IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received 6 May 2024, accepted 20 June 2024, date of publication 22 August 2024, date of current version 3 September 2024.

Digital Object Identifier 10.1109/ACCESS.2024.3428691

RESEARCH ARTICLE

Cognitive Load Estimation Using a Hybrid Cluster-Based Unsupervised Machine Learning Technique

IQBAL HASSAN[1], MONICA ZOLEZZI[2], HANAN KHALIL[2], RAFIF MAHMOOD AL SAADY[2], SHONA PEDERSEN[2], AND MUHAMMAD E. H. CHOWDHURY[3], (Senior Member, IEEE)

[1]Sozo Laboratory, Kyushu Institute of Technology, Fukuoka 808-0196, Japan
[2]College of Medicine, QU Health, Qatar University, Doha, Qatar
[3]Department of Electrical Engineering, Qatar University, Doha, Qatar

Corresponding author: Muhammad E. H. Chowdhury (mchowdhury@qu.edu.qa)

This work was supported in part by the Vice President (VP) Grant of the Health Cluster of Qatar University, and Open Access funding provided by the Qatar National Library.

This work involved human subjects or animals in its research. Approval of all ethical and experimental procedures and protocols was granted by the AGH University of Science and Technology Ethics Committee under IRB No. 3/2022.

ABSTRACT The increasing prevalence of non-invasive, portable Electroencephalography (EEG) sensors for neuro-physiological measurements has propelled EEG-based assessments of cognitive load (CL) into the spotlight. In this study, we harnessed the capabilities of a four-channel, wearable EEG device that captured brain activity data during two distinct CL states: *Baseline* (representing a non-CL, resting state) and the *Stroop Test* (a CL-inducing state). The primary objective of this study is to estimate the CL index through an innovative approach that employs a hybrid, cluster-based, unsupervised learning technique seamlessly integrated with a 1D Convolutional Neural Network (CNN) architecture tailored for automated feature extraction, rather than conventional supervised algorithms, which facilitated in the acquisition of latent complex patterns without the need for manual categorization. The approach was rigorously evaluated using stratified cross-validation, with several assessment criteria assessing both its quality and predictive capability to estimate the CL index. The results obtained (e.g., homogeneity score of 0.7, adjusted rand index of 0.78, silhouette coefficient of 0.5, and an accuracy rate of 93.2%) demonstrate that our module exhibits superiority over supervised approaches. These results are indicative that the adoption of multi-channel wearable EEG devices may facilitate real-time CL estimation, minimizing the need for extensive human intervention, and reducing potential bias, paving the way for more objective and efficient CL assessments.

INDEX TERMS Cognitive load, unsupervised machine learning, electroencephalography (EEG), brain–computer interface (BCI).

I. INTRODUCTION

Complex tasks demand the integration of diverse cognitive faculties, including the assimilation of task-related data, utilization of working memory, sustained attention, and nuanced decision-making. Previous inquiries [1], [2], [3] have introduced the CL theory, positing the finite nature of human cognitive resources for information processing and retention. When cognitive demands exceed human working memory capacity, performance decline ensues, characterized by cognitive overload and task incapacity [4]. Recognizing the limited cognitive capacity of the human cognitive system [5], scholars in Brain-Computer Interface (BCI), cognitive science, psychology, and Human-Computer Interaction (HCI) systems must acquaint themselves with user cognitive processing capacity, memory workload intricacies, and task engagement dynamics [6].

The associate editor coordinating the review of this manuscript and approving it for publication was Md. Moinul Hossain.

The cognitive workload has been scrutinized from diverse perspectives. Within [1], CL is defined as the constraining force on working memory during cognitive task execution, with Brouwer et al. [7] contextualizing cognitive workload as the burden on working memory within an n-back task. Mills et al. [8] strategically utilized true-false questions for minimal workload and open-ended queries for heightened workload, necessitating more precise memory retention. Additional research underscores the role of skill acquisition in simulating cognitive stress [3]. As emphasized by Logan et al. [9], expertise acquisition and task automation lead to a reduction in cognitive workload [10]. Consequently, task complexity influences cognitive burden and an individual's capacity for task execution. A highly intricate task by an inexperienced individual results in elevated CL, whereas a straightforward task by a skilled individual yields a lower cognitive burden. Cognitive demands are categorized into intrinsic and extrinsic CL [11], with intrinsic CL stemming from cumulative cognitive demands in a learning task. In contrast, extrinsic CL involves the demands from the design of a learning assignment [3].

The evaluation of CL is essential due to its intricate connection with the learning process. Its significance extends to various domains, including education [12] and the design of human-computer interfaces [13]. In education, CL assessment is commonly employed to gauge the effectiveness of learning materials and environments, aiming to improve student learning outcomes. This evaluation supports users in maintaining an optimal CL level, especially in diverse environments and challenging tasks [14]. Heightened emphasis has been placed on maintaining an ideal CL level during learning tasks, considering the brain's limited capacity for cognitive processing [15], [16]. Both insufficient and excessive CL can lead to adverse consequences, emphasizing the need to minimize such situations, particularly in critical decision-making scenarios [5].

In the realm of HCI, challenges arise in conceptualizing and quantifying CL in perceptually rich learning environments. Contrary to the prevalent notion that lower CL is always favorable, recent studies challenge this understanding in digital learning, indicating instances where higher CL leads to improved performance [17], [18], [19], [20], [21]. Skulmowski et al. [22] posit this as an inherent principle of perceptually rich virtual environments. Design elements such as interactivity, immersion, and realism, while contributing to extraneous CL, can also be essential components of a learning task. Interactive simulations, despite introducing extraneous CL, may lead to superior performance, especially if the assessment aligns with the interactive learning approach [22], [23], [24]. Hence, when assessing CL, considering the objective of the task is crucial, and the presence of extraneous CL should be evaluated based on whether it genuinely hinders learners from achieving their task objectives [22].

CL assessment encompasses five distinct types: overall, accumulated, average, peak, and instantaneous load [6]. Instantaneous load specifically refers to the CL imposed at each moment during a cognitive task [6]. Previous studies have employed both analytical and empirical approaches for CL assessment [5], [25]. There are two main methods for assessing CL: subjective and objective [26], [27], [28], [29]. Historically, subjective methods, relying on self-reporting and behavioral data like reaction times and error rates, dominated CL assessment. Instruments like the NASA task load index are popular subjective measures [30]. However, subjective methods have limitations in providing real-time assessments and may introduce bias [31], [32].

Physiological signals have emerged as a viable alternative to address these challenges. Traditionally, it is widely acknowledged that physiological signals constitute the most effective indicators of cognitive load due to their ability to offer an efficient temporal resolution for long-term monitoring [33] and greater feasibility for estimating CL when compared to subjective rating methods [34]. Objective physiological measurements possess the advantage of continuously assessing CL throughout a cognitive task using online capabilities [6]. These physiological techniques are based on the premise that changes in cognitive functioning are reflected in physiological variables such as heart rate variability [4], task-evoked brain activity [35], skin conductance [36], and eye movement patterns [37]. Changes in behavioral signals may result from alterations in physiology or from different mental processing strategies. These measurements can be utilized to visualize intricate trends and patterns of CL, encompassing instantaneous, peak, and accumulated load [38]. Specifically, EEG has been validated as an effective, noninvasive method for the detection, estimation, or prediction of the human brain activities [39], [40]. EEG measurements directly capture brain activity rather than relying on indirect measurements of other physiological responses triggered by the brain. In recent years, numerous empirical investigations have explored the connection between cognitive demands and EEG activity across different frequency bands and brain regions. These investigations have employed EEG as a standalone method or in conjunction with other subjective and objective measures to evaluate participants' cognitive workload in diverse settings. Examples include studies involving arithmetic tasks [51] and engagement in virtual reality human-computer interaction environments [52], [53]. EEG, as a neurophysiological measure with a high temporal resolution (approximately 1 millisecond), presents itself as a suitable choice for assessing instantaneous CL. This capability opens up the opportunity to monitor the dynamic changes in cognitive load on working memory during a cognitive task as numerous portable EEG devices are readily accessible for real-time CL assessment [41]. This method is characterized by its objectivity, non-invasiveness, and relative lack of constraints in comparison to other conventional methods [6]. The various rhythms generated by electrical brain activity, including delta (1–3 Hz), theta (4–7 Hz), alpha (8–13 Hz), beta (14–30 Hz), and gamma (31–50 Hz), hold paramount significance in the realm of CL recognition [42]. Notably, the theta and alpha ranges appear

to be closely associated with higher brain functions, reflecting task difficulty or CL across diverse task demands [43], [44], [45], [46], [47], [48], [49], [50].

While prior research has employed a combination of subjective and objective measures to assess CL, it has predominantly relied on a range of supervised learning techniques. However, when dealing with EEG data, the utilization of supervised strategies may entail certain drawbacks. Our study addresses limitations of supervised learning for Cognitive Load (CL) assessment using EEG data. Supervised methods require labeled data, which can be subjective, expensive, and time-consuming to obtain for real-time applications.

As a solution to these challenges, our study proposes unsupervised learning as a promising alternative. It eliminates labeling bias and facilitates feature extraction, making it suitable for complex EEG data. Unsupervised methods can find hidden patterns, group similar brain states, and identify anomalies, offering valuable insights. Additionally, they are adaptable to various EEG data without extensive adjustments.

In our study, we utilized a 1D CNN for unsupervised feature extraction from EEG data streams. 1D CNNs are adept at learning sequential patterns and extracting features at different levels of complexity. This allows them to adapt to the data's intricacies, especially when the data is structured in a way that makes sense in terms of time or order, like continuous EEG signal streams. These networks automatically learn hierarchical representations of sequential patterns, capturing both local and global dependencies in the data. 1D CNNs focus on short-term dependencies, providing benefits such as local pattern recognition, translation invariance, parameter sharing, and efficient processing of sequences. Shared weights in convolutional layers reduce parameters, enhancing computational efficiency and preventing overfitting, especially with limited data.

This study explores the viability of using unsupervised learning techniques for real-time assessment of Cognitive Load (CL) from EEG data. This approach offers several advantages over traditional methods that rely on supervised learning with labeled data. The potential of our proposed method are as follows:

- Unsupervised learning eliminates the subjectivity and potential bias inherent in labeling EEG data for supervised learning. This can lead to more accurate and reliable CL assessments.
- By not requiring pre-labeled data, unsupervised learning becomes much faster and more cost-effective. This opens doors for real-time CL assessment in various practical scenarios i.e., personalized learning, task optimization, and healthcare advancements.
- Unsupervised methods excel at uncovering hidden patterns and structures within EEG data. This can lead to a deeper understanding of how the brain functions under different cognitive loads. Additionally, unsupervised clustering can group similar brain states together, revealing new insights into brain activity.
- Unsupervised methods are adaptable to various EEG datasets without extensive adjustments. This makes them a more flexible and efficient approach compared to supervised learning, which often requires significant re-labeling for new data.

From these perspectives, our study introduces the following contributions and highlights their significance within the existing body of literature:

- The study proposes a method for quantifying CL in real-world settings, minimizing biases associated with controlled laboratory environments. This opens doors for practical applications.
- The introduction of a 1D CNN offers a streamlined and automated tool for extracting relevant features from EEG data streams. This is crucial for unsupervised learning as it eliminates the need for manual feature selection.
- The study proposes a unique approach for real-time CL assessment using unsupervised clustering techniques. This is a significant contribution as it offers a new way to estimate CL without relying on pre-labeled data.

Overall, this study presents a promising approach which has been proven in Section V by showing a comparative analysis with existing supervised literature. This real-time, objective CL assessment approach has the potential to revolutionize various fields like education, healthcare, and workplace optimization.

The remainder of this paper is structured as follows: Section II offers an overview of related studies, Section III provides insights into the dataset's structure, Section IV outlines our proposed methodology, Section V delves into the discussion of achieved results, and finally, Section VI concludes the paper while also suggesting avenues for future research.

II. LITERATURE STUDY

In this section of the literature, we present a review of previous research relevant to our work.

