Journal of Neural Engineering

EEG workload estimation and classification: a systematic review

To cite this article before publication: Jahid Hassan *et al* 2024 *J. Neural Eng.* in press https://doi.org/10.1088/1741-2552/ad705e

Journal of Neural Engineering

IOP Publishing

J. Neural Eng. (2024)

Topical Review

EEG Workload Estimation and Classification: A Systematic Review

Jahid Hassan[1,*], Md. Shamim Reza[1], Syed Udoy Ahmed[1], Nazmul Haque Anik[1], and Md. Obaydullah Khan[1]

[1] Department of Electrical and Electronic Engineering, Pabna University of Science and Technology, Pabna-6600, Bangladesh

[*] Authors to whom any correspondence should be addressed.

E-mail: kamranhassanjahid@gmail.com

Received xxxxxx

Accepted for publication xxxxxx

Published xxxxxx

Abstract

Objective. Electroencephalography (EEG) has evolved into an indispensable instrument for estimating cognitive workload in various domains. ML and DL techniques have been increasingly employed to develop accurate workload estimation and classification models based on EEG data. The goal of this systematic review is to compile the body of research on EEG workload estimation and classification using ML and DL approaches. *Methods.* The PRISMA procedures were followed in conducting the review, searches were conducted through databases at SpringerLink, ACM Digital Library, IEEE Explore, PUBMED, and Science Direct from the beginning to the end of February 16, 2024. Studies were selected based on predefined inclusion criteria. Data were extracted to capture study design, participant demographics, EEG features, ML/DL algorithms, and reported performance metrics. *Results.* Out of the 125 items that emerged, 33 scientific papers were fully evaluated. The study designs, participant demographics, and EEG workload measurement and categorization techniques used in the investigations differed. SVM, CNN, and hybrid networks are examples of ML and DL approaches that were often used. Analyzing the accuracy scores achieved by different ML/DL models. Furthermore, a relationship was noted between sample frequency and model accuracy, with higher sample frequencies generally leading to improved performance. The percentage distribution of ML/DL methods revealed that SVMs, CNNs, and RNNs were the most commonly utilized techniques, reflecting their robustness in handling EEG data. *Significance.* The comprehensive review emphasizes how ML may be used to identify mental workload across a variety of disciplines using EEG data. Optimizing practical applications requires multimodal data integration, standardization efforts, and real-world validation studies. These systems will also be further improved by addressing ethical issues and investigating new EEG properties, which will improve human-computer interaction and performance assessment.

Keywords: Deep learning (DL), electroencephalogram (EEG), machine learning (ML), Mental Workload (MWL).

J. Neural Eng. (2024)

1. Introduction

Due to human error being higher than that of contemporary machinery, human factors are responsible for a large number of industrial and manufacturing mishaps [1–3]. Many scholars are interested in researching mental workload (MWL) to prevent these tragedies. The term "MWL" describes the psychophysiological load that operators experience while executing particular cognitive activities, most of which are connected to emotion, situation awareness, and alertness [4]. Three categories apply to the latest MWL classification techniques: (1) Subjective evaluation of some psychological markers (exhaustion, exertion, and anxiety) [5]- it is simple to assess, but incorrect inferences would be made, particularly if the operator is mentally unstable.; (2) Assessment of task execution [6]-it is a type of objective measuring technique. It corrects some of the subjective assessment's flaws. If the operator performs the work more effectively, it is thought that they are in a better functional condition; if not, they are stated as being in a highly hazardous or susceptible functional condition. However, even though they are already mentally exhausted, certain proficient or well-trained operators may persevere and finish the jobs with seeming good performance. The high-level MWL cannot be measured with this method. (3) Data-driven evaluation [7]- it is a type of precise and objective measurement technique. Various physiological measurements are mostly used to continuously estimate the body's physical responses. These measurements include heart rate, respiration rate, eye and voice activity, and brain activity [8]. Current developments within the discipline of BCI (brain-computer interfaces) indicate an advantageous pattern in the use of mental effort for MWL [9]. BCI frequently seeks to recognize, map, support, enhance, or restore human cognitive or sensory-motor abilities [10]. The techniques most frequently utilized are magnetoencephalography (MEG), electroencephalography (EEG), functional magnetic resonance imaging (fMRI), and functional near-infrared spectroscopy (fNIRS) to extract brain signals [11–14]. Because of its portability and continuous data-capturing capabilities, EEG is the most widespread and reliable method for monitoring and analyzing the activity of the brain [15]. Due to the significant relationship underlying MWL and related brain function as well as the requirement to non-invasively and in real-time monitor changes in cerebral activation with a high temporal resolution, researchers concentrated on EEG data [16,17]. To record the brain oscillations caused by neuronal activity, electrodes are applied to the scalp, as demonstrated in figure 1. The electrodes are located in the occipital (O), parietal (P), temporal (T), and frontal (F) lobes, as indicated by the letter; The location of the electrodes in the brain's center is indicated by the letter Z [18]. Numerous frequency ranges comprising, gamma (31–100 Hertz), beta (16–31 Hertz), alpha (8–15 Hertz), theta (4–7 Hertz), and delta (1-4 Hertz), are used to classify EEG signals [19–21]. Researchers can evaluate and decipher neural activity associated with specific tasks or emotional states by using these frequency bands, which are tied to a range of psychological states and processes. The properties of EEG signals are nonstationary, noisy, and weak between persons. Given this, identifying strong features in EEG remains difficult. Conventional analytical techniques rely on statistical tests to verify differences between features, including power changes within particular ranges of frequency [22], which might not have considerable modeling ability [23], or to identify the cause-and-effect link [24]. Numerous ML techniques are put forth in the literature to address those issues [24].

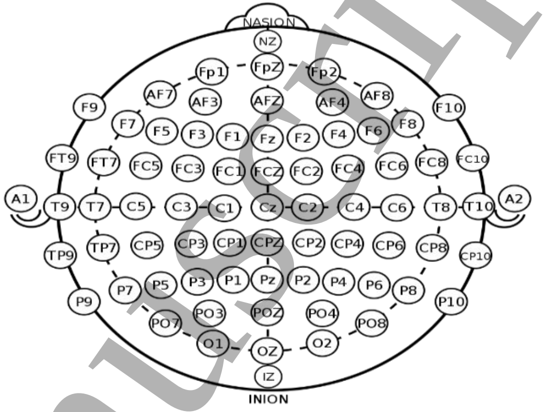

Figure 1. International 10-20 system for EEG-MCN. This image has been obtained by the author(s) from the Wikimedia website, where it is stated to have been released into the public domain. It is included within this article on that basis [25].

As a framework for data science and artificial intelligence, ML is currently one of the fastest-expanding industries. It tackles the issue of creating computer systems that automatically pick up new skills and become more efficient with data [26]. ML is widely thought of as "the problem of improving some measure of performance when executing some task, through some type of training experience" [26]. A developing body of research has demonstrated that ML can effectively extract significant data from noisy, higher-dimensional EEG signals [27,28]. The Berlin BCI Group applied for both BCI and arousal monitoring with EEG waves [28] employs LDA and a spatial filter. This demonstrates that ML is an essential approach for exploring these sectors.

A specific kind of ML called "deep learning" enables computer components to understand representations in the hierarchy of incoming data by applying a series of straightforward but non-linear transformations [29]. In domains such as clinical image analysis, audio identification, machine vision, and processing of natural languages, DL substantially outperforms conventional ML algorithms [29] attributable to more training data and increased computer processing power in the past few years. It has been utilized specifically in the discipline of BCI to identify motor imagery signals [30], emotions [31], sleep stages [32], affect states [33], epilepsy detection [34], and P300 detection [35]. Better inferences from EEG data have been made possible by recent developments in the field of DL, as neural networks can retain the spatial, spectral, and temporal structure of the recordings of the EEG.

The primary objective is to locate and compile all pertinent research papers, conference proceedings, and other academic works that have examined the use of ML and DL algorithms

J. Neural Eng. (2024)

for EEG workload assessment and classification. To evaluate and summarize the procedures and techniques utilized in the identified research, including preliminary processing, feature extraction, ML/DL algorithms, and assessment techniques, as well as the acquisition of EEG data. To evaluate several ML/DL-based techniques for EEG workload estimation and categorization, including performance metrics like accuracy. To recognize the gaps, difficulties, and restrictions in the body of current research, including variations in approaches, numbers of participants, and performance results as well as possible sources of bias or inconsistent data. To converse about the newest developments, potential applications, and areas for prospective studies in the area of EEG workload assessment and estimation using ML and DL algorithms.

The framework of the paper is as outlined below: firstly, the systematic review approach's procedure is outlined in the second section along with the research questions (RQs). The third section summarizes the review findings. A thorough overview of the articles that make up the review, arranged according to different ML and DL modalities, is also provided in the third segment. The discussion, conclusion, and recommendations for the future are covered in the fourth and fifth sections respectively.

2. Methods

PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analysis) was adhered to in the execution of this systematic study [36,37]. This methodology made use of the subsequent procedures:

2.1. Selecting Research Questions (RQs)

The following three research topics served as the basis for this systematic review:

RQ1. What are the main approaches to EEG workload? "What are the accuracy and performance measurements of studies that predict and classify using ML and DL?

This question seeks to identify and compare the main ML and DL approaches utilized in EEG-based WL estimation and classification. It aims to assess their respective accuracy and performance metrics to understand their effectiveness across various experimental conditions and participant demographics.

RQ2. What is the contribution of employing ML and DL algorithms in EEG-based workload assessment and classification methods to our comprehension of the dynamics of cognitive workload across different tasks and environments?

The insights offered by ML and DL algorithm-based EEG-based workload assessment and categorization techniques are the main subject of this question. It investigates how these techniques improve our comprehension of cognitive stress in various tasks and settings.

RQ3. What are the present obstacles, constraints, and potential biases linked to ML and DL techniques in EEG workload assessment and categorization, and what measures might be suggested to overcome them?

This inquiry focuses on the present difficulties, restrictions, and potential biases related to ML and DL approaches for EEG workload classification and estimate. Its ultimate goal is to advance the validity and reliability of EEG-based workload assessment methods.

2.2. Databases

A comprehensive electronic search was conducted from the beginning to the end of February 16, 2024, among the following five repositories: IEEE Explore, PUBMED, Science Direct, SpringerLink, and ACM Digital Library.

2.3. Search Strategy

The subsequent Boolean operators and keywords were combined and appended to all datasets: ("EEG" OR "electroencephalography") AND ("workload estimation" OR "workload classification") AND ("machine learning" OR "deep learning").

Only English-language searches were allowed to appear in the results. The allusions to the chosen major full-text publications were examined in more detail to find pertinent works. The criterion for eligibility listed in Table 1 was applied to further reduce the choices.

2.4. Data Extraction

There were 122 publications found in total from the preliminary electronic database searches: 46 in ScienceDirect, 24 in ACM Digital Library, 14 in IEEE Xplore, 13 in PubMed, 25 in SpringerLink, and 7 more articles identified through other sources. After duplicate records (n = 4) were eliminated, 125 pertinent reports were retained. This phase involved two separate investigators screening abstracts and titles against the inclusion and exclusion criteria. 45 papers were chosen for full-text review out of the 125 articles that were slated for screening based on abstracts and titles. 33 papers were included in the systematic review after the exclusion procedure, according to the eligibility requirements figure 2 demonstrates the process and outcomes of the review procedure.

Two authors conducted a formal evaluation of the study quality to guarantee the strength of the body of evidence and reduce the possibility of biases (please refer to figure 3). This tool evaluates seven criteria, of which investigations must include an explicit and straightforward description for at least five of them. It is adapted from a standard framework established via prior studies [38,39]. The aforementioned requirements include (i) a rationale or theoretical framework; (ii) goals and objectives; (iii) setting; (iv) sampling; (v) technique; (vi) analyzing data; and (vii) enough originally collected information to act as a mediator between the data and the interpretation that follows. Every study that was included in the review met at least five requirements. As a result, there was little chance of bias in the bulk of the investigations.

3. Review Findings

3.1. Distribution of ML and DL Methods Over Time for EEG Workload Estimation

Figure 4 illustrates the chosen articles that were published between 2015 and 2024. The bar graph shows a consistent rise

Table 1. Criterion for Eligibility.

