SIN-CHN
SCIENTIFIC PRESS PTE.LTD.

Molecular & Cellular Biomechanics 2025, 22(3), 892.
https://doi.org/10.62617/mcb892

Article

Cognitive load detection in English language learning using wearable biosensors: A machine learning approach

Lili Qin, Weixuan Zhong*

College of Foreign language, Hechi University, Yizhou 546300, China
*** Corresponding author:** Weixuan Zhong, 15523230444@sohu.com

CITATION

Qin L, Zhong W. Cognitive load detection in English language learning using wearable biosensors: A machine learning approach. Molecular & Cellular Biomechanics. 2025; 22(3): 892. https://doi.org/10.62617/mcb892

ARTICLE INFO

Received: 23 November 2024
Accepted: 9 December 2024
Available online: 19 February 2025

Abstract: To increase the use of wearable biosensors in language learning environments, approaches for accurately extracting small signs of cognitive load are necessary. However, assessing subjective cognitive states, such as the load experienced during language acquisition, provides significant obstacles. This research uses data from physiological sensors worn on the wrist, such as skin conductance, skin temperature, heart rate, and R-R intervals, to organize a machine learning (ML) challenge to develop techniques for quantifying cognitive load in English learners. Participants used data from respondents who completed English language tasks of various difficulty levels. A robust evaluation of preprocessing approaches such as Z-score normalization, signal detrending, and moving average filtering, as well as feature extraction methods such as time-domain and frequency-domain analysis, demonstrated that robust models efficiently used biosensor data. Classical classifiers, such as Adaptive Random Forest (ARF), performed better when optimized with Barnacle Mating Optimization (BMO) for hyperparameter tuning. The proposed method of BMO-ARF has attained accuracy at 95.89%, F1-score in the cognitive load of low at 0.95, medium at 0.90 and high at 0.97, sensitivity in the cognitive load of low at 80.3%, medium at 88.5% and high at 93.0% and specificity in the cognitive load of low at 87.5%, medium at 91.8% and high at 95.1%. The results show that cognitive load classifications were more accurate for higher-difficulty tasks and particular learners, potentially impacted by model overfitting and the subjective nature of physiological responses. The research highlights the need for more sophisticated annotation techniques to improve cognitive load monitoring in language learning environments and handle student response variability.

Keywords: cognitive load assessment; biosensor technology; English learners; English language tasks; Adaptive Random Forest (ARF); Barnacle Mating Optimization (BMO)

1. Introduction

Cognitive load (CL) refers to the intellectual effort required to process and store information during learning or problem-solving tasks. It is stimulated by using the complexity of the task, the learner's prior knowledge, and the way information is presented. High cognitive load can avert learning with the aid of overwhelming working memory, even as low cognitive load can facilitate comprehension and retention. The concept is crucial in educational psychology, guiding instructional design to optimize learning efficiency by means of balancing the demands located on cognitive resources. CL has emerged as a critical concept for operational learning approaches in language acquisition and the learning process in general [1]. CL means the number of mental resources that are necessary to be spent for information processing and is important for learners in any kind of learning, including foreign language acquisition [2]. Expressed in the context of second language learning, cognitive load is the key factor that defines the extent to which a student can comprehend, remember, and use newly acquired information. The pre-eminently used techniques of evaluating CL, including questionnaires, self-reports, or performance data, have limitations connected to objectivity, reproducibility, and feasibility [3]. In regards to these constraints, the development of new biosensors, which could provide objective real-time, and non-intrusive CL monitoring, presents the most promising approach to mitigate these shortcomings. Biosensors are therefore instruments, which are intended to detect signals of biological origin that are related to feelings and thoughts [4]. Using biosensors, it is possible to monitor the amount of mental effort that learners are applying to different aspects of English: Reading, listening, developing vocabulary, and grammar. Biosensors in cognitive load measurement, which assesses the brain activity. The identification of the level of cognitive load was possible when learners were engaged in English learning tasks as detected by Electroencephalography (EEG). For instance, if a subject is in a state of high cognitive workload, then there is a relationship between the amount of theta waves and amount of mental effort and information processing [5]. Conversely, lower CL is associated with more alpha waves in the brain, which are in turn related to relaxation. This level of detail is useful to the researchers and educators using the framework because it provides additional insight into not only how much CL a learner is experiencing but when they can be approaching overload or under load conditions [6]. Besides, eye-tracking or sensors are also used to estimate CL, and they demonstrate a high correlation to CL values. The learner's movements of the eyes and the time spent on a particular area when reading English text or instructional aides [7]. Higher rate of fixations or longer fixation time on certain words or phrases can be an indication that a learner is having difficulty in understanding the text and, therefore, has a high cognitive load [8]. Heart Rate Variability (HRV) sensors, however, capture the time interval based on which the heartbeats repeatedly occur and which can help to figure out the reaction of the autonomic nervous system to cognitive stress [9]. Biosensor technology as a method can be regarded as a more accurate tool in gauging cognitive load in learners, English in particular, when compared to conventional means of evaluation [10]. Incorporating biosensors and the basic physiological signs when designing languages and educational technologies is expected to give useful information that could help future innovations in this kind of learning method [11]. Knowledge of the physical reactions linked with CL makes it possible for instructors to present the best researching environment, hence enhancing accelerated mastery of English by their learners and greater success in the course of researching for the English language [12]. The purpose is to examine the usability of the biosensor technology to evaluate the learners' CL during the learning of English language.

Contribution of this research

- The research contributes to language learning research by utilizing wearable biosensors to accurately quantify cognitive load, enabling better understanding of learners' mental states during task performance.

- It demonstrates the effectiveness of combining Adaptive Random Forest (ARF) with Barnacle Mating Optimization (BMO) for hyperparameter tuning, improving the accuracy of cognitive load classification.
- The findings highlight the need for more sophisticated annotation methods to address the subjective nature of physiological responses and enhance the robustness of cognitive load monitoring systems.

The rest of the paper is articulated as follows: Phrase 2 highlights the literature review, Phrase 3 brings out the methods that were used in this research, and Phrase 4 presents the performance analysis and discussion, Phrase 5 provides the conclusion of this research.