Prior research endeavors [57] have illuminated the potential of Electroencephalography (EEG) as an invaluable resource for delving into an individual's Cognitive Load (CL). However, the nuanced outcomes of these investigations are intricately entwined with a myriad of factors. These include making smart choices about the number of EEG channels used, the size of the training and testing datasets for machine learning (ML) algorithms, the tasks that need to be done and how they change over time, choosing the right time-domain and frequency-domain features, and deciding whether to use personalised or generalised models. Importantly, within the realm of evaluating CL, the deployment of wearable EEG systems featuring four channels or fewer is relatively scarce [58], [59], [60]. Furthermore, there is a discernible paucity of studies conducted on CL assessment using dry electrode measurements [61]. Notably, a study utilizing four electrodes for stress classification

primarily focuses on discerning perceived stress rather than instantaneous stress levels [62]. A pivotal contribution of our research resides in the inventive utilization of an economical wearable device equipped with merely four dry electrodes to comprehensively assess CL. This groundbreaking approach establishes a foundational method for real-world monitoring of an individual's cognitive performance across diverse everyday scenarios.

The landscape of cognitive state evaluation has traditionally leaned on statistical analysis techniques, amalgamating subjective, behavioral, and physiological measures [45], [53], [54]. Nevertheless, contemporary research has progressively redirected its focus towards embracing advanced machine learning and deep learning methodologies to elevate the precision of CL assessments [51], [55], [56], [72], [81], [82]. For instance, [51] conducted subject-independent cognitive load estimation from EEG signal streams, achieving optimal results with K Nearest Neighbour (KNN) and ensembles of decision trees. In a notable instance, [55] introduced Riemannian geometry-based classifiers (RGC), showcasing a capacity to compete effectively with other deep learning-based approaches, boasting the highest mean accuracy of 72.73%. In the domain of real-time cognitive state classification, [63] presented a lightweight Convolutional Neural Network (CNN) model that reduces trainable parameters while sustaining high performance. Another robust avenue involves the extraction of spectral components from EEG data, as evidenced by the work of [64] and [65]. Through the application of machine learning techniques, these authors [66] unveiled correlations between CL and alpha-to-theta ratios and theta-to-alpha ratios. Moreover, the comprehensive study by [67] achieved CL prediction by amalgamating handcrafted spatial, temporal, and spectral EEG features. This involved calculating EEG-based functional connectivity, microstates, and power spectral densities (PSD) across three different CL levels. Experimental outcomes showcased associations between heightened CL and increased theta power, reduced alpha power, and noteworthy alterations in inter-channel connectivity and microstates. The predictive model exhibited accuracy exceeding 80% in cross-validation, real-time, and over-time prediction scenarios. Furthermore, leveraging 24 EEG channels, [68] achieved a classification accuracy of 75.9% for personalized models, underscoring the versatility and efficacy of machine learning in advancing CL assessment methodologies.

Typically, the effectiveness of machine learning and deep learning classifiers relies on various factors, including the chosen training paradigm. Many previous significant studies have predominantly employed supervised learning strategies. However, in contrast to these approaches, our research takes an unsupervised route to instantly estimate real-world CL from EEG data recorded through four channels. By opting for this unsupervised approach, we aim to reduce the subjective influence inherent in supervised techniques and establish a fundamental framework for real-time monitoring of CL from EEG signals.

III. DATASET DESCRIPTION

In our research, we employed the freely accessible open-access CogWear database [69]. The subsequent section provides a summary of the dataset's organization.

A. DATASET STRUCTURE

Data gathering occurred in two phases: a pilot study and a survey gamification experiment. Each of these stages encompassed two shared data recording sessions: one for establishing a baseline (representing a non-CL resting state) and another for conducting the Stroop test (representing a CL state). The survey gamification experiment consisted of two stages: pre and post-gamification sessions, where pre is the recording session before gamification and post is the gamified session. The baseline and Stroop test recording procedures were the same in both of these sessions. Each session lasted for roughly about 3-4 minutes. According to the dataset authors, throughout these data collection sessions, minimal interventions were applied to the participants without introducing any external stimuli, ensuring a natural and uninterrupted progression of activities. The motivation of the original dataset was to see if gamifying a survey structure has any significant impact on cognitive load. For our study, the goal is to estimate cognitive load states through the EEG signals. Thus, the gamified structure was ignored during our analysis, and the sample sets were treated as CL and non-CL states.

B. PARTICIPANT INFORMATION

The pilot study involved 11 participants, while the survey gamification experiment included 13 participants. Consequently, we are provided with a total of 11 recording sessions during the pilot study and 26 recording sessions during the survey gamification experiment, accounting for both pre-session and post-session measurements ($13 \times (1$ pre-session $+ 1$ post-session$) = 26$ sessions).

C. MODALITIES

Throughout the entire experiment, participants' physiological signals were recorded using three different wearable devices: the Empatica E4 watch, the Samsung Galaxy Watch4, and the Muse S (Gen2).

The Empatica E4 device captured various physiological parameters, including Blood Volume Pulse (BVP), which served as the basis for deriving Heart Rate Variability (HRV) and Inter Beat Interval (IBI) measurements. It also recorded skin temperature (TEMP), Electrodermal Activity (EDA), and 3-axis acceleration data (ACC). The Empatica E4 operated in online mode, recording BVP at a frequency of 64Hz, EDA at 4Hz, and TEMP at 4Hz.

The Samsung Galaxy Watch4 was utilized to collect the photoplethysmogram (PPG) signal, which had a frequency of 25 Hz.

The Muse S EEG headband measured brain activity through dry electrodes placed on specific locations on the scalp corresponding to prefrontal (AF7 and AF8 sensors)

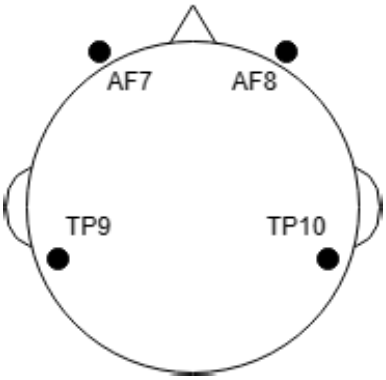


FIGURE 1. Electrode placement areas of the used EEG device.

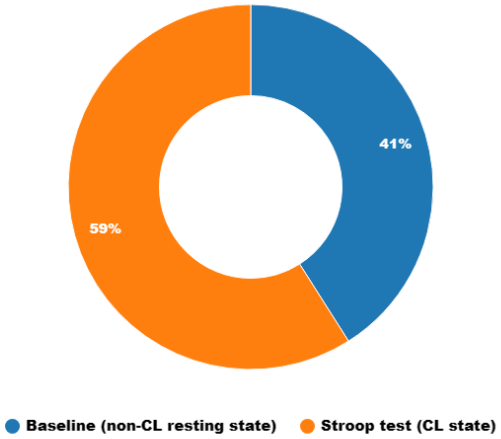


FIGURE 2. Cognitive load states sample distribution in the dataset.

and temporal (TP9 and TP10 sensors) brain regions. Fig. 1 depicts the electrode placement of the used device. The Fpz sensor served as a reference electrode, and data was sampled at a rate of 256 Hz. Electrode placement followed the standardized 10-20 system, which allows for the identification of specific brain regions (frontal, central, parietal, temporal, and occipital areas) using a standardized nomenclature for the sensors. This standard ensures the subject's analysis outcomes can be compiled and reproduced.

In our research, we only utilized the data collected by the Muse S EEG Headband, which included 4-channel EEG data.

D. SAMPLE DISTRIBUTION

In the recorded session data we received, not every participant had consistent EEG recordings throughout the entire experimental process. Consequently, we had to exclude one participant from the pilot study and five participants from the survey gamification experiment. This resulted in a total of 26 usable recorded sessions containing both baseline (representing non-cognitive load) and Stroop test (representing cognitive load) data. The fig. 2 illustrates the distribution of the final sample set.

IV. METHODOLOGY

In our study, the raw stream of the dataset at time step n can be expressed as $d_n = \{p_i, tp_9, af_7, af_8, tp_{10}\}$, where, p_i represents participant ID; tp_9 represents raw EEG recording of channel TP9; af_7 represents raw EEG recording of channel AF7; af_8 represents raw EEG recording of channel AF8; tp_{10} represents raw EEG recording of channel TP10.

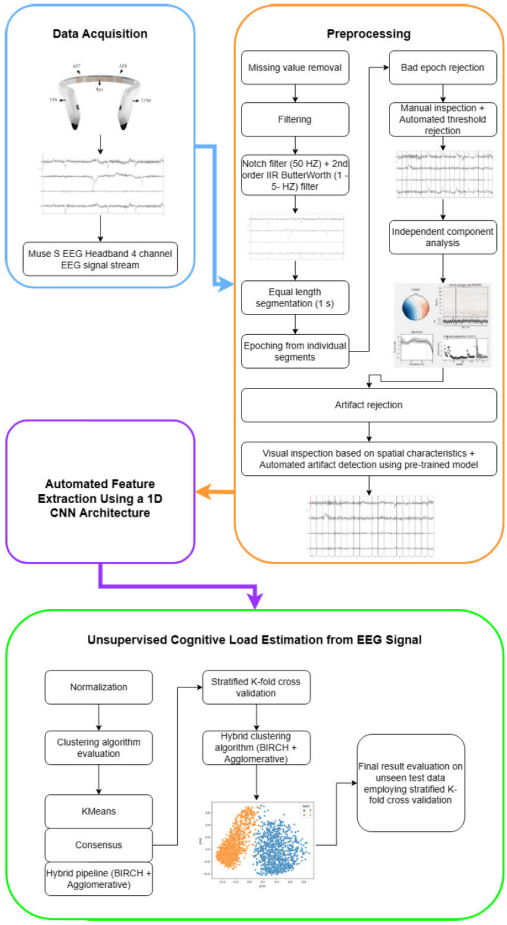
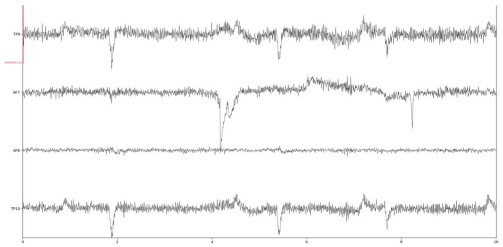


FIGURE 3. Workflow of the proposed pipeline.

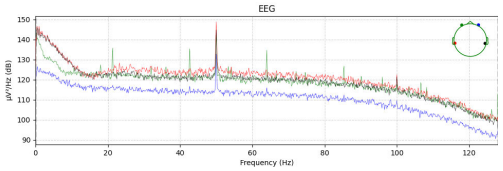
For cognitive load estimation from 4 channel EEG stream, we have to design a learning classifier that relates raw data d_n with cognitive load index cli_n using the function $R_\psi : d_n \rightarrow cli_n$, where, ψ is the parameter of the classifier. Constructing this system involved several steps, including data pre-processing, feature extraction, and unsupervised cognitive load (CL) estimation. These individual steps are discussed in detail in the following subsections. Fig. 3 illustrates the entire training phase process for this system. During the testing phase, the system remains largely identical, with the exception that the model is evaluated in the final step instead of undergoing training.

A. PRE-PROCESSING

Prior to training our estimator, it is essential to condition the data for the model. This data preparation procedure is detailed in this subsection. Initially, the data underwent cleaning by removing any missing values. Subsequently, various steps were implemented, encompassing filtering, segmentation and epoch formation, the rejection of problematic epochs (manual

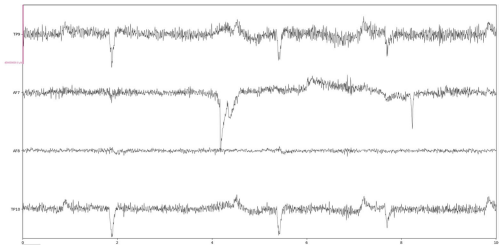


(a) Time domain plot of the signal snippet.