Criteria for Inclusion	Criteria for Exclusion
Articles subjected to peer review that are published.	Studies published in non-peer-reviewed sources
Articles composed and published after 2000.	Papers composed outside of English.
Studies involving human participants of any age group.	Studies exclusively involving non-human participants (e.g., animal studies).
Studies that concentrate on effort estimation or categorization methods based on EEG.	Studies that not concentrating on effort estimation or categorization methods based on EEG.
Studies that estimate and classify EEG workloads using ML or DL techniques.	Studies that do not estimate and classify EEG workloads using ML or DL techniques.

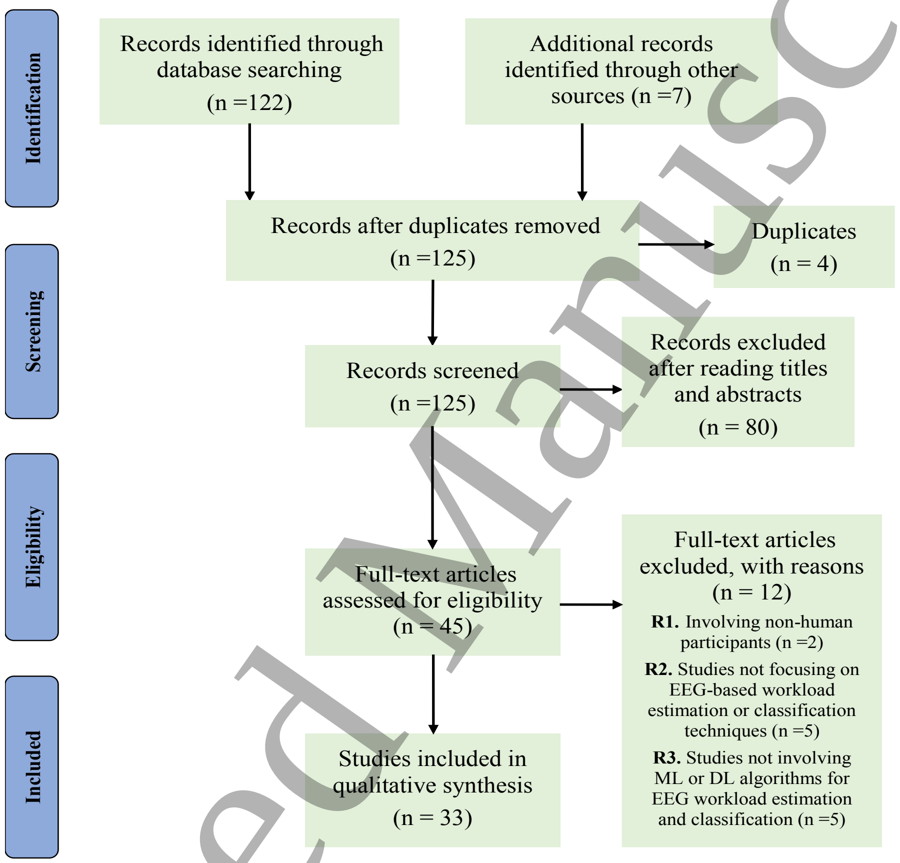

Figure 2. Diagram illustrating the steps involved in the systematic review's inclusion and removal of evidence for EEG workload estimation and classification.

in the application of DL techniques for EEG workload assessment, especially in the last few years. This pattern shows how DL approaches are becoming more and more popular as a means of managing complicated EEG data and increasing the accuracy of workload assessments. Conversely, the use of ML techniques seems to have leveled off or perhaps decreased in a few years, indicating a possible move toward more sophisticated DL strategies.

32 *Overview of ML Phases in EEG-Based MWL Identification and Categorization*

First, splitting general ML into DL and classical ones (see figure 5), We outline the key phases of ML techniques used in EEG-based MWL identification.

As illustrated in figure 5, (A) is denoted the data acquisition and preprocessing phase. Combining (A) and (B) shows the traditional machine ML stages, such as data preprocessing, extraction of EEG features, selection of EEG features, classification techniques, and performance assessment. (C) is a DL technique that can automatically learn features. The data pretreatment procedures might not be required for DL techniques since they can utilize raw data as inputs in this manner. For this area, DL techniques often employ computed features to learn spatial, spectral, and temporal information.

J. Neural Eng. (2024)

Study	I	II	III	IV	V	VI	VII	Total
(Salimi et al., 2019)	+	+	+	+	-	+	+	6
(Zhang et al., 2022)	+	+	+	+	+	+	+	7
(Samima & Sarma, 2019)	+	+	+	+	+	+	+	7
(Islam et al., 2019)	+	+	+	+	+	+	+	7
(Gu et al., 2019)	+	+	+	+	+	+	+	7
(Taheri Gorji et al., 2023)	+	+	+	+	+	+	+	7
(Afzal et al., 2023)	+	+	+	+	+	+	+	7
(Saha et al., 2018)	+	+	+	+	-	+	+	6
(Yedukondalu et al., 2023)	+	+	+	+	-	+	+	6
(Karacan et al., 2023)	+	+	+	+	+	+	+	7
(Samima & Sarma, 2023)	+	+	+	+	+	+	+	7
(Aygun et al., 2022)	+	+	+	+	+	+	+	7
(Shao et al., 2023)	+	+	+	+	+	+	+	7
(Aksu et al., 2024)	+	+	+	+	-	+	+	6
(Ke et al., 2015)	+	+	+	+	+	+	+	7
(S. K. Pandey et al., 2023)	+	+	+	+	-	+	+	6
(Bilalpur et al., 2018)	+	+	+	-	+	+	+	6
(Zhao et al., 2022)	+	+	+	+	+	+	+	7
(Yin & Zhang, 2017)	+	+	+	+	+	+	+	7
(Cao et al., 2022)	+	+	+	+	+	+	+	7
(A. Gupta et al., 2021)	+	+	+	+	+	+	+	7
(Das Chakladar et al., 2020)	+	+	+	+	+	+	+	7
(Qiao & Bi, 2020)	+	+	+	-	+	+	+	6
(Yin et al., 2019)	+	+	+	+	+	+	+	7
(Yu et al., 2023)	+	+	+	-	+	+	+	6
(Lobo et al., 2016)	+	+	+	+	-	+	+	6
(Ved & Yildirim, 2021)	+	+	+	+	+	+	+	7
(V. Pandey et al., 2020)	+	+	+	+	+	+	+	7
(Vishnu et al., 2022)	+	+	+	+	+	+	+	7
(Zhang & Li, 2017)	+	+	+	+	+	+	+	7
(Islam et al., 2020)	+	+	+	+	+	+	+	7
(Hefron & Borghetti, 2017)	+	+	+	-	+	+	+	6
(S. S. Gupta et al., 2021)	+	+	+	+	+	+	+	7
Total	33	33	33	29	27	33	33	

Figure 3. The reviews' studies' potential for bias [38,39].

Notes. *(+) study satisfies criteria; (−) either the study satisfies the criteria or it is ambiguous. (i) A clear explanation of the conceptual framework and/or the addition of a literature review that explains the intervention's justification. (ii) Clearly defined goals and objectives. (iii) A concise explanation of the background, including information on elements crucial to understanding the findings [a]. (iv) A precise explanation of the sample. (v) A detailed explanation of the approach, which includes organized techniques for gathering data [b]. (vi) Multiple researchers analyzing the data [c]. (vii) Including enough original data to act as a mediator between the data and interpretations [d] .a The study had to meet several requirements in order to meet this criterion, including who (i.e., who collected the data and was involved in patient care); where (i.e., where respondents were recruited and data collected); whether or not the data was anonymous; when (i.e., at what phase of treatment); and how (i.e., how the questionnaire was distributed, was the study approved ethically).b The study has to report on particular facets of the data gathering and analysis methodologies in order to satisfy this requirement. Data collection: Particulars include how the satisfaction of patients is measured. particulars of surveys, such as the design of the questionnaire, the questions asked, and the types of responses. Data analysis: Particulars about surveys: p-levels, proper statistical tests for the used measurement level, and the specification of potential response category dichotomization or aggregate. The study had to contain enough original data in order to meet this criterion. c Figures and legends are clear, and data supports the conclusions. d The use of suitable measurements of central tendency as well as variability indices for the degree of measurement are specific components of quantitative research.

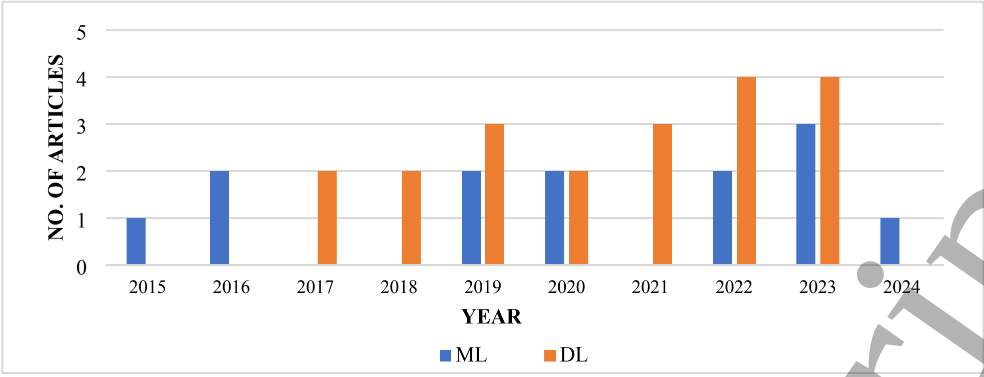

Figure 4. Evolution of ML Models in EEG Workload Estimation Over Time.

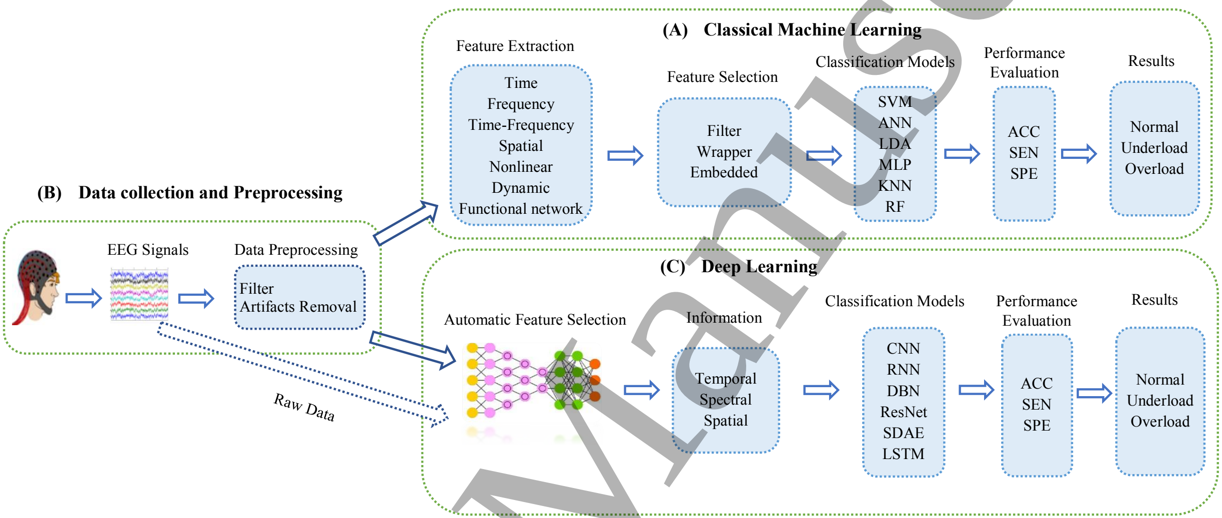

Figure 5. Typical procedures for ML techniques that use EEG to recognize MWL. (A) represents the steps involved in gathering EEG data and preprocessing it, such as filtering and removing artifacts. (B) consists of traditional machine learning procedures, such as feature selection and extraction from different domains, classification techniques (e.g. ANN, SVM, LDA), and performance metrics such as accuracy. (C) refers to the deep learning techniques that primarily use neural networks, such as CNN, RNN, and DBN. These techniques may not require data processing steps if the model is fed raw data. Most of the time, temporal, spectral, and spatial information is acquired by deep learning techniques using calculated features. The final output is a prediction of the cognitive workload with normal load, underload, and overload states, using a three-class classification task as an example.

3.3 *ML Techniques for Workload Estimation and Categorization*

This section comprises related studies on ML-based EEG workload estimation and categorization, including techniques, models, findings, and experimental data gathering.

