Research Article

Hierarchical Deep Neural Network for Child Brain Health Assessment

Noah Vassilev

Metropolitan International School, Walter-Gropius-Allee 3, Viernheim, Hessen, 68519, Germany

ABSTRACT

Early assessment of mental health in youth is vital for prevention and timely intervention, yet current clinical practice relies on subjective, labor-intensive questionnaires. This study presents a hierarchical deep neural network that predicts dimensional psychopathology scores from electroencephalography (EEG). Task-specific encoders are trained on power spectral density (PSD) and Hjorth features, demographics, and task annotations, and a second-stage Long Short-Term Memory (LSTM) fuses per-task embeddings to produce subject-level scores. Using the Healthy Brain Network EEG (HBN-EEG) dataset, evaluation occurs under a leave-one-release-out protocol with 5-fold cross-validation. The proposed model achieves a higher R^2 and lower root mean square error (RMSE) than the baselines with statistically significant gains in most comparisons, and shows robustness to heterogeneity in task completion, run counts, and session durations. These results demonstrate considerable robustness against dataset subject- and task-level heterogeneity. This study highlights great potential for an effective and comprehensive AI-driven evaluation method for mental healthcare.

Keywords: Electroencephalography; Machine Learning; Deep Learning; Psychopathology; Mental Health Assessment; Long Short-Term Memory; Power Spectral Density; Hjorth Parameters

INTRODUCTION

Psychiatric and learning disorders are among the most common and disabling conditions, with more than 75% of cases emerging before age 24 (1). Early identification is therefore crucial in children and adolescents (2, 3). High comorbidity is common—about 50% of patients who meet criteria for one disorder also meet criteria for another (4). Categorical approaches are sometimes too discrete to describe the complex overlap in comorbidity. To address this, dimensional approaches incorporating multiple psychopathological factors are utilized for a more continuous and comprehensive evaluation of the disorders (5). The general psychopathology factor (*p-factor*) captures shared variance across symptoms, while specific factors such as internalizing, externalizing, and attention provide additional resolution (6-9). These factors hold predictive value but are currently derived from questionnaires completed by children, parents, or teachers (10), a process that is subjective, time-consuming, and requires expert evaluation (11, 12). Therefore, decoding psychopathology factors from brain imaging utilizing advanced artificial intelligence methods and large-scale population neural data shows great potential for more unbiased and efficient evaluations in clinical mental healthcare.

Electroencephalography (EEG) is one of the most

Corresponding author: Noah Vassilev, E-mail: vassilevnoah@gmail.com.

Accepted October 3, 2025
https://doi.org/10.70251/HYJR2348.35607614

popular brain imaging techniques with attractive advantages: non-invasive, inexpensive, and directly measures neural activity (13). A wide range of machine learning approaches has demonstrated feasibility in decoding EEG signals for applications in mental and neurodevelopmental disorders. Traditional classifiers such as support vector machines (SVM), k-nearest neighbors (KNN) (14), and random forests (RF) have achieved considerable accuracy in detecting conditions including major depressive disorder, bipolar disorder, attention deficit hyperactivity disorder (ADHD), and schizophrenia across 2–3 class classification tasks (15-19). Similarly, Ahire *et al*. [2025] reported 96% accuracy in ADHD classification using a Naive Bayes model (20). More advanced methods, such as artificial neural networks (ANN) and convolutional neural networks (CNN) (21), have further improved the decoding accuracy (22-24). While these results are impressive, existing models have largely focused on categorical classification of specific disorders. Such tasks are less complex than the regression of continuous psychopathology scores and fail to address the high comorbidity between disorders.

In this work, EEG data from publicly available and de-identified Healthy Brain Network EEG (HBN-EEG) Dataset Releases 1–5 was used, sourced from the Healthy Brain Network and formatted according to the Brain Imaging Data Structure (BIDS) standard with annotated Hierarchical Event Descriptors (HED) (2, 7). These releases include high-density (128-channel) EEG recordings from 1125 participants (aged 5–21) across six cognitive tasks, comprising both passive (Resting State (RS), Surround Suppression (SuS), Movie Watching (MW)) and active (Contrast Change Detection (CCD), Sequence Learning (SL), Symbol Search (SyS)) paradigms. Each participant is accompanied by four psychopathology scores: internalizing, externalizing, attention, and p-factor —derived from a bifactor model based on Child Behavior Checklist (CBCL) questionnaire responses. Unlike conventional single-task EEG datasets, HBN-EEG is heterogeneous: not all participants completed every task or run, session durations vary, and task-specific annotations do not directly align with clinical scores. These challenges make existing methods unsuitable and motivate more robust modeling approaches (25).

