Empowering Computer Science Students in Electroencephalography (EEG) Analysis: A Review of Machine Learning Algorithms for EEG Datasets

Nathan Koome Murungi*
nmurung1@swarthmore.edu
Swarthmore College
Swarthmore, PA, USA

Michael Vinh Pham*
mpham1@swarthmore.edu
Swarthmore College
Swarthmore, PA, USA

Xufeng Caesar Dai*
xdai1@haverford.edu
Haverford Colleg
Haverford, PA, USA

Xiaodong Qu*
xqu1@swarthmore.edu
Swarthmore College
Swarthmore, PA, USA

ABSTRACT

In this paper, we present a systematic literature review that explores the utilization of machine learning (ML) algorithms for analyzing datasets from Electroencephalography (EEG) based Brain-Computer Interfaces (BCIs). Our primary aim is to provide computer science students with a comprehensive and accessible overview of the role of machine learning in EEG analysis. By synthesizing and organizing recent research from 2020 onwards, our objective is to empower the target audience to develop a solid foundational understanding of the current state of ML-EEG research. Through this work, we intend to enhance the accessibility and comprehension of ML-EEG studies and contribute to advancing BCI technology.

CCS CONCEPTS

• **Human-centered computing** → **Human computer interaction (HCI)**; • **Computing methodologies** → **Machine learning**.

KEYWORDS

Review, machine learning, deep learning, SVM, CNN, Attention, Transformer, Computer Science students, EEG, time series, spatiotemporal data.

ACM Reference Format:
Nathan Koome Murungi, Michael Vinh Pham, Xufeng Caesar Dai, and Xiaodong Qu. 2023. Empowering Computer Science Students in Electroencephalography (EEG) Analysis: A Review of Machine Learning Algorithms for EEG Datasets. In *Proceedings of KDD Undergraduate Consortium (KDD-UC '23).* ACM, New York, NY, USA, 7 pages. https://doi.org/XXXXXXX.XXXXXXX

*Koome, Michael, and Xufeng are undergraduate students. They contributed equally to this research. Professor Xiaodong Qu is the research advisor.

KDD-UC '23, August 06–10, 2023, Long Beach, CA

ACM ISBN 978-1-4503-XXXX-X/23/08…$15.00
https://doi.org/XXXXXXX.XXXXXXX

1 INTRODUCTION

ML has become a pivotal tool in the analysis of EEG data, playing a vital role in the rapidly expanding field of BCI [6, 11, 13, 26, 28, 29]. The availability of abundant EEG datasets has opened up exciting opportunities for researchers to employ ML techniques, driving innovative approaches to EEG data analysis. However, the surging research output has resulted in an overwhelming number of papers, making it challenging for newcomers with a Computer Science (CS) background to navigate the field effectively.

In response to this issue, our paper provides a systematic review of the current literature, with a specific focus on interpreting EEG using ML techniques, highlighting the prevailing trends as of 2023. By synthesizing and organizing these findings, we aim to facilitate a deeper understanding of the current state of BCI research and offer guidance to identify promising directions for future investigations. To further aid readers, we have included Table 1, presenting a comprehensive list of acronyms used throughout this paper.

This comprehensive review aims to empower CS students interested in BCI with the knowledge and insights to contribute meaningfully to this exciting and rapidly evolving field.

Abbreviation	Definition
AE	Autoencoder
ANN	Artificial Neural Network
CNN	Convolutional Neural Network
CV	Computer Vision
DBN	Deep Belief Network
DNN	Deep Neural Network
GAN	Generative Adversarial Network
KNN	K-Nearest Neighbor
LSTM	Long Short-Term Memory
RF	Random Forest
RNN	Recurrent Neural Network
SVM	Support Vector Machine
ViT	Vision Transformer

Table 1: List of Algorithm Acronyms

1.1 Research Questions

In this paper, we address the following research questions to empower CS students in their exploration of EEG analysis using ML:

1. Which machine learning algorithms are recommended for CS students to begin their exploration of EEG analysis? By identifying the most suitable algorithms, this paper equips CS students with essential tools and knowledge, enabling effective navigation of the BCI ML field and building a solid foundation in this domain.

2. What are the key EEG datasets that serve as optimal starting points for CS students in EEG-ML research? This paper provides an insightful overview of current trends, emphasizing popular datasets and their corresponding ML algorithms. By staying updated with cutting-edge research directions, CS students can identify promising avenues for their investigations and stay ahead in the field.

By answering these research questions, this paper aims to provide CS students with a seamless entry into BCI research, facilitating their understanding of fundamental concepts and methodologies. By equipping them with the right knowledge and resources, we aim to nurture the next generation of BCI researchers and foster advancements in this rapidly evolving field.

2 RELATED WORK

Previous research has provided valuable insights into the patterns and trends of EEG data analysis within the context of BCI and ML techniques [6, 19, 27]. While these works serve as a good starting point, it is important to acknowledge the rapid development of deep learning in the last four years, which might have introduced new trends and methodologies in EEG-ML research.

One common observation in the existing literature is that many review papers tend to approach the subject from a neuroscience or biomedical engineering perspective. While these perspectives are crucial in understanding the neurophysiological aspects of EEG data, there is a growing need for reviews that cater specifically to CS students entering the BCI and EEG analysis field. Our paper aims to address this gap by providing a systematic review of the current literature on EEG interpretation using ML, with a focus on empowering CS students with the knowledge and tools necessary for the effective exploration of the EEG-ML landscape.