Hefron et al. [40] proposed a new feature generation methodology that looks at the variability of the power distribution in the clinical frequency bands over a 10-second sliding temporal window, in addition to analyzing average power, from multi-day data collected from a MATB workload study. Training three conventional classifiers- LDA, Random Forest, and KNN- on the data from the first four days of the study allowed for the prediction of high vs low workload levels on day five of the study. It was suggested that the variance in the frequency-domain power distribution was a notable feature since it was statistically significant between circumstances. The cross-day workload categorization accuracy was 5.8% higher with variance included as a feature than with mean power alone in models. In addition, the distinct classifiers were integrated into a time-smoothed composite classification algorithm, which leveraged variations in features chosen in the models to enhance the total classification accuracy beyond 80%.

In this study, an MI-based feature set construction methodology combining vehicular signals and EEG is presented by Riyanul et al. [41]. To assess drivers' MWL in terms of labels and scores, the feature set was put into use. For the evaluation tasks, multiple ML models were trained. There was almost no difference between the predicted score produced using MI-based features and EEG features, according to the values of MAE in MWL score prediction. Alternatively, in terms of ML model performance metrics, it was found that RF classifiers outperformed other classifiers in labeling MWL and events; however, statistical tests revealed that SVM outperformed all other classifiers by a significant margin. With EEG-based features and MI-based features, the maximum classification accuracy for MWL was 88% and 82%, respectively. Also, with a 94% classification accuracy, MI-based features performed better than EEG-based features in two particular events (a person crossing the street and a car entering traffic).

Ayca et al. [42] present the findings of a research investigation that aimed to address this question by obtaining electroencephalogram (EEG) and eye gaze data from participants in an interactive multi-modal driving task. By adding supplementary tasks during driving, such as dialogue, blocking events, and tactile stimuli, different levels of cognitive workload were produced. Our findings demonstrate that pupil diameter, as opposed to EEG, is a more trustworthy indication for workload prediction. Crucially, all five ML models that combined the extracted EEG and pupil diameter features failed to demonstrate any improvement in workload classification when compared to eye gaze alone. This indicates that eye gaze is an adequate modality for evaluating cognitive workload in humans in multi-modal, interactive settings.

Jesus et al. [43] study a technique to categorize various MWL levels based on synchronized eye tracking and EEG data. The authors took advantage of the information gathered during the experiments' second phase, which involved 21 participants. The datasets are multiclass and labeled; that is, they display eye-tracking and EEG data samples at three distinct workload classes (LW, MW, and HW). There is an imbalance in all the datasets, with roughly 15% of LW and HW samples and 70% of MW samples.

Hao et al. [44] developed a hybrid MWL classification system, which blends SVM and ELM. To uncover hidden information in high-dimensional EEG data, the former is employed as the member classifier. The latter is employed to combine the member classifier's outputs. Lastly, we illustrate the efficacy of the suggested ELM-SVM model by contrasting it with the traditional MWL classifiers. With a training accuracy of 1 and a test accuracy of 0.8328, the hybrid suggested model performs better.

In [45] study intends to compare MS patients with a healthy group undertaking the same tasks to categorize the MWL level of MS patients as low, medium, or high based on EEG signals during cognitive tasks in computer and virtual reality environments. In this study, three cognitive activities in computer and virtual reality environments are used to measure the mental workload level of 45 volunteers using EEG data and the NASA-Raw Task Load Index questionnaire. With the SVM classifier, the three-level mental workload classification accuracy for MS patients is 96.08% for computer environments and 94.12% for virtual reality environments. In computer and virtual reality environments, the categorization accuracy for participants in good health is 95.24% and 94.05%, respectively.

Samima et al. [46] suggested a study that illustrated how the workload of operators engaged in lengthy operations is evaluated. Workload can be distinguished with ease using the suggested feature vector. Furthermore, the data is classified by the ANN architecture with a 96.6% accuracy rate using the suggested feature vector. A brain connection study has been done to validate the suggested method. Strength values are higher during more demanding tasks than during low-demanding tasks. Additionally, when doing high-demanding activities compared to low-demanding ones, the clustering coefficient displayed lower values. The outcomes show that our suggested strategy is effective.

Cao et al. [47] presented a novel framework to address multi-level mental workload classification. It is based on the features of hybrid EEG–functional near-infrared spectroscopy (EEG–fNIRS), reinforced by machine-learning features. We also suggest employing bivariate functional brain connectivity (FBC) features in the temporal and frequency domains of three bands: delta (0.5–4 Hz), theta (4–7 Hz), and alpha (8–15 Hz) in place of the widely utilized univariate power spectral density (PSD) for EEG recording. The FBC technique greatly enhanced classification performance at a 77% accuracy for 0-back vs. 2-back and 83% for 0-back vs. 3-back using a public dataset with the help of the fNIRS oxyhemoglobin and deoxyhemoglobin (HbO and HbR) indicators.

Compounding workload value for each individual is the main goal of the proposed task which was developed by Shabnam et al. [48]. The study suggests a cognitive workload graph to meet the requirements; the graph measures the cognitive burden of the "current" instance by referencing the cognitive workload of the "former" instance. Workload measurement irregularities will be eliminated by measuring workload in this way. Additionally, the shift in a person's workload can be simply visualized with the help of graphic interpretation. The relative value of the PSD of the EEG rhythms utilizing the "idle" and "experiment" mental states was used in this study to propose a workload-defining vector for the graph's design. Using an artificial neural network classifier, the defined feature vector can accurately categorize the load states with 98.66% accuracy.

Yufeng et al. [49] proposed regression and SVM classifier models were investigated in both within-task (trained and tested on the same task) and cross-task (trained on one task and tested on another task) scenarios for proficient n-back verbal and spatial tasks. With written informed consent, 17 participants (four females and thirteen males) aged between 19 and 24 (M = 21.9 years, SD = 1.6) voluntarily took part. The duration of the subjects' participation in the experimental tasks was included in the cropping of all the raw, continuous EEG data. Six characteristics of each channel for all 30 channels were obtained by extracting the PSD sums in six frequency bands (δ: 0.5–3 Hz, θ: 3–8 Hz, α: 8–13 Hz, β1: 13–20 Hz, β2: 20–30 Hz, and γ1: 30-45 Hz).

The study collected EEG data from 10 collegiate aviation students in a live-fight scenario in a single-engine aircraft to better understand MWL in aviation. Every pilot held a commercial pilot certificate issued by the FAA and a class I or class II medical certificate issued by the FAA. Every pilot executed a conventional battle profile, which mirrored a typical instrument flight training sequence. The four primary sub-bands of the collected EEG signals- delta, theta, alpha, and beta-were utilized for data processing. After calculating PSD and log energy entropy of each sub-band throughout a 20-electrode set, two feature selection algorithms- RFE and Lasso CV- as well as a stacking EML technique made up of SVM, RF, and LR were applied by Hamed et al. [50]. Following the feature selection process, fifteen features were identified that

J. Neural Eng. (2024)

Table 2. An overview of the evaluated literature on ML techniques

Article	Dataset Source	Subject/ Participant	Paradigm/ Stimuli	No. of channel	Sample Frequency	EEG Workload Estimation Method	Workload Access Level	Data Acquisition Method / Mental Workload Test	Classification Technique	Accuracy
[40]	A prior study completed in 2011	Eight participants	Multi-Attribute Task Battery (MATB)	19	256Hz	PSD	Low, Medium, and High	D.A.M: Electrode placement M.W.T: NASA Task Load Index (NASA-TLX)	LDA, RF, KNN	72.8% by LDA, 72.6% by KNN, 79.2% by RF
[41]	The Declaration of Helsinki of 1975	Twenty males participants (24.9 ± 1.8 years old)	Real Driving Simulation	15	256Hz	PSD	Low, High	D.A.M: Electrode Placement M.W.T: NASA Task Load Index (NASA-TLX)	LgR, MLP, SVM, and RF	68% by LgR, 70% by MLP, 70% by SVM and 87% by RF
[42]	The Local Community	80	Interactive Multi-Modal Driving Task	8	500Hz	PSD	4 levels which are the combination of Dialogue, Braking, and DRT.	Event-Based Cognitive Workload Assessment (EEG and Eye Gaze used for data acquisition to assess cognitive workload)	*k*-NN, NB, RF, SVM, and NNM	40.31% by k-NN, 43.7% by NB, 41.2% by RF, 41.07% by SVM, and 50.7% by NNM (Pupil diameter).
[43]	The Suor Orsola Benincasa University	20	Dual-Task Paradigm and Syntactic Transformation Task	8	-N/A-	Combining synchronized EEG with eye-tracking	Low, Medium, High	Performance Metrics (Reaction Times, Lexical Complexity)	k-NN	-N/A-
[44]	University of Shanghai for Science and Technology	8	Dual-Task Paradigm	11	500Hz	PSD	Low, High	D.A.M: Electrode Placement M.W.T: Performance Metrics (Reaction Times, Accuracy Rates)	ELM-SVM	83.28%
[45]	Kutahya Dumlupinar University Neurotechnology Education Application and Research Center	17 MS (11F, 6M) patients and 28 (8F, 20M) healthy subjects. 31.11±8.27, and the mean age of the healthy group is 24.42±4.20	Attention and Memory Task, Information Processing Task, and Car Simulation Task	32	500Hz	PSD	Low, High	D.A.M: Electrode Placement M.W.T: NASA Task Load Index (NASA-TLX)	SVM	94.87%
[46]	IIT, Kharagpur	Twenty healthy volunteers (25±5 years old)	Working Memory Test Battery (Verbal Tasks and Visuo-Spatial Tasks)	64	256Hz	Time-frequency Domain, spectral powers (SP)	Low, Medium, High	D.A.M: Electrode Placement M.W.T: NASA Task Load Index (NASA-TLX)	ANN	96.6%
[47]	Technical University Berlin	26	n-Back Tasks	30	200Hz	PSD	-N/A-	D.A.M: Electrode Placement M.W.T: NASA Task Load Index (NASA-TLX)	SVM	77% (0-back vs. 2-back) 83% (0-back vs. 3-back) 59% (2-back vs. 3-back)
[48]	IIT, Kharagpur	10 healthy male volunteers (25±5 years old)	Working Memory Test Battery (Verbal Tasks and Visuo-Spatial Tasks)	64	256 Hz	PSD	Low, Medium, High	D.A.M: Electrode Placement M.W.T: NASA Task Load Index (NASA-TLX)	ANN	98.66%
[49]	NeuroScan, Inc., Charlotte, NC	17	n-back task	30	1000 Hz	PSD	-N/A-	Electrode Placement	SVM	Above 95%
[50]	University of North Dakota	10 collegiate aviation students	n-back Task	20	256 Hz	PSD and log energy entropy	Low, Medium, and High	D.A.M: Electrode Placement	Stacking ML model	91.67% (± 0.11)

*-N/A-= No Available, D.A.M = Data Acquisition Method, M.W.T.= Mental Workload Test.

may be used as a gauge for the MWL states of pilots. Following the application of these features to the stacking ensemble algorithm, the RFE algorithm produced the best results when it used the chosen features, with 91.67% (± 0.11), 93.89% (± 0.09), recall, 91.67% (± 0.11), F-score, and mean ROC-AUC of 0.93 (± 0.06).

3.4 *Summary of Classical ML Techniques in EEG Workload Estimation and Categorization*

Table 2 provides a detailed summary of the factors related to ML methods used in EEG workload prediction. The data source, subject/participant demographics, paradigm/ stimuli, number of channels, frequency of samples, EEG workload estimating method, workload access level, data acquisition,

J. Neural Eng. (2024)

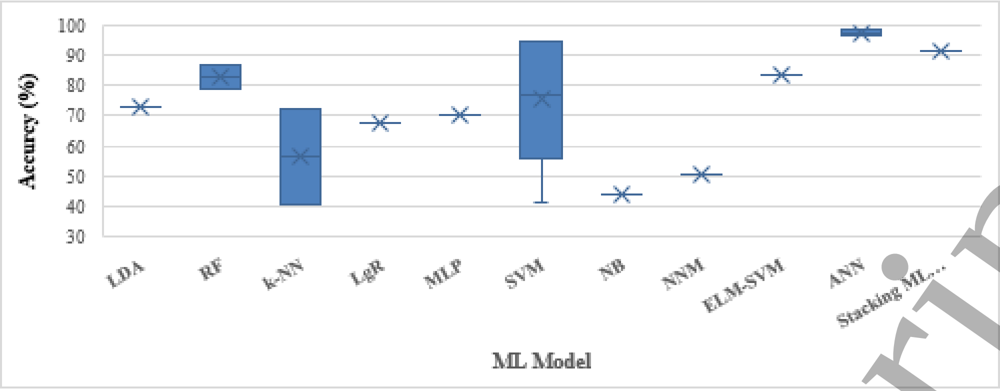

Figure 6. Illustrating the distribution of accuracy scores achieved by different ML models in EEG workload estimation.