This study proposes a Hierarchical Deep Neural Network (DNN) Model to automatically predict four psychopathology dimensions — p-factor, internalizing, externalizing, and attention from EEG in children, aiming to automate and scale the assessment of brain health beyond traditional pediatric psychopathology grading methods.

METHODS AND MATERIALS

Data Preprocessing & Segmentation

EEG recordings were first band-pass filtered between 1 and 40Hz using a zero-phase, 10-second finite impulse response (FIR) filter to remove slow drifts and high-frequency noise. Power line interference was suppressed using a 60Hz infinite impulse response (IIR) notch filter with a 3Hz bandwidth. After filtering, signals were downsampled from 500Hz to 250Hz and re-referenced to the 'Cz' electrode, such that all channels reflected voltage differences relative to this reference. Finally, the recordings were segmented into non-overlapping 2-second windows for subsequent feature extraction.

Segmentation was performed in a task-specific manner. For the resting state, data were segmented between eye-open and eye-closed intervals, with eye status encoded as a feature. For movie watching, segments were taken from movie onset to offset, with movie type added as a feature. In surround suppression, the 2.4-second period following each stimulus change was extracted, with foreground contrast, background contrast, and stimulus condition provided as features. Contrast change detection was segmented trial by trial, including participant feedback as an additional feature. In sequence learning, recordings were split by presented sequences, with sequence length and percentage recall encoded as features. For symbol search, segments extended from the search onset to the participant's response. EEG recordings from participants with incomplete demographics (sex, age, handedness) or psychopathology scores were not used for the model.

Feature Extraction

Demographic information (sex, age, handedness) was used alongside features extracted from EEG signals to capture complementary sources of variance. From each task's EEG, the Hjorth parameters—*Activity*, *Mobility*, and *Complexity* —which summarize time-domain amplitude, spectral spread, and waveform regularity (26), and power spectral density (PSD) using Welch's method to quantify oscillatory content (27) were computed. To improve numerical stability and comparability across sessions, PSDs were log-scaled and all features were z-scored within the training data;

the same transform was then applied to validation/ test folds to avoid leakage. Because PSD and Hjorth features are high-dimensional and strongly collinear (across channels, bands, and runs), principal component analysis (PCA) was applied to obtain 20 orthogonal components that retain the dominant variance while reducing noise and overfitting. PCA was fitted only on the combined training set in each inner fold (and for each release split) and then used to transform held-out data.

Hierarchical DNN Model

This study proposes a hierarchical DNN model to predict the four psychopathology scores from EEG recordings. For each participant and task (Figure 1). 20 principal components were used from the extracted features (PSD and Hjorth parameters from the EEG signals, demographic information, and task-specific annotations) as input. In the first stage of the framework, a separate multilayer perceptron (MLP) encoder was trained for each task to map these features to the four target scores, thereby learning task-specific representations optimized with mean squared error loss. In the second stage, the pretrained encoders were leveraged by loading their checkpoints, freezing the weights of these encoders, and extracting the latent embeddings from the penultimate hidden layer of each model. These embeddings, representing task-level abstractions, were concatenated into a sequence and passed into a long short-term memory (LSTM) network, which modeled cross-task dependencies and integrated information across tasks. The output of the LSTM was then trained to predict the four psychopathology scores at the subject level.