By presenting the latest advancements in deep learning applied to EEG analysis and highlighting the most relevant trends as of 2023, our research aims to offer a fresh and relevant perspective for CS students seeking to engage in BCI research. In doing so, we aim to bridge the gap between the neuroscientific and computational aspects of EEG data analysis and contribute to the holistic understanding and application of ML techniques in BCI research.

3 METHODS

3.1 Keywords

We conducted a systematic review using the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) method to identify relevant ML-EEG papers, similar to previous studies [6, 12, 27]. The search was conducted over a five-month period, from February to July 2023, encompassing multiple research paper resources, including Google Scholar, Paperwithcode, arXiv, and PubMed.

The following keywords were used for the search: (*'Machine Learning' OR 'SVM' OR 'KNN' OR 'Random Forest' OR 'Deep Learning' OR 'CNN' OR 'RNN' OR 'LSTM' OR 'DNN' OR 'Autoencoder' OR 'GAN' OR 'Attention' OR 'Transformer' OR 'Vision Transformer' OR 'Classification') AND ('Dataset' OR 'Time Series' OR 'Spatio-Temporal') AND ('Seizure' OR 'Emotion' OR 'Motor Imagery') AND ('EEG' OR 'Electroencephalography' AND 'Survey' or 'Review')* .

The search strategy aimed to identify papers relevant to surveys and reviews in the context of EEG data analysis using machine learning techniques. Figure 1 visually represents the search process, illustrating the number of papers found at each step and the number of papers excluded based on predefined inclusion and exclusion criteria. Considering the target audience's limited time, we narrowed down the recommended papers to provide new researchers with a manageable selection to familiarize themselves with the current trends within the field.

3.2 Paper Selection Criteria

To ensure the relevance and quality of the research papers included in our review, we applied the following selection criteria:

- **Dataset Focus:** Selected papers must thoroughly highlight human EEG datasets, providing detailed explanations of their creation and acquisition.
- **Publication Time frame:** To maintain up-to-date insights, we considered literature published after 2020.

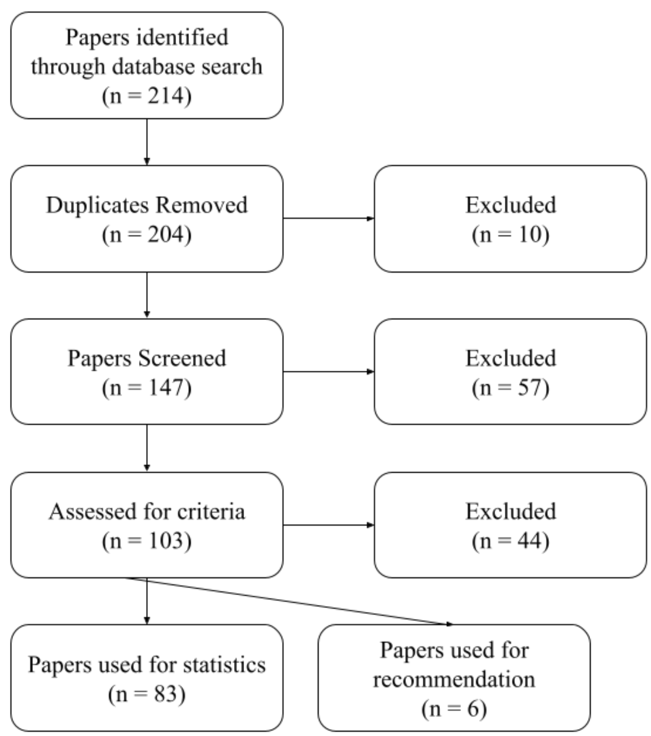

Figure 1: Selection process for the papers

- **ML/DL Focus:** Papers were highly prioritized if they extensively covered Machine Learning and/or Deep Learning approaches for EEG data processing.
- **Reproducibility:** We included papers that provided essential reproducible components, such as data sources, results, code, feature extraction, and pre-processing details.
- **Target Audience:** The selected papers are particularly relevant for undergraduate students majoring or minoring in computer science or data science, with an interest in machine learning. While prior knowledge in Python programming is beneficial for thorough comprehension and replication of our work, Sci-kit learn and PyTorch skills are considered helpful but not mandatory.

4 RESULTS

4.1 EEG Tasks

Our review reveals that EEG experiments are primarily conducted in clinical settings, focusing on specific tasks like Motor Imagery (MI), Seizure Detection, and Emotion Detection. Non-clinical experiments have also gained popularity in recent years [6, 26, 38].

For CS students, we recommend exploring these common EEG tasks due to their high accuracy and reasonable computational cost. Table 2 presents the distribution of EEG tasks in BCI research literature, with Motor Imagery, Seizure Detection, and Emotion Detection highlighted as the top three tasks.

Figure 2 showcases the most prevalent machine learning algorithms used for Motor Imagery analysis. For Emotion Detection, commonly used algorithms include CNN, RNN, SVM, and KNN, while for Seizure Detection, frequently employed algorithms are CNN, RNN, Transformer, and KNN.