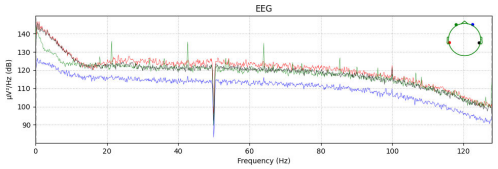


(b) Frequency domain PSD plot of the signal snippet.

FIGURE 4. Raw signal snippet.



(a) Time domain plot of the signal snippet.



(b) Frequency domain PSD plot of the signal snippet.

FIGURE 5. Notch filtered signal snippet.

inspection and automated rejection based on minimum and maximum peak-to-peak amplitude thresholds), independent component analysis, and the identification and removal of artifacts (visual inspection based on spatial characteristics and automated artifact detection using a pre-trained model). These steps are elaborated upon in the subsequent portion of this subsection.

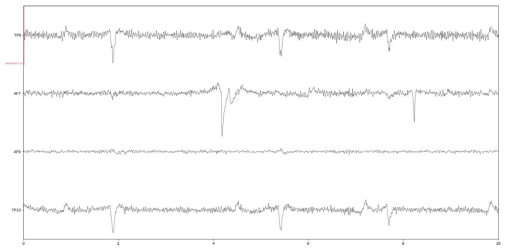
1) FILTERING

The EEG signal often suffers from interference caused by powerline noise, which can lead to distortion of the genuine brain wave patterns observed in the signal stream. To address this issue and eliminate powerline noise at 50Hz, we employed a notch filter. Fig. 4 shows the raw signal snippet of participant 2, and fig. 5 depicts the signal snippet after applying the notch filter.

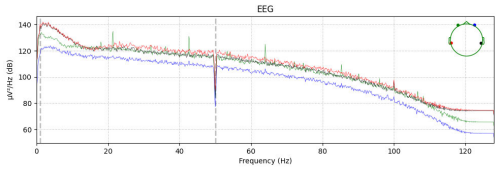
In a recent investigation, researchers in [52] examined the feasibility of passively monitoring cognitive workload using EEG while individuals engaged in a traditional n-back task within an interactive virtual reality (VR) environment. They extracted EEG spectral powers from eight electrode positions (Fz, F3, F4, C3, C4, P3, P4, and Pz) across four frequency bands (theta, alpha, beta, and gamma). The outcomes of their study indicated a positive correlation between alpha activity in the parietal area and levels of workload. In another experimental scenario, [44] employed theta and alpha band power to assess cognitive workload in a multitasking context. Participants were tasked with completing a challenge commonly utilized in airline pilot recruitment, which involved an increasing number of concurrent sub-tasks as the task progressed. EEG analysis was conducted on five electrodes located in the frontal area (Fz, F3, F4, F7, and F8) for the theta rhythm and five electrodes situated in the parietal area (Pz, P3, P4, P7, and P8) for the alpha rhythm. In addition to these EEG features, the researchers collected performance data, subjective assessments (NASA-TLX), and pupillometry measurements as comprehensive indicators of overall cognitive workload. The results indicated that both theta and alpha band powers increased with task complexity, highlighting the direct influence of these bands on CL. In another study, [45] assigned different levels of CL based on the linguistic complexity of the presented content. Their findings showed that theta oscillations are potentially an objective indicator of CL. While various indicators have been proposed in previous literature, it is imperative to explore the most optimal indices for assessing CL. The noteworthy research findings suggest that alpha and theta band power serve as direct indicators of fluctuations in CL. These compelling pieces of evidence have motivated our decision to place a strong emphasis on these aspects of EEG signals in our own research. The theta band power encompasses frequencies from 4 to 7.5 Hz, while the alpha band power falls within the range of 8 to 13 Hz. So the frequency range of 4 Hz to 13 Hz, with some saturating to both extremes, contains the most significant information about cognitive load, but substantially larger (e.g., >50 Hz) or smaller (e.g., 1 Hz) frequencies potentially do not include much valuable information about cognitive load states. Consequently, after implementing the notch filter to eliminate powerline noise, we applied a second-order Infinite Impulse Response (IIR) Butterworth filter. This filter featured a low cutoff frequency of 1 Hz and a high cutoff frequency of 50 Hz. This configuration was chosen to minimize redundant brain activity information while preserving the integrity of the essential frequency range. Fig. 6 visualizes the signal after applying the second-order IIR Butterworth filter.

One of the notable advantages of this filter design is its ability to maintain a maximally flat frequency response within the passband. This characteristic ensures minimal distortion of the signal within the passband, which is crucial for preserving signal characteristics. Additionally, Butterworth filters are inherently stable, guaranteeing reliability in their performance. Another noteworthy feature is the smooth and gradual roll-off from the passband to the stopband, which further contributes to the preservation of signal characteristics.

For these reasons, we incorporated this filter architecture into our study.



(a) Time domain plot of the signal snippet.



(b) Frequency domain PSD plot of the signal snippet.

FIGURE 6. Second-order IIR Butterworth filter applied signal snippet.

2) SEGMENTATION AND EPOCH CREATION

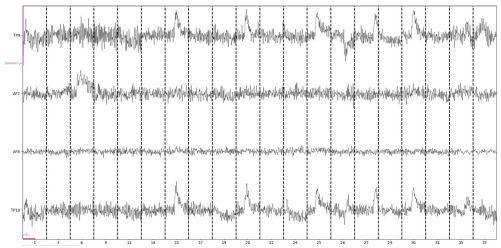
Due to the transient nature of EEG brain wave events, it is necessary to divide the raw data stream into appropriate segments to ensure that important information is not misaligned. To achieve this, each continuous stream from every session was divided into 1-second segments.

Once the continuous stream was segmented into equal lengths, we created epochs by taking these individual segments. Epochs provide a means of representing and analyzing continuous data by breaking it into time-locked, equally-sized trial chunks. They are particularly valuable for various statistical methods in neuroscience and allow for a quick overview of what happens during each trial. Epochs are commonly used to represent data that is time-locked to repeated experimental events, such as stimulus onsets or subject button presses. However, they can also be employed to store sequential or overlapping frames of continuous signals, for instance, in the analysis of resting-state and cognitive load activity. Segmenting the data into epochs facilitated the identification of problematic data streams and enabled us to rectify these issues through the rejection of faulty epochs.

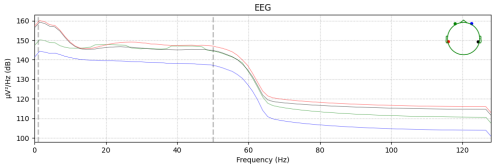
3) EPOCH REJECTION

Following the creation of epochs, the data undergoes a process of preparing for the identification and removal of problematic epochs. This involves two approaches: manual inspection to identify consecutive segments with no events and automated rejection based on minimum and maximum peak-to-peak amplitude thresholds. To establish these amplitude thresholds, we applied the automated rejection threshold estimation algorithm as proposed in [70].

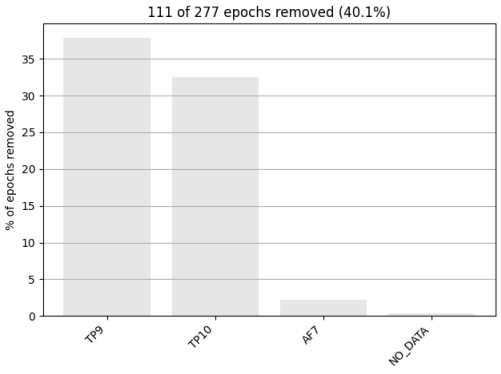
This algorithm functions by dividing the data stream into K folds for each threshold candidate. For each fold, a specific threshold rejection is applied. After rejection, the mean and median of the signal trial are calculated, and the error is assessed using the Frobenius norm. Subsequently, the error is linked to the corresponding thresholds. After evaluating all the candidates, the threshold associated with the least amount



(a) Time domain plot of the signal snippet.



(b) Frequency domain PSD plot of the signal snippet.



(c) Statistics for bad epoch rejection.

FIGURE 7. Bad epoch rejected signal snippet.

of error is selected. The automated nature of this method reduces the need for manual inspection, thus enhancing scalability and reliability.

Fig. 7 visualizes the statistics of the signal snippet after bad epoch rejection.

4) INDEPENDENT COMPONENT ANALYSIS

EEG signal recordings often suffer from interference caused by factors like eye blinks, muscle movements, and other anomalies [76], [77], [78], [79], [80]. Removing these artifacts is essential before using the data for analysis, as they introduce misleading non-brain-related information that can distort the results.

One widely adopted method for detecting and eliminating these artifacts is Independent Component Analysis (ICA). ICA is a computational technique applied in EEG data processing to break down the recorded EEG signals into a collection of statistically independent components. This decomposition proves valuable for artifact removal and the isolation of meaningful brain signals from sources of noise or disturbances.

The fundamental concept behind ICA is to identify a linear transformation of the mixed EEG signals that generates a set of independent components. These independent components possess statistical independence, indicating that they lack

shared information or correlations. ICA effectively disentangles the mixed EEG signals into distinct components, some of which represent genuine brain activity, while others correspond to various artifacts.

One significant advantage of ICA is that it doesn't necessitate prior knowledge about the sources of artifacts; it blindly separates the mixed signals solely based on their statistical independence. Subsequently, the separated independent components are scrutinized and analyzed to pinpoint which ones are associated with artifacts.

After ICA analysis, a total of 4 independent components were determined from the epoched EEG streams. Fig. 8 shows the 4 independent components obtained from the independent component analysis.

5) ARTIFACT REJECTION

After identifying the independent components, we distinguished the artifact components through a combination of visual inspection based on spatial characteristics and automated artifact detection using a pre-trained model outlined in [71]. This automated approach reduces the necessity for manual review, thereby improving scalability and reliability. Once the components associated with artifacts were detected, they were eliminated from the EEG data by subtracting them from the original recordings. This process effectively purifies the EEG data, leaving only the components related to brain activity. Following the removal of artifacts, the cleaned EEG data was reconstructed by reversing the ICA transformation, resulting in a set of EEG signals that were devoid of the identified artifacts.

Fig. 9 demonstrates the final clean signal snippet after artifact rejection process.

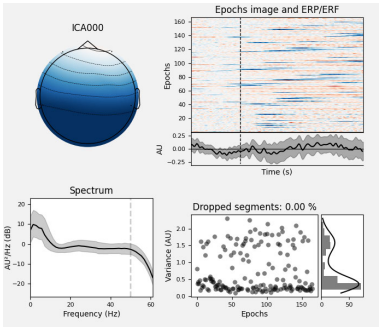
B. FEATURE EXTRACTION

In order to estimate cognitive load accurately, it is crucial to derive meaningful features from the raw EEG data. These extracted features enable us to examine intricate patterns, providing a deeper understanding of the characteristics of brain waves. The process of feature extraction is briefly outlined in the subsequent section of this document.

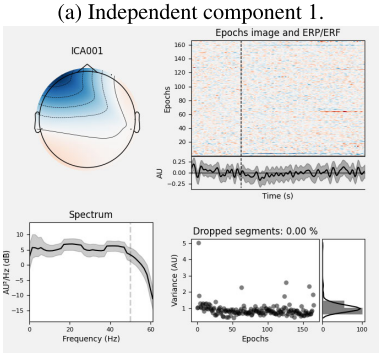
1) EXTRACTION STRATEGY

In our study, we chose an automated feature extraction approach using a 1D Convolutional Neural Network (CNN) pipeline instead of traditional manual feature engineering. Utilizing 1D CNNs for feature extraction in EEG data offers numerous advantages over the conventional handcrafted feature extraction methods.

