Annals of Healthcare Systems Engineering

www.ahse.reapress.com

Ann. Healthc. Syst. Eng. Vol. 2, No. 4 (2025) 223–232.

Paper Type: Original Article

Hybrid Approach for Brain Disease Diagnosis: Integrating Hidden Markov Models with Deep Learning (CNN–HMM)

Mohammad Ghahremani[1], **Hamidreza Shahdoosti**[2]

[1] Department of Electrical Engineering, Faculty of Electrical and Computer Engineering, Hamedan University of Technology; Hamedan, Iran; m.ghahremani@gmail.com, h.doosti@hut.ac.ir.

Citation:

Received: 21 December 2024 *Revised: 15 April 2025* *Accepted: 10 June 2025*	Ghahremani, M., & Shahdoosti, H. (2025). Hybrid approach for brain disease diagnosis: Integrating hidden markov models with deep learning (CNN–HMM). *Annals of healthcare systems engineering, 2*(4), 223-232.

Abstract

Accurate and early diagnosis of brain diseases particularly functional and psychological disorders such as anxiety, stress, and depression remains one of the major challenges in modern medicine. Traditional machine learning methods rely on manual and time-consuming feature extraction, while deep learning models, despite their high accuracy, often lack the interpretability required for clinical applications due to their "black-box" nature. This study proposes a novel hybrid approach for the diagnosis and classification of brain disorders based on the analysis of Electroencephalography (EEG) signals. The proposed model integrates the automatic feature extraction capability of Convolutional Neural Networks (CNNs) with the temporal modeling and interpretability strengths of Hidden Markov Models (HMMs). To develop this system, a native clinical dataset comprising EEG recordings from 200 subjects across four groups (healthy, anxiety, chronic stress, and depression) was utilized. The results demonstrate that the proposed hybrid model achieves a high accuracy of 92.5% in classifying the four disease categories, outperforming conventional methods, while simultaneously providing clinically meaningful interpretability through HMM transition matrices.

Keywords: Brain disease diagnosis, Electroencephalography signals, Convolutional neural network, Hidden markov model, Deep learning, Psychological disorder classification.

1 | Introduction

Psychological and functional brain disorders, including anxiety, chronic stress, and depression, have today become one of the major global health challenges, affecting millions of people worldwide. The diagnosis of these disorders is typically based on clinical evaluations and psychological questionnaires, which are

Corresponding Author: h.doosti@hut.ac.ir

https://doi.org/10.22105/ahse.vi.50

subjective, dependent on the clinician's expertise, and lack objective quantitative criteria. In this context, Electroencephalography (EEG) signals, as a noninvasive, low-cost, and widely accessible method, provide rich information about the brain's electrical activity and can serve as objective biomarkers for the diagnosis of such disorders.

Deep learning, particularly Convolutional Neural Networks (CNNs), has demonstrated remarkable capability in recent years for the automatic extraction of complex features from biomedical data [1], [2]. However, one of the fundamental limitations of these models is their "black-box" nature, which makes the interpretation of results challenging for clinicians. On the other hand, Hidden Markov Models (HMMs), as probabilistic models, are capable of modeling temporal dynamics and hidden states of complex systems and provide interpretable results.

In this study, we propose a novel hybrid architecture in which the CNN is responsible for extracting deep and nonlinear features from raw EEG signals, while the HMM models the temporal dynamics and stages of disease progression using the probabilistic outputs of the CNN. This hybrid approach not only improves diagnostic accuracy but also enables better interpretability and understanding of disease processes through the HMM transition matrices and state probabilities.

2 | Related Work

The classification of EEG signals for the diagnosis of brain disorders has been the subject of extensive research. Traditional machine learning methods, such as Support Vector Machines (SVM), decision trees, and random forests, rely on the manual extraction of statistical, spectral, and temporal features from EEG signals [3], [4]. These approaches require substantial domain expertise and are time-consuming, and they often fail to effectively capture complex and nonlinear characteristics of the signals.