4.2 ML Algorithms

In line with previous research [6, 27], Convolutional Neural Networks (CNNs) continue to be the most frequently used algorithm in EEG analysis. However, an interesting shift in the landscape has been observed since 2020, with Transformers gaining popularity and replacing many Recurrent Neural Networks (RNNs). In fact, Transformers are on track to become the most utilized algorithm in EEG analysis due to their ability to achieve similar or even higher prediction accuracy while significantly reducing runtime.

Notably, Deep Belief Networks (DBNs), which constituted a substantial portion of earlier BCI research reviews by 2019 as highlighted in [6, 27], are now becoming less popular in EEG analysis.

Table 3 provides a breakdown of the most commonly used algorithms in non-review papers over the last four years, with CNNs, RNNs, and Transformers standing out as the most prevalent choices.

Task	Paper Count
Motor Imagery	26
Emotion	19
Seizure	16

Table 2: Most Common Task Breakdown for Non-Review Papers

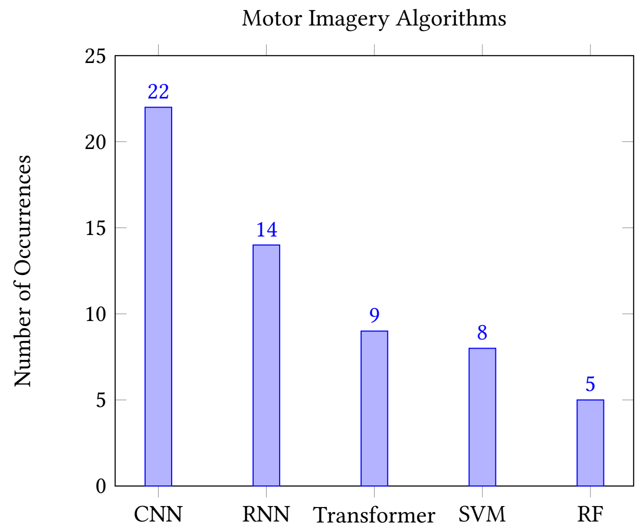

Figure 2: Breakdown of Algorithms used for MI

4.2.1 CNN. The Convolutional Neural Network (CNN) is a prevalent deep learning architecture used in both Computer Vision (CV) and BCI. In EEG-based BCI research, CNNs and their variations account for nearly fifty percent of the ML algorithms employed, as highlighted in previous review papers [6, 27].

For instance, a CNN model was used to classify stages of epileptic seizures based on EEG data[23]. CNN's strength lies in its ability to extract essential features and spatial relationships from EEG spectrograms, representing EEG signals' frequency over time.

The CNN architecture consists of convolutional, pooling, and fully connected layers, enabling effective feature extraction and pattern comprehension. CNNs are well-suited for EEG analysis, as they process EEG spectrogram data like image-like data.

In summary, CNNs have emerged as a powerful tool for EEG analysis, facilitating feature extraction and accurate classification. Their application in EEG-based BCI research opens new possibilities for real-world applications and deeper insights into brain dynamics.

4.2.2 RNN. Recurrent Neural Networks (RNNs) are essential neural networks designed for sequential data, making them ideal for processing time-series data like EEG. The Long Short-Term Memory (LSTM) variant is commonly used in EEG classification to address the vanishing gradient problem and learn long-term dependencies

Algorithm	Paper Count	MI	Seizure	Emotion
CNN	42	22	5	6
RNN	33	14	6	8
Transformer	16	9	2	3
SVM	15	8	3	3
RF	14	5	4	4
KNN	12	4	3	2

Table 3: Algorithm Breakdown for Non-Review Papers

in the data. LSTM incorporates "gates" to retain relevant information for longer periods.

RNN models, particularly LSTM, are often combined with other architectures like CNNs to create a powerful feature extraction and sequence modeling pipeline. This combination has been successfully applied in EEG classification, as shown in [3].

However, one limitation of RNN models is their higher computational cost compared to other architectures like CNNs and Transformers. Researchers must consider this trade-off when dealing with large-scale EEG data or real-time applications.

In conclusion, RNNs, especially LSTM, are valuable for capturing temporal dependencies in EEG data. When combined with CNNs, they yield powerful results for EEG classification. However, the computational cost should be taken into account based on the research requirements and available resources.

4.2.3 Transformers. As of 2023, transformers [36, 37, 39, 40, 42, 43] are increasingly popular for EEG classification tasks and similar time-series data analysis. Originally designed for extensive sequential data like NLP, transformers have gained attention for their versatility and power, exemplified by chatGPT.

Transformers use self-attention to capture relationships in a sequence, making them effective in representing data. They consist of encoder and decoder layers with multi-head self-attention, position-wise feed-forward networks, normalization layers, and residual connections.

Transformers offer advantages over RNNs, such as parallelizability, enabling faster training and inference on larger datasets with complex models. Vision Transformers (ViT) adapt transformers to CV and show promise as an alternative to traditional CNNs in EEG analysis. Pretrained transformers demonstrate great potential in EEG tasks through transfer learning and fine-tuning techniques.

In conclusion, transformers are a transformative architecture for EEG tasks, offering improved performance and parallelizability. They will likely play a central role in advancing EEG-based BCI and facilitating cutting-edge research in the field.