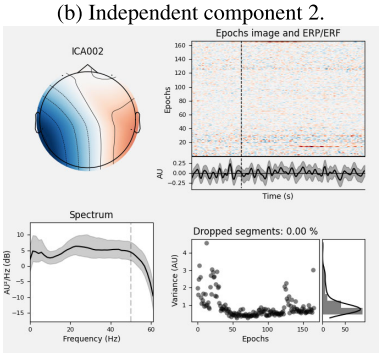
1D CNNs possess the ability to autonomously learn pertinent features from the raw EEG data, eliminating the need for domain-specific knowledge or manual feature crafting. This adaptability proves particularly valuable when dealing with intricate EEG patterns that might not be effectively captured by handcrafted features. Additionally, CNNs operate in a hierarchical manner, with early layers identifying basic patterns like edges and simple oscillations, while the deeper layers acquire more intricate and abstract representations. This hierarchical approach enables the discovery of informative EEG patterns that may elude traditional, handcrafted feature extraction methods. Furthermore, CNNs can automatically reduce data dimensionality by selecting the most pertinent features during training, addressing the



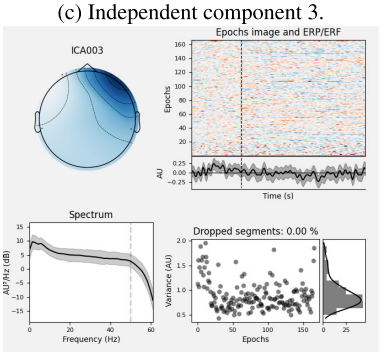
(a) Independent component 1.



(b) Independent component 2.

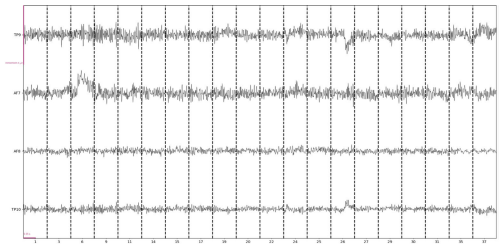


(c) Independent component 3.

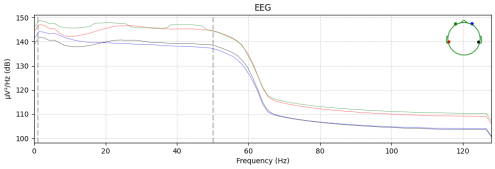


(d) Independent component 4.

FIGURE 8. 4 independent components after independent component analysis.



(a) Time domain plot of the signal snippet.



(b) Frequency domain PSD plot of the signal snippet.

FIGURE 9. **Artifact rejected clean signal snippet.**

challenges posed by the curse of dimensionality and yielding more efficient and proficient models.

In terms of generalization, properly regularized and diversified CNNs have the potential to perform well on unseen EEG data. This contrasts with handcrafted features, which may struggle to generalize effectively due to their reliance on specific assumptions about data distribution. Notably, manual feature extraction often involves subjective decisions by experts, introducing human bias and variability into the process. Manual feature extraction may also struggle to capture complex patterns and relationships within the data. Some patterns may be subtle or non-intuitive, making it challenging for human experts to identify and encode them as features. Furthermore, manual feature extraction may be sensitive to noise and outliers in the data and may not scale well with high-dimensional data. Considering these rationales, we made the decision to employ a 1D CNN architecture for automated feature extraction.

2) EXTRACTOR MODEL ARCHITECTURE

Our initial 1D CNN structure consisted of eight consecutive 1D convolutional layers. We used the original recording session types to categorize the input samples into CL and non-CL states to train the 1D CNN architecture for feature extraction. No additional data labeling was used for this purpose. In contrast to the earlier layers in the architecture, the intermediate layers incorporated a higher number of filters. We adopted this configuration to enhance the level of abstraction. Deeper layers are expected to capture more intricate and complex features, and by increasing the number of filters, we enabled the network to learn a broader spectrum of patterns and features within the data stream. Additionally, CNNs follow a hierarchical feature-learning process. The initial layers detect low-level features, while the deeper ones amalgamate these foundational features to identify more intricate objects or patterns. Enlarging the neuron count in the deeper layers allows for richer combinations of low-level features. Moreover, augmenting the number of neurons grants the network an increased capacity to represent complex relationships within the data, enhancing its capacity to learn from the training data.

TABLE 1. **Cluster quality evaluation of feature vectors from 2 different CNN architectures.**

Architecture Type	Homogeneity	Completeness	V Measure	Adjusted Rand Index	Adjusted Mutual Information	Silhouette Coefficient
Single Max Pooling Layer Architecture	0.097	0.098	0.098	0.176	0.097	0.185
Multi Max Pooling Layer Architecture	0.205	0.200	0.203	0.302	0.202	0.301

Following the final convolutional layer, we introduced a max-pooling layer, followed by a dropout layer, aiming to diminish dimensionality and enhance regularization. By extracting the feature vector from the layer just before the ultimate output layer, we obtained our feature representation. We employed the leaky ReLU activation function and the Adam optimizer for our feature extractor. The justification for using leaky ReLU is that it introduces a small, non-zero slope for negative inputs, ensuring that even when the input is negative, there is still a small gradient that allows weight updates. Additionally, leaky ReLU facilitates a more continuous gradient flow during back-propagation. This leads to improved training dynamics and faster convergence. This also helps to mitigate the saturation problem associated with standard ReLU. Saturation occurs when a large portion of the input space leads to the same output, resulting in a loss of gradient information. Leaky ReLU introduces a non-zero slope, making it less prone to saturation for negative inputs. On the other hand, as an optimizer, Adam includes an element of RMSprop, which involves dividing the learning rate for a parameter by the square root of the moving average of the squared gradients for that parameter. This element helps normalize the updates, especially when dealing with features with varying scales. Also, Adam performs bias correction to account for the fact that the moving averages of the gradients (first and second moments) are initialized with zeros. The bias correction helps in the early stages of training when the estimates of moments are less accurate. Furthermore, the adaptivity of the learning rates in Adam makes it robust to noisy or sparse gradients. This led us to employ Adam in our network architecture. While constructing our CNN architecture, we observed an intriguing phenomenon. Including an additional max-pooling layer after each convolutional layer, rather than just at the end, significantly improved the predictive capabilities for estimating cognitive load. To assess and compare prediction performance, we trained simple KMeans clustering algorithms using feature vectors from both architectures. The results of these two architectures are presented in the table 1 and 2. Notably, the addition of max-pooling layers after each convolutional layer substantially improved prediction performance, leading us to adopt this architecture as our final feature extraction pipeline.

Upon feature extraction using the proposed 1D CNN, we obtained a feature vector of size 128. Fig.10 visualizes the general structure of the feature extractor architecture.

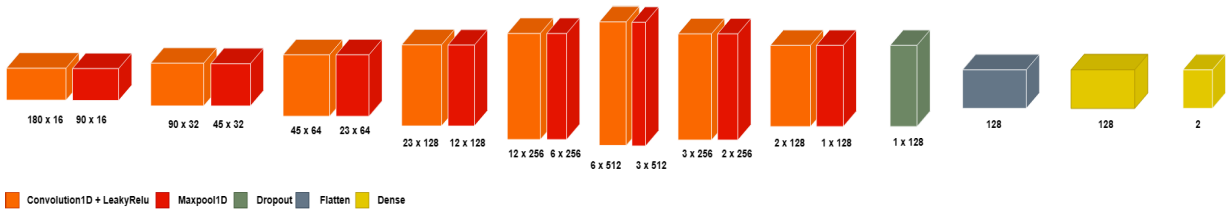


FIGURE 10. General structure of the feature extractor architecture.

TABLE 2. Cognitive load estimation performance evaluation of feature vectors from 2 different CNN architecture.

Architecture Type	Precision	Recall	F1 Score	Accuracy
Single Max Pooling Layer Architecture	68	68	68	72
Multi Max Pooling Layer Architecture	74	75	74.5	77.5

C. UNSUPERVISED CL ESTIMATION

For the task of estimating cognitive load from 4-channel EEG data, we opted for an unsupervised learning approach based on clustering, rather than a supervised one. There are several compelling reasons behind this choice. Unsupervised learning eliminates the need for labeled data during training, which can be particularly challenging and subjective in EEG cognitive load estimation. It enables us to derive patterns and relationships directly from the EEG data without the necessity of labor-intensive and potentially biased labeling efforts. Unsupervised learning techniques, such as clustering, have the capability to unveil concealed patterns and structures within the EEG data, often leading to insights and discoveries that might remain hidden in a supervised setting. Additionally, unsupervised learning methods exhibit adaptability to various EEG data types and tasks without the requirement to redefine labels or annotations for each specific task. This adaptability proves advantageous when working with diverse EEG datasets.

The details of our proposed unsupervised approach are elaborated upon in the subsequent section of this document.

1) NORMALIZATION

While participants generally exhibit common patterns across various features, the specific numerical values often vary from one individual to another, particularly in the case of brain wave patterns. Consequently, our prediction system focuses solely on the magnitude of these features, disregarding their units. This approach results in significant fluctuations in outcomes across individuals, as features with larger magnitudes disproportionately influence the distance estimation between sample points compared to features with smaller magnitudes. Consequently, our model may exhibit an erroneous bias toward features with high magnitudes. To mitigate this effect, it is essential to normalize all features to a consistent magnitude level. This is where feature normalization proves valuable.

In our specific case, we followed a two-step normalization process. This process has a comparatively higher computational overhead but provides scale invariance. This ensures that the normalized vectors have a constant norm, making them invariant to scaling. Initially, we subtracted the minimum value within each feature from every feature value and then divided the result by the feature's range. We employed the L2 norm, which involves taking the square root of the sum of squared values. By squaring the values, greater weight is assigned to larger values, while smaller values have less impact. The reason for using the L2 norm is that it scales the vectors to have a unit norm, which simplifies the interpretation of the data. Normalized vectors lie on the surface of a unit hypersphere, and their magnitudes represent relative proportions rather than absolute values. This can simplify computations and improve the interpretability of the learned weights. Thus the approach ensures computationally efficient solutions during the calculation process.

2) EVALUATION CRITERIA

For our analysis, the evaluations were conducted using cross-validation to ensure that no sample data was omitted from either the training or testing tasks. For our cross-validation approach, we adopted a stratified K-fold cross-validation algorithm. This algorithm offers several advantages. In stratified K-fold cross-validation, each fold or data partition maintains the same class distribution as the original dataset, guaranteeing that each fold comprises a representative sample of every class. By preserving the class distribution in each fold, stratified K-fold cross-validation effectively mitigates bias in model evaluation. It ensures that the model undergoes testing on all classes under consistent conditions, resulting in a more precise assessment of its generalization performance. When you keep the class distribution, the evaluation results are more reliable and consistent. This makes it easier to compare different models or hyperparameter configurations because the evaluation does not depend as much on picking random data points within each fold. Stratified K-fold cross-validation also gives more accurate estimates when figuring out performance metrics or confidence intervals for model evaluation because the class distribution is kept the same. This makes the evaluation results less variable.