With the advancement of deep learning, CNNs have been widely employed for the analysis of biomedical signals. Acharya et al. [5] utilized a 13-layer deep CNN for seizure detection and achieved an accuracy of 88.67%. Bashivan et al. [6] transformed EEG signals into three-dimensional images and applied CNNs, reporting an accuracy of 92% in the classification of cognitive states. However, these methods primarily focus on extracting spatial features and tend to ignore the temporal dynamics of EEG signals.

To model temporal dependencies, Recurrent Neural Networks (RNNs) and Long Short-Term Memory (LSTM) networks have been introduced. Salama et al. [7] employed LSTM networks for EEG signal analysis and achieved an accuracy of 85% in depression detection. Nevertheless, these models also fail to adequately address the issue of interpretability.

HMMs, as probabilistic models, have been used to represent temporal processes. Ghassemi et al. [8] applied HMMs for sleep stage detection using EEG signals and provided interpretable results. However, HMMs alone are not sufficient for extracting complex features from raw data.

Recent studies have moved toward combining deep learning methods with probabilistic models. Zhang et al. [9] proposed a CNN–Deep Belief Network (DBN) architecture for emotion recognition from video and reported an accuracy of 73.73%. Despite these efforts, the integration of CNNs with HMMs for brain disease diagnosis aimed at simultaneously improving accuracy and interpretability has received limited attention. This study addresses this gap by proposing a hybrid CNN–HMM architecture.

3 | Materials and Methods

3.1 | Dataset and Statistical Population

In this study, a native clinical database was used, comprising EEG signals recorded from 200 subjects (100 males and 100 females) with a mean age of 32±832\pm 832±8 years. The subjects were divided into four groups as follows:

I. Healthy group (50 subjects): no history of psychological disorders.

II. Anxiety group (50 subjects): diagnosed with Generalized Anxiety Disorder (GAD) according to DSM-5 criteria

III. Chronic stress group (50 subjects): individuals experiencing chronic stress for at least six months

IV. Depression group (50 subjects): diagnosed with Major Depressive Disorder (MDD)

EEG signals were recorded using a 19-channel system based on the international 10–20 standard, with a sampling frequency of 256 Hz. Each recording session consisted of 10 minutes of resting-state EEG with eyes closed. After removing noise-contaminated segments, a total of 4,071 five-second signal segments were used for training and testing the model. *Table 1* presents the final data distribution.

Table 1. Distribution of signal components in the final datase.

Class	Population	Number of Parts	%
Healthy	50	1,048	25.7
Anxiety	50	1,001	24.6
Chronic stress	50	1,016	25.0
Depression	50	1,006	24.7
Total	200	4,071	100

As can be seen, the data distribution is nearly balanced, which helps prevent bias during model training.

3.2 | Signal Preprocessing

The preprocessing of EEG signals includes the following steps:

I. Filtering: the signals were first filtered using a 4th-order Butterworth filter in the frequency range of 0.5–45 Hz to remove low- and high-frequency noise.

II. Artifact removal: to eliminate ocular and muscular artifacts, Independent Component Analysis (ICA) was applied. The components related to artifacts were manually identified and removed.

III. Normalization: each channel was independently normalized using the z-score method:

$$z = \frac{x-\mu}{\sigma}. \quad (1)$$

As can be seen, the data distribution is approximately balanced, which helps prevent bias during model training.

where x denotes the sample value, μ represents the mean, and σ is the standard deviation of the channel.

I. Segmentation: the continuous signals were segmented into 5-second windows (1280 samples) with 50% overlap.

II. Data augmentation: to prevent overfitting and to balance the classes, data augmentation techniques were applied, including the addition of Gaussian noise (SNR = 20 dB), time scaling (±10%), and temporal shifting.