4.2.4 SVM, KNN, and RF. Classic ML methods, such as SVM, KNN, and Random Forests, play a significant role in EEG analysis due to their ease of implementation, interpretability, and reasonable prediction accuracies with low computational cost.

SVM is versatile, capable of linear and non-linear separation using kernel functions. It is used in EEG analysis, as shown in diagnosing autism in [15].

KNN relies on nearest neighbor concept, suitable for EEG analysis due to time continuity effects [25].

Ensemble methods like Random Forests and XGBoost combine weaker classifiers for higher prediction accuracies, with faster runtime than deep learning approaches.

In summary, basic algorithms (SVM and KNN) and ensemble methods (Random Forests and XGBoost) are reliable baselines for EEG benchmark tasks. Deep learning approaches (CNN and RNN) offer higher accuracies at higher computational cost. The emergence of Transformers presents an attractive option, achieving similar accuracies with lower computational demands.

4.3 Datasets

Previous research has summarized numerous public EEG datasets [1, 6, 12, 27]. Each dataset was collected for different research purposes and contains diverse labels representing various mental information, which are subsequently used for classification tasks.

Given the overwhelming number of available datasets, it can take time for new researchers to decide where to begin. To address this issue, we carefully selected a subset of datasets that we believe are well-suited for our target audience, computer science (CS) students.

Table 4 provides an overview of the selected datasets along with their key characteristics. Notably, the DEAP dataset is currently the most highly cited dataset in the field. On the other hand, the EEGEyeNet dataset is a new addition with baseline and benchmark data, making it an excellent choice for CS students to start their exploration. This dataset aligns with our research team's new member training, covering various aspects of ML-EEG analysis.

4.3.1 DEAP. The DEAP dataset [17] offers a valuable human affective state analysis resource, combining EEG and peripheral physiological signals from 32 participants. During data collection, participants watched 40 one-minute music video excerpts and rated them based on arousal, valence, like/dislike, dominance, and familiarity.

This dataset provides rich EEG data, allowing researchers to explore brain activity patterns linked to emotional responses. Additionally, peripheral physiological signals offer insights into physiological changes during affective experiences.

Notably, frontal face recordings of 22 participants enhance the dataset, enabling investigations into facial expressions and their emotional connections.

With diverse data modalities and comprehensive affective annotations, the DEAP dataset supports emotion recognition, affective computing, and EEG-based analysis studies. It facilitates machine learning model development, brain response understanding, and physiological signal correlation with emotions.

4.3.2 CHB-MIT. This Scalp EEG dataset[30] comprises EEG recordings from 22 pediatric subjects with intractable seizures. The dataset was designed for monitoring subjects after anti-seizure medication withdrawal to assess their suitability for surgical intervention.

Focused on epilepsy research, the dataset provides insights into brain electrical activity during seizures, with 182 seizure onsets and ends. Its pediatric nature makes it valuable for understanding pediatric epileptic conditions and advancing pediatric neurology.

Researchers can use the CHB-MIT Scalp EEG dataset to develop seizure prediction algorithms, seizure detection methods, and explore novel epilepsy management approaches.

Dataset	Task	Year	Cited
DEAP [18]	Emotion	2011	3439
CHB-MIT [30]	Seizure	2009	887
BCI Competition IV [35]	MI	2012	837
SEED [10]	Emotion	2013	659
EEGEyeNet [16]	EEG-Eye	2021	20

Table 4: Dataset Breakdown for Non-Review Papers

4.3.3 BCI Competition IV. The BCI Competition IV dataset [35] serves as a vital resource for validating signal processing techniques in Brain-Computer Interfaces (BCIs), focusing on Motor Imagery (MI) tasks. It offers a standardized benchmark for evaluating BCI algorithms.

The dataset comprises two parts: Part 1 (Calibration) and Part 2 (Evaluation).

Part 1 includes spontaneous brain activity recordings from 7 subjects using 64 EEG channels (1000Hz). It provides data for left hand, right hand, foot (with an idle state) MI tasks, facilitating BCI system calibration.

Part 2 is divided into two subparts:

Part 2a has EEG motor-imagery data from 9 subjects (22-electrode EEG), with two sessions containing 288 four-second trials of imagined movements. It enables performance evaluation in a controlled setting.

Part 2b has EEG motor-imagery data from 9 subjects (3-electrode EEG), with five sessions incorporating online feedback for an interactive BCI paradigm.

Researchers can use the BCI Competition IV dataset to benchmark signal processing techniques, classification algorithms, and BCI control strategies, exploring various aspects of BCI system development.

4.3.4 SEED. The SEED dataset [5] investigates emotions and their neural correlates using EEG and eye movement data. It includes recordings from 12 participants with both EEG and eye movement data, along with additional EEG data from 3 participants. Participants watched film clips inducing positive, negative, and neutral emotions.

The dataset's multimodal nature allows researchers to study the relationship between emotions, brain activity, and eye movements comprehensively. Variations like SEED-IV, SEED-VIG, SEED-V, SEED-FRA, and SEED-GER provide diverse cultural and emotional contexts for exploration.

With a focus on affective computing and cognitive neuroscience, the SEED dataset has been widely utilized in emotion recognition and brain response studies. Researchers can leverage this dataset to develop emotion classification algorithms and deepen their understanding of emotional processing mechanisms.