2. Literature review

The vocabulary instruction suggested by Zai [13], using bio-sensing technologies with biosensors can detect knowledge and comprehension gaps in pupils by tracking physiological signals. The findings demonstrate that incorporating bio-sensing technology into vocabulary instruction in Spanish can greatly enhance students' interest and learning outcomes. With its potential to transform conventional teaching methods and enhance educational outcomes in Spanish language training, the research presented viewpoints and ideas for using bio-sensing technology in language instruction. Lee [14] provided individualized help by employing biosensor-based feedback to alleviate the presentation anxiety that English language learner's encounter. By categorizing the severity and traits of learners' anxiety related to public speaking and foreign languages, the method enabled teachers to provide individualized emotional and educational support. Two more studies were suggested to assess the prototype and determine the emotional and instructional support needs of instructors, which helps to create guidelines for the use of biosensor-based feedback in the classroom. Hijazi et al. [15] recommended an inventive technique for locating the spatiotemporal causes of understanding issues in digital content was called Intelligent Biofeedback Augmented Content understanding (Tell Back). It evaluated cognitive stress using non-intrusive biofeedback devices and employed SVM (Support Vector Machine) to achieve 83.00%. Hijazi et al. [16] aimed to measure how well individuals understand and engage with code, providing insights into the effectiveness of the learning process and the quality of the code itself.

To gauge student participation in a sizable college lecture hall, skin biosensors, most especially GSR (galvanic skin response), were employed. When compared to the standard lecture portion, the active-learning portion demonstrated more engagement and self-reported benefits. During group projects, GSR activity rose, while during listening exercises it fell. Although there were barriers to widespread adoption, McNeal et al. [17] indicated that GSR is a promising instrument for gauging student participation in college courses. A range of behavioral in the performances of subjective and physiological metrics was to calculate CL during online learning. Predicting both intrinsic content difficulty and subjective CL and difficulty was the main goal. According to a 21-person research suggested by Herbig et al. [18], an intrinsic content difficulty was more effective for quizzes. The best results were obtained from features based on the eyes, followed by measures based on the heart and skin. Performance was enhanced by combining several modalities. A novel method was employed for predicting comprehension issues in digital information by utilizing eye-tracker data and wearable technology. 'Hotspots' and high cognitive load areas were identified using contextual information and physiological responses. The method's accuracy was 72.11% ± 2.21, its precision was 0.77, its recall was 0.70, and its F1 score was 0.73. Hijazi et al. [19] created opportunities to improve language comprehension through the development of intelligent, cognitively aware interfaces. The undergraduate students' creativity, attentiveness, and cognitive load were affected by virtual reality (VR) technology and vibrotactile feedback. 149 students with different spatial skills were split up into a control group and an experimental group. According to the results, students with greater spatial ability had lower cognitive loads and the experimental group performed better in terms of originality and attention value. Wu et al. [20] used the multimodal biometric techniques such as galvanic skin response, and electrocardiography, 30 student volunteers participated in a cognitive load. Data was gathered by participant activity measures, self-report questionnaires, and biometric assessments. Muke et al. [21] employed input diagrammatic reasoning and Sternberg memory tests. Wearable technology is being used in higher education to improve engineering curriculum instruction. Students' involvement, focus, and spatial awareness were improved by the gadgets, which were mostly worn on the head, wrist, and chest. In addition to offering suggestions for enhancing the use of these tools in higher education, the difficulties encountered by researchers when integrating learning technologies for improved engineering education were addressed by Khosravi et al. [22]. The biosensor module and a specially designed signal processing algorithm were used to present a contactless, user-friendly RR (respiration rate) monitoring technique suggested by Pavithran et al. [23]. The approach that combined data collection, noise reduction, and feature extraction was especially helpful during pandemics.

The recognition of CL through physiological reactions based on information about the heart and eyes was presented by Ahmad et al. [24]. The following stages provided a framework for detecting CL. During a task, it first gathered physiological measures using cutting-edge, commercially available sensor technology. Second, it utilized feature removal in conjunction with supervised ML methods. The framework was used in an experiment when participants were asked to identify words and phrases that were accurate and wrong while their heart and eyes were being measured. The Deep Neural Network (DNN)-based higher education student mental well-being forecast model were suggested by Li [25]. To identify college students' cognitive stress levels while they were on research tours the model had developed based on biosensors. To classify the state of mental health as usual, adverse, or positive, the model used the best Biosensor-based and Deep Neural Network (BDNN) based on biosensor data, such as EEG and biomechanical metrics. Therefore, the collected sentimental and biomechanical data was classified using BDNN, and the mental state of learners in college was ascertained from the classification results. Marin et al. [26] detected the significant differences, electrical muscle contractions from five facial and cervix muscles were recorded while performing communication tasks. Participants in the research ranged in age from 18 to 40 and included 55 stutterers and 30 non-stutterers A greater typical amplitude was found in the zygomaticus significant muscles of stuttering participants when the 5–15 Hz range of frequencies was analyzed. Those findings imply that delicate aspects of time and organization in activation of muscles may be involved in the neurological and muscular structures behind the stutter. Individuals' foot pressure and surface electromyography were recorded by Yang et al. [27]. The quality and dependability of the data were first ensured by filtering and denoising the signals that had been gathered. Determine the time field and frequency range includes to record the most important gait data. To accurately recognize human gait and estimate joint angle, the model's output includes both cycle of gait and joint angle.

3. Methodology

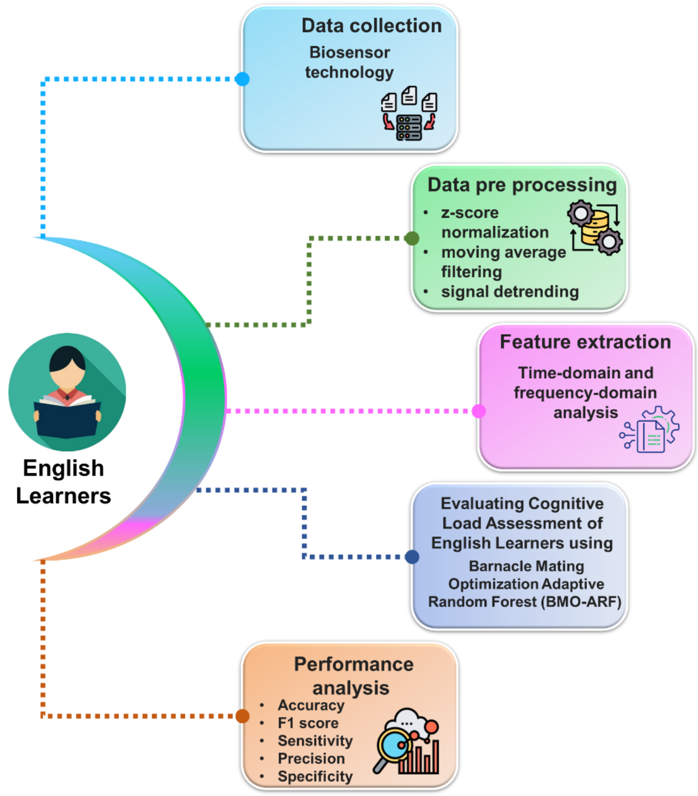

Figure 1. Flow of the research.