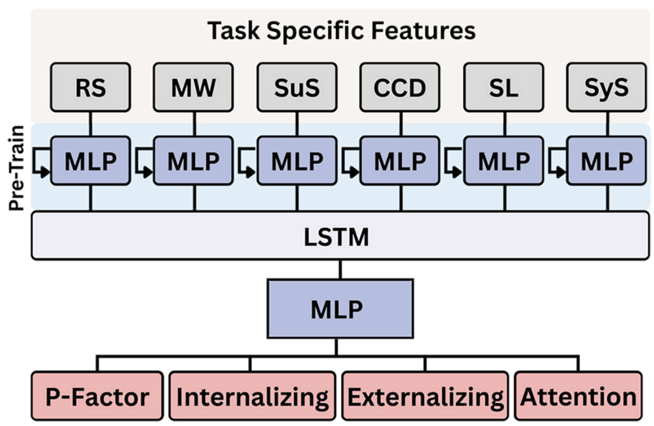

Figure 1. Architecture of the proposed hierarchical DNN model trained and evaluated on 128-Channel high density EEG recordings of 1125 participants (aged 5–21). Hjorth parameters, power spectral density and task-specific features were extracted from 6 tasks: Resting State (RS), Movie Watching (MW), Surround Suppression (SuS), Contrast Change Detection (CCD), Sequence Learning (SL) and Symbol Search (Sys). In the first training stage, separate task-specific multilayer perceptron (MLP) encoders were trained on the features of each task. In the second stage, encoder weights were frozen and all encoder embeddings were passed on to a long short-term memory (LSTM) layer. LSTM output then trained a final encoder to predict the four psychopathology scores: p-factor, internalizing, externalizing and attention on a subject level.

This two-stage design has several advantages over direct end-to-end training: by first optimizing task-specific encoders independently, each model can learn stable representations tailored to its unique feature distributions and labels, while the second-stage LSTM exploits shared structure across tasks without being dominated by missing data or uneven trial counts. The task-specific encoders were trained on each participants available EEG recordings, as not all participants completed every task or run. This ensured the use of all available recordings per task, without needing to exclude participants with incomplete task-coverage. Participants were included in the subject-level model as long as one task was available. In this way, the framework improves generalization, mitigates the impact of heterogeneity across tasks, and enables robust prediction of continuous psychopathology scores in the HBN-EEG dataset.

Experiment Design

To evaluate model generalization across heterogeneous data sources, a Leave-One-Release-Out strategy was employed. Specifically, from the five available HBN-EEG releases [1–5], one release was held out as an independent test set, while the remaining four were used for model training and validation. Within the four training releases, all available data was combined and performed 5-fold cross-validation. This procedure yielded five trained models, each of which was subsequently evaluated on the held-out release.

This process was repeated so that each release served once as the test set. For evaluation, the mean and standard deviation of the performance across folds, considering two regression metrics were reported:

the coefficient of determination R^2 and root mean squared error (RMSE), calculated for each of the four psychopathology scores (*p*-factor, internalizing, externalizing, and attention).

The metrics are defined as follows:

$$R^2 = 1 - \frac{\sum_{i=1}^{n} (y_i - \hat{y}_i)^2}{\sum_{i=1}^{n} (y_i - \bar{y})^2}$$

where are the true labels, the predicted values, and the mean of the true labels.

$$RMSE_{(y,\hat{y})} = \sqrt{\frac{1}{\mathrm{n}} \sum_{i=1}^{n} (y_i - \hat{y}_i)^2}$$

where is the number of samples.

To contextualize the performance of the proposed deep learning framework, three classical regression baselines were implemented: Linear Regression (LR), Random Forest (RF), and k-Nearest Neighbors (kNN). Each baseline model was trained under the same settings with the task-specific encoders, and built an ensemble model across 6 tasks for each of the baselines.

RESULTS & DISCUSSION

The full Leave-One-Release-Out evaluation results are shown in Table 1 and Table 2, RMSE of different models are also plotted in Figure 2.

Independent-sample t-tests were conducted to compare the mean performance of the hierarchical DNN with each baseline. For all four psychopathology dimensions, the proposed hierarchical DNN consistently outperformed the classical machine learning baselines. Compared to kNN and random forest, the proposed DNN achieved significantly higher R^2 values and lower RMSEs for *p*-factor, internalizing, and attention (all p < 0. 95%; all 95% CI of mean R^2 differences = 0.066 – 0.571; all 95% CI of mean RMSE differences = −0.436 – −0.075), confirming its ability to capture

Table 1. Prediction performance of different models for the psychopathology scores p-factor and internalizing

Model	P-Factor		Internalizing	
	R^2	RMSE	R^2	RMSE
Linear Regression	-6.277 ± 8.116	2.256 ± 1.430	-216.535 ± 483.179	5.874 ± 10.960
Random Forest	-0.147 ± 0.065	1.010 ± 0.069	-0.121 ± 0.050	0.854 ± 0.039
kNN	-0.364 ± 0.123	1.094 ± 0.074	-0.308 ± 0.033	0.912 ± 0.026
Our Hierarchical DNN	0.029 ± 0.010	0.749 ± 0.013	0.035 ± 0.006	0.731 ± 0.009

The metrics shown are mean ± standard deviation for the R^2 and root mean square error (RMSE). The models shown include linear regression, random forest, k-Nearest Neighbors (kNN) and our proposed Hierarchical deep neural network (DNN). For both psychopathology scores, our proposed hierarchical DNN achieved the highest predictive performance.