4.3.5 EEGEyeNet. The EEGEyeNet dataset [16] is a pioneering resource that combines EEG and EyeTracking (ET) recordings from 356 participants. This novel dataset and benchmark advance research at the intersection of brain activities and eye movements.

The EEGEyeNet benchmark presents three gaze prediction tasks of increasing difficulty: left-right, angle-amplitude, and absolute position. Researchers can evaluate their models using solid baselines from classical machine learning and large neural networks.

With complete code and data release, along with an easy-to-use interface, collaboration and innovation in gaze prediction research are encouraged, fostering a better understanding of brain-eye interactions.

5 DISCUSSION

In this discussion section, we provide concrete recommendations for undergraduate CS student researchers to get started in the BCI field. Based on the results, we suggest focusing on the top three tasks, algorithms, datasets, and papers to build a strong foundation for their research.

Tasks: We recommend students start with Motor Imagery (MI) recognition, Seizure detection, and Emotion classification. These tasks are well-established in BCI research, with abundant datasets and studies available, facilitating reproducibility and high accuracy.

Algorithms: For EEG analysis, we recommend students explore Convolutional Neural Networks (CNNs), Recurrent Neural Networks (RNNs), and Transformers. CNNs are widely used and excel in processing EEG spectrogram data. RNNs, particularly Long Short-Term Memory (LSTM) variants, are effective in capturing temporal dependencies in sequential EEG data. Transformers are an emerging and powerful option, offering parallel processing and competitive accuracy.

Datasets: We suggest students consider the DEAP dataset for human affective state analysis, BCI Competition IV dataset for Motor Imagery tasks, and the CHB-MIT Scalp EEG dataset for seizure analysis. These datasets are widely used, well-documented, and provide diverse EEG signals for comprehensive research.

Literature Review: We highlight the growing importance of Transformers in BCI research. Transformers were originally designed for sequential data like natural language processing (NLP) and have shown promise in EEG analysis due to their self-attention mechanism. Researchers can explore recent papers like [8] and [31] to understand the application of Transformers in EEG classification tasks.

In conclusion, we aim to provide CS student researchers with practical advice and focused recommendations for their BCI and EEG analysis research. By concentrating on the recommended tasks, algorithms, datasets, and exploring the potential of Transformers, undergraduate researchers can confidently embark on their journey in this exciting and rapidly evolving field.

In addition to our previous recommendations, we provide a further step-by-step guide to CS students conducting EEG analysis:

1. Conduct a Thorough Literature Review: Perform a comprehensive literature review to understand the existing research landscape and gain insights into successful approaches. Analyze review papers and individual experiments to shape and guide your research.

2. Mapping Research Questions to Algorithms: Refer to table 5 for insights into mapping research questions to appropriate machine learning algorithms for EEG analysis. This mapping will help students choose the most suitable algorithms for their specific investigations, ensuring efficient and effective research.

3. Utilize Relevant Datasets: Identify and utilize datasets that align with your research questions. The systematic review has

Dataset	Questions	Algorithms
DEAP [18]	Classification	SVM, CNN, RNN
CHB-MIT [30]	Classification	SVM, KNN, CNN
BCI Competition IV [35]	Classification	SVM, CNN, LSTM
SEED [10]	Classification	SVM,KNN, CNN
EEGEyeNet [16]	Regression	CNN, Transformer

Table 5: Dataset Breakdown for Non-Review Papers

highlighted datasets suitable for tasks such as motor imagery recognition, seizure detection, and emotion classification. Working with relevant datasets ensures meaningful and insightful results.

4. Enhance the Quality of Work: By streamlining the research process and leveraging appropriate algorithms and datasets, students can enhance the overall quality of their work. Focusing on well-defined tasks and relevant data ensures a solid foundation for their research.

5. Begin with Supervised Learning: Starting with supervised learning is a common approach in classic machine learning research. We recommend focusing on classification tasks within supervised learning for CS students beginning their EEG ML journey. Tasks like motor imagery recognition or emotion classification offer clear objectives and straightforward methodologies, making them ideal starting points for a smooth introduction to EEG ML research.

Later in your research journey, you can further expand your exploration into other areas of machine learning. For instance, you may delve into regression tasks within supervised learning, which involve predicting continuous values, or explore unsupervised learning tasks like clustering, where the objective is to group similar data points together based on their similarities. This gradual progression will enable you to comprehensively understand various machine learning techniques and their applications in EEG analysis.

6. Embrace Peer-Review Process: Seek feedback from peers and experts in the field through the peer-review process. Engaging in constructive criticism and incorporating valuable feedback can significantly improve the quality and rigor of your research.

7. Submit Research to Conferences: Consider submitting your research to computer science conferences, especially those relevant to EEG analysis and BCI research. Presenting your work at conferences provides valuable exposure, fosters collaboration, and opens up opportunities for further research.

By following these additional suggestions, CS students can confidently approach EEG analysis, focus on relevant research questions, and utilize appropriate algorithms and datasets to contribute meaningfully to the field of Brain-Computer Interfaces.

We recommend the following math courses for CS students pursuing ML-EEG research:

Linear Algebra: This course provides a foundation for machine learning calculations, which is particularly useful when working with EEG data represented as matrices or high-dimentional vectors. Understanding linear algebra is essential for tasks like dimensionality reduction and feature engineering in EEG analysis.