The use of biosensor technology to measure cognitive load in English learners during language acquisition tasks is examined. The wearable biosensors such as heart rate sensors, skin conductance sensors, and skin temperature sensors are used to assess the overall mental load of learners in different levels of difficulty of English language tasks. The aim is to establish a reliable process for measuring cognitive load at runtime, with bio-signal data analyzed through ML methods. This research responds to the problem of predicting changes in CL, as these alterations can be subtle and usually involve self-report. Preprocessing techniques as well as feature extraction methods were also discussed extensively in the research as crucial to enhancing the primary classification outcome. This research seeks to improve language learning by developing learning environments that employ cognitive load.

The CLAS dataset were used, to preprocess and extracted the feature the z-score normalization and time-domain and frequency-domain analysis was applied. As proposed method classical classifiers, such as Adaptive Random Forest (ARF), performed better when optimized with Barnacle Mating Optimization (BMO) used for hyper parameter tuning, it was shown in **Figure 1**.

3.1. Data collection

The CLAS dataset is a multimodal resource created to aid in the development of automated technologies for the recognition of particular mental states. It emphasizes negative feelings, mental stress, and intense mental work. The dataset consists of 62 healthy volunteers' accelerometer readings, physiological signals, and metadata. It contains perceptive and interactive tasks to assess particular mental states. The findings are essential for intelligent human-machine interaction interfaces and effective human-robot collaborations [28].

Data collection process

Data were gathered from 62 healthy volunteers the usage of accelerometer sensors, physiological tracking devices (consisting of heart price and skin conductance), and different metadata. Volunteers were subjected to numerous tasks to induce mental states, including pressure and cognitive load. Data from those obligations have been synchronized and recorded for in addition evaluation, ensuring a comprehensive view of the participants' responses throughout distinct modalities.

Physiological sensors inclusive of skin conductance, pores and skin temperature, heart rate, and R-R intervals are commonly used to monitor autonomic nervous system responses. Skin conductance measures sweat gland activity, indicating arousal levels. Skin temperature reflects thermal regulation and stress. Heart rate provides insights into cardiovascular activity, even as R-R intervals, the time between successive heartbeats, provide unique information on heart rate variability (HRV), that is a key indicator of pressure and cognitive load, assisting examine physiological states in various situations.

3.2. Data pre-processing

Data preprocessing involved techniques such as *Z*-score normalization, signal detrending, and moving average filtering to clean and standardize the sensor data. These steps ensured that the data was suitable for effective feature extraction and analysis.

3.2.1 *Z*-score normalization

Z-score normalization is the process of transforming any output descriptors into their normalized counterparts by computing the normalized mean and standard deviation for each parameter over a number of assessments in English learners' datasets. A mean and standard deviation are provided for each characteristic. Following are the replacement details provided by generalized Equation (1):

$$c' = \frac{c - \mu_Y}{\sigma_Y} \tag{1}$$

where the standard deviation of the attribute is denoted by the letter’s discrepancies in values while the attribute’s σ_Y is denoted by the letters means. As a consequence, there is no volatility and no significance for any characteristic in the assessment of English learners’ dataset. Each training sample in the English learners’ dataset is initially placed through the Z-score normalization procedure before creating a trainee collection and beginning the training approach.

3.2.2. Signal detrending

Signal detrending is a data preprocessing technique used in the removal or trend from the signal, which is not required for analysis. This process enables identification of patterns that can mask other more significant patterns or short-term changes relevant to the research, such as the variation in the cognitive load in the physiological signals. These tendencies are useful when the base data holds slow trends or biases that can obscure the analysis of sharper or quicker trends. Some ways of detrending include detrending by subtracting the mean value and using other polynomial fitting methods. By excluding such trends, enhance the significance of the shorter trends that are vital for interpretation of the signal. Detrending, when used in terms of cognitive load assessment, guarantees that the recorded physiological signs account for fluctuations associated with cognitive load only and exclude other contributing factors. Finally, the detrending helps to enhance the signal for ML techniques by removing noise to enrich the future boosts.

3.2.3. Moving average filtering

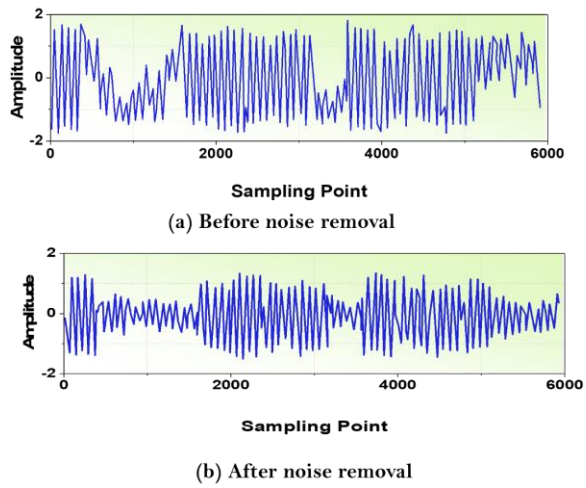

Figure 2. Noise removing through moving average filtering. **(a)** Before noise removal **(b)** After noise removal.

As for cognitive load assessment of English learners, the moving average filter can be used as a method to eliminate noise and average levels of fluctuation data that include skin conductance, heart rate, and skin temperature. These sensors monitor other physical parameters that are affected when executing language tasks by the cognitive load. However, the sensor data comes with ambient noises of movement, and environment from the cognitive load. Using the window size, it is possible to achieve different level of smoothing. A small window gives a fast and lively feedback about the cognitive load in the learner’s brain, and a large window gives a broad and general feedback about the physiological state of the learner. **Figure 2** shows the before and after comparison of noise removing through moving average filtering.

In cognitive load assessment, this filter helps in identifying the mental effort during the English language tasks of different level of complexity and shows how cognitive load changes over time and could result in more accurate interpretations of learner's activity and performance. The moving average filtering was described from Equations (2) to (5).

$$g[m] = \sum_{l=0}^{N} b_l \, \delta(m - l) \tag{2}$$

$$z[m] = w(m) \times (m) = \sum_{l=0}^{N} g[l] w[m - l] \tag{3}$$

$$z_j = \frac{1}{N} \sum_{l=0}^{N-1} w(j - l) \tag{4}$$

$$z_j = \frac{1}{N} \sum_{l=-o}^{0} w(j - i) \tag{5}$$

3.3. Feature extraction using time-domain and frequency-domain analysis

Time-domain feature extraction involves analyzing sensor data directly to extract characteristics like amplitude, duration, and peak count, which reflect immediate changes in the body's response. These features provide valuable insights into the physiological fluctuations and help in understanding how the body reacts to physical exertion. Frequency-domain analysis manipulates the frequency characteristics of the signal in an attempt to extract features with respect to frequency in an attempt to extract patterns and coefficients for mental effort, possibly in the power spectra densities and few frequency bands. By combining, these techniques provide the assessment of cognitive load from two different perspectives: temporal and frequency-related aspects of the data.