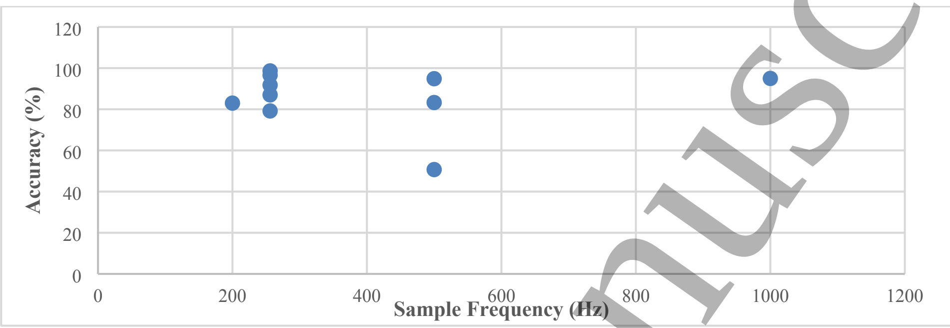

Figure 7. Relationship between sample frequency and accuracy in ML models for EEG workload estimation.

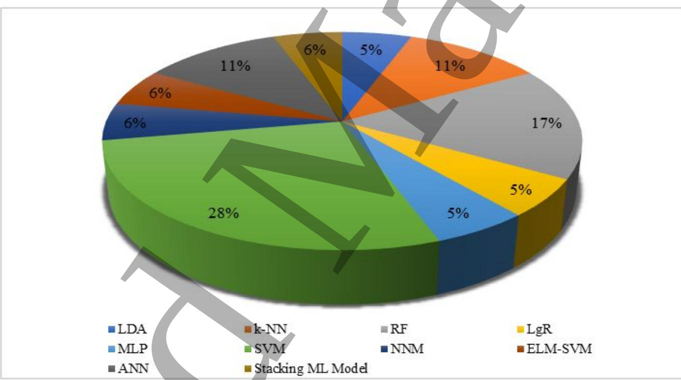

Figure 8. Illustrating the percentage distribution of ML methods employed in EEG workload estimation.

method / mental workload test classification approach, and the matching accuracy scores are among the criteria. This section comprises related studies on ML-based EEG workload estimation and categorization, including techniques, models, findings, and experimental data gathering.

3.5 Analysis of ML Techniques in EEG Workload Estimation and Categorization

The box plot (see figure 6) illustrates how different ML techniques perform differently when it comes to estimating EEG workload. The narrow distribution of the boxes and high median values show that techniques like ANN and stacking ML models regularly exhibit great accuracy. On the other hand, there is greater variation in accuracy scores for techniques such as k-NN and SVM, indicating less consistent performance across investigations.

Figure 7 shows a scatter plot that illustrates the correlation between sample frequency and accuracy in ML. An ML model is represented by each point on the plot, where the y-coordinate shows the model's accuracy and the x-coordinate shows the sample frequency. Plotting the data reveals patterns and trends that shed light on how sample frequency affects the efficiency of ML models used to estimate EEG workloads.

SVM, which makes up the majority of publications, is the most widely employed ML approach in EEG workload estimation research, according to the pie chart (see figure 8). The supremacy of SVM can be ascribed to its adaptability and efficiency while managing intricate datasets. While not as much, other ML techniques like RF and ANN also show significant representation.

3.6 DL Techniques for Workload Estimation and Categorization

This section comprises related studies on DL-based EEG workload estimation and categorization, including techniques, models, findings, and experimental data gathering.

Jianhua et al. [51] suggested a DL strategy for the MWL classification problem that is based on RBMs. First, entropy-based node-pruning techniques and trial-and-error were used to determine the DBN structure. When the suggested MAD approach was compared to the other two feature extraction algorithms, the MWL multi-class classification accuracy improved statistically significantly ($p<0.05$). Next, the most pertinent EEG channels were chosen using the linking weights between the input layer and the first hidden layer in the DBN. The findings of comparing the chosen EEG channels with the random channel selection method demonstrated that they may be used for MWL classification more quickly without sacrificing accuracy. In many application cases, the suggested EEG channel selection approach may also be utilized to lower the computing cost of a BCI system.

Mengyuan et al. [52] evaluated changes in cognitive effort in the Chinese context for two scenarios of human-machine interaction: unclear and clear pronoun resolution. We found that in ambiguous conditions as opposed to unambiguous controls, there was a contradiction between an increase in reaction time and a decrease in theta power, which was correlated with an increase in cognitive workload. This suggests that additional working memory resources were used when participants realized that, in the case of unambiguous pronoun resolution, the cognitive task had a standard answer. To discover high-level abstractions to model the workload, the power features displayed in spatial feature maps could be retrieved. When the GoogleNet and EfficentNet outputs were combined, the overall workload classification rate varied between 55 to 63% depending on the individual.

Shankar et al. [53] investigate how visual elements, such as geometric shapes and fixed shapes with numerous colors, can be used to cross-task and classify cognitive burdens. To characterize cognitive workload, this study uses band power, curve length, and approximation entropy variables from the statistical, morphological, and nonlinear domains, respectively. For variable duration segments, the current study computes traits from the statistical, morphological, and non-linear domains while considering four clinical sub-bands. The most similar characteristics between the two jobs are found via the neighborhood component feature selection algorithm. Deep structure is used for binary classification of cross-tasks, utilizing LSTM and BLSTM. Compared to standard classifiers, RNN gets the highest classification accuracy of 92.8% for the cross-task.

To identify MWL data, two DL models are investigated: LSTM followed by an MLP, and SDAE followed by a multilayer perceptron (MLP). MLP is utilized for classification, and SDAE, LSTM, and feature extraction are employed by A. Saha et al. [54]. The classification accuracy of SDAE and LSTM are 89.51% and 86.33% respectively.

Using EEG data, Afzal et al. [55] successfully attempted to classify cognitive workload in this study. By applying the BDGN model, we achieved an astounding 98% accuracy rate. To improve its appropriateness, the STEW dataset was preprocessed and put through power spectral density (PSD) and temporal domain analysis. The BDGN model was implemented to accomplish real-time cognitive workload classification using the MQTT protocol. These findings highlight the potential of using the BDGN model for EEG-based cognitive workload classification. The excellent accuracy attained validates the efficacy of this method in accurately determining cognitive exertion levels. Notably, our comparison of the suggested strategy with cutting-edge approaches demonstrates its superiority.

Using an open-source EEG dataset, Vishnu et al. [56] present an LSTM-based RNN to address the three-level MWL cross-section classification problem. The approximate entropy of every trial and the average spectral power across all EEG frequency bands were utilized as input features for the LSTM classification network. On an unobserved session of the same subject, the proposed framework yielded a mean validation accuracy of 87.38% and a test accuracy of 43.79%. Despite demonstrating generic predicting skills, we hypothesize that our network may have tuned to the subject-specific features based on the effectiveness of our approach.

Using EEG signals, an adaptive DL model based on SDAE is created for cross-session MW categorization by Zhong et al. [57]. To track the modification of the statistical properties in the EEG power features between two consecutive days, the weights of the first hidden layer connected to the input layer in the deep learning models are adjusted iteratively. Regarding the latter, more performance comparisons are made using the noise corruption paradigm and various feature selections between the adaptive SDAE and classical MW classifiers. In addition, an analysis is conducted on the ideal step length and data augmentation strategy to present a workable model selection framework for the adaptive SDAE.

In [58] presents a new hybrid approach for automatic feature extraction from the EEG signals and demonstrated with MWL classification. To compare the DL-based and traditional feature extraction techniques, several classifiers have been deployed. Results have shown that the highest value of AUC-ROC is 0.94, achieved using the features extracted by CNN-AE and SVM.

To learn the temporal-spatial-spectral aspects of EEG data, dynamic hierarchical attention based on a transformer (DHAT) is presented by Zhang et al. [59]. Using a multi-head self-attention computation, the DHAT begins with hierarchical attention from extracted features. By automatically giving the dimension vector a distinct weight size, the inferred representation of the vector is computed. To improve performance, feed-forward neural networks, dropout components, and position encoding are all injected at the same time. The temporal information gleaned from the hierarchical learning was then dynamically fused using the gate recurrent unit (GRU) with residual add. 48 subjects from the STEW database were used for validation, and the mean accuracy of MWL classification across subjects was 76.81%.

Maneesh et al. [60] evaluate MWL using three methods: (a) user perception of sound-data mapping accuracy for various acoustic parameters; (b) user reports of MWL impressions; and (c) implicit EEG responses collected throughout the mapping task. Our main conclusions are as follows: (i) higher mapping accuracy is associated with low cognitive load-inducing (i.e., more intuitive) acoustic parameters; (ii) EEG spectral power analysis reveals higher α band power for low cognitive load parameters, suggesting a congruent relationship between explicit and implicit user responses; and (iii) MWL classification using EEG features achieves a peak F1-score of

0.64, confirming that user EEG data collected using wearable sensors can be used to reliably estimate workload.

D. Das Chakladar et al. [61] estimate the workload that human subjects experience when engaging in multitasking mental activities. The "STEW" dataset is used to estimate mental effort. There are two tasks in the dataset: "Simultaneous capacity (SIMKAP)-based multitasking activity" and "No task." With the help of a composite architecture made up of a deep neural network and the Grey Wolf Optimizer (GWO), two jobs with varying workload levels have been assessed. To choose the best features of mental activity, GWO has been employed. When compared to the convergence rate of GWO, other optimization methods like Genetic Algorithm (GA) and particle swarm optimization (PSO) are often slower. A deep hybrid model based on BLSTM and LSTM has been developed to classify workload levels. For "No task" and "SIMKAP-based multitasking activity," the suggested deep model obtains classification accuracy of 86.33% and 82.57%, respectively.

A technique developed by Shiliang et al. [62] based on the time-frequency and spatial domains of EEG signals is put forth to raise MWL's classification accuracy. Additionally, a hybrid deep learning model is introduced. First, various brain areas' spatial domain properties are suggested. Using the wavelet method, EEG time-frequency domain data is simultaneously acquired. Two categories of DL models are fed the spatial and time-frequency domain variables for MWL classification. This proposed method's performance is verified using the Simultaneous Task EEG Workload public database. When compared to current techniques, the suggested method exhibits superior classification accuracy. It offers a fresh way to evaluate MWL.

To classify different levels of MWL, Yinhu et al. [63] introduced a novel deep learning model called latent space coding capsule network (LSCCN). The VAE, convolution, and capsule modules are the three main modules that make up the LSCCN. To improve the robustness of feature representation, the VAE module generates latent variables from the fused features of the EEG. For a more accurate classification of mental effort, the capsule module can capture the relationships between the latent variables. In terms of classification performance, the suggested model was contrasted with LSTM, CNN, RF, the single VAE classifier, and the single capsule network. The comparison's findings showed that the suggested model performed better than each of the methods that were examined. To learn more about mental effort, the authors looked into band powers, connectivity strengths, and latent variables.

Vishal et al. [64] compare ML techniques that are used to calculate workload using EEG data. Based on an open-access EEG dataset obtained from a "no task" and a "simultaneous capacity (SIMKAP) experiment," models are developed and verified for the binary categorization of workload as present or absent. The deployment of several categorization algorithms that forecast workload using EEG data is shown in the article. This paper reports on the implementation of the Random Forest classifier (57.19%), MLP network classifier (58.2%), CNN+LSTM network classifier (58.68%), LSTM network classifier (61.08%), and KNN classifier (57.3%). The research can be expanded to investigate operator workload in real-time for any type of task in a real-world application by utilizing a brain-computer interface paradigm.

To capture the individual differences and dynamical aspects of EEG data, Z. Yin et al. [65] suggest a novel transfer dynamical autoencoder (TDAE). The feature filter, abstraction filter, and transferred MWL classifier are the three successively connected modules that make up the TDAE. To abstract the EEG features over adjacent time steps to prominent MWL indications, the feature and abstraction filters establish a dynamic deep network. To increase the stability of the model training, the transferred MWL classifier makes use of a huge volume of EEG data from a source-domain EEG database that was captured in response to emotional stimuli. The authors used two target EEG databases to test their algorithms. According to the classification performance, TDAE performs noticeably better than both the current shallow and deep MWL classification models. The outcomes demonstrate that for two situations, TDAE may obtain binary classification accuracies of 0.86 and 0.90.