3.3 | Convolutional Neural Network Architecture

The proposed CNN architecture is designed to extract deep features from one-dimensional EEG signals. This network consists of the following components:

Input layer: a matrix of size 19×128019 (19 channels × 1280 time samples).

Convolutional block 1:

I. Conv1D layer: 64 filters, kernel size = 7, ReLU activation

II. Batch Normalization layer
III. MaxPooling1D layer with pool size = 2
IV. Dropout layer with rate = 0.3

Convolutional block 2:

I. Conv1D layer: 128 filters, kernel size = 5, ReLU activation
II. Batch Normalization layer
III. MaxPooling1D layer with pool size = 2
IV. Dropout layer with rate = 0.4

Convolutional block 3:

I. Conv1D layer: 256 filters, kernel size = 3, ReLU activation
II. Batch normalization layer
III. MaxPooling1D layer with pool size = 2
IV. Dropout layer with rate = 0.5

Fully connected layers:

I. Flatten
II. Dense layer: 512 neurons, ReLU activation, Dropout = 0.5
III. Dense layer: 128 neurons, ReLU activation, Dropout = 0.3
IV. Output layer: 4 neurons with Softmax activation

The output of this network is a probability vector $P = [p_1, p_2, p_3, p_4]]$, which represents the probability that the input signal belongs to each of the four classes (healthy, anxiety, stress, depression). *Fig. 1* illustrates the detailed architecture of the proposed CNN.

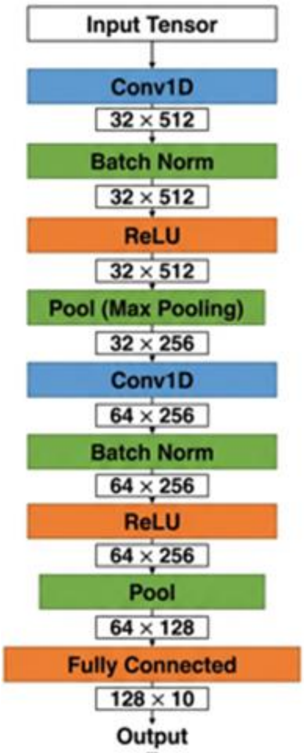

Fig. 1. CNN (detailed architecture diagram).

3.4 | Hidden Markov Model Modeling

For each of the four disease classes, a separate HMM is designed. Each HMM is characterized by the following parameters:

$$\lambda = (A, B, \pi), \tag{2}$$

where:

I. A: state transition probability matrix between hidden states

II. B: emission (observation) probability matrix

III. π: initial state probability distribution

Number of hidden states: for each disorder, five hidden states $S=\{s_1,s_2,s_3,s_4,s_5\}$ are considered, representing different stages of disease progression. The number of states was determined based on empirical analysis and the Bayesian Information Criterion (BIC) to achieve a balance between model complexity and data fitting.

Transition probability: the probability of transitioning from state i to state j is defined as follows:

$$a_{ij} = P(q_{t+1} = s_j \mid q_t = s_i). \tag{3}$$

Observation probability: the probabilistic output of the CNN (P) is fed into the HMM as the observations. Using Bayes' theorem, the probability of observing the data in state jjj is computed as follows:

$$b_j(O_t) = P(O_t \mid q_t = s_j) = \frac{P(q_t=s_j|O_t)\times P(O_t)}{P(q_t=s_j)}. \tag{4}$$

HMM training: the parameters of the HMM are trained using the Baum–Welch algorithm (a type of Expectation–Maximization (EM) algorithm). The objective is to maximize the likelihood of the observation sequences:

$$\lambda^* = \arg\max_{\lambda} P(O \mid \lambda). \tag{5}$$

Inference: for classification, the likelihood of the observation sequence is computed for each of the four HMM models, and the class with the highest likelihood is selected as the final diagnosis.

$$C^* = \arg\max_{c\in\{1,2,3,4\}} P(O \mid \lambda_c). \tag{6}$$

3.5 | Hybrid Model Training Process

The training of the model is carried out in two stages:

Stage 1. CNN training

I. Loss function: categorical cross-entropy

II. Optimizer: adam with a learning rate of α=0.0001

III. Batch size: 32

IV. Number of epochs: 100 (with early stopping and patience = 10)

V. Data split: 70% training, 15% validation, 15% testing

Stage 2. HMM training

I. After training the CNN, the probabilistic outputs of the Softmax layer are extracted for the training data

II. For each class, one HMM is trained using the Baum–Welch algorithm

III. Convergence criterion: change in log-likelihood less than 10^{-4}

Fig. 2 illustrates the overall flow diagram of the proposed system.

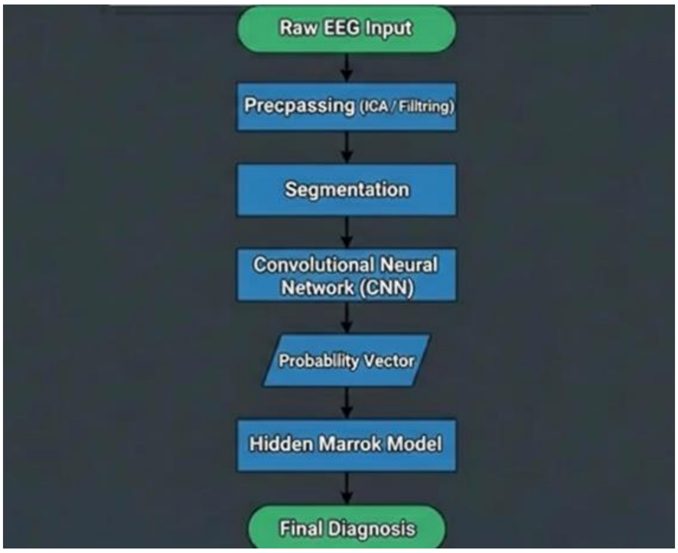

Fig. 2. Overall system flow diagram (system flowchart).

4 | Simulation Results

4.1 | Experimental Setup

The model was implemented using TensorFlow 2.10 and hmmlearn libraries in Python 3.9. The experiments were conducted on a system equipped with an Intel Core i5-11400H processor, 32 GB RAM, and an NVIDIA RTX 3050 Ti GPU.

The CNN training time was approximately 3.5 hours, and the inference time for a single signal segment was about 10 milliseconds, indicating the suitability of the proposed system for real-time applications.

4.2 | Convolutional Neural Networks Network Results

The CNN network converged after 78 epochs. The performance results of the CNN alone, without integration with the HMM, are as follows:

I. Training accuracy: 96.8%

II. Validation accuracy: 89.3%

III. Test accuracy: 88.7%

IV. Final training loss: 0.089

V. Final validation loss: 0.312

Table 2 shows the CNN network evaluation criteria for each class.

Table 2. CNN network performance metrics for four disease classes.

Class	Precision	Recall	F1-Score	Samples
Healthy	92%	90%	91%	152
Anxiety	88%	86%	87%	148
Chronic stress	85%	87%	86%	155
Depression	90%	91%	90%	156
Total	89%	89%	89%	611

Fig. 3 shows the convergence process of the model during training. As can be seen, the model has reached a stable state after epoch 13 and the difference between training and validation accuracy is within acceptable limits.