In the analysis, we assessed not only the ultimate prediction performance but also the quality of the clusters themselves, preceding the computation of the final prediction scores. To assess the quality of our cluster models, we employed a variety of metrics, including the homogeneity score, com-

pleteness score, V-measure, adjusted Rand index, adjusted mutual information, and silhouette coefficient. These metrics offer insights into different aspects of clustering performance. The following portion of the literature will give a brief overview of all these metrics:

- Homogeneity: Homogeneity score measures the extent to which each cluster contains only data points from a single class, irrespective of the specific class labels. It is bounded between 0 and 1, with low values indicating low homogeneity. The term is defined with the following attribute:

$$h = 1 - \frac{H(X|K)}{H(X)} \tag{1}$$

  where h is the homogeneity score, functions $H(X|K)$ and $H(X)$ are defined by Shannon's entropy, X is the information of the number of samples labelled x in cluster k and K is information for the total number of samples in cluster k.
- Completeness: The completeness score evaluates whether all data points belonging to the same class are grouped together in a single cluster. This score is also bounded between 0 and 1. The following attribute is used to calculate the metric:

$$c = 1 - \frac{H(K|X)}{H(K)} \tag{2}$$

  where c is the completeness score, functions $H(K|X)$ and $H(K)$ are defined by Shannon's entropy, X is the information of the number of samples labelled x in cluster k and K is information for the total number of samples in cluster k.
- V Measure: The v measure is a harmonic mean between homogeneity and completeness. It is defined by the following attribute:

$$v = \frac{(1+\beta) * h * c}{\beta * h + c} \tag{3}$$

  where v is the v measure, h is the homogeneity score, c is the completeness score, and β is used to adjust the weight of h and c. V measure ranges from 0 to 1.
- Adjusted Rand Index: The Rand Index quantifies the similarity between two clusterings by considering how samples are assigned to clusters in both the predicted and true clusterings. It takes into account the fact that some agreement between two clusters can occur by chance, and it adjusts the rand index to account for this possibility. The score is calculated from the following attributes:

$$A_{ri} = \frac{(ri - eri)}{(max(ri) - eri)} \tag{4}$$

  where A_{ri} is the adjusted rand index, ri is the rand index, and eri is the expected rand index.
- Adjusted Mutual Information: Adjusted mutual information is an adjustment of the mutual information score to account for chance. It accounts for the fact that the mutual information is generally higher for two clusterings with a larger number of clusters, regardless of whether there is actually more information shared. For two clusterings A and B, the adjusted mutual information is given as:

$$ami_{A,B} = \frac{mi_{A,B} - emi_{A,B}}{avg(H(A), H(B)) - emi_{A,B}} \tag{5}$$

  where $ami_{A,B}$ is the adjusted mutual information, $mi_{A,B}$ is the mutual information, $H(A)$ and $H(B)$ are defined by Shannon's entropy for cluster A and B, and lastly $emi_{A,B}$ is the expected mutual information.
- Silhouette Coefficient: the Silhouette Coefficient is computed using both the mean intra-cluster distance and the mean nearest-cluster distance for each sample. A value of 1 represents ideal clustering, while negative values suggest misclassification. The silhouette coefficient ranges from -1 to 1 and is calculated using the following formula:

$$s = \frac{b - a}{max(a, b)} \tag{6}$$

  where s is the silhouette coefficient, b is the mean nearest-cluster distance, and a is the mean intra-cluster distance.

In our evaluation of prediction performance, while accuracy is a widely used metric, we also employed a range of other performance measures to obtain a more comprehensive understanding of the effectiveness of our proposed approach. These additional metrics include precision, recall, and the F1 score. A brief overview of these metrics is given in the following portion of the literature:

- Precision: Precision talks about out of all the predicted positive, how many of them are actual positive. It is calculated by the following attribute:

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

  where TP is the true positive and FP is the false positive.
- Recall: Recall is the true positive rate and is calculated with the following attribute:

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

  where TP is the true positive and FN is the false negative.
- F1 Score: F1-score is a metric which takes into account both precision and recall and is defined as follows:

$$f1 = 2 * \frac{precision * recall}{precision + recall} \tag{9}$$

- Accuracy: Accuracy represents the number of correctly classified data instances over the total number of data instances and is calculated using the following attribute:

$$accuracy = \frac{TN + TP}{TP + FP + TN + FN} \tag{10}$$

  where TN is the true negative, TP is the true positive, FN is the false negative, and FP is the false positive.

TABLE 3. Cluster quality evaluation of 3 different cluster algorithms.

Model	Homogeneity	Completeness	V Measure	Adjusted Rand Index	Adjusted mutual Information	Silhouette Coefficient
Kmeans	0.209	0.206	0.208	0.310	0.207	0.602
Consensus	0.322	0.315	0.319	0.444	0.318	0.417
BIRCH	0.650	0.652	0.651	0.723	0.651	0.410

TABLE 4. Cognitive load estimation performance evaluation of 3 different cluster algorithms.

Model	Precision	Recall	F1 Score	Accuracy
Kmeans	74.5	75	75	78
Consensus	81	82	82	84
BIRCH	90.5	90.5	90.5	91.5

3) CLUSTER ALGORITHM EVALUATION

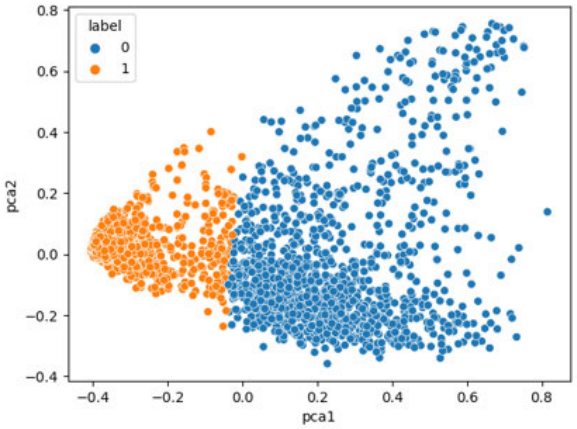
In our study, we conducted tests using three distinct clustering algorithms: KMeans clustering, Consensus clustering, and a hybrid approach involving the BIRCH algorithm in conjunction with agglomerative clustering.

Consensus clustering, regarded as an ensemble technique, is employed to consolidate predictions from multiple cluster models, resulting in improved final prediction performance compared to individual algorithms. Initially, we trained several KMeans cluster models. Subsequently, we constructed a cluster similarity matrix based on these initial KMeans clusters. Following this step, we applied a spectral clustering algorithm to amalgamate the outcomes of these initial KMeans cluster predictions, guided by the cluster similarity matrix.

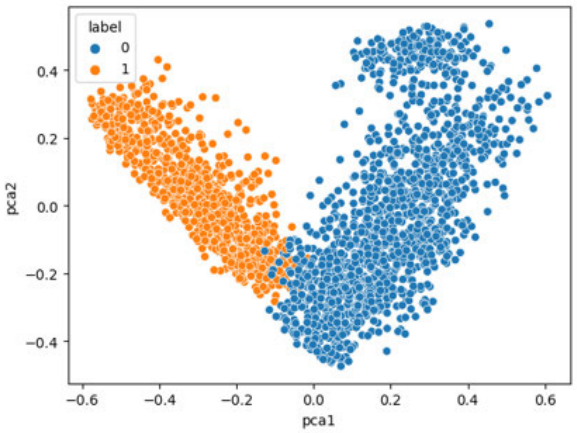
Regarding the hybrid pipeline, which combines BIRCH and the agglomerative algorithm, our approach entailed training a BIRCH model on the training data. BIRCH is an efficient, memory-friendly, online learning algorithm that constructs a tree data structure with cluster centroids derived from the leaf nodes. After generating these centroids, we utilised them as input for an agglomerative clustering model. The model was fit by treating the sub-clusters as new samples, and the initial data points were mapped to the label of the nearest sub-cluster.

To assess the performance of these three distinct clustering algorithms, we first created stratified folds from the training data. Stratified K-fold cross-validation was chosen for its advantage in maintaining the original category distribution within each fold, closely resembling the distribution in the original data. This approach ensures that no fold lacks any categories or classes during model training, thereby reducing bias toward any specific class.

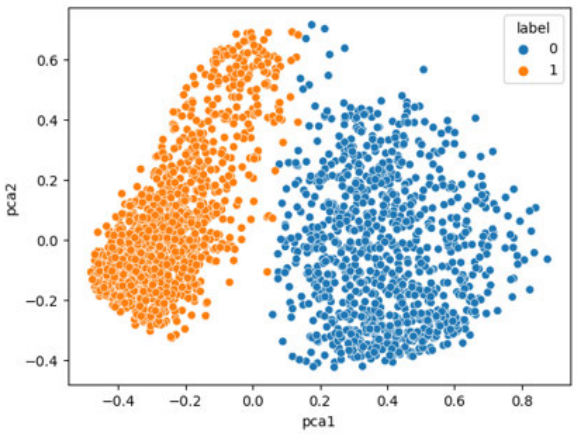
Table 3 and 4 display the evaluation results of the three different clustering algorithms used for a single fold. Fig. 11 depicts cluster plots of the 3 different clustering algorithms from the sample distribution's primary component analysis. From the figure, we can clearly see that the hybrid pipeline cluster has the best quality. Based on these results, we selected the hybrid pipeline of BIRCH and agglomerative clustering as our final estimator due to its superior performance.



(a) PCA cluster plot for the K-Means algorithm.



(b) PCA cluster plot for the Consensus algorithm.



(c) PCA cluster plot for the Hybrid (BIRCH + agglomerative) algorithm.

FIGURE 11. PCA cluster plot for the 3 different cluster algorithms.

4) EXECUTION ENVIRONMENT

For this work we used a Ryzen 7 5800H series cpu, 32 Gigabytes of ram, and a NVIDIA Geforce RTX 3050 Ti graphics unit with 4 gigabytes of vram. The component models and our pipeline training and testing time are provided in table 5.

TABLE 5. Pipeline execution environment details.

CPU	AMD Ryzen 7 5800H
GPU	NVIDIA GeForce RTX 3050 Ti
RAM	32 GB
Training Time	2 hrs 45 mins (approximately)
Testing Time	20 mins (approximately)

TABLE 6. Cluster quality evaluation of the final cluster algorithm with 10 fold stratified cross validation.

Fold	Homogeneity	Completeness	V Measure	Adjusted rand Index	Adjusted Mutual Information	Silhouette coefficient
1	0.8	0.839	0.818	0.873	0.818	0.499
2	0.635	0.638	0.636	0.705	0.636	0.476
3	0.67	0.665	0.667	0.748	0.667	0.45
4	0.708	0.737	0.721	0.786	0.721	0.481
5	0.664	0.654	0.658	0.724	0.658	0.5
6	0.689	0.691	0.689	0.757	0.689	0.442
7	0.715	0.715	0.705	0.787	0.704	0.501
8	0.727	0.745	0.735	0.816	0.735	0.5
9	0.691	0.7	0.695	0.773	0.695	0.46
10	0.685	0.709	0.697	0.761	0.696	0.481
Mean	0.6984	0.7093	0.7021	0.773	0.7019	0.479
STD	0.044523	0.057176	0.050194	0.047451	0.050201	0.021995

V. RESULT AND DISCUSSION

The findings from our study exhibited some impressive results establishing our model as a serious competitor to conventional supervised techniques. Our hybrid cluster model achieved an average homogeneity score of 0.7, indicating that the majority of samples within a cluster belong to a single class. For completeness measure, our model achieved an average score of 0.71, signifying that data points from the same class tend to be placed within the same cluster. Our achieved V-measure score is 0.7, indicating a balance between homogeneity and the completeness of clustering quality. The adjusted Rand index is designed to have a value close to 0.0 for random labeling, exactly 1.0 for identical clusterings, and values below -0.5 for highly dissimilar clusterings. For our model, the average adjusted Rand index is 0.77, suggesting a high degree of similarity between predicted and true clusterings. The adjusted mutual information returns a value of 1 when two partitions are identical and typically hovers around 0 for random partitions. Our model achieved an average adjusted mutual information score of 0.7, indicating a meaningful level of agreement between partitions. Lastly, our model achieved an average Silhouette coefficient score of 0.5, indicating generally well-separated clusters with a low amount of overlap. These metrics collectively provide a comprehensive evaluation of the quality and effectiveness of our cluster models. Table 6 shows the overall summary of these metrics for each fold in a 10-fold cross-validation process.

The results of our model's prediction performance across 10-fold cross-validation are presented in Table 7. Precision, which relates to the ability to minimize false positives, was achieved at an average rate of 93% with our model. Higher recall scores indicate lower false-negative rates, and our model achieved an average recall of 91.6%. The F1 score, which considers both precision and recall, averaged 92%. Additionally, our unsupervised predictor model demonstrated an average accuracy of 93.2%. The fig. 12 depicts the confusion matrix of the proposed architecture.

TABLE 7. Cognitive load estimation performance evaluation of the final cluster algorithms with 10-fold stratified cross-validation.