To effectively classify the participants' MWL, Anmol et al. [66] investigated the feasibility of combining deep learning with several model-free functional connectivity indicators. To achieve this, 19 participants' 64-channel EEG data were gathered as they completed the standard n-back task. The functional connectivity features, Phase Transfer Entropy (PTE), Mutual Information (MI), and Phase Locking Value (PLV), were extracted from these data (post-processing). These three were selected to provide a thorough analysis of model-free functional connectivity measures that are directed and non-directed (allowing for faster computations). Three DL classifiers- CNN, LSTM, and Conv-LSTM were employed to identify the cognitive effort as low (1-back), medium (2-back), or high (3-back) based on these features. According to the results, the combination of MI and CNN achieves the highest multi-class classification accuracy at 80.87%. This is followed by the combinations of PLV and CNN, which achieved 75.88%, and MI and LSTM, which achieved 71.87%.

Jammisetty et al. [67] present a novel method for subject-wise cognitive load identification using multiple T-F conversions and different optimizers-based Bi-LSTM. Preprocessing techniques were used using filters to get rid of artifacts. The dataset was produced by dividing long EEG recordings into 4-second intervals. The data was normalized using the Z-score normalization technique. Spectrogram, scalogram, and constant Q-Gabor transform were among the T-F conversion approaches that came after the 2D-EEG spectral entropy characteristics. Subsequently, the data was classified into the two cognitive load classes using the Bi-LSTM network and other optimizers, including Adam, SGDM, and RMSprop. Using 5- and 8-fold data splits, 10-fold cross-validation was used to assess the performance of the suggested model. The findings demonstrated remarkable accuracy rates for subject-wise cognitive load identification, with 99.55% (5:5), 99.88% (8:2), and an average Ac (%) of 85.22% (8:2).

In order to assess cognitive status, Weizheng et al. [68] suggest using a deep hybrid network known as the Ternary-task Convolutional Bidirectional Neural Turing Machine (TT-CBNTM). Initially, TT-CBNTM is made up of CNN and BNTM. CNN is used to maintain the EEG's spatial, spectral representations, while BNTM is used to extract features from the CNN and learn temporal representations. Second, we provide a novel approach to the EEG database overfitting called the ternary-task regularization framework. The primary objective is to evaluate EEG-based MWL by categorizing EEG signals. Identification and verification are the auxiliary tasks that help us improve the EEG-based cognitive workload categorization by increasing the inter-class variances and decreasing the intra-class variations. The model has a 96.3 % classification accuracy which is 5% better than the most advanced versions.

Harshita et al. [69] used DL and conventional ML methods to categorize VR MWL levels based on EEG brainwaves. Based on the results, SVM performed the best among machine learning classifiers, with an accuracy rate of 33%. This can be attributed to the margin maximization's superior generalization properties. Roughly speaking, it is comparable to random classification, which is consistent with other studies that produced 33% accuracy. Regarding DL architectures, their CNN model with spectrograms produced results that were marginally better than the baseline models (accuracy of 35%) but not significantly better.

Using data instances as brief as 1.1s, a subject-specific mental workload classifier has been developed by Nima et al. [70] that can accurately classify two levels of MWL. An ensemble learner is the suggested model, which automatically extracts features from EEG channel data. Every base model in the ensemble classifier was customized to acquire the spectral-temporal characteristics of a specific channel. By integrating the outputs of the channel-specific base classifiers using a majority voting mechanism, the model was able to incorporate spatial information into its decision. Several data instance durations were used to test the ensemble classifier. For every instance duration that was examined, the suggested ensemble classifier outperformed a single classifier that used all channel data as input.

Saroj et al. [71] present an end-to-end system for classifying EEG signals of individuals with epileptic disorder into three categories: preictal, normal, and seizure. It does this by combining two deep learning models, CNNs and LSTM. The widely used and publicly accessible Bonn University dataset is used to produce the experimental results. The processes of feature extraction, selection, and classification in this CNN-LSTM classification model are carried out automatically without the need for manually developed feature extraction techniques. Using the tenfold cross-validation method, the CNN–LSTM model's performance is analyzed and assessed in terms of specificity, sensitivity, and accuracy. The conducted tests and the outcomes demonstrate 99.33% accuracy, 99.33% sensitivity, and 99.66% specificity in that order.

In [72] we propose a three-dimensional convolutional neural network (3D CNN) employing a multilevel feature fusion algorithm for mental workload estimation using EEG signals. The author used the Sternberg task to gauge each participant's MWL, which was categorized as either low or high workload condition based on the task's level of difficulty to validate our network's performance. Our network is trained on this dataset and the accuracy of our network is 90.8 %. The authors also tested their approach with the publicly available EEG dataset, and we obtained an accuracy of 93.9%.

3.7. Analysis of ML Techniques in EEG Workload Estimation and Categorization

Table 3 provides a detailed summary of the factors related to DL methods used in EEG workload prediction. The data source, subject/participant demographics, paradigm/stimuli number of channels, frequency of samples, EEG workload estimating method, workload access level, data acquisition method / mental workload test, data augmentation methods classification approach, and matching accuracy scores are among the criteria.

3.8. Analysis of DL Techniques in EEG Workload Estimation and Categorization

The box plot (see figure 9) illustrates how various DL models perform differently when it comes to estimating EEG workload. The high median values and narrow box spread of models like BiLSTM and ConvLSTM show that they have consistently excellent accuracy. On the other hand, models with greater fluctuation in accuracy scores, such as CNN-AE and EfficientNet, indicate less consistent performance across studies.

Figure 10 shows a scatter plot that illustrates the correlation between sample frequency and accuracy in DL models. Every point on the figure represents a DL model; the sampling frequency is shown by the x-coordinate, and the accuracy attained by the model is indicated by the y-coordinate. Plotting the data shows patterns and trends that shed light on how sample frequency affects the efficiency of DL models for estimating EEG workload.

In EEG workload estimation research, the relative prevalence of various DL approaches is displayed in a pie chart (see figure 11). With 14% of the examined studies using DL methods like CNN and Conv-LSTM, it seems that these are the most often utilized techniques. While less often used, other techniques like DBN and EfficientNet nevertheless make significant contributions to the literature.

4. Discussion

In this section, we delve into the key findings and implications drawn from the systematic review of EEG-based MWL estimation and classification. This review synthesizes current research trends, identifies contradictions, establishes consensus points, and explores the feasibility of implementing ML and DL techniques in practical applications outside laboratory settings. By analyzing the methodologies and outcomes of various studies, this discussion aims to provide insights into the evolving landscape of EEG-based MWL assessment, highlighting both the advancements made and the challenges that persist in this dynamic field.

4.1. Trends, Contradictions, and Consensus in EEG-Based MWL Estimation

Several significant trends have emerged in EEG-based MWL estimation methods. There is a noticeable shift towards employing advanced machine ML and DL techniques, such as SVM, KNN, RF, and ANN. DL models, particularly CNN and

J. Neural Eng. (2024)

Table 3. An overview of the evaluated literature on DL techniques

Article	Dataset Source	Subject/ Participant	Paradigm/Stimuli	No. of Channel	Sample Frequency	EEG Workload Estimation Method	Workload Access Level	Data Acquisition Method / Mental Workload Test	Data Augmentation Methods	Classification Technique	Accuracy
[51]	East China University of Science and Technology	six volunteer participants (22–24 years)	aCAMS control task in space capsule	15	500Hz	Time-domain feature	Low, Medium, High	D.A.M: Electrode Placement	-NA-	DBN	87%-91.2%
[52]	University of Shanghai for Science and Technology (USST)	seventeen volunteers	Behavioral stimuli	14	128Hz	PSD	High	EEG Acquisition with EMOTIV Epoc+ EEG headset.	-NA-	CNNs- LeNeT-5, GoogleNet, and EffifcientNet	55–63% by GoogLeNet, and EfficentNet
[53]	SGGS institute	44 (20.7 ± 2.2 years)	Visual tasks (identifying geometric shapes and colors of balloons)	32	500Hz	Entropy features	level 1 (Easy level) and level 4 (Difficult level)	EEG Acquisition with ENOBIO System	Noise addition and time windowing	Deep RNN, k-NN, and ESkNN	92.8%, 85.3%, and 87.8% is obtained with deep RNN, kNN, and ESkNN, respectively
[54]	RMS India	Two healthy male and two healthy female volunteers (25 to 30 years)	Paragraph reading tasks	64	2-32Hz	Neural potentials	Easy, Moderate, Difficult	EEG-based CL Estimation with Paragraph Reading	-NA-	SDAE and LSTM	89.51 % by SDAE and 86.33% LSTM
[55]	Nanyang Technological University	48	SIMKAP Task	14	128Hz	Time Domain Analysis, PSD	Low, Average, High	Emotiv EEG capture electrical signals from the brain	-NA-	BDGN	98%
[56]	EmotivEPOC EEG device	30 average of 25 years	Multi-Attribute Task Battery (MATB)	64	500Hz	PSD	Low, Medium, High	D.A.M: Electrode placement M.W.T: NASA Task Load Index (NASA-TLX)	-NA-	RNN	87.38%
[57]	University of Shanghai for Science and Technology	Seven healthy volunteers (21–24 years	AutoCAMS involving complex control tasks	11	500Hz	PSD	Low, High	AutoCAMS Task Management under Varying NOFS Conditions	Noise addition	SDAE	85.79%
[58]	University of Bologna, Italy	20 participants mean age of 24 (±1.8) years	Driving Simulation	12	256Hz	PSD	Normal & Rush, ROAD-HOUR driving, Easy-Normal and Hard-Rush	BEmicro systems by EBNeuro	-NA-	CNN-AE	94%
[59]	EmotivEPOC EEG device	48	SIMKAP Task	14	128Hz	PSD	Low, Medium, High	Emotiv EEG capture electrical signals from the brain	Time Windowing	DHAT	76.81%
[60]	consumer-grade Epoc device	20 participants (16 male) with an average age of 28.9 ± 4.9 years	Visual, Noise, Pitch, Roughness, Combined Roughness and Noise, Combined Image plus Acoustics Stimuli	14	128Hz	PSD	Low	NASA Task Load Index (NASA-TLX)	Noise Addition and Time Warping	CNN	64%
[61]	"STEW" dataset	48	SIMKAP Task	14	128Hz	PSD	Low, moderate, and high	SIMKAP	Noise Injection and Time Warping	BLSTM-LSTM	97.80 by train/86.33 by test
[62]	STEW dataset	48	SIMKAP Task	14	128Hz	Time-frequency Domain,	low, moderate, and high	SIMKAP, NASA-TLX	-NA-	BiLSTM +ResNet	82.85%
[63]	National University of Singapore	7 healthy participants	Simulated aircraft manipulation tasks in VR	62	256 Hz	PSD	Low, Medium, High	Electrode Placement	-NA-	LSCCN	88.34% ± 4.77%

Table 3. (*Continued*)

