Jurnal Kejuruteraan 38(7) 2026: 3639-3657
https://doi.org/10.17576/jkukm-2026-38(7)-36

Multiclass EEG Spectrogram Classification Using Spatial Channel Grouping and CNN for ICU Monitoring

Ahmed J. Abougarair[a*], Hamzah M. Marhoon[b], Hari Maghfiroh[c], Ahmed F. Zrigan[d] & Moad Shlebk[e]

[a,e]*Dept. of Electrical and electronics Engineering , University of Tripoli , Tripoli, Libya*

[b]*Dept .of Automation and AI, Al-Nahrain University, Baghdad, Iraq*

[c]*Dept. of Electrical l Engineering ,Universitas Sebelas Maret, Surakarta, Indonesiac*

[d]*Dept. of Electrical and Electronics Technologies, Higher Institute of Science and Technologies, Tripoli ,Tripoli, Libyab*

[*]*Corresponding author: a.abougarair@uot.edu.ly*

Received 18 February 2026, Received in revised form 6 July 2026
Accepted 6 August 2026, Available online 30 October 2026

ABSTRACT

The continuous recording of the Electroencephalogram (EEG) by Intensive Care Units (ICU) clinicians results in massive data volumes that are challenging to interpret and can take significant time, depending on the experience of the individual interpreting the data. In this paper, we propose a novel deep learning method of automatic EEG spectrogram classification into different classes for ICU brain monitoring. The classification process involves assigning an EEG pattern into one of the six clinically relevant classes, namely seizures, Ateralized Periodic Discharges (LPD), Generalized Periodic Discharges (GPD), Lateralized Rhythmic Delta Activity (LRDA), eneralized Rhythmic Delta Activity (GRDA), and Other. The primary innovation in the current research study includes a clinically motivated spatial electrode grouping method, which groups the EEG channels into four regions of the brain: left lateral, right lateral, left parasagittal, and right parasagittal. In this way, the information about the hemispheres is preserved, and the distinction between lateralized and generalized anomalies can be improved. A convolutional neural network (CNN) was trained on augmented spectrogram data with class weights, which helps in handling the high levels of class imbalance present in the HMS Harmful Brain Activity Classification Dataset. Our model resulted in a classification accuracy of 90.8% and a macro-F1 score of 0.86 with a high seizure detection F1 score of 0.86. Furthermore, a Grad-CAM visualization tool was used to provide a visualization for the predictions, while the Monte Carlo dropout technique was used to measure prediction uncertainty. In addition, our model had an Expected Calibration Error (ECE) of just 0.04, ensuring accurate confidence measures.

Keywords: EEG; CNN; seizure detection; spatial channel grouping; Grad-CAM

INTRODUCTION

Electroencephalography (EEG) is still an important clinical tool for monitoring brain function, especially in intensive care units (ICUs) where finding abnormal brain activity early can have a big impact on patient outcomes. As more and more critical care facilities start using continuous EEG (cEEG) monitoring, there is now an unprecedented amount of neurophysiological data that needs to be interpreted by experts (Craik et al. 2019; Roy et al. 2019).However, going through cEEG recordings by hand takes a lot of time, is very different from one rater to the next, and can make clinicians tired, all of which raise the risk of missing important clinical events like seizures or periodic discharges (Alkhachroum et al. 2022; Khemakhem et al. 2024).

The problems with manual interpretation have led to a lot of research into automated EEG analysis systems. Traditional methods that use handcrafted features and standard machine learning classifiers like support vector machines or random forests have had some success in controlled binary classification tasks, especially seizure

detection (Amin et al. 2019; Khemakhem et al. 2024). But these methods don't work well with a wide range of patients and recording conditions because EEG signals are not stationary and it's hard to design features that capture the full complexity of pathological patterns (Chaddad et al. 2023; Jiang et al. 2019). More importantly, they have trouble with the subtle multiclass discrimination needed in clinical practice, like telling the difference between lateralized and generalised periodic discharges, which has huge diagnostic and therapeutic implications (Bai et al. 2025).

Recent advances in deep learning, especially convolutional neural networks (CNNs), have changed biomedical signal processing by allowing end-to-end learning of discriminative features directly from data that has been minimally processed (Craik et al. 2019; Roy et al. 2019). Time-frequency spectrograms have become very useful as input representations because they show both how brain activity changes over time and its spectral properties. This makes pathological patterns stand out as visual features that can be analyzed using CNNs (Chaddad et al. 2023; Jiang et al. 2019). Many studies have successfully used spectrogram-based CNNs to classify EEG data for a variety of tasks, such as sleep staging (Huang et al. 2024), recognizing emotions (Ali et al. 2025), and decoding motor imagery (Amin et al. 2019). Most of this work, on the other hand, has been focused on binary classification or specific application domains. There hasn't been much focus on the full range of clinically important patterns seen in ICU monitoring, such as seizures, lateralized periodic discharges (LPD), generalized periodic discharges (GPD), lateralized rhythmic delta activity (LRDA), generalized rhythmic delta activity (GRDA), and background activity (Bai et al. 2025).