3.3.1. Frequency-domain analysis

For frequency-domain approach in cognitive load estimation for the English learners, the physiological signal data collected is transformed into its frequency using techniques such as Fourier Transform. To detect major frequency characteristics associated with mental load during language activities. They help to complement the time-domain features because they reveal wider and more stable patterns in the data. Finally, the frequency-domain analysis increases the validity and reliability of cognitive load measurement when combined with the time-domain analysis for improved understanding of the detailed picture of the learner's state during language learning activities. The frequency-domain analysis was as described from Equations (6) to (8).

$$DF = \frac{\sum_{j=1}^{M} Power_j \times Frequency_j}{\sum_{j=1}^{M} Power_j} \tag{6}$$

$$SE = \sum_{e_j=e_1}^{e_2} O_m\,(e_j) \times log\left(\frac{1}{O_m(e_j)}\right) \tag{7}$$

$$NSE = \frac{SE}{log(M)} \tag{8}$$

3.3.2. Time-domain analysis

The method of time-domain analysis for cognitive load assessment of English learners involves identification of features from signals collected from physiological sensors like heart rate, skin conductance, or skin temperature. These characteristics are related to variations in cognitive load and can signal moments of effort or stress. This assertion shows that time domain analysis enables capturing of short-term fluctuations in cognitive load, hence making it easier to monitor the level of engagement of learners. This method is especially applicable when monitoring fast resynchronizations during task switching or changes in task demand. Time-domain analysis is therefore an underpinning framework by which learner's prompt cognitive load can be appreciated and the data generated could help in improving language learning approaches.

3.4. Barnacle Mating Optimization Adaptive Random Forest (BMO-ARF)

The BMO improves the performance of the ARF model in assessing the cognitive load by fine-tuning the latter's parameters. The nature-inspired optimization algorithm known as Balance Model Optimization or BMO, enhances the accuracy of classification decision-making in the ARF with suitable fine-tuning. The integration between BMO and ARF leads to effective models for ML for cognitive monitoring. This technique does well in managing the diversity and variation in functionality of the learners' responses. The proposed BMO-ARF model enhances feature selection and model adaptability. Unlike the standard Random Forest (RF), BMO-ARF effectively optimizes decision tree, improving accuracy, robustness, and adaptability to diverse cognitive load patterns, leading to higher performance in assessing English learners' cognitive load.

3.4.1. Adaptive Random Forest (ARF)

An ARF is a type of ensemble learning device that produces hypotheses by fusing the vast majority of predictions from numerous different base models to assess English learners. The estimators in each tree leaf are fitted using evaluation nodes to assess English learners. Each tree encounters a random division of the views regarding the English learner's data by assigning each point to the design or estimation component. The distribution of the random vectors W, Z was used to generate the testing instances for the classifiers $g_1(w), g_2(w), \ldots, g_l(w)$. The monetization feature is formatted as shown in the Equation (9).

$$Nh(W,Z) = bu_l J(g_l(W) = Z) - \frac{max}{i \neq Z} bu_l J(g_l(W) = i) \tag{9}$$

here $J(w)$ is the measured value. The error's root cause is as shown in the Equation (10).

$$OF^{\times} = O_{W,Z}(mg(W,Z) < 0) \tag{10}$$

The probability of occurrence in the wZ dimension is shown by the location of the W, Z space. The margin feature for an ARF is shown in the Equations (11) and (12).

$$nq(W,Z) = O_{\Theta}(g(W \times \Theta) = Z) - \frac{max}{i \neq Z} O_{\Theta}(g(W,\Theta)\} \tag{11}$$

Additionally, the value of the classifiers in the set $g(W.\Theta)$

$$T = F_{W,Z} mr(W,Z)I \tag{12}$$

3.4.2. Barnacle Mating Optimization (BMO)

BMO is a biological optimization algorithm that can be applied to tune hyperparameters of the machine learning models. In cognitive load assessment, BMO recreates the process of courtship in barnacles to analyze a wide selection of possible solutions, with the goal of optimizing the algorithm. Through the integration of the features of the selected solutions, BMO optimizes the remaining options with a focus on the feature choice and learning rates. This optimization process helps to increase the accuracy of the proposed model for estimating cognitive load in EFL. Such variability is incorporated through BMO in ways that enhance the model's efficacy in capturing learners' responses and making the necessary improvements. The unique characteristics of barnacles known as hermaphrodite microorganisms, which include both male and female reproductive systems as well as extended penises for copulation, served as the inspiration for BMO. They reproduce by sperm-cast pairing and regular copulation. Barnacles use random motions with their penis range to find a mate, and then they discharge their sperm into the mantle cavity of their spouse. The opposite is true for sperm-cast mating, which happens when sperm discharged into the ocean fertilize barnacle eggs. The development of BMO is being inspired by these two procedures for drilling and mining.

Initialization

BMO begins with an initialization procedure in which potential solutions are created at random, much like other nature-based optimization methods. Following the evaluation process, an organizational method places each applicant at the top of the entire sample based on their initial efficiency.

Selection for mating

In BMO, two barnacles are randomly selected to mate based on their size or length, which is the primary variable that can be adjusted to achieve optimal results. Due to their hermaphroditic nature, barnacles are capable of both providing and receiving sperm from their partners. Although barnacles can receive sperm from multiple mates, for simplicity, the algorithm assumes that each barnacle can only be fertilized once during each mating process. If the selected barnacles fall outside the predetermined size range, sperm-cast mating occurs, leading to the generation of new offspring.