Article	Dataset Source	Subject/ Participant	Paradigm/ Stimuli	No. of Channel	Sample Frequency	EEG Workload Estimation Method	Workload Access Level	Data Acquisition Method / Mental Workload Test	Data Augmentation Methods	Classification Technique	Accuracy
[64]	STEW dataset	48	SIMKAP Task	14	-NA-	-NA-	Low and High	Emotiv EEG captures electrical signals from the brain.	Noise Injection, Artifact Injection, and Channel Dropout	Random Forest (RF), KNN, MLP, ConvLSTM, LSTM	RF-58.52% KNN-61.01% MLP-58.68% ConvLSTM-57.19% LSTM-57.%
[65]	DEAP database and two target domain sets	Eight and six healthy participants for target domain (TD)-1 and target domain 2 respectively	AutoCAMS involving complex control tasks	11 for TD-1 15 for TD-2	500 Hz	PSD	Low and High	Electrode Placement	-NA-	TDAE	86% and 90% for TD-1 and TD-2
[66]	An eegoTMmylab amplifier (ANT Neuro, Enschede, The Netherland)	19	n- back Task	64	256 Hz	Connectivity (PLV)	Low, Medium, High	D.A.M: Electrode Placement Systems	Spectral augmentation	CNN, LSTM, and Conv-LSTM	Conv-LSTM, CNN 97.92%
[67]	EEG-MAT dataset	36 students, aged 18–26	Multi-arithmetic Task (MAT)	-NA-	500 Hz	Time-frequency (T–F)	-NA-	D.A.M: Signal Processing Techniques	-NA-	Bi-LSTM	Accuracy rates, with 99.55% (5:5) and 99.88% (8:2)
[68]	The dataset was collected in a modified Sternberg memory experiment.	15	Modified Sternberg memory experiment	64	500 Hz	Spatial, spectral, and temporal	-N/A-	D.A.M: Electrode Placement Systems	Spectral augmentation	Ternary-task Convolutional Bidirectional Neural Turing Machine (TT-CBNTM)	96.3 %
[69]	The dataset was collected from [73]	15	n- back Task	8	256 Hz	PSD	Low, Medium, High	VR-Based n-back Task with EEG Recording	Spectral augmentation	DNN without Dropout, DNN with Dropout, LSTM, CNN	DNN without Dropout-34%, DNN with Dropout-29%, LSTM-31%, CNN-35%
[70]	EEG data recorded by [74]	26 right-handed humans	n-back Task	28	-N/A-	Spectral-temporal features	-N/A-	D.A.M: Electrode Placement Systems	Temporal Augmentation	CNN	70-90%
[71]	Bonn University dataset	5 healthy persons	Epileptic Seizure Detection	100	173.6 Hz	-NA-	-NA-	-NA-	Signal Transformation	Hybrid CNN-LSTM model	99.33%
[72]	Institutional Review Board (IRB) of Korea University	84 participants	Sternberg task	64	500	PSD	Low and High	Starstim R32 (Neuroelectrcis, Spain)	Noise Injection	CNN	90.8%

hybrid models like Conv-LSTM and Bi-LSTM show potential in deriving hierarchical features from EEG data, enhancing classification performance. Additionally, many studies emphasize developing real-time MWL estimation systems for dynamic environments like air traffic control and driving, highlighting the need for efficient and robust algorithms. Furthermore, integrating EEG data with other physiological and behavioral measures (e.g. heart rate variability, eye-tracking) is becoming more common, aiming to enhance robustness and accuracy.

Despite advancements, several contradictions and challenges persist. A significant issue is the inconsistency in validation practices across studies. Different datasets, experimental protocols, and evaluation metrics make direct comparisons of results difficult, hindering generalizability. Many studies also suffer from limited sample sizes, affecting the robustness of conclusions. Methods performing well in controlled laboratory settings may not be as effective in real-world environments with higher noise and variability. Additionally, advanced ML and DL models require significant computational resources and expertise, presenting a trade-off between achieving high accuracy and maintaining usability.

There is consensus on several key points. There is widespread agreement on the need for standardized protocols and benchmarks for validating MWL estimation methods, facilitating more reliable comparisons. Researchers also agree on the importance of validating MWL estimation methods in real-world scenarios, and testing systems in diverse environments, and with varied user populations to ensure robustness and generalizability. Addressing ethical issues related to data privacy, consent, and responsible use of MWL estimation technologies is crucial for broader acceptance and adoption.

42 Practical Considerations for ML and DL Methods in Non-Laboratory Scenarios

The article comprehensively covers a wide range of ML and

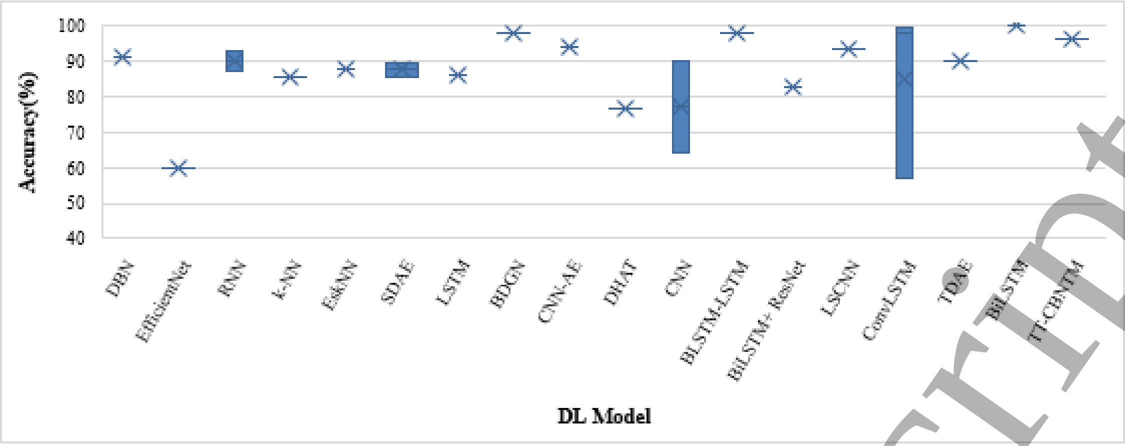

Figure 9. Illustrating the distribution of accuracy scores achieved by different DL models in EEG workload estimation.

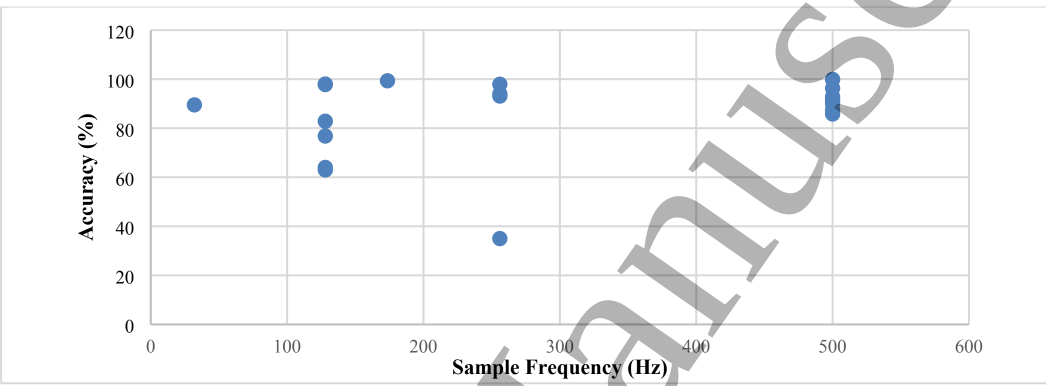

Figure 10. Illustrating the distribution of accuracy scores achieved by different DL models in EEG workload estimation.

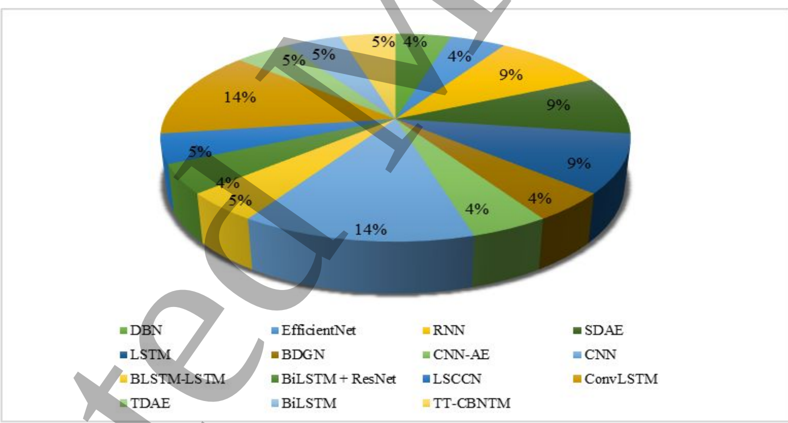

Figure 11. Illustrating the percentage distribution of DL methods employed in EEG workload estimation.

DL techniques for EEG-based workload estimation including ML methods such as DT, RF, KNN, LgR, MLP, SVM, NB, and NNM as well as DL methods including DBN, various CNN like LeNet-5, GoogleNet, and EfficientNet, Deep RNN, k-NN, ESkNN, SDAE, LSTM, BDGN, RNN, CNN-AE, and many others. This section analyzed the discussion, results, and practical suggestions for applying these methods in non-laboratory scenarios.

While these methods demonstrate effectiveness in controlled laboratory settings, their application in non-laboratory environments presents several challenges. Traditional ML techniques like RF, KNN, LgR, MLP, SVM, NB, and NNM typically require well-processed data and controlled conditions for optimal performance. In real-world scenarios, data variability, noise, and environmental factors can significantly impact their accuracy. Future research should focus on developing robust preprocessing techniques and adaptive algorithms that can handle the unpredictability of real-world data. DL models such as DBN, CNN, Deep RNN, SDAE, and LSTM although powerful, are computationally intensive and require large amounts of labeled data for training. Their deployment in real-world applications necessitates access to sufficient computational resources and strategies for continuous learning from streaming data. Additionally, the interpretability of these models in real-world decision-making processes remains a critical challenge.

The variability in user demographics, cognitive states, and task demands is much greater in non-laboratory settings. Developing personalized models that can be adapted to individual users can significantly improve the accuracy and reliability of workload estimation. Techniques such as transfer learning and domain adaptation can help models generalize better across different user groups and tasks. Implementing dynamic frameworks that can adjust to real-time changes in the environment and user context is crucial. These frameworks should be capable of self-tuning and learning from new data to maintain performance.

For MWL estimation methods to be practical in real-world applications, integration with user-friendly and widely accessible technologies is essential. Incorporating these methods into wearable EEG devices or integrating them with existing portable technology such as smartphones and smartwatches can facilitate continuous monitoring in everyday settings. Ensuring ease of use, minimal intrusion, and user comfort are key factors for adoption. Developing intuitive user interfaces that provide real-time feedback based on MWL estimates can enhance user experience and support decision-making. Visualizations and alerts should be designed to be easily interpretable and actionable.

Implementing these methods in non-laboratory environments involves ethical and privacy challenges. Ensuring the security and privacy of user data is paramount. Developing protocols that protect sensitive neurophysiological data and comply with data protection regulations is essential for ethical deployment. Transparent communication with users about how their data will be used and obtaining informed consent is critical. Ethical considerations should be embedded in the development and deployment process.

To bridge the gap between laboratory research and real-world application, future research should conduct longitudinal studies to validate these methods in various real-world scenarios and gather comprehensive performance data. Developing adaptive algorithms that can handle the dynamic and unpredictable nature of real-world environments is necessary. Collaboration with industry partners to pilot these technologies in occupational, educational, and healthcare settings will help evaluate their practicality, usability, and impact. Addressing ethical and privacy concerns by developing robust data protection frameworks and ensuring transparent user communication is essential. By focusing on these practical considerations, future studies can enhance the applicability and effectiveness of ML and DL methods for EEG-based workload estimation in real-world scenarios, making significant contributions to the field.

4.3 Feasibility and Applicability of ML and DL Techniques for Workload Estimation

Workload estimation using EEG signals plays a pivotal role in assessing cognitive states across different tasks and environments. This section delves into the feasibility and applicability of ML and DL techniques for EEG-based workload estimation, highlighting their effectiveness under varying conditions.

4.3.1. Feasibility Under Different Conditions

Task Complexity: For simple tasks with clear classification boundaries, such as distinguishing between low and high workload states, ML techniques like LDA, RF, KNN, LgR, MLP, SVM, NB, and NNM are highly effective. Their simplicity and efficiency make them suitable for real-time applications in scenarios like classroom settings or basic cognitive assessments. While ML techniques can handle simpler tasks, DL models such as DBN and simpler CNNs (LeNeT-5, GoogleNet, EfficientNet) can also be used when higher accuracy and feature representation are required. For more complex tasks requiring nuanced pattern recognition, advanced ML models like ANN and hybrid techniques combining various ML algorithms may be employed. Deep learning models such as CNNs (LeNeT-5, GoogleNet, EfficientNet), ConvLSTM, Deep RNN, BLSTM-LSTM, BiLSTM+ResNet, and LSCCN excel in these scenarios. These models capture intricate patterns and temporal dependencies in EEG data, making them suitable for multi-level workload classification or real-time monitoring in dynamic environments like air traffic control.

Real-Time Processing: In terms of efficiency, ML techniques such as KNN, SVM, and RF offer faster inference times compared to more complex DL models. Their efficiency makes them suitable for applications requiring immediate feedback, such as driver monitoring or adaptive learning systems. While DL models like CNNs and ConvLSTM provide higher accuracy, they demand significant computational resources for real-time processing. Optimized hardware and algorithms are crucial for deploying these models effectively. Regarding computational demands, ML methods such as LDA, NB, and simpler algorithms require less computational power, making them ideal for scenarios with limited resources. In contrast, DL models, especially CNN architectures and hybrid models (e.g., CNN-LSTM), require substantial computational resources. Efficient hardware is essential to support the training and inference of these models in real-time settings.