Table 2. Prediction performance of different models for the psychopathology scores externalizing and attention

Model	Externalizing		Attention	
	R^2	RMSE	R^2	RMSE
Linear Regression	-0.475 ± 0.332	0.890 ± 0.118	-0.098 ± 0.115	1.080 ± 0.479
Random Forest	-0.111 ± 0.042	0.776 ± 0.036	-0.157 ± 0.067	0.886 ± 0.027
kNN	-0.313 ± 0.028	0.844 ± 0.048	-0.410 ± 0.121	0.976 ± 0.024
Our Hierarchical DNN	0.007 ± 0.004	0.802 ± 0.007	0.011 ± 0.007	0.780 ± 0.012

The metrics shown are mean ± standard deviation for the R^2 and root mean square error (RMSE). The models shown include linear regression, random forest, k-Nearest Neighbors (kNN) and our proposed Hierarchical deep neural network (DNN). For both psychopathology scores, our proposed hierarchical DNN achieved the highest predictive performance.

stable cross-task representations in a heterogeneous dataset. Linear regression performed especially poorly on the internalizing score as it assumes a single global linear mapping between features and labels, while the coupling between internalizing and task-level EEG features is weak and highly nonlinear, leading to severe underfitting. For the externalizing score, the proposed hierarchical DNN achieved a significantly higher R^2 than all baselines, although the RMSE improvement was not significant against random forest, which achieved slightly lower error variance. The performances of the linear regression baselines had high variability, as they often produced extreme negative R2 values, underscoring their inability to handle the high dimensionality and non-linear structure of HBN-EEG features. These findings demonstrate that the two-stage framework—task-specific MLP encoders followed by a cross-task LSTM—can better integrate information across the six tasks than traditional models designed for single-task EEG classification. These results showed the model's superior performance in generalizing across tasks and subjects. While absolute effect sizes remain modest, reflecting the intrinsic difficulty of predicting questionnaire-derived psychopathology factors from EEG, the observed improvements highlight the potential of hierarchical deep models to automate mental health assessment in children and adolescents using large-scale, heterogeneous neurophysiological datasets.

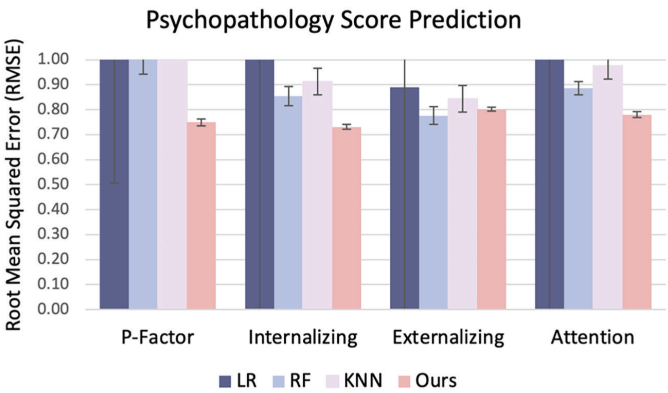

Figure 2. Root mean square error (RMSE) values of the predicted psychopathology scores p-factor, internalizing, externalizing and attention. The models shown are linear regression (LR), random forest (RF), k-Nearest Neighbors (kNN) and our proposed Hierarchical DNN (Ours). Our proposed hierarchical DNN achieved the lowest RMSE for all scores except for externalizing.

Table 3 shows the performance of task-specific models compared to the full Hierarchical DNN Model.