Fold	Precision	Recall	F1 Score	Accuracy
1	97.5	94	95.5	96.5
2	91	90	89.5	89.5
3	90.5	91.5	91	92.5
4	94.5	91.5	92.5	94
5	90	91.5	90.5	91.5
6	92	92	92	92.5
7	93.5	91	92	94
8	94.5	92.5	93.5	95
9	92.5	91.5	92	93.5
10	93.5	90.5	91.5	93
Mean	93	91.6	92	93.2
STD	2.254009	1.100505	1.649916	1.917753

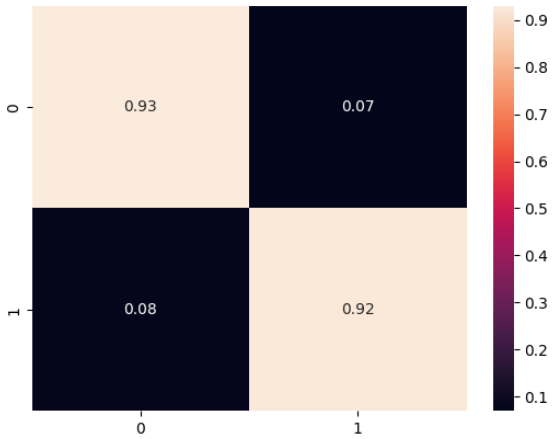


FIGURE 12. Normalized confusion matrix of the proposed model.

The scores obtained from the assessment of our proposed methodology exhibit considerable promise; however, certain caveats merit attention. Notably, while the quality of our clustering algorithm remains commendable, the homogeneity and v-measure of the clusters are in 0.7 mark in the range of 0 to 1. So there is a gap of roughly 0.3 for these scores with the highest standard. This limitation arises primarily due to the inherent complexity of our utilized dataset, which comprises solely resting baseline cognitive states and CL states, lacking more refined categories. Moreover, the participants experienced minimal interventions during recording sessions, a deliberate attempt to simulate a natural environment. Consequently, there exists the potential for the inclusion of cognitive load-inducing cogitation data with the non-CL resting state data. This overlap contributes to a reduction in both homogeneity and v-measure within the clusters associated with CL states. To address these challenges effectively, it is imperative to acquire datasets featuring precisely marked events for enhanced data granularity and reliability.

Despite encountering these challenges, our model asserts its competitive standing within the domain of supervised EEG-based cognitive load (CL) detection pipelines. A thorough performance comparison between our proposed method

TABLE 8. Performance comparison between other EEG-based supervised approaches and our proposed unsupervised method.

Method	Peak Precision	Peak Recall	Peak F1 Score	Peak Accuracy
Autoencoding + Transfer Learning [74]	NA	NA	71.17	74.55
Bi-LSTM Attention [72]	88	87	87	87
CNN [55]	NA	NA	NA	72.73
Linear Mixed Models + ERD/ERS [75]	NA	NA	NA	90
1D CNN Transfer Learning [73]	NA	NA	91.6	NA
Our Proposed Method	93	91.6	92	93.2

and recent works is meticulously presented in table 8. To ensure clarity and transparency, the comparison was performed between the original results of these works and our proposed method. Upon scrutinizing the results, it becomes evident that many of these prior works lack the comprehensive set of evaluation criteria employed in our study, wherein our model surpasses all counterparts. Noteworthy is the work by [73], which, utilizing the same dataset as the foundation for our research, approached our results closely in terms of the F1 score. However, it is crucial to highlight that their optimal performance was attained only subsequent to employing a model pre-trained on an additional dataset. In contrast, our methodology achieves these commendable results by employing a singular dataset. The robustness of our performance evaluation scores unequivocally affirms that our innovative unsupervised cluster-based approach to cognitive load estimation stands not only as a contender but, in fact, can outperform recent and prominent works reliant on supervised learning methodologies.

VI. CONCLUSION

This investigation endeavors to assess the real-time cognitive load instantaneously using 4-channel EEG data as an objective metric. Departing from prevalent practices that predominantly employ supervised methods in cognitive load studies, necessitating human intervention in data labeling and posing challenges in collecting diverse data at scale, our approach leverages an innovative unsupervised learning strategy employing a hybrid cluster-based model. This model proficiently estimates various cognitive load states without relying on continuous human pre-labeled data. Addressing the challenge of selecting optimal evaluation metrics, we departed from conventional practices and employed a diverse set of metrics for a comprehensive evaluation, including but not limited to the F1 score and accuracy. The intricate process of feature selection was streamlined through the utilization of automated feature extraction facilitated by a 1D CNN architecture. Rigorous validation ensured through cross-validation employing a spectrum of evaluation criteria, revealing promising results with an adjusted Rand index of 0.77, a silhouette coefficient of 0.5, and an accuracy rate of 93.2%. Despite these achievements, certain limitations persist, notably the restricted diversity of our dataset containing only two cognitive load states. To address this, we are actively in the process of developing a more varied dataset. Additionally, we acknowledge the need for broader scenario representation and aim to compensate for data scarcity by creating a diversified dataset in the future. Considering a homogeneity score of 0.7, indicating room for performance enhancement, we plan to explore a hybrid cognitive load index merging both subjective and objective measures. Envisioning a hybrid approach that integrates aspects of both supervised and unsupervised learning, we foresee the potential of unsupervised learning informing data exploration, feature extraction, or anomaly detection, subsequently integrated into a supervised model for cognitive load estimation. Through this study, we seek to establish a foundation for future research, laying the groundwork for advancements in real-time instantaneous cognitive load estimation, and contributing to the evolution of the field.

REFERENCES

[1] F. Paas, A. Renkl, and J. Sweller, "Cognitive load theory: Instructional implications of the interaction between information structures and cognitive architecture," *Instructional Sci.*, vol. 32, nos. 1–2, pp. 1–8, Jan. 2004.
[2] J. Sweller, "Element interactivity and intrinsic, extraneous, and germane cognitive load," *Educ. Psychol. Rev.*, vol. 22, no. 2, pp. 123–138, Jun. 2010, doi: 10.1007/s10648-010-9128-5.
[3] J. Sweller, J. J. G. van Merrienboer, and F. G. W. C. Paas, "Cognitive architecture and instructional design," *Educ. Psychol. Rev.*, vol. 10, no. 3, pp. 251–296, 1998, doi: 10.1023/a:1022193728205.
[4] F. Paas, A. Renkl, and J. Sweller, "Cognitive load theory and instructional design: Recent developments," *Educ. Psychologist*, vol. 38, no. 1, pp. 1–4, Jan. 2003.
[5] F. Paas, J. E. Tuovinen, H. Tabbers, and P. W. M. Van Gerven, "Cognitive load measurement as a means to advance cognitive load theory," *Educ. Psychologist*, vol. 38, no. 1, pp. 63–71, Jan. 2003.
[6] P. Antonenko, F. Paas, R. Grabner, and T. van Gog, "Using electroencephalography to measure cognitive load," *Educ. Psychol. Rev.*, vol. 22, no. 4, pp. 425–438, Dec. 2010.
[7] A.-M. Brouwer, M. A. Hogervorst, J. B. F. van Erp, T. Heffelaar, P. H. Zimmerman, and R. Oostenveld, "Estimating workload using EEG spectral power and ERPs in the n-back task," *J. Neural Eng.*, vol. 9, no. 4, Aug. 2012, Art. no. 045008, doi: 10.1088/1741-2560/9/4/045008.
[8] C. Mills, I. Fridman, W. Soussou, D. Waghray, A. M. Olney, and S. K. D'Mello, "Put your thinking cap on: Detecting cognitive load using EEG during learning," in *Proc. 7th Int. Learn. Analytics Knowl. Conf.*, Vancouver, BC, Canada, Mar. 2017, pp. 80–89.
[9] G. D. Logan, "Skill and automaticity: Relations, implications, and future directions," *Can. J. Psychol./Revue canadienne de psychologie*, vol. 39, no. 2, pp. 367–386, Jun. 1985, doi: 10.1037/h0080066.
[10] G. Borghini, P. Aricò, G. Di Flumeri, G. Cartocci, A. Colosimo, S. Bonelli, A. Golfetti, J. P. Imbert, G. Granger, R. Benhacene, S. Pozzi, and F. Babiloni, "EEG-based cognitive control behaviour assessment: An ecological study with professional air traffic controllers," *Sci. Rep.*, vol. 7, no. 1, p. 1, Apr. 2017, doi: 10.1038/s41598-017-00633-7.
[11] J. Sweller, J. J. G. van Merriënboer, and F. Paas, "Cognitive architecture and instructional design: 20 years later," *Educ. Psychol. Rev.*, vol. 31, no. 2, pp. 261–292, Jun. 2019.
[12] J. Sweller, "Measuring cognitive load," *Perspect. Med. EduC.*, vol. 7, no. 1, pp. 1–2, Jan. 2018, doi: 10.1007/s40037-017-0395-4.
[13] J. Zagermann, U. Pfeil, and H. Reiterer, "Measuring cognitive load using eye tracking technology in visual computing," in *Proc. 6th Workshop Beyond Time Errors Novel Eval. Methods for Visualizat.*, Oct. 2016, pp. 78–85, doi: 10.1145/2993901.2993908.
[14] A. Baddeley, "Working memory: The interface between memory and cognition," *J. Cognit. Neurosci.*, vol. 4, no. 3, pp. 281–288, Jul. 1992.
[15] N. Kumar and J. Kumar, "Measurement of cognitive load in HCI systems using EEG power spectrum: An experimental study," *Proc. Comput. Sci.*, vol. 84, pp. 70–78, 2016.
[16] G. S. Halford, R. Baker, J. E. McCredden, and J. D. Bain, "How many variables can humans process?" *Psychol. Sci.*, vol. 16, no. 1, pp. 70–76, Jan. 2005.
[17] T. Seufert, F. Wagner, and J. Westphal, "The effects of different levels of disfluency on learning outcomes and cognitive load," *Instructional Sci.*, vol. 45, no. 2, pp. 221–238, Apr. 2017.

[18] A. Skulmowski, "Is there an optimum of realism in computer-generated instructional visualizations?" *Educ. Inf. Technol.*, vol. 27, no. 7, pp. 10309–10326, Aug. 2022.

[19] A. Skulmowski and G. D. Rey, "The realism paradox: Realism can act as a form of signaling despite being associated with cognitive load," *Human Behav. Emerg. Technol.*, vol. 2, no. 3, pp. 251–258, Jul. 2020.

[20] A. Skulmowski and G. D. Rey, "Realism as a retrieval cue: Evidence for concreteness-specific effects of realistic, schematic, and verbal components of visualizations on learning and testing," *Human Behav. Emerg. Technol.*, vol. 3, no. 2, pp. 283–295, Apr. 2021.

[21] U. Tugtekin and H. F. Odabasi, "Do interactive learning environments have an effect on learning outcomes, cognitive load and metacognitive judgments?" *Educ. Inf. Technol.*, vol. 27, no. 5, pp. 7019–7058, Jun. 2022.

[22] A. Skulmowski and K. M. Xu, "Understanding cognitive load in digital and online learning: A new perspective on extraneous cognitive load," *Educ. Psychol. Rev.*, vol. 34, no. 1, pp. 171–196, Mar. 2022.

[23] A. Skulmowski and G. D. Rey, "Subjective cognitive load surveys lead to divergent results for interactive learning media," *Human Behav. Emerg. Technol.*, vol. 2, no. 2, pp. 149–157, Apr. 2020.

[24] M. C. Johnson-Glenberg and C. Megowan-Romanowicz, "Embodied science and mixed reality: How gesture and motion capture affect physics education," *Cognit. Research, Princ. Implications*, vol. 2, no. 1, p. 2, Dec. 2017, doi: 10.1186/s41235-017-0060-9.

[25] F. G. W. C. Paas and J. J. G. Van Merriënboer, "Variability of worked examples and transfer of geometrical problem-solving skills: A cognitive-load approach," *J. Educ. Psychol.*, vol. 86, no. 1, pp. 122–133, 1994.

[26] R. Sarailoo, K. Latifzadeh, S. H. Amiri, A. Bosaghzadeh, and R. Ebrahimpour, "Assessment of instantaneous cognitive load imposed by educational multimedia using electroencephalography signals," *Frontiers Neurosci.*, vol. 16, Aug. 2022, doi: 10.3389/fnins.2022.744737.