Data Requirements and Generalization: DL models such as DBN, CNNs, and LSTM-based architectures benefit from large-scale datasets, enabling them to learn complex patterns and generalize across diverse conditions. This advantage is significant in scenarios with varied participant demographics and environmental factors. On the other hand, ML algorithms like SVM, RF, and KNN perform well with moderate-sized datasets and generalize effectively across different tasks and settings. They require less data preprocessing and are less sensitive to variations in data distribution, making them robust for diverse applications.

4.3.2 Applicability in Practical Scenarios

Education and Training: In educational settings, ML techniques like SVM and RF can effectively monitor student engagement and cognitive workload, guiding personalized learning strategies. DL models such as CNNs provide detailed insights into cognitive states during complex learning tasks, enhancing

J. Neural Eng. (2024)

educational outcomes through adaptive learning systems.

Healthcare and Clinical Applications: In healthcare, DL models like LSTM, DBN, and CNNs are instrumental in diagnosing cognitive disorders and assessing cognitive function in clinical populations. They leverage extensive datasets to identify subtle EEG patterns indicative of neurological conditions, supporting early intervention and personalized treatment plans.

Industrial and Professional Environments: For safety-critical tasks, real-time workload estimation using ML techniques such as SVM, KNN, and RF ensures optimal performance and safety in high-stakes professions such as air traffic control or surgical procedures. Advanced DL models like BiLSTM+ResNet and hybrid CNN-LSTM models enhance decision-making by continuously monitoring cognitive load and alerting operators to potential risks, thereby improving safety and efficiency in critical tasks.

This comprehensive analysis underscores the diverse capabilities of ML and DL techniques in EEG-based workload estimation, emphasizing their adaptability to different task complexities, real-time processing demands, and practical applications across various domains. Tailoring model selection to specific application requirements is crucial for optimizing performance and enhancing the utility of workload estimation systems in real-world scenarios.

5. Conclusion and Future Direction

EEG-based workload estimation and classification hold significant potential across various domains, including human-computer interaction, healthcare, and cognitive neuroscience. This paper presents a thorough assessment of the research on ML for EEG-based MWL identification. We systematically found and filtered pertinent papers by applying the PRISMA framework methodically. This allowed us to minimize bias in our review process and ensure the inclusion of high-quality research. The findings demonstrated that EEG parameters, which are indicative of the diversity in data collection techniques and experimental setups, differed significantly between investigations, including the number of channels and sample frequency. PSD, time-frequency domain, entropy features, spectral-temporal features, and PLV were among the features and methodologies used in the EEG workload estimating process. Furthermore, to accommodate various application scenarios, the research used both offline and real-time mechanisms for workload access. EEG-based workload assessment and categorization were greatly aided by ML and DL approaches, which frequently used methods like SVMs, CNNs, and hybrid models (e.g. Conv-LSTM). Finally, the article discusses the trends, contradictions, and consensus in EEG-based MWL estimation, highlights practical considerations for applying ML and DL methods in non-laboratory scenarios, and evaluates the feasibility and applicability of these techniques for workload estimation.

Future directions for investigation in the field of EEG workload classification and estimate show promise for the field's advancement. To provide standardized protocols, datasets, and evaluation criteria, standardization, and benchmarking activities are essential. This will allow for more thorough comparisons of approaches and improve comprehension of their relative efficacy. Furthermore, the accuracy and dependability of workload estimating systems may be increased by including multimodal data, such as merging EEG signals with physiological measurements like heart rate variability or eye tracking. Validation studies in the real world carried out in association with industry partners, are crucial for evaluating the usefulness and practical applicability of EEG-based workload assessment in various operational situations. In addition, the creation of adaptive systems that may modify workload estimation models according to specific cognitive states and job requirements may eventually result in assessments that are more precise and customized. It is possible to find new biomarkers linked to cognitive workload by investigating novel EEG features using methods from graph signal processing or deep learning. This could improve system performance. Along with efforts to comprehend user acceptability and opinions through focused studies, ethical concerns about privacy, permission, and data security must be properly addressed. In conclusion, research into clinical applications such as tracking cognitive burden in patients with neurological illnesses or evaluating cognitive exhaustion in healthcare professionals may provide insightful information and enhance patient outcomes. All of these future efforts are intended to develop EEG workload estimation and categorization, leading to improvements in performance evaluation, well-being, and human-computer interaction in a variety of fields.

Data availability statement

The data used in this study is publicly available in the definite online repository named IEEE Explore, PUBMED, Science Direct, SpringerLink, ACM Digital Library, and seven more articles identified through data sources.

Acknowledgments

Certain images in this publication have been obtained by the author(s) from the Wikimedia website, where they were made available under a Creative Commons license or stated to be in the public domain. Please see individual figure captions in this publication for details. To the extent that the law allows, IOP Publishing disclaims any liability that any person may suffer as a result of accessing, using, or forwarding the image(s). Any reuse rights should be checked and permission should be sought if necessary from Wikimedia and/or the copyright owner (as appropriate) before using or forwarding the image(s).

Funding

This research received no external funding.

J. Neural Eng. (2024)

Conflict of interest statement

The authors have no conflicts of interest to declare.

Ethical statement

No new data were created or analyzed in this study, leading to no ethical protocol requirement.

Appendix A. List of acronyms

ANN	Artificial Neural Network
BLSTM	Bidirectional Long Short-Term Memory
BCI	Brain-Computer Interface
BDGN	Brain-Derived Growth Factor Network
CNN	Convolutional Neural Network
CNN-AE	Convolutional Neural Network with Autoencoder
DBN	Deep Belief Network
DL	Deep Learning
DHAT	Dynamic Hierarchical Attention based on Transformer
EEG	Electroencephalography
ELM	Extreme Learning Machine
fMRI	Functional Magnetic Resonance Imaging
fNIRS	Functional Near-Infrared Spectroscopy
GA	Genetic Algorithm; GRU, Gated Recurrent Unit
GWO	Grey Wolf Optimizer
KNN	K-Nearest Neighbors
LDA	Linear Discriminant Analysis
LgR	Logistic Regression
LSCCN	Latent Space Coding Capsule Network
LSTM	Long Short-Term Memory
MAE	Mean Absolute Error
MEG	Magnetoencephalography
ML	Machine Learning
MLP	Multilayer Perceptron
MWL	Mental Workload
MS	Multiple Sclerosis
PTE	Phase Transfer Entropy
PLV	Phase Locking Value
PRISMA	Preferred Reporting Items for Systematic Reviews and Meta-Analyses
PSD	Power Spectral Density
PSO	Particle Swarm Optimization
RBMs	Restricted Boltzmann Machines
RF	Random Forest
RFE	Recursive Feature Elimination
ResNet	Residual Network
ROC-AUC	Receiver Operating Characteristic - Area Under the Curve
RNN	Recurrent Neural Network
SDAE	Stacked Denoising Autoencoder
SIMKAP	Simultaneous Capacity
SVM	Support Vector Machine
STEW	Simultaneous Task EEG Workload
TDAE	Transfer Dynamical Autoencoder
T-F	Time-Frequency
TT-CBNTM	Ternary-task Convolutional Bidirectional Neural Turing Machine
VAE	Variational Autoencoder

ORCID iDs

Jahid Hassan https://orcid.org/0009-0004-1716-2441

References

[1] A, Goldman C V., Deutsch O and Tsimhoni O 2017 On human–machine relations Cognition, Technology and Work 19 211–31
[2] Holden R J 2011 Cognitive performance-altering effects of electronic medical records: An application of the human factors paradigm for patient safety Cognition, Technology and Work 13 11–29
[3] Roberts R, Flin R, Gas J C-S O E O and and 2013 undefined 2013 Drillers' Situation Awareness onepetro.org
[4] Kamioka D, Murai K, Kitamura K and Wakida S I 2017 Construction of port-coordinator duties list for mental workload assessment 2017 IEEE International Conference on Systems, Man, and Cybernetics, SMC 2017 2017-January 3409–14
[5] DiDomenico A and Nussbaum M A 2008 Interactive effects of physical and mental workload on subjective workload assessment Int J Ind Ergon 38 977–83
[6] Smith-Jackson T, Psychology K K-J of environmental and 2009 undefined Open-plan offices: Task performance and mental workload Elsevier
[7] Paxion J, Galy E and Berthelon C 2014 Mental workload and driving Front Psychol 5
[8] Shriram R, Sundhararajan M, Sundhararajan M and Daimiwal N EEG based cognitive workload assessment for maximum efficiency researchgate.netR Shriram, M Sundhararajan, N DaimiwalInt. Organ. Sci. Res. IOSR, 2013•researchgate.net 34–8
[9] Kosmyna N and Lécuyer A 2019 A conceptual space for EEG-based brain-computer interfaces PLoS One 14
[10] Krucoff M O, Rahimpour S, Slutzky M W, Edgerton V R and Turner D A 2016 Enhancing nervous system recovery through neurobiologics, neural interface training, and neurorehabilitation Front Neurosci 10
[11] Liao C Y, Tai S K, Chen R C and Hendry H 2020 Using EEG and Deep Learning to Predict Motion Sickness under Wearing a Virtual Reality Device IEEE Access 8 126784–96
[12] Tremmel C, Herff C, Sato T, Rechowicz K, Yamani Y and Krusienski D J 2019 Estimating Cognitive Workload in an Interactive Virtual Reality Environment Using EEG Front Hum Neurosci 13 459823
[13] Benerradi J, Maior H A, Marinescu A, … J C-P of the and 2019 undefined 2019 Exploring machine learning approaches for classifying mental workload using fNIRS data from HCI tasks dl.acm.orgJ Benerradi, H A. Maior, A Marinescu, J Clos, M L. WilsonProceedings of the Halfway to the Future Symposium 2019, 2019•dl.acm.org
[14] Appriou A, Cichocki A, CHI F L-E A of the 2018 and 2018 undefined 2018 Towards robust neuroadaptive HCI: exploring modern machine learning methods to estimate mental workload from EEG signals dl.acm.orgA Appriou, A Cichocki, F LotteExtended Abstracts of the 2018 CHI Conference on Human Factors in Computing, 2018•dl.acm.org
[15] Scarapicchia V, Brown C, Mayo C and Gawryluk J R 2017 Functional magnetic resonance imaging and functional near-infrared spectroscopy: Insights from combined recording studies Front Hum Neurosci 11
[16] Coyle S, Ward T, neural C M-J of and 2007 undefined 2007 Brain–computer interface using a simplified functional near-infrared spectroscopy system iopscience.iop.orgSM

Coyle, TE Ward, CM MarkhamJournal of neural engineering, 2007•iopscience.iop.org 4 219–26

[17] Hill N J, Lal T N, Schröder M, Hinterberger T, Widman G, Elger C E, Schölkopf B and Birbaumer N 2006 Classifying event-related desynchronization in EEG, ECoG and MEG signals Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics) 4174 LNCS 404–13

[18] Kannathal N, Choo M L, Acharya U R and Sadasivan P K 2005 Entropies for detection of epilepsy in EEG Comput Methods Programs Biomed 80 187–94

[19] Pfeifer M D, Scholkmann F and Labruyère R 2018 Signal processing in functional near-infrared spectroscopy (fNIRS): Methodological differences lead to different statistical results Front Hum Neurosci 11

[20] Kuanar S, Athitsos V, Pradhan N, Mishra A and Rao K R 2018 Cognitive Analysis of Working Memory Load from Eeg, by a Deep Recurrent Neural Network ICASSP, IEEE International Conference on Acoustics, Speech and Signal Processing - Proceedings 2018-April 2576–80

[21] Aghajani H, Garbey M and Omurtag A 2017 Measuring mental workload with EEG+fNIRS Front Hum Neurosci 11

[22] Smith M E, Gevins A, Brown H, Karnik A and Du R 2001 Monitoring Task Loading with Multivariate EEG Measures during Complex Forms of Human-Computer Interaction https://doi.org/10.1518/001872001775898287 43 366

[23] Lemm S, Blankertz B, Dickhaus T, Neuroimage K M- and 2011 undefined 2011 Introduction to machine learning for brain imaging ElsevierS Lemm, B Blankertz, T Dickhaus, KR MüllerNeuroimage, 2011•Elsevier 56 387–99