The results underline the proposed model's clear heterogeneity and predictive utility. The tasks Resting State, Movie Watching, and Contrast Change Detection yield small but positive R^2 with competitive RMSE, which indicates moderate alignment with the questionnaire-derived psychopathology scores. RS and MW are passive tasks which are less complex and have reduced motor demands. Similarly, CCD only probes

Table 3. R^2 of Task Specific Model (Mean ± SD)

Tasks	P-Factor	Internalizing	Externalizing	Attention
RS	0.003 ± 0.009	-0.007 ± 0.011	-0.032 ± 0.014	0.003 ± 0.005
MW	0.007 ± 0.004	0.016 ± 0.015	-0.004 ± 0.009	0.001 ± 0.012
SuS	-0.318 ± 0.146	-0.266 ± 0.157	-0.578 ± 0.251	-0.962 ± 0.861
CCD	0.013 ± 0.006	0.006 ± 0.003	-0.008 ± 0.022	-0.047 ± 0.051
SL	-0.097 ± 0.036	-0.903 ± 0.547	-2.320 ± 1.271	-1.216 ± 0.566
SyS	-1.620 ± 2.621	-5.264 ± 0.479	-3.383 ± 0.871	-0.930 ± 0.645
Full	**0.029 ± 0.010**	**0.035 ± 0.006**	**0.007 ± 0.004**	**0.011 ± 0.007**

Prediction performance of the task-specific models for the psychopathology scores p-factor, internalizing, externalizing and attention. Mean R^2 ± standard deviation is shown. Included tasks are Resting State (RS), Movie Watching (MW), Surround Suppression (SuS), Contrast Change Detection (CCD), Sequence Learning (SL) and Symbol Search (SyS). RS, MW and CCD performed the best, yielding small positive R^2 values, whereas SuS, SL and SyS often produced large negative R^2 values. The full hierarchical DNN learned robust task-embeddings allowing it to outperform all individual tasks.

basic sensory processing and attention mechanisms that avoid excessive cognitive load. Therefore, due to the simplicity and consistency of these tasks, EEG signals likely capture robust patterns related to general cognitive engagement. These patterns most likely provide higher predictive power for the questionnaire-derived psychopathology dimensions, resulting in a stronger and more stable performance.

In contrast, Surround Suppression, Sequence Learning, and Symbol Search often produce large negative R^2 (worse than mean prediction) and high variance. SuS assesses primarily visual perception and low-level sensory processing. SL and SyS primarily assess working memory and processing efficiency. For all 3 tasks, the probed sensory and motor processes might not strongly correlate with psychopathology scores. Additionally, trial-specific metrics might not directly map to the scores, due to weak coupling between task-evoked EEG and questionnaire-derived traits, sparse/misaligned annotations, and cross-release co-variate shift. Differences in task availability might further compound this, as SL and SyS have the lowest number of available recordings across all tasks, as there might be insufficient training data leading to poor generalizations or overfitting. This results in the observed weak performance and predictive instability.

This negative transfer is mediated by the proposed hierarchical DNN by learning robust per-task embeddings which are fused via a cross-task LSTM, delivering consistent gains over all single tasks. This underlines the model's effectiveness in integrating heterogenous information from across multiple tasks. Results highlight the importance of multitask representation learning for large scale EEG datasets.

Future work should use task-aware gating or attention mechanisms to further stabilize performance.

CONCLUSION

This paper introduced a hierarchical DNN that learns task-specific EEG representations and fuses them with a cross-task sequence model to predict dimensional psychopathology in children. Unlike existing models that categorize EEG signals into pre-existing classes of psychiatric disorders, the proposed model performs regression of four continuous psychopathology scores. This addresses the high comorbidity between disorders by providing the general p-factor, as well as the three additional scores internalizing, externalizing, and attention provide additional resolution. The proposed model was tested under a leave-one-release-out protocol on HBN-EEG and compared with three classical baselines. Results showed that the proposed hierarchical DNN consistently outperformed strong classical baselines, demonstrating robustness to heterogeneous tasks, uneven runs, and cross-release covariate shift. RS, MW and SuS consistently yielded small but positive R^2 with competitive RMSE. This indicates moderate predictive power for the questionnaire-derived psychopathology scores, most likely due to the simplicity and consistencies of the tasks allowing the model to capture robust and generalizable patterns. SuS, SL and SyS often produced large negative R^2 with high variance, most likely due to insufficient training data, weak coupling between task-evoked EEG and questionnaire-derived traits, sparse/misaligned annotations, and cross-release co-variate shift, leading to weak and performance and predictive instability. Beyond this dataset, the framework offers a general recipe for scalable, objective brain-health assessment: learn reliable per-context encoders, then integrate them with temporally aware fusion that down-weights detrimental contexts. This could support low-cost pre-screening, monitoring, and triage in pediatric mental health, and extend to other biosignals and settings. Future work should consider building an end-to-end self-supervised pretraining method. This should involve learning shared representations that generalize across tasks and support fair, reliable, clinically useful deployment. Therefore, whilst absolute effect sizes are far from effective clinical applications, this study highlights observed improvements in the potential of hierarchical deep models to automate mental health assessment in children and adolescents using large-scale, heterogeneous neurophysiological datasets.