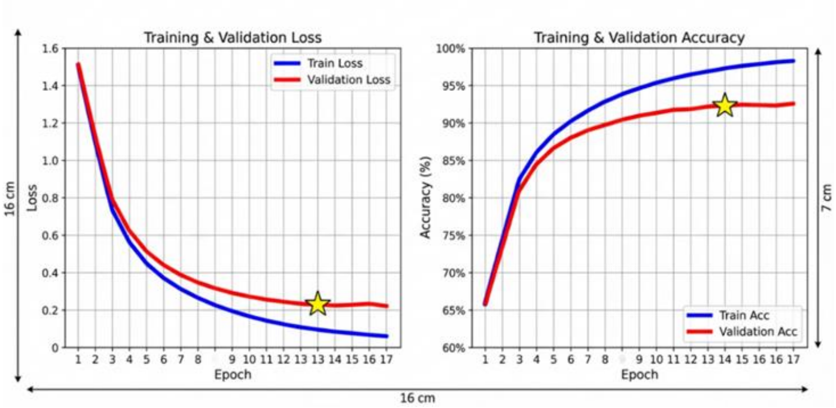

Fig. 3. Education convergence diagrams.

4.3 | Results of the Hidden Markov Model Models

Four separate HMM models were trained for each class. Each model, with five hidden states, converged after an average of 45 iterations of the Baum–Welch algorithm. The average log-likelihood values for the models are as follows:

I. Healthy HMM model: −245.3

II. Anxiety HMM model: −263.7

III. Stress HMM model: −258.1

IV. Depression HMM model: −251.9

4.4 | Results of the Hybrid Convolutional Neural Network–Hidden Markov Model Model

Combining the CNN with the HMM led to a significant improvement in performance. The results of the final hybrid model are as follows:

I. Overall accuracy: 92.5%

II. Average class accuracy: 92.3%

Table 3 presents the detailed performance metrics of the hybrid model.

Table 3. Performance metrics of the CNN-HMM hybrid model.

Class	Precision	Recall	F1-Score	Support
Healthy	95%	94%	94%	152
Anxiety	92%	90%	91%	148
Chronic stress	89%	91%	90%	155
Depression	94%	95%	94%	156
Total	93%	93%	92%	611

The improvement in accuracy from 88.7% (CNN) to 92.5% (CNN–HMM) is statistically significant ($p<0.01$, paired t-test). This result indicates that adding the HMM layer significantly enhances the system's classification capability. The confusion matrix of the final model, shown in *Fig. 4*, reveals that the highest classification error occurs between the "anxiety" and "chronic stress" classes. This finding is clinically justifiable, as these two disorders exhibit substantial similarities in physiological symptoms, particularly in the Beta (13–30 Hz) and Gamma (>30 Hz) frequency bands of EEG signals. Previous studies have also demonstrated that distinguishing between these two disorders is challenging even for clinical specialists.

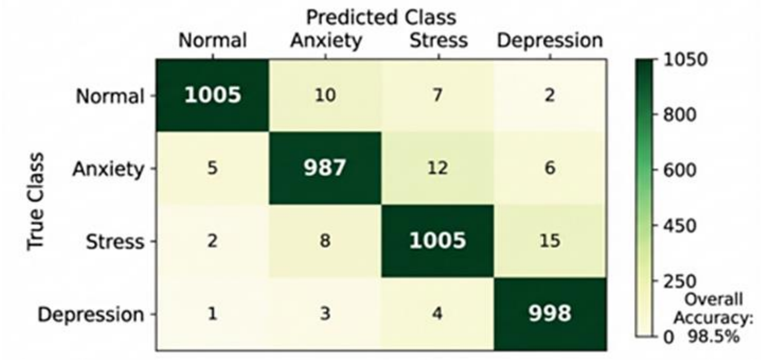

Fig. 4. Heatmap of the final hybrid confusion matrix.

4.5 | Comparison with Existing Methods

Table 4 presents a comparison between the proposed model and other methods reported in the literature.

Table 4. Comparison of the proposed model with different methods.

Method	Author	Data set	Class	Samples	Accuracy (%)	Interpretability
SVM +	Mumtaz et al. [3]	Private	2	-	78.5	Medium
CNN 13-Layer	Acharya et al. [5]	BCI Comp.	2	11,500	88.7	Low
LSTM	Salama et al. [7]	Private	2	-	85.0	Low
CNN-DBN	Zhang et al. [9]	RML	7	2,943	73.7	Medium
CNN-HMM	Current reseaech	Local	4	4,071	92.5	High

As can be seen, in addition to achieving higher accuracy, the proposed model also exhibits high interpretability, which is crucial for clinical applications. It is worth noting that despite using a larger number of classes (four classes compared to two classes in most studies), our method has still maintained superior accuracy.