[24] Sciaraffa N, Aricò P, Borghini G, Flumeri G Di, Florio A Di and Babiloni F 2019 On the Use of Machine Learning for EEG-Based Workload Assessment: Algorithms Comparison in a Realistic Task Communications in Computer and Information Science 1107 170–85

[25] Anon File:International 10-20 system for EEG-MCN.svg - Wikimedia Commons

[26] Jordan M, Science T M- and 2015 undefined Machine learning: Trends, perspectives, and prospects science.orgMI Jordan, TM MitchellScience, 2015•science.org

[27] Sciaraffa N, Aricò P, Borghini G, Flumeri G Di, Florio A Di and Babiloni F 2019 On the Use of Machine Learning for EEG-Based Workload Assessment: Algorithms Comparison in a Realistic Task Communications in Computer and Information Science 1107 170–85

[28] Müller K, Tangermann M, … G D-J of neuroscience and 2008 undefined Machine learning for real-time single-trial EEG-analysis: from brain–computer interfacing to mental state monitoring ElsevierKR Müller, M Tangermann, G Dornhege, M Krauledat, G Curio, B BlankertzJournal of neuroscience methods, 2008•Elsevier

[29] LeCun Y, Bengio Y, nature G H- and 2015 undefined Deep learning nature.comY LeCun, Y Bengio, G Hintonnature, 2015•nature.com

[30] Lu N, Li T, Ren X, neural H M-I transactions on and 2016 undefined A deep learning scheme for motor imagery classification based on restricted Boltzmann machines ieeexplore.ieee.org

[31] Xu H, Workshop K P-2016 I 18th I and 2016 undefined Affective states classification using EEG and semi-supervised deep learning approaches ieeexplore.ieee.orgH Xu, KN Plataniotis2016 IEEE 18th International Workshop on Multimedia Signal, 2016•ieeexplore.ieee.org

[32] Dong H, Supratak A, Pan W, Wu C, Matthews P M and Guo Y 2018 Mixed Neural Network Approach for Temporal Sleep Stage Classification IEEE Transactions on Neural Systems and Rehabilitation Engineering 26 324–33

[33] Xu H, Academic K P-2016 D M I & and 2016 undefined EEG-based affect states classification using deep belief networks ieeexplore.ieee.orgH Xu, KN Plataniotis2016 Digital Media Industry & Academic Forum (DMIAF), 2016•ieeexplore.ieee.org

[34] Vidyaratne L, Glandon A, … M A-… J C on and 2016 undefined Deep recurrent neural network for seizure detection ieeexplore.ieee.orgL Vidyaratne, A Glandon, M Alam, KM Iftekharuddin2016 International Joint Conference on Neural Networks (IJCNN), 2016•ieeexplore.ieee.org

[35] Kawasaki K, … T Y-2015 7th I and 2015 undefined Visualizing extracted feature by deep learning in p300 discrimination task ieeexplore.ieee.orgK Kawasaki, T Yoshikawa, T Furuhashi2015 7th International Conference of Soft Computing and Pattern, 2015•ieeexplore.ieee.org

[36] Moher D, Liberati A, Tetzlaff J, Altman D G, Antes G, Atkins D, Barbour V, Barrowman N, Berlin J A, Clark J, Clarke M, Cook D, D'Amico R, Deeks J J, Devereaux P J, Dickersin K, Egger M, Ernst E, Gøtzsche P C, Grimshaw J, Guyatt G, Higgins J, Ioannidis J P A, Kleijnen J, Lang T, Magrini N, McNamee D, Moja L, Mulrow C, Napoli M, Oxman A, Pham B, Rennie D, Sampson M, Schulz K F, Shekelle P G, Tovey D and Tugwell P 2009 Preferred Reporting Items for Systematic Reviews and Meta-Analyses: The PRISMA Statement PLoS Med 6 e1000097

[37] Page M J, McKenzie J E, Bossuyt P M, Boutron I, Hoffmann T C, Mulrow C D, Shamseer L, Tetzlaff J M, Akl E A, Brennan S E, Chou R, Glanville J, Grimshaw J M, Hróbjartsson A, Lalu M M, Li T, Loder E W, Mayo-Wilson E, McDonald S, McGuinness L A, Stewart L A, Thomas J, Tricco A C, Welch V A, Whiting P and Moher D 2021 The PRISMA 2020 statement: an updated guideline for reporting systematic reviews BMJ 372

[38] Dancet E A F, Nelen W L D M, Sermeus W, de Leeuw L, Kremer J A M and D'Hooghe T M 2010 The patients' perspective on fertility care: a systematic review Hum Reprod Update 16 467–87

[39] Shepherd J, Harden A, … R R-H education and 2006 undefined Young people and healthy eating: a systematic review of research on barriers and facilitators academic.oup.comJ Shepherd, A Harden, R Rees, G Brunton, J Garcia, S Oliver, A OakleyHealth education research, 2006•academic.oup.com

[40] Hefron R G and Borghetti B J 2017 A new feature for cross-day psychophysiological workload estimation Proceedings - 2016 15th IEEE International Conference on Machine Learning and Applications, ICMLA 2016 785–90

[41] Islam M R, Barua S, Ahmed M U, Begum S, Aricò P, Borghini G and Flumeri G Di 2020 A novel mutual information based feature set for drivers' mental workload evaluation using machine learning Brain Sci 10 1–23

[42] Aygun A, Lyu B, Nguyen T, Haga Z, Aeron S and Scheutz M 2022 Cognitive Workload Assessment via Eye Gaze and EEG in an Interactive Multi-Modal Driving Task ACM International Conference Proceeding Series 337–48

[43] Lobo J L, Ser J Del, De Simone F, Presta R, Collina S and Moravek Z 2016 Cognitive workload classification using eye-tracking and EEG data Proceedings of the International Conference on Human-Computer Interaction in Aerospace, HCI-Aero 2016

[44] Gu H, Yin Z and Zhang J 2019 EEG based mental workload assessment via a hybrid classifier of extreme learning machine and support vector machine Chinese Control Conference, CCC 2019-July 8398–403

[45] Karacan S Ş, Saraoğlu H M, Kabay S C, Akdağ G, Keskinkılıç C and Tosun M 2023 EEG-based mental workload estimation of multiple sclerosis patients Signal Image Video Process 17 3293–301

[46] Samima S and Sarma M 2019 EEG-Based Mental Workload Estimation Proceedings of the Annual International Conference of the IEEE Engineering in Medicine and

Biology Society, EMBS 5605–8

[47] Cao J, Garro E M and Zhao Y 2022 EEG/fNIRS Based Workload Classification Using Functional Brain Connectivity and Machine Learning Sensors 22 7623

[48] Samima S and Sarma M 2023 Mental workload level assessment based on compounded hysteresis effect Cogn Neurodyn 17 357–72

[49] Ke Y, Qi H, Zhang L, Chen S, Jiao X, Zhou P, Zhao X, Wan B and Ming D 2015 Towards an effective cross-task mental workload recognition model using electroencephalography based on feature selection and support vector machine regression International Journal of Psychophysiology 98 157–66

[50] Taheri Gorji H, Wilson N, VanBree J, Hoffmann B, Petros T and Tavakolian K 2023 Using machine learning methods and EEG to discriminate aircraft pilot cognitive workload during flight Scientific Reports 2023 13:1 13 1–13

[51] Zhang J and Li S 2017 A deep learning scheme for mental workload classification based on restricted Boltzmann machines Cognition, Technology and Work 19 607–31

[52] Zhao M, Ji Z, Zhang J, Zhu Y, Ye C, Wang G and Yin Z 2022 Assessing Distinct Cognitive Workload Levels Associated with Unambiguous and Ambiguous Pronoun Resolutions in Human–Machine Interactions Brain Sci 12

[53] Gupta S S, Taori T J, Ladekar M Y, Manthalkar R R, Gajre S S and Joshi Y V. 2021 Classification of cross task cognitive workload using deep recurrent network with modelling of temporal dynamics Biomed Signal Process Control 70 103070

[54] Saha A, Minz V, Bonela S, Sreeja S R, Chowdhury R and Samanta D 2018 Classification of EEG signals for cognitive load estimation using deep learning architectures vol 11278 LNCS (Springer International Publishing)

[55] Afzal M A, Gu Z, Afzal B and Bukhari S U 2023 Cognitive Workload Classification in Industry 5.0 Applications: Electroencephalography-Based Bi-Directional Gated Network Approach Electronics 2023, Vol. 12, Page 4008 12 4008

[56] Vishnu K N, Madhavan S and Gupta C N 2022 Cross-session Classification of Mental Workload Levels using Recurrent Neural Networks Proceedings of the 2022 IEEE Symposium Series on Computational Intelligence, SSCI 2022 201–5

[57] Yin Z and Zhang J 2017 Cross-session classification of mental workload levels using EEG and an adaptive deep learning model Biomed Signal Process Control 33 30–47

[58] Islam M R, Barua S, Ahmed M U, Begum S and Di Flumeri G 2019 Deep Learning for Automatic EEG Feature Extraction: An Application in Drivers' Mental Workload Classification Communications in Computer and Information Science 1107 121–35

[59] Zhang J, Hua Y, Gu J, Chen Y and Yin Z 2022 Dynamic hierarchical learning of temporal-spatial-spectral EEG features with transformers for cognitive workload estimation Chinese Control Conference, CCC 2022-July 7112–7

[60] Bilalpur M, Kankanhalli M, Winkler S and Subramanian R 2018 EEG-based Evaluation of Cognitive Workload Induced by Acoustic Parameters for Data Sonification 315–23

[61] Das Chakladar D, Dey S, Roy P P and Dogra D P 2020 EEG-based mental workload estimation using deep BLSTM-LSTM network and evolutionary algorithm Biomed Signal Process Control 60 101989

[62] Shao S, Han G, Wang T, Lin C, Song C and Yao C 2023 EEG-Based Mental Workload Classification Method Based on Hybrid Deep Learning Model Under IoT IEEE J Biomed Health Inform PP 1–10

[63] Yu Y, Bezerianos A, Cichocki A and Li J 2023 Latent Space Coding Capsule Network for Mental Workload Classification IEEE Transactions on Neural Systems and Rehabilitation Engineering 31 3417–27

[64] Pandey V, Choudhary D K, Verma V, Sharma G, Singh R and Chandra S 2020 Mental Workload Estimation Using EEG Proceedings - 2020 5th International Conference on Research in Computational Intelligence and Communication Networks, ICRCICN 2020 83–6

[65] Yin Z, Zhao M, Zhang W, Wang Y, Wang Y and Zhang J 2019 Physiological-signal-based mental workload estimation via transfer dynamical autoencoders in a deep learning framework Neurocomputing 347 212–29

[66] Workload S C, Using C and Learning D 2021 EEG-Based Functional Connectivity and Deep Learning 1–19

[67] Yedukondalu J, Sharma D and Sharma L D 2023 Subject-Wise Cognitive Load Detection Using Time–Frequency EEG and Bi-LSTM Arab J Sci Eng

[68] Qiao W and Bi X 2020 Ternary-task convolutional bidirectional neural turing machine for assessment of EEG-based cognitive workload Biomed Signal Process Control 57 101745

[69] Ved H and Yildirim C 2021 Detecting Mental Workload in Virtual Reality Using EEG Spectral Data: A Deep Learning Approach Proceedings - 2021 4th IEEE International Conference on Artificial Intelligence and Virtual Reality, AIVR 2021 173–8

[70] Salimi N, Barlow M and Lakshika E 2019 Mental Workload Classification Using Short Duration EEG Data: An Ensemble Approach Based on Individual Channels 2019 IEEE Symposium Series on Computational Intelligence, SSCI 2019 393–8

[71] Pandey S K, Janghel R R, Mishra P K and Ahirwal M K 2023 Automated epilepsy seizure detection from EEG signal based on hybrid CNN and LSTM model Signal Image Video Process 17 1113–22

[72] Kwak Y, Kong K, Song W J, Min B K and Kim S E 2020 Multilevel Feature Fusion with 3D Convolutional Neural Network for EEG-Based Workload Estimation IEEE Access 8 16009–21

[73] Tremmel C, Herff C, Sato T, Rechowicz K, Yamani Y and Krusienski D J 2019 Estimating Cognitive Workload in an Interactive Virtual Reality Environment Using EEG Front Hum Neurosci 13 459823

[74] Shin J, Lühmann A Von, Kim D, data J M-S and 2018 undefined Simultaneous acquisition of EEG and NIRS during cognitive tasks for an open access dataset nature.com