[27] F. G. Paas, "Training strategies for attaining transfer of problem-solving skill in statistics: A cognitive-load approach.," *J. Educ. Psychol.*, vol. 84, no. 4, pp. 429–434, 1992.

[28] M. Klepsch, F. Schmitz, and T. Seufert, "Development and validation of two instruments measuring intrinsic, extraneous, and germane cognitive load," *Frontiers Psychol.*, vol. 8, p. 2, Nov. 2017, doi: 10.3389/fpsyg.2017.01997.

[29] J. Leppink, F. Paas, C. P. M. Van Der Vleuten, T. Van Gog, and J. J. G. Van Merriënboer, "Development of an instrument for measuring different types of cognitive load," *Behav. Res. Methods*, vol. 45, no. 4, pp. 1058–1072, Dec. 2013.

[30] S. G. Hart and L. E. Staveland, "Development of NASA-TLX (task load index): Results of empirical and theoretical research," *Adv. Psychol.*, vol. 52, pp. 139–183, 1988.

[31] Ø. Anmarkrud, A. Andresen, and I. Bråten, "Cognitive load and working memory in multimedia learning: Conceptual and measurement issues," *Educ. Psychologist*, vol. 54, no. 2, pp. 61–83, Apr. 2019, doi: 10.1080/00461520.2018.1554484.

[32] R. Brunken, J. L. Plass, and D. Leutner, "Direct measurement of cognitive load in multimedia learning," *Educ. Psychologist*, vol. 38, no. 1, pp. 53–61, Jan. 2003.

[33] J. Frey, C. Mühl, F. Lotte, and M. Hachet, "Review of the use of electroencephalography as an evaluation method for human-computer interaction," 2013, *arXiv:1311.2222*.

[34] T. Q. Tran, R. L. Boring, D. D. Dudenhoeffer, B. P. Hallbert, M. D. Keller, and T. M. Anderson, "Advantages and disadvantages of physiological assessment for next generation control room design," in *Proc. IEEE 8th Human Factors Power Plants HPRCT 13th Annu. Meeting*, Aug. 2007, pp. 259–263, doi: 10.1109/HFPP.2007.4413216.

[35] Y. Shi, N. Ruiz, R. Taib, E. Choi, and F. Chen, "Galvanic skin response (GSR) as an index of cognitive load," in *Proc. Extend. Abstracts Hum. Fact. Comput. Syst.*, 2007, pp. 2651–2656.

[36] R. W. Backs and L. C. Walrath, "Eye movement and pupillary response indices of mental workload during visual search of symbolic displays," *Appl. Ergonom.*, vol. 23, no. 4, pp. 243–254, Aug. 1992.

[37] V. Skaramagkas, G. Giannakakis, E. Ktistakis, D. Manousos, I. Karatzanis, N. S. Tachos, E. Tripoliti, K. Marias, D. I. Fotiadis, and M. Tsiknakis, "Review of eye tracking metrics involved in emotional and cognitive processes," *IEEE Rev. Biomed. Eng.*, vol. 16, pp. 260–277, 2023, doi: 10.1109/RBME.2021.3066072.

[38] N. V. Thakor and S. Tong, "Advances in quantitative electroencephalogram analysis methods," *Annu. Rev. Biomed. Eng.*, vol. 6, no. 1, pp. 453–495, Aug. 2004.

[39] A. Gevins, "Neurophysiological measures of working memory and individual differences in cognitive ability and cognitive style," *Cerebral Cortex*, vol. 10, no. 9, pp. 829–839, Sep. 2000, doi: 10.1093/cercor/10.9.829.

[40] A. Murata, "An attempt to evaluate mental workload using wavelet transform of EEG," *Human Factors: J. Human Factors Ergonom. Soc.*, vol. 47, no. 3, pp. 498–508, Sep. 2005, doi: 10.1518/001872005774860096.

[41] J. Xu and B. Zhong, "Review on portable EEG technology in educational research," *Comput. Hum. Behav.*, vol. 81, pp. 340–349, Apr. 2018, doi: 10.1016/j.chb.2017.12.037.

[42] E. Başar, C. Başar-Eroglu, S. Karakaş, and M. Schürmann, "Gamma, alpha, delta, and theta oscillations govern cognitive processes," *Int. J. Psychophysiology*, vol. 39, nos. 2–3, pp. 241–248, Jan. 2001, doi: 10.1016/s0167-8760(00)00145-8.

[43] M. Mazher, A. Abd Aziz, A. S. Malik, and H. Ullah Amin, "An EEG-based cognitive load assessment in multimedia learning using feature extraction and partial directed coherence," *IEEE Access*, vol. 5, pp. 14819–14829, 2017, doi: 10.1109/ACCESS.2017.2731784.

[44] S. Puma, N. Matton, P.-V. Paubel, É. Raufaste, and R. El-Yagoubi, "Using theta and alpha band power to assess cognitive workload in multitasking environments," *Int. J. Psychophysiology*, vol. 123, pp. 111–120, Jan. 2018, doi: 10.1016/j.ijpsycho.2017.10.004.

[45] L. J. Castro-Meneses, J.-L. Kruger, and S. Doherty, "Validating theta power as an objective measure of cognitive load in educational video," *Educ. Technol. Res. Develop.*, vol. 68, no. 1, pp. 181–202, Feb. 2020, doi: 10.1007/s11423-019-09681-4.

[46] A. Gevins, "High-resolution EEG mapping of cortical activation related to working memory: Effects of task difficulty, type of processing, and practice," *Cerebral Cortex*, vol. 7, no. 4, pp. 374–385, Jun. 1997, doi: 10.1093/cercor/7.4.374.

[47] N. A. Herweg, E. A. Solomon, and M. J. Kahana, "Theta oscillations in human memory," *Trends Cognit. Sci.*, vol. 24, no. 3, pp. 208–227, Mar. 2020, doi: 10.1016/j.tics.2019.12.006.

[48] P. Sauseng, B. Griesmayr, R. Freunberger, and W. Klimesch, "Control mechanisms in working memory: A possible function of EEG theta oscillations," *Neurosci. Biobehavioral Rev.*, vol. 34, no. 7, pp. 1015–1022, Jun. 2010, doi: 10.1016/j.neubiorev.2009.12.006.

[49] D. L. Schacter, "EEG theta waves and psychological phenomena: A review and analysis," *Biol. Psychol.*, vol. 5, no. 1, pp. 47–82, Mar. 1977, doi: 10.1016/0301-0511(77)90028-x.

[50] M. B. Sterman, C. A. Mann, D. A. Kaiser, and B. Y. Suyenobu, "Multiband topographic EEG analysis of a simulated visuomotor aviation task," *Int. J. Psychophysiology*, vol. 16, no. 1, pp. 49–56, Feb. 1994, doi: 10.1016/0167-8760(94)90041-8.

[51] M. Plechawska-Wójcik, M. Tokovarov, M. Kaczorowska, and D. Zapala, "A three-class classification of cognitive workload based on EEG spectral data," *Appl. Sci.*, vol. 9, no. 24, p. 5340, 2019, doi: 10.3390/app9245340.

[52] C. Tremmel, C. Herff, T. Sato, K. Rechowicz, Y. Yamani, and D. J. Krusienski, "Estimating cognitive workload in an interactive virtual reality environment using EEG," *Frontiers Human Neurosci.*, vol. 13, p. 401, Nov. 2019, doi: 10.3389/fnhum.2019.00401.

[53] S. Baceviciute, A. Mottelson, T. Terkildsen, and G. Makransky, "Investigating representation of text and audio in educational VR using learning outcomes and EEG," in *Proc. CHI Conf. Human Factors Comput. Syst.* New York, NY, USA: Association for Computing Machinery, Apr. 2020, pp. 1–13, doi: 10.1145/3313831.3376872.

[54] C. Scharinger, A. Schüler, and P. Gerjets, "Using eye-tracking and EEG to study the mental processing demands during learning of text-picture combinations," *Int. J. Psychophysiology*, vol. 158, pp. 201–214, Dec. 2020, doi: 10.1016/j.ijpsycho.2020.09.014.

[55] A. Appriou, A. Cichocki, and F. Lotte, "Modern machine-learning algorithms: For classifying cognitive and affective states from electroencephalography signals," *IEEE Syst., Man, Cybern., Mag.*, vol. 6, no. 3, pp. 29–38, Jul. 2020, doi: 10.1109/MSMC.2020.2968638.

[56] R. F. Rojas, E. Debie, J. Fidock, M. Barlow, K. Kasmarik, S. Anavatti, M. Garratt, and H. Abbass, "Electroencephalographic workload indicators during teleoperation of an unmanned aerial vehicle shepherding a swarm of unmanned ground vehicles in contested environments," *Frontiers Neurosci.*, vol. 14, p. 40, Feb. 2020, doi: 10.3389/fnins.2020.00040.

[57] A. Asif, M. Majid, and S. M. Anwar, "Human stress classification using EEG signals in response to music tracks," *Comput. Biol. Med.*, vol. 107, pp. 182–196, Apr. 2019.