One major problem with current spectrogram-based methods is that they treat multichannel EEG as a set of separate images, ignoring the important spatial relationships between electrodes that are fundamental to clinical EEG interpretation (Tveitstøl et al. 2024). For instance, the practice of using standard montages by neurologists, like the "double banana" bipolar montage, enables them to identify if the abnormalities are localized to just one side or if they affect both sides. Spatial localization is key for proper differential diagnosis. For example, it makes sense to diagnose LPD rather than GPD (Tveitstøl et al. 2024). Similarly, differentiating between lateralized and generalized rhythmic abnormalities, such as LRDA and GRDA, provides important diagnostic and prognostic information. Therefore, the reliable discrimination of lateralized and generalized EEG patterns represents a key clinical objective for automated EEG analysis systems. Recently, a few methods have begun addressing this challenge through region-based pooling techniques (Tveitstøl et al. 2024) and multichannel fusion networks (Sun et al. 2025). However, to the best of our knowledge, no previous study has explicitly incorporated a clinically inspired montage structure into a CNN framework for multiclass ICU EEG pattern recognition. The proposed spatial channel grouping strategy was developed to address this gap by preserving anatomically meaningful spatial information and improving the discrimination between lateralized and generalized EEG abnormalities.

AI-based systems also need to be transparent and have reliable methods of quantifying uncertainty if they are to be utilized successfully in clinical settings, in addition to making accurate predictions. These are factors that have often been overlooked in previous EEG classification research (Abdar et al. 2021; Shen et al. 2024). Clinicians need to know why models make predictions and how much they trust automated outputs, especially in high-stakes situations where false negatives (like missed seizures) can have serious effects. Gradient-weighted Class Activation Mapping (Grad-CAM) and other explainable AI (XAI) methods can show visually which parts of the input are most important for the model's decisions (Shen et al. 2024). At the same time, uncertainty estimation methods like Monte Carlo (MC) Dropout can find cases that are unclear and need expert review. This builds trust and makes it easier to safely integrate into clinical practice (Abdar et al. 2021).

This paper fills in these related gaps by presenting a complete automated system for multiclass EEG spectrogram classification that makes the following new contributions:

1. A clinically-informed spatial channel grouping pipeline: This combines EEG electrodes into four neuroanatomical regions that are similar to the standard "double banana" bipolar montage. This preprocessing method clearly encodes spatial information, which lets the model tell the difference between lateralized (LPD, LRDA) and generalized (GPD, GRDA) brain activity patterns. This is something that generic CNN architectures can't do.
2. A framework that is easy to understand and takes uncertainty into account: combining Grad-CAM visual explanations with MC-Dropout-based confidence estimation. As far as we know, this is the first time that both techniques have been used together to classify multiclass EEG spectrograms in the ICU setting.
3. Full empirical validation: on the public HMS dataset, showing better performance (macro F1-score: 0.86, seizure detection F1: 0.90) than strong baselines, along with a thorough examination of model calibration, robustness, and clinical plausibility.

Even though remarkable strides have been made by the EEG classification methods employing deep learning, there are still some notable shortcomings in the field. Most of the previous works either concentrate on binary classification or deal with EEG signals from individual channels, thus ignoring clinically relevant spatial correlations between brain regions which neurologists employ while reading EEGs. Also, there is a lack of research related to applying clinically motivated spatial representation, explainability, uncertainty assessment, and calibration evaluation within one methodology for multiclass classification of ICU EEGs. This paper attempts to bridge these knowledge gaps through the application of the proposed clinically motivated CNN approach using EEG channels grouped according to clinical standard EEG montages to improve differentiation between lateralized and generalized EEG patterns. Moreover, explainability, uncertainty estimation, and calibration analysis using such techniques as Grad-CAM, Monte Carlo Dropout, and histogram visualization were incorporated into the methodology.

The rest of this paper is set up like this. Section 2 gives important clinical information about EEG patterns and montages. Section 3 looks at other work that is related to ours and groups it by theme to put our contributions in context. Our methodology is explained in detail in Section 4. It includes a description of the dataset, the construction of the spatial channel, the architecture of the CNN, and the training configuration. Section 5 shows the results of the experiments and talks about how well they worked, how easy they were to understand, and how to estimate uncertainty. Section 6 wraps up the paper and talks about what needs to be done in the future.