Statistics and Probability: A strong grasp of statistics and probability is crucial for ML-EEG research. Many classical machine learning models, such as Naive Bayes and regressions, rely on statistical principles. This knowledge is particularly useful for EEG artifact removal and data normalization.

Calculus: Calculus is widely used in machine learning training procedures, such as optimization and backpropagation. A solid understanding of calculus is essential for implementing and fine-tuning ML algorithms in EEG analysis.

By taking these math courses, undergraduate CS students can equip themselves with the necessary mathematical tools and concepts to excel in the ML-EEG research domain, enabling them to tackle complex challenges and make meaningful contributions to the field.

Limitations: While our systematic review provides a comprehensive overview of ML-EEG research, there are a few limitations to acknowledge. Firstly, the field of ML-EEG is rapidly evolving, and new datasets, algorithms, and research papers may emerge after our review. It is crucial for researchers to stay up-to-date with the latest advancements in the field. Additionally, the selection of datasets and algorithms recommended in this review is not exhaustive, and other valuable resources may exist.

Future Work: Exciting avenues for further research in ML-EEG include exploring subject-task relations within and across subjects, investigating methods like knowledge graphs (KG) utilized in recommendation systems [14, 41, 45], and applying deep learning and transfer learning on time-series data for health applications and beyond[2, 4, 7, 9, 20–22, 24, 32–34, 44]. Additionally, novel combinations of algorithms, such as hybrid approaches of CNNs and Transformers, could potentially lead to even higher accuracy and efficiency in EEG classification. Moreover, incorporating interpretability and explainability techniques for ML models in EEG analysis holds promise in enhancing the understanding and trustworthiness of the results. Finally, extending the focus to unsupervised and reinforcement learning paradigms may offer valuable insights into the brain's underlying dynamics. CS students ' active exploration of these directions can significantly contribute to the advancement of ML-EEG research and its applications in Brain-Computer Interfaces and neuroscience.

6 CONCLUSION

In this paper, we conducted a systematic review of EEG analysis using machine learning for computer science students. Our aim was to provide valuable starting points by presenting commonly used tasks, algorithms, and datasets.

We highlight the importance of tasks such as motor imagery, seizure detection, and emotion classification, which offer abundant datasets and well-established studies with high accuracy. These tasks serve as ideal entry points for new researchers to gain practical experience and replicate previous results.

Additionally, we identified key algorithms, namely SVM, CNN, and Transformers, which are gaining popularity in EEG-ML research. Leveraging these algorithms can enhance the accuracy and efficacy of EEG analysis.

By following our recommendations, CS student researchers can establish a solid foundation and confidently contribute to the rapidly evolving field of Brain-Computer Interfaces (BCI) research. Our systematic overview and suggested starting points aim to empower newcomers, facilitating their exploration of novel techniques and encouraging valuable contributions to the advancements in ML-EEG research.

As the field continues to progress, we anticipate exciting opportunities for CS students to make meaningful contributions and drive the future of BCI research. Embracing these recommendations, they can embark on a rewarding journey of discovery and innovation in the fascinating realm of EEG analysis using machine learning.

REFERENCES

[1] Mohit Agarwal. 2021. On the interplay between brain-computer interfaces and machine learning algorithms: A systems perspective. (2021).

[2] Sizhe An, Ganapati Bhat, Suat Gumussoy, and Umit Ogras. 2023. Transfer learning for human activity recognition using representational analysis of neural networks. *ACM Transactions on Computing for Healthcare* 4, 1 (2023), 1–21.

[3] Pouya Bashivan, Irina Rish, Mohammed Yeasin, and Noel Codella. 2015. Learning representations from EEG with deep recurrent-convolutional neural networks. *arXiv preprint arXiv:1511.06448* (2015).

[4] Lulu Chen, Yingzhou Lu, Chiung-Ting Wu, Robert Clarke, Guoqiang Yu, Jennifer E Van Eyk, David M Herrington, and Yue Wang. 2021. Data-driven detection of subtype-specific differentially expressed genes. *Scientific reports* 11, 1 (2021), 332.

[5] Yucel Cimtay and Erhan Ekmekcioglu. 2020. Investigating the use of pretrained convolutional neural network on cross-subject and cross-dataset EEG emotion recognition. *Sensors* 20, 7 (2020), 2034.

[6] Alexander Craik, Yongtian He, and Jose L Contreras-Vidal. 2019. Deep learning for electroencephalogram (EEG) classification tasks: a review. *Journal of neural engineering* 16, 3 (2019), 031001.

[7] Ranadeep Deb, Ganapati Bhat, Sizhe An, Holly Shill, and Umit Y Ogras. 2021. Trends in technology usage for parkinson's disease assessment: A systematic review. *MedRxiv* (2021), 2021–02.

[8] Alexey Dosovitskiy, Lucas Beyer, Alexander Kolesnikov, Dirk Weissenborn, Xiaohua Zhai, Thomas Unterthiner, Mostafa Dehghani, Matthias Minderer, Georg Heigold, Sylvain Gelly, et al. 2020. An image is worth 16x16 words: Transformers for image recognition at scale. *arXiv preprint arXiv:2010.11929* (2020).

