue. However, *Extreme Fatigue* has FAS values within the limits of *Moderate Fatigue*. This would indicate a performance drop in some classification metrics when evaluating a model's performance in 3-Class fatigue level prediction.

TABLE III: Subject's level of class depending on filtered data.

Device	2-Class Fatigue Level		3-Class Fatigue Level		
	No	*Substantial*	*No*	*Moderate*	*Extreme*
EEG	4	7	4	4	3
EEG & Empatica	4	6	4	3	3

After removing outliers, combined features were created (see Table I) considering four EEG channels and six spectral signals (Alpha, Beta, Gamma, Theta, Delta), resulting in 24 EEG base features. On the other hand, considering BVP, EDA, and 9 statistical functions (Mean, Standard Deviation, Median, Maximum, Minimum, Interquartile Range, Kurtosis, Skewness, Coefficient Variation), resulted in 18 Empatica features; however, we also included the mean temperature, so there were 19 Empatica features. The combinations of EEG features resulted in 810 total features for model A and $2,994$ for model B.

Afterwards, *MinMaxScaler* was used to re-scale features' values. Afterwards, the MDI criterion in Eq. 5 was used to evaluate each feature's importance according to the FAS score, a discrete numerical variable. Normalized feature importance for the best 10 features is shown in Table IV.

TABLE IV: Best 10 features according to normalized feature importance by MDI using random forest regressor.

Device	Feature	Feature Importance
EEG	Alpha_A2-D-Alpha_A1	0.319703
	Alpha_A2-D-Alpha_C4	0.153771
	Gamma_A1	0.113352
	Gamma_A1-I	0.056444
	Theta_A1-D-Theta_C3	0.044882
	Alpha_A1-D-Alpha_A2	0.039322
	Gamma_A1-L	0.030554
	Theta_A2-D-Theta_C3	0.017749
	Gamma_A1-D-Gamma_A2	0.016955
	Beta_A2-D-Beta_A1	0.013921
EEG & Empatica	Alpha_C3-D-EDA_Kurtosis	0.081134
	Beta_A1-D-Beta_A2	0.053186
	Beta_A2-D-Beta_A1	0.042182
	EDA_Minimum-M-Gamma_A1	0.034817
	Delta_A1-D-BVP_Mean	0.034447
	Alpha_A1-D-Delta_C4	0.034355
	Gamma_A1-M-EDA_Median	0.033587
	Alpha_A1-D-Alpha_A2	0.029386
	Delta_A1-M-BVP_CoefficientVariation	0.028752
	Gamma_A2-D-Temp_Mean	0.021396

Based on the best 10 features for each data, we created an MLR model, which predicted the FAS score and then encoded it into 2-Class and 3-Class fatigue levels. Results are displayed in Fig. 2, with four classification metrics (Precision, Recall, F1-Score, Accuracy) and one regression metric (R^2). As the predicted scores are equal regarding the number of classes to which the FAS score is encoded, we obtained a sole R^2 metric for each model.

Interesting insights could be drawn regarding the models' performance: 2-Class prediction had better results in classification metrics (91% accuracy for model A and 80% for model B), as compared to 3-Class prediction (64% accuracy for model A and 80% for model B). Additionally, only using EEG features performed better than using EEG & Empatica features, as both R^2 and classification metrics were best (especially in 2-Class prediction). This could be due to the difference in the number of initially-generated combined features, as model B had nearly three times more combined features than model A. Even though they both used 10 features in the last model, the MDI criterion could have missed essential features, as it was trying to evaluate each feature's performance.

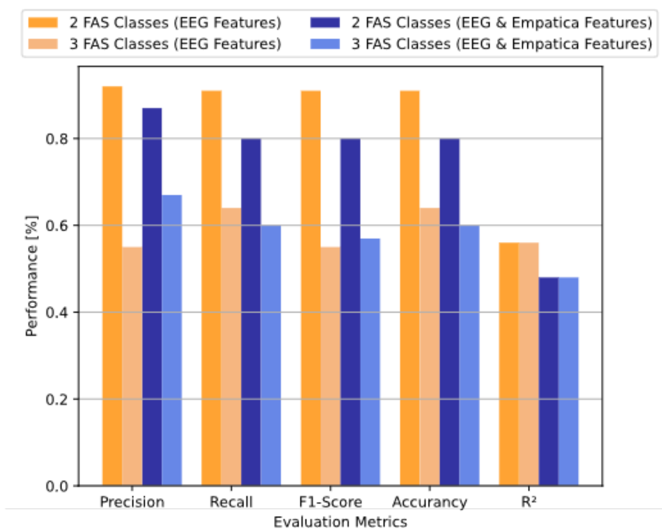
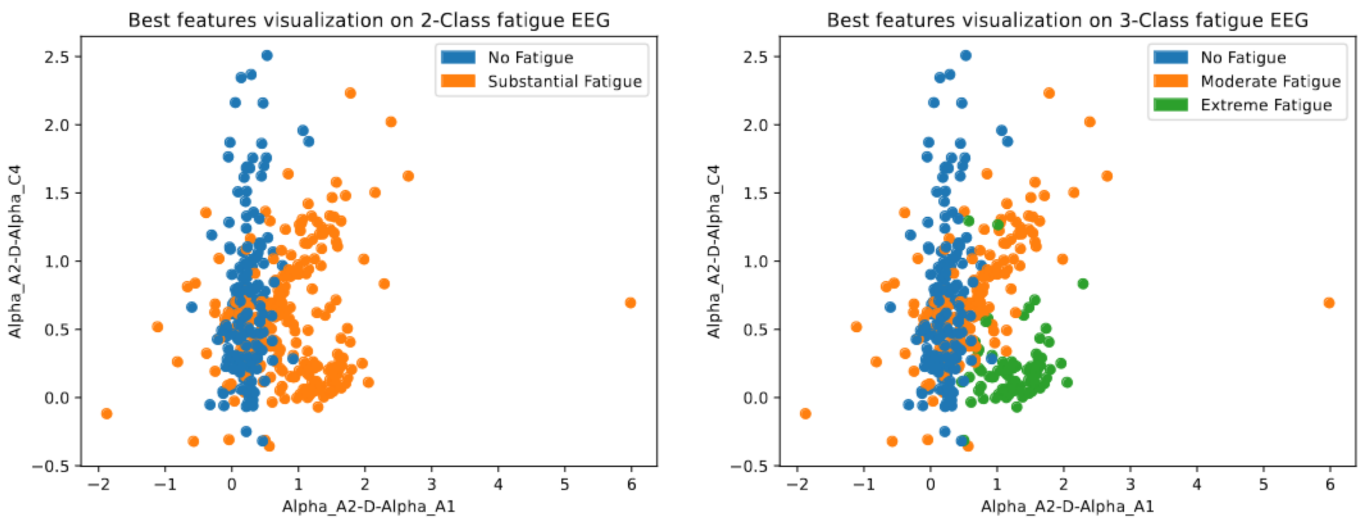