[58] R. Katmah, F. Al-Shargie, U. Tariq, F. Babiloni, F. Al-Mughairbi, and H. Al-Nashash, "A review on mental stress assessment methods using EEG signals," *Sensors*, vol. 21, no. 15, p. 5043, Jul. 2021.
[59] J. W. Ahn, Y. Ku, and H. C. Kim, "A novel wearable EEG and ECG recording system for stress assessment," *Sensors*, vol. 19, no. 9, p. 1991, Apr. 2019.
[60] F. Zeng, P. Siriaraya, D. Choi, and N. Kuwahara, "Textile EEG cap using dry-comb electrodes for emotion detection of elderly people," *Int. J. Adv. Comput. Sci. Appl.*, vol. 11, no. 4, 2020.
[61] P. Arpaia, N. Moccaldi, R. Prevete, I. Sannino, and A. Tedesco, "A wearable EEG instrument for real-time frontal asymmetry monitoring in worker stress analysis," *IEEE Trans. Instrum. Meas.*, vol. 69, no. 10, pp. 8335–8343, Oct. 2020.
[62] A. Arsalan, M. Majid, A. R. Butt, and S. M. Anwar, "Classification of perceived mental stress using a commercially available EEG headband," *IEEE J. Biomed. Health Informat.*, vol. 23, no. 6, pp. 2257–2264, Nov. 2019.
[63] P. Pandey and K. P. Miyapuram, "Brain2depth: lightweight CNN model for classification of cognitive states from EEG recordings," in *Proc. Annu. Conf. Med. Image Understand. Anal.* Cham, Switzerland: Springer, 2021, pp. 394–407, doi: 10.1007/978-3-030-80432-9_30.
[64] L. E. Ismail and W. Karwowski, "Applications of EEG indices for the quantification of human cognitive performance: A systematic review and bibliometric analysis," *PLoS ONE*, vol. 15, no. 12, Dec. 2020, Art. no. e0242857.
[65] L. Longo, C. D. Wickens, G. Hancock, and P. A. Hancock, "Human mental workload: A survey and a novel inclusive definition," *Frontiers Psychol.*, vol. 13, p. 3, Jun. 2022.
[66] B. Raufi and L. Longo, "An evaluation of the EEG alpha-to-theta and theta-to-alpha band ratios as indexes of mental workload," *Frontiers Neuroinform.*, vol. 16, p. 3, May 2022.
[67] Y. Liu, Y. Yu, Z. Ye, M. Li, Y. Zhang, Z. Zhou, D. Hu, and L.-L. Zeng, "Fusion of spatial, temporal, and spectral EEG signatures improves multilevel cognitive load prediction," *IEEE Trans. Hum.-Mach. Syst.*, vol. 53, no. 2, pp. 357–366, Apr. 2023, doi: 10.1109/THMS.2023.3235003.
[68] L. Pang, L. Guo, J. Zhang, X. Wanyan, H. Qu, and X. Wang, "Subject-specific mental workload classification using EEG and stochastic configuration network (SCN)," *Biomed. Signal Process. Control*, vol. 68, Jul. 2021, Art. no. 102711.
[69] M. K. Grzeszczyk, R. Blanco, P. Adamczyk, M. Kus, S. Marek, R. Precikowski, and A. Lisowska. (2023). *CogWear: Can We Detect Cognitive Effort With Consumer-Grade Wearables?*. version 1.0.0. PhysioNet. [Online]. Available: https://doi.org/10.13026/5f6t-b637
[70] M. Jas, D. A. Engemann, Y. Bekhti, F. Raimondo, and A. Gramfort, "Autoreject: Automated artifact rejection for MEG and EEG data," *NeuroImage*, vol. 159, pp. 417–429, Oct. 2017, doi: 10.1016/j.neuroimage.2017.06.030.
[71] A. Li, J. Feitelberg, A. P. Saini, R. Höchenberger, and M. Scheltienne, "MNE-ICALabel: Automatically annotating ICA components with ICLabel in Python," *J. Open Source Softw.*, vol. 7, no. 76, p. 4484, Aug. 2022, doi: 10.21105/joss.04484.
[72] G. Yoo, H. Kim, and S. Hong, "Prediction of cognitive load from electroencephalography signals using long short-term memory network," *Bioengineering*, vol. 10, no. 3, p. 361, Mar. 2023, doi: 10.3390/bioengineering10030361.
[73] M. K. Grzeszczyk, P. Adamczyk, S. Marek, R. Pręcikowski, M. Kuś, M. P. Lelujko, R. Blanco, T. Trzciński, A. Sitek, M. Malawski, and A. Lisowska, "Can gamification reduce the burden of self-reporting in mHealth applications? A feasibility study using machine learning from smartwatch data to estimate cognitive load," 2023, *arXiv:2302.03616*.
[74] D. Pulver, P. Angkan, P. Hungler, and A. Etemad, "EEG-based cognitive load classification using feature masked autoencoding and emotion transfer learning," in *Proc. Int. Conf. MULTIMODAL Interact.*, New York, NY, USA, Oct. 2023, pp. 190–197, doi: 10.1145/3577190.3614113.
[75] C. Saitis, M. Z. Parvez, and K. Kalimeri, "Cognitive load assessment from EEG and peripheral biosignals for the design of visually impaired mobility aids," *Wireless Commun. Mobile Comput.*, vol. 2018, pp. 1–9, Feb. 2018.
[76] S. Mahmud, M. S. Hossain, M. E. H. Chowdhury, and M. B. I. Reaz, "MLMRS-net: Electroencephalography (EEG) motion artifacts removal using a multi-layer multi-resolution spatially pooled 1D signal reconstruction network," *Neural Comput. Appl.*, vol. 35, no. 11, pp. 8371–8388, Apr. 2023.
[77] M. S. Hossain, M. E. H. Chowdhury, M. B. I. Reaz, S. H. M. Ali, A. A. A. Bakar, S. Kiranyaz, A. Khandakar, M. Alhatou, R. Habib, and M. M. Hossain, "Motion artifacts correction from single-channel EEG and fNIRS signals using novel wavelet packet decomposition in combination with canonical correlation analysis," *Sensors*, vol. 22, no. 9, p. 3169, Apr. 2022.
[78] Md. S. Hossain, M. B. I. Reaz, M. E. H. Chowdhury, S. H. M. Ali, A. A. A. Bakar, S. Kiranyaz, A. Khandakar, M. Alhatou, and R. Habib, "Motion artifacts correction from EEG and fNIRS signals using novel multiresolution analysis," *IEEE Access*, vol. 10, pp. 29760–29777, 2022.
[79] G. S. Spencer, J. A. Smith, M. E. H. Chowdhury, R. Bowtell, and K. J. Mullinger, "Exploring the origins of EEG motion artefacts during simultaneous fMRI acquisition: Implications for motion artefact correction," *NeuroImage*, vol. 173, pp. 188–198, Jun. 2018.
[80] M. B. Hossain, M. A. Iftekhar, and M. E. H. Chowdhury, "Comparative study of different pulse artefact correction techniques during concurrent EEG-FMRI using FMRIB," *J. Multidisciplinary Res.*, vol. 4, no. 3, pp. 93–99, 2016.
[81] K. Kyriaki, D. Koukopoulos, and C. A. Fidas, "A comprehensive survey of EEG preprocessing methods for cognitive load assessment," *IEEE Access*, vol. 12, pp. 23466–23489, 2024, doi: 10.1109/ACCESS.2024.3360328.
[82] Q. Wang, D. Smythe, J. Cao, Z. Hu, K. J. Proctor, A. P. Owens, and Y. Zhao, "Characterisation of cognitive load using machine learning classifiers of electroencephalogram data," *Sensors*, vol. 23, no. 20, p. 8528, Oct. 2023, doi: 10.3390/s23208528.

IQBAL HASSAN received the B.Sc. degree from the Ahsanullah University of Science and Technology, in 2022. He is currently pursuing the M.Sc. degree with the Department of Human Intelligence Systems, Kyushu Institute of Technology, Fukuoka, Japan. He was a Research Engineer with United International University before joining Qatar University's Collaborative Machine Learning Group as a Research Engineer. He is a Graduate Research Assistant with Sozo Laboratory, Kyushu Institute of Technology. He is being funded by Japan Science and Technology Agency (JST) for his research work. His current research interests include human behavior analytics, behavioral pattern recognition, healthcare and wellbeing technology, and human activity recognition.

MONICA ZOLEZZI received the bachelor's degree in pharmacy and biochemistry from Catholic University Santa Maria in Arequipa, Peru, in 1985, the Master of Science degree from the Faculty of Pharmacy and Pharmaceutical Sciences, University of Alberta, Edmonton, Canada, in 1988, and the Ph.D. degree from the School of Pharmacy, The University of Auckland, New Zealand. Then, she completed an accredited Hospital Pharmacy Residency from the Calgary General Hospital, in 1989. In 2013, she received the Advanced Prescribing Authorization from the Alberta College of Pharmacists, Canada. She is currently an Associate Professor and the Director of the Doctor in Pharmacy Program, College of Pharmacy, Qatar University. Throughout her career, she has participated in several research projects and has authored a number of peer-reviewed publications, presentations, and research posters in the area of clinical and hospital pharmacy practice. Her research focuses on rational use of medications in people with mental illness and in the geriatric population, implementation of pharmacist-led programs for the prevention and management of cardiovascular disease, and using technology for the advancement of pharmacy practice and education.

HANAN KHALIL received the Ph.D. degree from Cardiff University, U.K. She is an Associate Professor of neurological rehabilitation with the Department of Rehabilitation Sciences, College of Health Sciences, Qatar University. She is a Researcher in the area of rehabilitation and exercise and related outcome measures in people with long term neurological conditions (LTNCs) namely multiple sclerosis (MS), dementia and Parkinson's disease (PD). She successfully received funding (as a principal investigator) and led clinical several studies (both observational and interventional) in people with neurological diseases and mental health issues. Further, her research interests include cross-cultural adaptation and validation of outcome measures in Arabic language; motor learning; and the impact of exercise on improving non-motor symptoms in these populations, such as sleep, cognition, an mood. More recently, she is also involved in the area of mental health, specifically in the area of trauma-informed care for traumatized individuals. As part of the Ph.D. degree, she has been actively involved in these areas of research, for ten years. She was a member of the International Parkinson's Disease and Movement Disorders Society (MDS) task force for the Middle East and the MDS Working Group for the Middle East. Through this, she led two multi-site projects for translating and validating standardized outcome measures in PD in Arabic language. She also led several MDS educational courses in the area of rehabilitation for people with neurological diseases for the allied health professions. As a recognition of her research and scholarly activities, she received the L'Oréal UNESCO for Women In Science Levant Fellowship, in 2018. Further, she was selected for the TechWomen Emerging Leader Program Award and the International Parkinson's Disease and Movement Disorders' Society (MDS) LEAP Program for Young Leaders of 2019. Further, she was selected to participate in a number of WHO advisory committees in the area of brain health and neurodegeneration.

RAFIF MAHMOOD AL SAADY received the M.B.Ch.B. degree from Basrah Medical College, Iraq, in 1991. She started her teaching career with Basrah Medical College, as a Lecturer for undergraduate and postgraduate medical and pharmacy students. In addition, she was a Consultant Pathologist with Basrah Teaching Hospital; and the Head of the Cytology Laboratory, Basrah Centre of Early Detection of Breast Cancer; and participating in Cancer Control Program in Basrah. In 2005, she moved to Qatar and established the Histopathology and Cytology Laboratory, Al Ahli Hospital. She joined College of Medicine, Qatar University, in 2020, as an Assistant Professor of pathology. She received the fellowship in pathology from the Iraqi Board for Medical Specialization, in 1998. She received the European Board of Pathology, in 2008; and a Diplomate membership from the Royal College of Pathologists, U.K., in 2019. She has published many peer-reviewed academic works in pathology. Her main interests include early diagnosis of cancer, putting strategies for cancer screening programs and control, targeted therapy, and cancer biology and behavior, in addition to understanding disease pathology and different diagnostic modalities and artificial inelegance in medicine.

SHONA PEDERSEN received the Ph.D. degree in medical biochemistry from Aalborg University, Denmark, in 2004. Following the Ph.D. study, she was a Senior Scientist with the Department of Clinical Biochemistry, Aalborg University Hospital, from 2004 to 2021, where she was an instrumental in founding and developing the Proteomics and Metabolomics Research Centre. From 2009 to 2015, she maintained a dual affiliation with the Department of Biomedicine and Clinical Medicine, Aalborg University, Denmark, and was appointed as an Associate Professor, in 2016. In August 2021, she joined Qatar University as an Associate Professor of biochemistry with the Department of Basic Medical Science, College of Medicine. Her academic contributions include over 60 scientific publications and two patents. She has also been an active supervisor and an examiner of many undergraduate and master's national and international students and has guided five Ph.D. students to successful completion. Currently, she leads the Neuro-Cardio Networking Group, Qatar University, a role in which she has been acknowledged as a distinguished leader by the Vice President of the QU Health Cluster. She serves as a Board Member of the Clinical Cancer Research, Denmark. Her noteworthy contributions were recently honored with a Research Achievement Award for her groundbreaking work titled "Decoding Blood Mysteries: A Beacon of Hope for Early Detection of Multiple Myeloma." She serves as an Associate Editor for *Molecular Signaling*, a specialty section of *Frontiers in Molecular Neuroscience*.

MUHAMMAD E. H. CHOWDHURY (Senior Member, IEEE) received the Ph.D. degree from the University of Nottingham, U.K., in 2014. He was a Postdoctoral Research Fellow with the Sir Peter Mansfield Imaging Centre, University of Nottingham. He is currently an Assistant Professor and a Program Coordinator with the Department of Electrical Engineering, Qatar University. He has filed several patents and published more than 180 peer-reviewed journal articles, more than 30 conference papers, and several book chapters. He is currently running NPRP, UREP, and HSREP grants from Qatar National Research Fund (QNRF); and internal grants (IRCC and HIG) from Qatar University along with academic projects from HBKU and HMC. His current research interests include biomedical instrumentation, signal processing, wearable sensors, medical image analysis, machine learning and computer vision, embedded system design, and simultaneous EEG/fMRI. He is a member of British Radiology, ISMRM, and HBM. He received the COVID-19 Dataset Award, the AHS Award from HMC, and the National AI Competition Awards for his contribution to the fight against COVID-19. His team was the Gold-Medalist in the 13th International Invention Fair in the Middle East (IIFME). He has been listed among the top 2% of scientists in the World List, published by Stanford University. He is serving as a Guest Editor for *Polymers*, an Associate Editor for IEEE Access, and a Topic Editor and a Review Editor for *Frontiers in Neuroscience*.

Open Access funding provided by 'Qatar National Library' within the CRUI CARE Agreement