[9] Guangyao Dou, Zheng Zhou, and Xiaodong Qu. 2022. Time Majority Voting, a PC-Based EEG Classifier for Non-expert Users. In *HCI International 2022-Late Breaking Papers. Multimodality in Advanced Interaction Environments: 24th International Conference on Human-Computer Interaction, HCII 2022, Virtual Event, June 26–July 1, 2022, Proceedings*. Springer, 415–428.

[10] Ruo-Nan Duan, Jia-Yi Zhu, and Bao-Liang Lu. 2013. Differential entropy feature for EEG-based emotion classification. In *2013 6th International IEEE/EMBS Conference on Neural Engineering (NER)*. IEEE, 81–84.

[11] Shu Gong, Kaibo Xing, Andrzej Cichocki, and Junhua Li. 2021. Deep learning in EEG: Advance of the last ten-year critical period. *IEEE Transactions on Cognitive and Developmental Systems* 14, 2 (2021), 348–365.

[12] Khondoker Murad Hossain, Md Islam, Shahera Hossain, Anton Nijholt, Md Atiqur Rahman Ahad, et al. 2023. Status of deep learning for EEG-based brain–computer interface applications. *UMBC Student Collection* (2023).

[13] Essam H Houssein, Asmaa Hammad, and Abdelmgeid A Ali. 2022. Human emotion recognition from EEG-based brain–computer interface using machine learning: a comprehensive review. *Neural Computing and Applications* 34, 15 (2022), 12527–12557.

[14] Chao Jiang, Victoria Ngo, Richard Chapman, Yue Yu, Hongfang Liu, Guoqian Jiang, and Nansu Zong. 2022. Deep Denoising of Raw Biomedical Knowledge Graph From COVID-19 Literature, LitCovid, and Pubtator: Framework Development and Validation. *Journal of medical Internet research* 24, 7 (2022), e38584.

[15] Jiannan Kang, Xiaoya Han, Jiajia Song, Zikang Niu, and Xiaoli Li. 2020. The identification of children with autism spectrum disorder by SVM approach on EEG and eye-tracking data. *Computers in biology and medicine* 120 (2020), 103722.

[16] Ard Kastrati, Martyna Beata Płomecka, Damián Pascual, Lukas Wolf, Victor Gillioz, Roger Wattenhofer, and Nicolas Langer. 2021. EEGEyeNet: a simultaneous electroencephalography and eye-tracking dataset and benchmark for eye movement prediction. *arXiv preprint arXiv:2111.05100* (2021).

[17] Sander Koelstra, Christian Muhl, Mohammad Soleymani, Jong-Seok Lee, Ashkan Yazdani, Touradj Ebrahimi, Thierry Pun, Anton Nijholt, and Ioannis Patras. 2011. Deap: A database for emotion analysis; using physiological signals. *IEEE transactions on affective computing* 3, 1 (2011), 18–31.

[18] Sander Koelstra, Christian Muhl, Mohammad Soleymani, Jong-Seok Lee, Ashkan Yazdani, Touradj Ebrahimi, Thierry Pun, Anton Nijholt, and Ioannis Patras. 2012. DEAP: A Database for Emotion Analysis ;Using Physiological Signals. *IEEE Transactions on Affective Computing* 3, 1 (2012), 18–31. https://doi.org/10.1109/T-AFFC.2011.15

[19] Fabien Lotte, Laurent Bougrain, Andrzej Cichocki, Maureen Clerc, Marco Congedo, Alain Rakotomamonjy, and Florian Yger. 2018. A review of classification algorithms for EEG-based brain–computer interfaces: a 10 year update. *Journal of neural engineering* 15, 3 (2018), 031005.

[20] Yingzhou Lu, Huazheng Wang, and Wenqi Wei. 2023. Machine Learning for Synthetic Data Generation: a Review. *arXiv preprint arXiv:2302.04062* (2023).

[21] Xiaoling Luo, Xiaobo Ma, Matthew Munden, Yao-Jan Wu, and Yangsheng Jiang. 2022. A multisource data approach for estimating vehicle queue length at metered on-ramps. *Journal of Transportation Engineering, Part A: Systems* 148, 2 (2022), 04021117.

[22] Xiaobo Ma. 2022. *Traffic Performance Evaluation Using Statistical and Machine Learning Methods.* Ph. D. Dissertation. The University of Arizona.

[23] WL Mao, HIK Fathurrahman, Y Lee, and TW Chang. 2020. EEG dataset classification using CNN method. In *Journal of physics: conference series*, Vol. 1456. IOP Publishing, 012017.

[24] Xiaodong Qu and Timothy J Hickey. 2022. EEG4Home: A Human-In-The-Loop Machine Learning Model for EEG-Based BCI. In *International Conference on Human-Computer Interaction*. Springer, 162–172.

[25] Xiaodong Qu, Peiyan Liu, Zhaonan Li, and Timothy Hickey. 2020. Multi-class time continuity voting for EEG classification. In *Brain Function Assessment in Learning: Second International Conference, BFAL 2020, Heraklion, Crete, Greece, October 9–11, 2020, Proceedings 2*. Springer, 24–33.

[26] Xiaodong Qu, Qingtian Mei, Peiyan Liu, and Timothy Hickey. 2020. Using EEG to distinguish between writing and typing for the same cognitive task. In *Brain Function Assessment in Learning: Second International Conference, BFAL 2020, Heraklion, Crete, Greece, October 9–11, 2020, Proceedings 2*. Springer, 66–74.