Offspring generation

In BMO, offspring generation involves pair breeding of two barnacles according to their size or length. The principal part of the mating process of barnacles is the exchange of some component, essentially the genetic material, which forms a new generation of barnacles with better characteristics. They partly determine the evolution of the population and therefore they influence how the algorithm evolves to the better versions. This reproductive mechanism assists in the preservation of diversity, and the consistent enhancement of the optimization procedure. Equations (13) and (14) shows the offspring generation in BMO of the fertilized eggs. **Figure 3** shows the architecture of BMO.

$$w_{child}^{M\_new} = ow_{dad}^{M} + rw_{mum}^{M}\ for \quad l \leq ok \tag{13}$$

$$w_{child}^{M\_new} = [rand] \times w_{mum}^{M} for \ \ l > ok \tag{14}$$

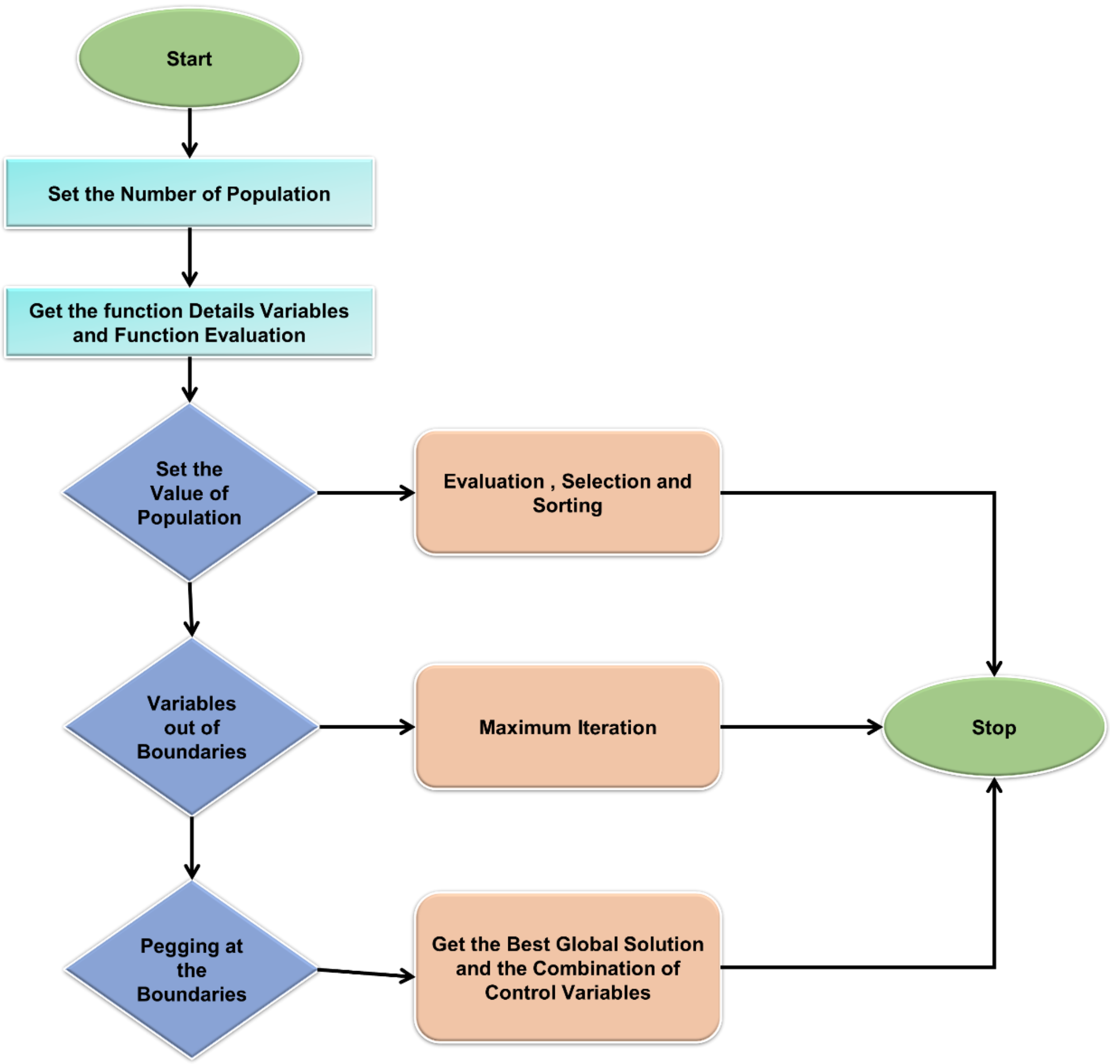

Figure 3. The architecture of BMO.

Algorithm 1 BMO-ARF

```
# Step 1: Adaptive Random Forest (ARF)
Initialize ARF model
For each sample:
  Extract features
  For each tree:
    if random_split:
      Train with selected English learners
  predictions = Aggregate predictions
  if error_rate > threshold:
    Adjust tree weights
# Step 2: Barnacle Mating Optimization (BMO)
Initialize barnacle population
While not terminated:
  For each barnacle:
    fitness = Evaluate ARF performance
if fitness > threshold:
selected_barnacle = barnacle
if size_in_range(selected_barnacle):
offspring = Copulation (exchange features)
else:
offspring = Sperm − cast mating (randomized features)
Update ARF with offspring
if new_fitness > fitness:
    Replace with fitter offspring
```

To evaluate the English learners' cognitive load, the ARF model is trained over bootstrapped subsets of the collected physiological sensor data. The fitness of each barnacle solution is then calculated to enhance the given model performance. After several 'Fixed Point Iterations,' two solutions of barnacles are evaluated for their fitness levels, following which they either have normal copulation or sperm cast mode of mating produces offspring from the two parent solutions. These newly formed configurations are used to train and cross-validate the ARF model, which makes a more precise measurement of cognitive load possible. Each generation of offspring replaces less fit solutions with fitter ones until a terminal criterion is met, such as the accuracy level acquired after the maximum number of iterations. The last result is the English learners' cognitive load prediction ARF model with the assistance of the hyper parameters set by BMO, as shown in **Algorithm 1**.

4. Result and discussion

The Python platform and a laptop's RAM of 8.00 GB are used in this research to retrieve data rapidly on Intel® Core i9 processors running Windows 11. The research has proposed BMO-ARF and existing methods such as random forest (RF) [24], and Naïve Bayes (NB) [24] and parameters such as accuracy, $F1$ score, sensitivity, specificity, and precision to evaluate the effectiveness of the cognitive load assessment of the English learners.

4.1. Accuracy

Accuracy, in a learning model, focuses on determining the cognitive load of English learners, despite the measure of the ability of a given model to accurately predict the amount of a learner's cognitive burden when performing language tasks. The confusion matrix as a measure is defined as the ratio between the true positive and true negative predictions out of the total number of predictions carried out by the model. Using high accuracy, the system is able to differentiate the different degrees of cognitive load, and thus the learner engagement and mental load levels can be accurately measured. In cognitive load studies, accuracy is paramount because biosensors and machine learning models' ability in sensing and interpreting variations in physiological signs can be justified. When accuracy is evaluated, it becomes possible to check whether the system offers valuable information on the learner's behavior for the improved delivery of personalized learning. The accurate identification of cognitive load effectively contributes to the development of proper instructional support interventions for language learning. The value of BMO-ARF in accuracy has attained 95.99%, which is higher than the NB [24] at 85.83% and RF [24] at 91.66%. The performance of accuracy is shown in **Figure 4** and **Table 1**.