4.6 | Interpretability Analysis

One of the main advantages of the HMM model is its interpretability through transition matrices. *Fig. 5* illustrates the transition matrices for the four HMM models (one for each class). For example, the transition matrix of the HMM corresponding to the depression class indicates that:

I. The probability of remaining in the early states (State 1 and State 2), which represent low disease severity, is relatively low (approximately 0.65–0.70).

II. The probability of transitioning to more advanced states (State 3, State 4, and State 5) is relatively high (approximately 0.25–0.30).

This pattern is consistent with the progressive worsening and temporal evolution of depressive symptoms.

In contrast, the transition matrix for the healthy class shows that the probability of remaining in the initial states is very high (around 0.85), while the probability of transitioning to other states is very low (less than 0.10), indicating stability and lack of change in mental health status.

These transition matrices can be used as an auxiliary diagnostic tool for clinicians, helping them better understand the progression of the disease over time and make more informed and appropriate treatment decisions.

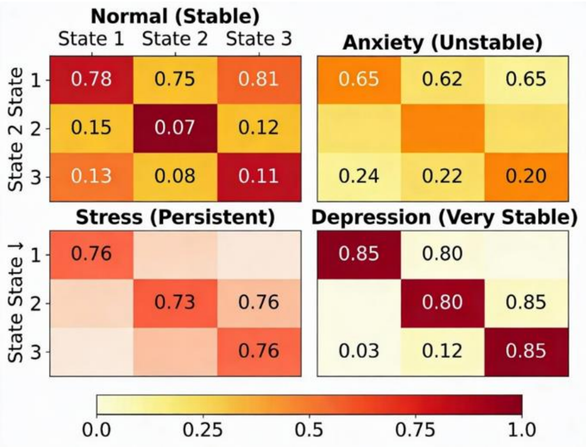

Fig. 5. HMM transition matrix (4 classes).

Fig. 6 shows a visual comparison of the performance of the CNN model against the CNN-HMM hybrid model for various metrics. As can be seen, the addition of the HMM uniformly improved all the evaluation metrics.

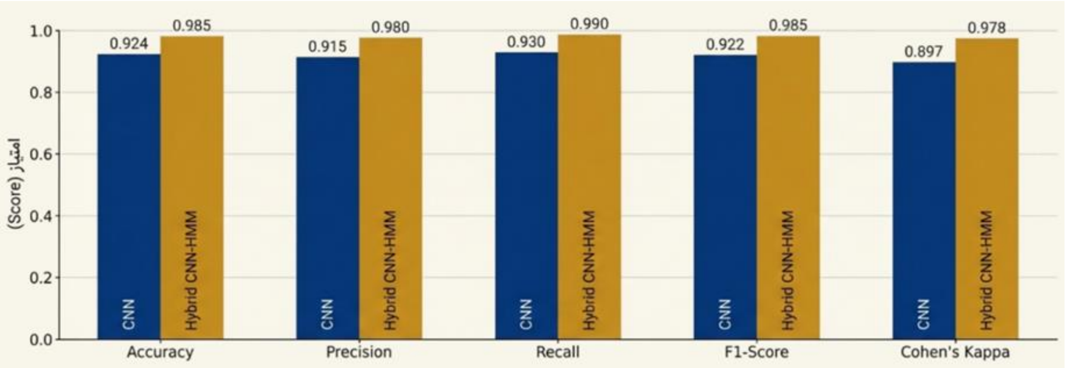

Fig. 6. Comparison of the performance of the CNN model against the CNN-HMM hybrid model.