[27] Yannick Roy, Hubert Banville, Isabela Albuquerque, Alexandre Gramfort, Tiago H Falk, and Jocelyn Faubert. 2019. Deep learning-based electroencephalography analysis: a systematic review. *Journal of neural engineering* 16, 5 (2019), 051001.

[28] Maham Saeidi, Waldemar Karwowski, Farzad V Farahani, Krzysztof Fiok, Redha Taiar, PA Hancock, and Awad Al-Juaid. 2021. Neural decoding of EEG signals with machine learning: a systematic review. *Brain Sciences* 11, 11 (2021), 1525.

[29] MNAH Sha'Abani, N Fuad, Norezmi Jamal, and MF Ismail. 2020. kNN and SVM classification for EEG: a review. In *InECCE2019: Proceedings of the 5th International Conference on Electrical, Control & Computer Engineering, Kuantan, Pahang, Malaysia, 29th July 2019*. Springer, 555–565.

[30] Ali Hossam Shoeb. 2009. *Application of machine learning to epileptic seizure onset detection and treatment.* Ph. D. Dissertation. Massachusetts Institute of Technology.

[31] Jiayao Sun, Jin Xie, and Huihui Zhou. 2021. EEG classification with transformer-based models. In *2021 IEEE 3rd Global Conference on Life Sciences and Technologies (LifeTech)*. IEEE, 92–93.

[32] Jieyuan Tan, Xiang Shen, Xiang Zhang, Zhiwei Song, and Yiwen Wang. 2022. Estimating Reward Function from Medial Prefrontal Cortex Cortical Activity using Inverse Reinforcement Learning. In *2022 44th Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC)*. IEEE, 3346–3349.

[33] Jieyuan Tan, Xiang Shen, Xiang Zhang, and Yiwen Wang. 2021. Multivariate Encoding Analysis of Medial Prefrontal Cortex Cortical Activity during Task Learning. In *2021 43rd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC)*. IEEE, 6699–6702.

[34] Yang Tang, Shuang Song, Shengxi Gui, Weilun Chao, Chinmin Cheng, and Rongjun Qin. 2023. Active and Low-Cost Hyperspectral Imaging for the Spectral Analysis of a Low-Light Environment. *Sensors* 23, 3 (2023), 1437.

[35] Michael Tangermann, Klaus-Robert Müller, Ad Aertsen, Niels Birbaumer, Christoph Braun, Clemens Brunner, Robert Leeb, Carsten Mehring, Kai J Miller, Gernot Mueller-Putz, et al. 2012. Review of the BCI competition IV. *Frontiers in neuroscience* (2012), 55.

[36] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Łukasz Kaiser, and Illia Polosukhin. 2017. Attention is all you need. *Advances in neural information processing systems* 30 (2017).

[37] Long Yi and Xiaodong Qu. 2022. Attention-Based CNN Capturing EEG Recording's Average Voltage and Local Change. In *International Conference on Human-Computer Interaction*. Springer, 448–459.

[38] Su Zhang, Ziyuan Zhao, and Cuntai Guan. 2023. Multimodal continuous emotion recognition: A technical report for abaw5. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 5763–5768.

[39] Zhengming Zhang, Renran Tian, and Zhengming Ding. 2023. TrEP: Transformer-based Evidential Prediction for Pedestrian Intention with Uncertainty. In *Proceedings of the AAAI Conference on Artificial Intelligence*, Vol. 37.

[40] Zhengming Zhang, Renran Tian, Rini Sherony, Joshua Domeyer, and Zhengming Ding. 2022. Attention-Based Interrelation Modeling for Explainable Automated Driving. *IEEE Transactions on Intelligent Vehicles* (2022).

[41] Huan Zhao, Hui Du, Senqi Yang, and Feixiang Yao. 2022. Rec-RN: User Representations Learning over the Knowledge Graph for Recommendation Systems. In *2022 4th International Conference on Machine Learning, Big Data and Business Intelligence (MLBDBI)*. IEEE, 228–233.

[42] Ziyuan Zhao, Kartik Chopra, Zeng Zeng, and Xiaoli Li. 2020. Sea-Net: Squeeze-And-Excitation Attention Net For Diabetic Retinopathy Grading. In *2020 IEEE International Conference on Image Processing (ICIP)*. IEEE, 2496–2500.

[43] Ziyuan Zhao, Kerui Zhang, Xuejie Hao, Jing Tian, Matthew Chin Heng Chua, Li Chen, and Xin Xu. 2019. Bira-net: Bilinear attention net for diabetic retinopathy grading. In *2019 IEEE International Conference on Image Processing (ICIP)*. IEEE, 1385–1389.

[44] Zheng Zhou, Guangyao Dou, and Xiaodong Qu. 2022. BrainActivity1: A Framework of EEG Data Collection and Machine Learning Analysis for College Students. In *International Conference on Human-Computer Interaction*. Springer, 119–127.

[45] Nansu Zong, Ning Li, Andrew Wen, Victoria Ngo, Yue Yu, Ming Huang, Shaika Chowdhury, Chao Jiang, Sunyang Fu, Richard Weinshilboum, et al. 2022. BETA: a comprehensive benchmark for computational drug–target prediction. *Briefings in bioinformatics* 23, 4 (2022), bbac199.