Table 1. Quantitative of accuracy.

Methods	Accuracy (%)
NB [24]	85.83
RF [24]	91.66
BMO-ARF [Proposed]	95.89

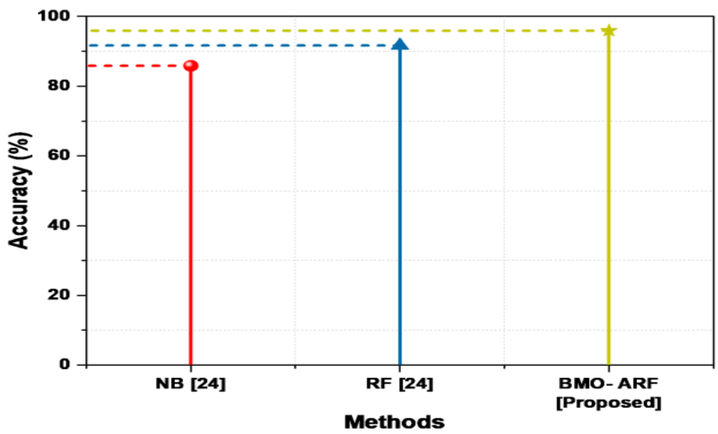

Figure 4. Graphical representation of accuracy.

4.2. F1 score

F1 score in the context of cognitive load assessment for English learners is a measure that balances precision and recall rates and is used to assess the ability of a model to classify cognitive load levels, such as high, medium, and low. It offers a fairly reasonable measure, primarily in situations where the data can be twisted, and some of the levels of cognitive load are rarer than others. For high, medium, and low cognitive categories, the F1 score takes both the true positive and false positive/negative into account, thereby excluding the biasness from any one of the load levels. A high F1 score means that the model distinguishes between all three levels of cognitive load with high accuracy and minimal cross-classification. This metric is more valuable in the learning process; identifying learner's states of mind accurately facilitates appropriate interventions and further learner interventions. **Table 2** presents the F1 score attained: BMO-ARF of 0.95 in cognitive load of low and other methods at 0.87 in NB [24] and 0.91 at RF [24], BMO-ARF of 0.90 in cognitive load of medium and other methods at 0.81 in NB [24] and 0.85 at RF [24], BMO-ARF of 0.97 in cognitive load of high and other methods at 0.90 in NB [24] and 0.95 at RF [24]. The outcome of the F1-score is shown in **Figure 5**.

Table 2. Quantitative of F1-score.

Cognitive load	Methods	F1-score
Low	NB [24]	0.87
	RF [24]	0.91
	BMO-ARF [Proposed]	0.95
Medium	NB [24]	0.81
	RF [24]	0.85
	BMO-ARF [Proposed]	0.90
High	NB [24]	0.90
	RF [24]	0.95
	BMO-ARF [Proposed]	0.97

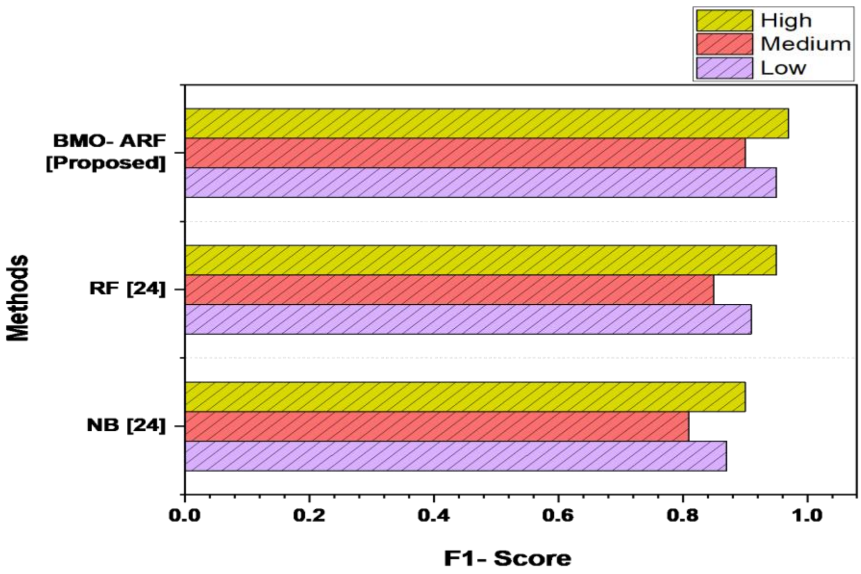

Figure 5. Outcome of F1-score.

4.3. Sensitivity

When applied to the assessment of cognitive load for learners learning English, sensitivity implies the effectiveness of a model or a system in identifying shifts in levels of cognitive load. The true positives indicate the proportion of correctly identified high cognitive load situations among all actual high cognitive load events. This means that the system measures high sensitivity to show that it is able to identify when learners exert a lot of mental effort while learning languages. Sensitivity enables tracking of learners' cognitive load to avoid frustration in the learning task and thus enhance learning. The purpose here is to increase the accuracy of load identification to enhance the usability of education interventions.

Consequently, sensitivity becomes central to the accurate assessment of cognitive status and improving learners' experiences. **Table 3** shows the cognitive load of low sensitivity at 80.3, medium at 88.5, and 93.0 at high in the proposed method.The result of sensitivity is shown in **Figure 6**.

Table 3. Quantitative sensitivity.

Cognitive load	Sensitivity (%)
Low	80.3
Medium	88.5
High	93.0

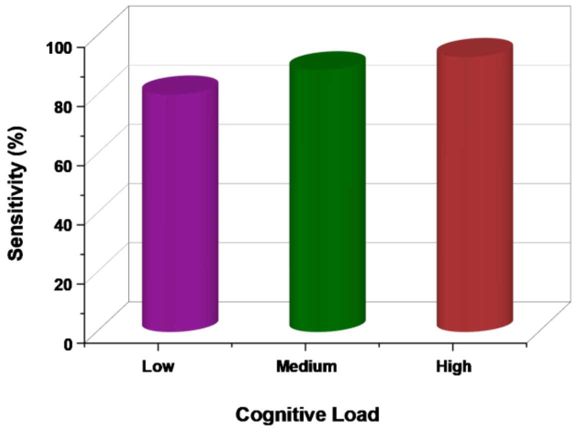

Figure 6. Graphical representation of sensitivity.