FUNDING SOURCES

No funding was received for the conduct of the research and the preparation of the article.

CONFLICT OF INTERESTS

The author declares no conflict of interests.

REFERENCES

1. Ronald Kessler, Patricia Berglund, Olga Demler, Robert Jin, Kathleen Merikangas, and Ellen Walters. Lifetime prevalence and age-of-onset distributions of

dsm-iv disorders in the national comorbidity survey replication. *Archives of general psychiatry*. 2005; 62: 593–602. doi: 10.1001/archpsyc.62.6.593.

2. Lindsay Alexander, Jasmine Escalera, Lei Ai, Charissa Andreotti, *et al*. An open resource for transdiagnostic research in pediatric mental health and learning disorders. *Scientific Data*. 2017; 4: 170181. doi:10.1038/sdata.2017.181.
3. Adriana Di Martino, Damien Fair, Clare Kelly, Theodore Satterthwaite, *et al*. Unraveling the miswired connectome: A developmental perspective. *Neuron*. 2014; 83: 1335–1353. doi: 10.1016/j.neuron.2014.08.050.
4. Terrie Moffitt, Avshalom Caspi, and Phil Silva. Comorbid mental disorders: Implications for treatment and sample selection. *Journal of abnormal psychology*. 1998; 107: 305–11. doi: 10.1037/0021-843X.107.2.305.
5. Helena Chmura Kraemer, Art Noda, and Ruth O'Hara. Categorical versus dimensional approaches to diagnosis: methodological challenges. *Journal of psychiatric research*. 2004; 38 (1): 17–25. https://doi.org/10.1016/S0022-3956(03)00097-9
6. Caspi A, Houts R, Daniel Belsky, Goldman-Mellow S, *et al*. The p factor: One general psychopathology factor in the structure of psychiatric disorders? *Association for Psychological Science*. 2013; 2: 119–137. https://doi.org/10.1177/2167702613497473
7. Seyed Yahya Shirazi, Alexandre Franco, Mauricio Hoffmann, Nathalia Esper, *et al*. Hbn-eeg: The fair implementation of the healthy brain network (hbn) electroencephalography dataset. 10-2024. https://doi.org/10.1101/2024.10.03.615261
8. Eoin McElroy, Jay Belsky, Natacha Carragher, and Praveetha Patalay. Developmental stability of general and specific factors of psychopathology from early childhood to adolescence: Dynamic mutualism or p-differentiation? *Journal of Child Psychology and Psychiatry*. 2017; 59. doi: 10.1111/jcpp.12849.
9. Benjamin B Lahey, Robert F Krueger, Paul J Rathouz, Irwin D Waldman, and David H Zald. Validity and utility of the general factor of psychopathology. *World Psychiatry*. 2017; 16 (2): 142. https://doi.org/10.1002/wps.20410
10. Isabel A Bordin, Marina M Rocha, Cristiane S Paula, Maria Cristina TV Teixeira, *et al*. Child behavior checklist (cbcl), youth self-report (ysr) and teacher's report form (trf): an overview of the development of the original and brazilian versions. *Cadernos de sa´ude p´ublica*. 2013; 29: 13–28. https://doi.org/10.1590/S0102-311X2013000100004
11. Avshalom Caspi, Renate M Houts, Helen L Fisher, Andrea Danese, and Terrie E Moffitt. The general factor of psychopathology (p): Choosing among competing models and interpreting p. Clinical Psychological Science. 2024; 12 (1): 53–82. https://doi.org/10.1177/21677026221147872
12. Ashley L Watts, Ashley L Greene, Wes Bonifay, and Eiko I Fried. A critical evaluation of the p-factor literature. *Nature Reviews Psychology*. 2024; 3 (2): 108–122. https://doi.org/10.1038/s44159-023-00260-2
13. Alice F Jackson and Donald J Bolger. The neurophysiological bases of eeg and eeg measurement: A review for the rest of us. *Psychophysiology*. 2014; 51 (11): 1061-1071. https://doi.org/10.1111/psyp.12283
14. Nazmus Sakib, Md Kafiul Islam, and Tasnuva Faruk. Machine learning model for computer-aided depression screening among young adults using wireless eeg headset. *Computational Intelligence and Neuroscience*. 2023; 2023 (1): 1701429. https://doi.org/10.1155/2023/1701429
15. M Margarette Sanchez, Lee Borden, Nafiah Alam, Amin Noroozi, *et al*. A machine learning algorithm to discriminating between bipolar and major depressive disorders based on resting eeg data. In 2022 44th Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC). IEEE. 2022; pages 2635-2638. https://doi.org/10.1109/EMBC48229.2022.9871453
16. Manjusha Deshmukh, Mahi Khemchandani, and Paramjit Mahesh Thakur. Contributions of brain regions to machine learning-based classifications of attention deficit hyperactivity disorder (adhd) utilizing eeg signals. *Applied Neuropsychology: Adult*. 2024; pages 1-15. https://doi.org/10.1080/23279095.2024.2368655
17. Kuk-In Jang, Sungkean Kim, Soo Young Kim, Chany Lee, and Jeong-Ho Chae. Machine learning-based electroencephalographic phenotypes of schizophrenia and major depressive disorder. *Frontiers in Psychiatry*. 2021; 12: 745458. https://doi.org/10.3389/fpsyt.2021.745458
18. Elif Sarisik, David Popovic, Daniel Keeser, Adyasha Khuntia, *et al*. Eeg-based signatures of schizophrenia, depression, and aberrant aging: a supervised machine learning investigation. *Schizophrenia Bulletin*. 2025; 51 (3): 804-817. https://doi.org/10.1093/schbul/sbae150
19. Lijuan Duan, Huifeng Duan, Yuanhua Qiao, Sha Sha, *et al*. Machine learning approaches for mdd detection and emotion decoding using eeg signals. *Frontiers in Human Neuroscience*. 2020; 14: 284. https://doi.org/10.3389/fnhum.2020.00284
20. Nitin Ahire, RN Awale, and Abhay Wagh. Electroencephalogram (eeg) based prediction of attention deficit hyperactivity disorder (adhd) using machine learning. *Applied Neuropsychology: Adult*. 2025; 32