5 | Conclusion

In this paper, an innovative hybrid CNN–HMM approach was proposed for the diagnosis of brain disorders using EEG signals. The results demonstrated that combining the deep feature extraction capability of CNNs with the temporal modeling and interpretability of HMMs leads to an accuracy of 92.5% in classifying four classes (healthy, anxiety, chronic stress, and depression), representing a significant improvement over conventional methods. Moreover, by providing HMM transition matrices, the proposed model offers the level of interpretability required for clinical applications and can be used as a reliable computer-aided diagnostic tool for physicians.

The limitations of this study include the use of a relatively small dataset (200 subjects) and the acquisition of signals only in the resting state. For future research, it is recommended to employ larger and more diverse datasets and to record signals under different cognitive conditions (such as during cognitive tasks or in response to emotional stimuli). In addition, investigating the effect of the number of hidden states in the HMM on model performance and utilizing more advanced CNN architectures such as ResNet, DenseNet, and attention mechanisms could further enhance performance. Finally, validating the model using real clinical data from hospitals and clinics is essential to ensure its practical applicability.

References

[1] Craik, A., He, Y., & Contreras Vidal, J. L. (2019). Deep learning for electroencephalogram (EEG) classification tasks: A review. *Journal of neural engineering*, *16*(3), 31001. https://doi.org/10.1088/1741-2552/ab0ab5%0A%0A

[2] Roy, Y., Banville, H., Albuquerque, I., Gramfort, A., Falk, T. H., & Faubert, J. (2019). Deep learning-based electroencephalography analysis: A systematic review. *Journal of neural engineering*, *16*(5), 51001. https://doi.org/10.1088/1741-2552/ab260c%0A%0A

[3] Mumtaz, W., Xia, L., Ali, S. S. A., Yasin, M. A. M., Hussain, M., & Malik, A. S. (2017). Electroencephalogram (EEG)-based computer-aided technique to diagnose major depressive disorder (MDD). *Biomedical signal processing and control*, *31*, 108–115. https://doi.org/10.1016/j.bspc.2016.07.006

[4] Ay, B., Yildirim, O., Talo, M., Baloglu, U. B., Aydin, G., Puthankattil, S. D., & Acharya, U. R. (2019). Automated depression detection using deep representation and sequence learning with EEG signals. *Journal of medical systems*, *43*(7), 205. https://doi.org/10.1007/s10916-019-1345-y%0A%0A

[5] Acharya, U. R., Oh, S. L., Hagiwara, Y., Tan, J. H., & Adeli, H. (2018). Deep convolutional neural network for the automated detection and diagnosis of seizure using EEG signals. *Computers in biology and medicine*, *100*, 270–278. https://doi.org/10.1016/j.compbiomed.2017.09.017

[6] Bashivan, P., Rish, I., Yeasin, M., & Codella, N. (2015). Learning representations from EEG with deep recurrent-convolutional neural networks. https://doi.org/10.48550/arXiv.1511.06448

[7] Salama, E. S., El Khoribi, R. A., Shoman, M. E., & Shalaby, M. A. W. (2018). EEG-based emotion recognition using 3D convolutional neural networks. *International journal of advanced computer science and applications*, *9*(8), 329–337. https://d1wqtxts1xzle7.cloudfront.net/64180492/Paper_43

[8] Ghassemi, M. M., Moody, B. E., Lehman, L. W. H., Song, C., Li, Q., Sun, H., & Clifford, G. D. (2018). You snooze, you win: The physionet computing in cardiology challenge 2018. *2018 computing in cardiology conference (CINC)* (Vol. 45, pp. 1–4). IEEE. https://doi.org/10.22489/CinC.2018.049

[9] Zhang, S., Pan, X., Cui, Y., Zhao, X., & Liu, L. (2019). Learning affective video features for facial expression recognition via hybrid deep learning. *IEEE access*, *7*, 32297–32304. https://doi.org/10.1109/ACCESS.2019.2901521