Fig. 2: Linear Regression Performance Metrics (weighted average).

Delving into the analysis of the best features, we can observe in Table IV that all of the features using EEG only features involved channels from the ear A (10): Alone (3), combined with central C channel bandpowers (3), and combined with other A channels (4). Also, γ_{A1} is repeated four times in the Table as base bandpower, inverse, natural logarithm, and combined with γ_{A2} through division. In addition, the latter (division) was the predominant combined feature (7), indicating that ratio bandpowers are the most related to predicting mental fatigue. On the other hand, regarding model B using EEG & Empatica features, the A channel bandpowers were also predominant (9). Empatica features also appear combined with EEG features (6): Multiplication (3) and division (3), with EDA being the most predominant feature (3),



(a) Best two features, using 2-Class level of FAS score.

(b) Best two features, using 3-Class level of FAS score.

Fig. 3: Best two features from model A with EEG only features, based on MDI normalized importance in Table IV. (a) 2-Class predictions with 91%, (b) 3-Class prediction with 64%, lower performance due to Extreme Fatigue samples in the border of being Moderate Fatigue, and so hard assignment clustering suffer in this regard.

followed by BVP (2) and Temperature (1).

To have a more visual approach to the results, we plotted the best 2 features on EEG only data in a 2D plot (see Fig. 3) with scaled features' values based on the *MinMaxScaler* for each subject. Where Fig. 3a represents encoding on 2-Class fatigue, and Fig. 3b has the 3-Class fatigue encoding. These features are involved in the best classification performance, which reported 91% accuracy in predicting 2-Class fatigue. The total number of samples using EEG features was 11, which means that the MLR trained algorithm could successfully predict the level of fatigue of 10 students.

The 2D scatter plot shows all valid samples from the 11 students. In Fig. 3a, there seems to be a clear division of data when using $\frac{\alpha_{A2}}{\alpha_{A1}}$ on the X axis, as $\frac{\alpha_{A2}}{\alpha_{A1}} > 1$ seems to be an area where *Moderate Fatigue* samples are more prevalent; otherwise, when $\frac{\alpha_{A2}}{\alpha_{A1}} < 1$, *No Fatigue* is more prevalent. On the other hand, in Fig. 3b, *Extreme Fatigue* seems to be in a cluster where $\frac{\alpha_{A2}}{\alpha_{A1}} > 1$ and $\frac{\alpha_{A2}}{\alpha_{C3}} < 0.5$, so the Y axis feature can differentiate *Extreme Fatigue* from *Moderate Fatigue*. Due to a clear trend visually differentiating the target variable when plotting these most essential features, one could implement a Support Vector Machine (SVM) with lesser features to reduce the model's complexity, as hyperplanes could be drawn to predict a subject's fatigue level. One limitation of this study is its sample size. However, there are reported studies that propose biometric tools with machine learning that have similar sample size tools to identify mental fatigue (8) [16], and drowsiness (10) [17]. Another consideration is that the study was conducted at a university during the COVID pandemic, so it was impossible to follow a large-scale approach. However, we are aware that increased sample size could translate to higher performance by the proposed model, which will be explored in future research.

IV. Conclusion

In summary, educational neuroscience has been growing as an area of research and practice over the last several decades [18]. Thus, our work focused on developing a system that observed biofeedback tools to assess the efficiency of different teaching techniques and developed biofeedback tools to assess different learning and teaching techniques. It is noteworthy that the proposed framework can enable "plug and play" applications for educational neuroscience [18].

Cognitive features, such as mental fatigue level, could be predicted using EEG bandpowers and Empatica measurement features such as BVP, EDA, or skin temperature. The best result was 91% accuracy using 2-Class FAS fatigue level and EEG-only features (model A), and the best two features were plotted; the best two features were plotted in 3-Class. LOO validation was the only type of validation implemented among the same subjects, with several $90{:}10$ splits to predict FAS scores, according to $n = 10$.

The validation could be further improved using cross-validations or a hold-out test dataset from new subjects' recordings unknown to the model. However, the current validation scheme provided valid results when predicting cognitive features using biometric data from EEG and Empatica features. Additionally, the proposed system can add other biometric devices in further experiments, thus, adding new features of interest into the predictive models, such as EMG features.

The proposed objective to provide real-time biofeedback was achieved (see Fig 1).

Our research methodology was concise and flexible. The system can be escalated and enhanced with additional biometrics and algorithms. The current COVID-19 pandemic has