(4): 966-977. https://doi.org/10.1080/23279095.2023.2247702

21. Jinpei Han, Xiao Gu, and Benny Lo. Semi-supervised contrastive learning for generalizable motor imagery eeg classification. In 2021 IEEE 17th International Conference on Wearable and Implantable Body Sensor Networks (BSN). IEEE. 2021; pages 1-4. https://doi.org/10.1109/BSN51625.2021.9507038
22. Zaeem Ahmed, Aamir Wali, Saman Shahid, Shahid Zikria, Jawad Rasheed, and Tunc Asuroglu. Psychiatric disorders from eeg signals through deep learning models. *IBRO Neuroscience Reports*. 2024; 17: 300-310. https://doi.org/10.1016/j.ibneur.2024.09.003
23. Yanting Xu, Hongyang Zhong, Shangyan Ying, Wei Liu, *et al*. Depressive disorder recognition based on frontal eeg signals and deep learning. *Sensors*. 2023; 23 (20): 8639. https://doi.org/10.3390/s23208639
24. Adil O Khadidos, Khaled H Alyoubi, Shalini Mahato, Alaa O Khadidos, and Sachi Nandan Mohanty. Machine learning and electroencephalogramsignal based diagnosis of depression. *Neuroscience letters*. 2023; 809: 137313. https://doi.org/10.1016/j.neulet.2023.137313
25. Xiao Gu, Fani Deligianni, Jinpei Han, Xiangyu Liu, *et al*. Beyond supervised learning for pervasive healthcare. *IEEE Reviews in Biomedical Engineering*. 2023; 17: 4262. https://doi.org/10.1109/RBME.2023.3296938
26. Bo Hjorth and AB Elema Schonander. Eeg analysis based on time domain properties, electro-encephalography and clinical neurophysiology. 1970. https://doi.org/10.1016/0013-4694(70)90143-4
27. Peter Welch. The use of fast fourier transform for the estimation of power spectra: A method based on time averaging over short, modified periodograms. *IEEE Transactions on audio and electroacoustics*. 2003; 15 (2): 70-73. https://doi.org/10.1109/TAU.1967.1161901