4.4. Precision

In the context of cognitive load assessment of English learners, precision can be understood as the measure of a model or a system's accuracy in sorting out instances of high or low cognitive load. A true positive rate with regard to cognitive load assessment of English learners can be defined as the ability of a model or a system that distinguish between high and low cognitive load without resulting in a false positive. It calculates the ratio of actual true positive, which is the correct assessment of the actual cognitive load of students, to the entire class where the assessment considers positive for a certain value of cognitive load. The kind of intervention that the learner is likely to need is thus well captured when considerable language learning accuracy is achieved. This is especially significant while dealing with biosensor data due to the fact that the varied physiological signals presented can be affected by people's variations and environmental factors. **Table 4** shows the value of precision in the cognitive load of low at 84.5%, medium at 89.0%, and high at 94.0%. The performance of precision is shown in **Figure7**.

Table 4. Quantitative values of precision.

Cognitive load	Precision (%)
Low	84.5
Medium	89.0
High	94.0

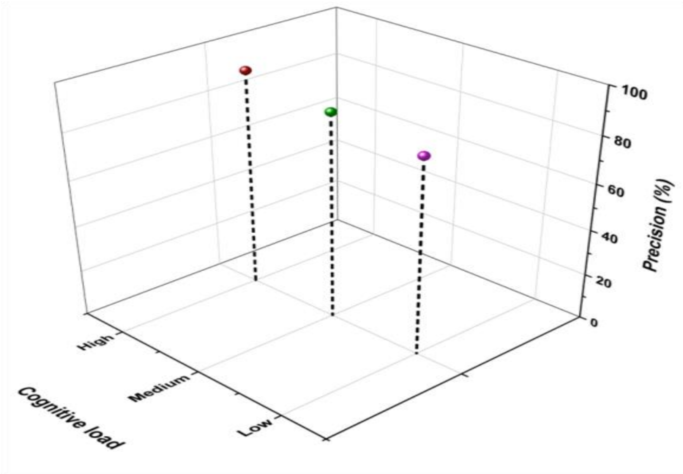

Figure 7. Graphical representation of precision.

4.5. Specificity

When applied to the context of cognitive load assessment for English learners, the specificity of the model lies in its correct identification of non-high cognitive load cases. The accuracy between low and high categories is usually important in cases where the model involves testing of cognitive load, as it will estimate how correctly the model differentiates learners' cognitive load from those who are not overloaded by learning material. A high specificity shows that the high CL would allow for an exclusion of false positives, thus indicating that only the subjects with the anticipated high cognitive load would be recognized. This is important when measuring cognitive load because it assists in finding more of the non-high-load situations that will aid in the creation of better learning interventions. The interventions or adjustments happen to the learners to the barest minimum to avoid disruptions or alterations. The specificity reduces the number of false identifications, making the assessment system more credible. The paper shows that by paying attention to specificity, investigators can gain increased accuracy in their cognitive load theories, thereby improving the learning strategies employed. **Table 5** shows the value of specificity in the cognitive load of low at 87.5%, medium at 91.8%, and high at 95.1%. The output of Specificity is shown in **Figure 8**.

Table 5. Quantitative values of specificity.

Cognitive load	Specificity (%)
Low	87.5
Medium	91.8
High	95.1

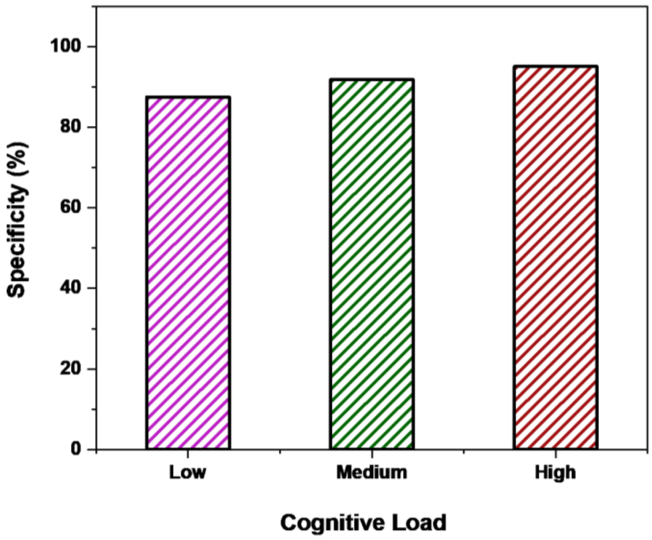

Figure 8. Graphical representation of specificity.

4.6. Discussion

The method RF, although better at dealing with large data and intricate structures, belongs to the risk of overfitting significantly when the extent of regression trees is too many or when data contains noise. Additionally, it may experience issues with interpretability. Since the model truly consists of decision trees, it is quite difficult to determine which decision was made on a particular element. Further, RF can be slow to operate and sometimes it consumes much time in analyzing large datasets that have many features. It can also have a small level of performance in the event of feature extraction from highly imbalanced data where the minority class is dominated by the majority class. NB is proven to be a simple and fast classifier, but NB classifies the features independently. This assumption could result in huge errors whenever relationships between features exist. Adequately with continuous attributes, especially when they are supplied in a flow form, it is fundamentally well suited to work with categorical variables. It also found that NB is sensitive to a skewed distribution of data and always leans towards the side with the most density. In addition, when the decision boundaries are complex, this is due to the fact that it models the class probability linearly, which is insufficient to capture the data complexity sufficiently well enough. BMO-ADF also eliminates bias in classification as it optimizes the decision-making part, providing improved accuracy in the classification of cognitive load. Furthermore, it combines the traits of the two classifiers, leaving their excess with various drawbacks, to produce a more precise model identifying cognitive load in English learners. The proposed method BMO-ADF guarantees a beneficial differential effect for different measures of practicing difficulty and learner characteristics and is more valid in estimating cognitive load throughout language acquisition.

5. Conclusion

The use of biosensor technology in monitoring cognitive load in English language learners is gaining traction through the assessment of physiological signals such as heart rate, skin conductance, and skin temperature. It also enables quantification of cognitive load involved in performing language learning tasks of varying complexity. The proposed method BMO-ARF has attained accuracy at 95.89%, F1 score in the cognitive load of low at 0.95, medium at 0.90and high at 0.97, sensitivity in the cognitive load of low at 80.3%, medium at 88.5% and high at 93.0%and specificity in the cognitive load of low at 87.5%, medium at 91.8% and high at 95.1%. This research also presents efficient ways of determining cognitive load, especially in learning cultures in English civilization, by using physical data from such relative devices as skin conductance, heart rate, and skin temperature. When provided in the evaluation, techniques of pre-processing and feature extraction methods are shown to be very important in model development. In cognitive load classification, the use of the classical classifier BMO-ARF is very difficult to achieve with the classification of user performance on more complex tasks. Nevertheless, the issues of overfitting the model, which is the intrinsic limitation of modern ML methods, and the heterogeneity of the psycho-physiological response require more sophisticated labeling solutions. Future work needs to make improvements to these cognitive load monitoring methods to enhance the reliability and generalizability of the systems to measure cognitive load in language learning given that there are both individual differences and intrinsic language learning issues.

Author contributions: Conceptualization, LQ and WZ; writing—original draft preparation, LQ and WZ; writing—review and editing, LQ and WZ. All authors have read and agreed to the published version of the manuscript.

Ethical approval: Not applicable.

Conflict of interest: The authors declare no conflict of interest.

References

1. Sari FM. Exploring English learners’ engagement and their roles in the online language course. Journal of English Language Teaching and Linguistics. 2020; 5(3), pp.349–361.
2. Yang H. Secondary-school Students' Perspectives of Utilizing TikTok for English learning in and beyond the EFL classroom. In 2020 3rd International Conference on Education Technology and Social Science (ETSS 2020) 2020. Vol. 1, pp. 162–83.
3. Sugarman J, Lazarín M., Educating English learners during the COVID-19 pandemic. Migration Policy Institute, 2020; 1, pp.1–30.
4. Rifiyanti H. Learners' perceptions of online English learning during the COVID-19 pandemic. Scope: Journal of English language teaching, 2020; 5(1), pp.31–35.
5. Syakur A, Fanani Z, Ahmadi R. 2020. The effectiveness of reading English learning process based on blended learning through “Absyak” website media in higher education. Budapest International Research and Critics in Linguistics and Education (BirLE) Journal, 3(2), pp.763–772.
6. Degirmenci R. 2021. The use of Quizizz in language learning and teaching from the teachers' and students’ perspectives: A literature review. Language Education and Technology, 1(1), pp.1–11.
7. Syafiq AN, Rahmawati A, Anwari A, Oktaviana T. 2021. Increasing speaking skills through YouTube videos as English learning material during online learning during the pandemic covid-19. Elsya: Journal of English Language Studies, 3(1), pp.50–55.

8. Rozal E, Ananda R, Zb A, et al. 2021. The effect of project-based learning through YouTube presentations on English learning outcomes in physics. AL-Ishlah: Jurnal Pendidikan, 13(3), pp.1924–1933.
9. Anwar K. 2021. The perception of using technology Canva application as a media for English teachers creating media virtual teaching and English learning in loeithailand. Journal of English Teaching, Literature, and Applied Linguistics, 5(1), pp.62–69.
10. Utami AR, Oktaviani L, Emaliana I. 2021. The Use of Video for Distance Learning During Covid-19 Pandemic: Students' Voice. JET (Journal of English Teaching) Adi Buana, 6(02), pp.153–161.
11. Mavrogordato M, White RS. 2020. Leveraging policy implementation for social justice: How school leaders shape educational opportunity when implementing policy for English learners. Educational Administration Quarterly, 56(1), pp.3–45.
12. Rahardjo A, Pertiwi S. 2020. Learning motivation and students' achievement in learning English. JELITA, 1(2), pp.56–64.
13. Zai X. 2024. Leveraging Bio-Sensing Technology and IoT for Optimizing Spanish Vocabulary Instruction Across Chinese and Western Cultures: A Biotechnological Approach. Journal of Commercial Biotechnology, 29(3), pp.305–314.
14. Lee H. 2020, October. Supporting instructors to provide emotional and instructional scaffolding for English language learners through biosensor-based feedback. In Proceedings of the 2020 International Conference on Multimodal Interaction (pp. 733–737).
15. Hijazi H, Couceiro R, Castelhano J, et al. 2021. Intelligent biofeedback augmented content comprehension (tellback). IEEE Access, 9, pp.28393–28406.
16. Hijazi H, Couceiro R, Castelhano J, et al. 2022, June. Intelligent Biofeedback Comprehension Assessment: Theory, Research, and Tools. In 2022 IEEE 21st Mediterranean Electrotechnical Conference (MELECON) (pp. 414–419). IEEE.
17. McNeal KS, Zhong M, Soltis NA, et al. 2020. Biosensors show promise as a measure of student engagement in a large introductory biology course. CBE—Life Sciences Education, 19(4), p.ar50.
18. Herbig N, Düwel T, Helali M, et al. 2020, July. Investigating multi-modal measures for cognitive load detection in e-learning. In Proceedings of the 28th ACM conference on user modeling, adaptation and personalization (pp. 88–97).
19. Hijazi H, Gomes M, Castelhano J, et al. 2024. Dynamically predicting comprehension difficulties through physiological data and intelligent wearables. Scientific Reports, 14(1), p.13678.
20. Wu L, Chau KT, Wan Y, et al. 2024. The Effect of Electroencephalogram Feedback in Virtual Reality Interactive System on Creativity Performance, Attention Value, and Cognitive Load. International Journal of Human-Computer Interaction, pp.1–18.
21. Muke PZ, Telec Z, Trawiński B. 2022, November. Multimodal Approach to Measuring Cognitive Load Using Sternberg Memory and Input Diagrammatic Reasoning Tests. In Asian Conference on Intelligent Information and Database Systems (pp. 693–713). Cham: Springer Nature Switzerland.
22. Khosravi S, Bailey SG, Parvizi H, Ghannam R. 2022. Wearable sensors for learning enhancement in higher education. Sensors, 22(19), p.7633.
23. Pavithran M, Ganeshmurthy MS, Periyasamy R. 2024, March. A Robust Algorithm for respiration rate monitoring during cognitive load using PPG Signals. In 2024 Tenth International Conference on Bio Signals, Images, and Instrumentation (ICBSII) (pp. 1–6). IEEE.
24. Ahmad MI, Keller I, Robb DA, Lohan KS. 2023. A framework to estimate cognitive load using physiological data. Personal and Ubiquitous Computing, pp.1–15.
25. Li K. 2024. Using biosensors and machine learning algorithms to analyze the influencing factors of research tours on students' mental health. Molecular & Cellular Biomechanics, 21(1), pp.328–328.
26. Marin E, Unsihuay N, Abarca VE, Elias DA. 2024. Identification of the Biomechanical Response of the Muscles That Contract the Most during Disfluencies in Stuttered Speech. Sensors, 24(8), p.2629.
27. Yang L, Shi Z, Jia R, et al. 2024. Multi-branch deep learning neural network prediction model for the development of angular biosensors based on sEMG. Frontiers in Bioengineering and Biotechnology, 12, p.1492232.
28. Markova, Valentina (2020), CLAS: A Database for Cognitive Load, Affect and Stress Recognition, Mendeley Data, V1, doi: 10.17632/8hm59ryzb8.1