CLINICAL EEG BACKGROUND

When doctors read EEGs in hospitals, especially in ICUs, they look for brain activity patterns that show nervous system problems. The most important patterns are the following:

1. Seizure: Characterized by excessive, synchronous, and rhythmic neuronal activity.
2. Periodic Discharges (LPD/GPD): Repeated patterns of spikes or sharp waves that can happen in one hemisphere (Lateralized Periodic Discharges, LPD) or both hemispheres (Generalized Periodic Discharges, GPD).
3. Rhythmic Delta Activity (LRDA/GRDA): This is an activity that happens in slow waves and can be focused or spread out. It usually means that the brain isn't working right.
4. Other: This designation means normal background activity and patterns that don't fit into any of the other groups.

A common way for neurologists to see how brain activity is linked in space is to use the "double banana" bipolar montage. This clinical approach led us to consider categorizing spatial channels. It digitally recreates this view by putting electrodes in four distinct parts of the brain: left lateral, right lateral, left parasagittal, and right parasagittal. This protects important information about space.

RELATED WORK

The review below groups the current research on automated EEG classification into central themes and highlights important methods and results. First, we look at traditional methods. Then, we move on to deep learning-based methods, focusing on spectrogram-based CNNs, the use of spatial information, and the growing importance of exploitability and uncertainty estimation in medical AI (Cao et al. 2024, Castiglioni et al. 2025).

Alternative approaches, such as visibility graph-based feature extraction, have also shown promise in capturing nonlinear dynamics of EEG signals (Belhadi et al. 2025).

EEG ANALYSIS: FROM HANDCRAFTED FEATURES TO DEEP LEARNING

Early efforts in automated EEG interpretation focused on extracting handcrafted features from time or frequency domains, subsequently fed into conventional machine learning classifiers. People often used temporal metrics like mean amplitude, variance, and line length, as well as frequency-domain features from Fourier or wavelet transforms, with classifiers like support vector machines (SVMs), k-nearest neighbors (k-NN), and random forests (Craik et al. 2019; Roy et al. 2019). These techniques were effective in controlled binary tasks, particularly in detecting seizures (Amin et al. 2019; Khemakhem et al. 2024). However, their efficacy significantly decreased in practical applications due to the intrinsic non-stationarity of EEG signals, vulnerability to noise and artefacts, and the challenges in developing features that generalize across varied patient populations and recording conditions (Chaddad et al. 2023; Jiang et al. 2019). Handcrafted feature methods also struggled with the nuanced distinctions between various classes essential in clinical contexts, such as differentiating between morphologically similar patterns like lateralized and generalized periodic discharges (Bai et al. 2025).

DL has completely changed biomedical signal processing by allowing hierarchical representations to be found automatically from raw or lightly processed data (Craik et al. 2019). Comprehensive studies by Craik et al. 2019 and Roy et al. 2019 have documented the successful application of various deep architectures specifically convolutional neural networks (CNNs), recurrent neural networks (RNNs), and autoencoders in multiple EEG contexts. CNNs are great at looking at time frequency representations because they can learn local patterns and features that stay the same when you change the translation (Craik et al. 2019).

SPECTROGRAM-BASED CNNS FOR EEG CLASSIFICATION

Time-frequency spectrograms have become the favored input representation for CNN-based EEG analysis as they concurrently capture temporal evolution and spectral content while presenting pathological patterns as unique visual features (Chaddad et al. 2023; Jiang et al. 2019). Researchers have effectively utilized spectrogram-based CNNs in various applications, including sleep stage classification (Huang et al. 2024), emotion recognition (Ali et al. 2025), and motor imagery decoding (Amin et al. 2019). For example (Li et al. 2017) created a hybrid CNN+LSTM network that uses common spatial patterns for multiclass motor imagery and is the best in its class (Sudhakar Jebaraj and Elango 2025).

Even with these improvements most studies that use spectrograms have only looked at certain application areas or binary classification tasks. They haven't paid much attention to the full range of patterns that are important for ICU monitoring (Bai et al. 2025). The six-class problem examined in this study comprising seizures, LPD, GPD, LRDA, GRDA and background activity represents the complete diagnostic spectrum observed in clinical practice and is inadequately represented in the literature (Bai et al. 2025).

INCORPORATING SPATIAL INFORMATION IN EEG MODELS

A critical limitation of conventional spectrogram-based approaches is their treatment of multichannel EEG as independent two-dimensional images, thereby discarding the spatial relationships between electrodes that are fundamental to clinical interpretation (Tveitstøl et al. 2024). Neurologists routinely employ standardized montages, most notably the (double banana) bipolar configuration to assess whether abnormalities are focal (confined to one hemisphere) or generalized (affecting both hemispheres). This spatial information is essential for accurate differential diagnosis; for instance, distinguishing LPD from GPD directly informs the localization of cerebral dysfunction and guides treatment decisions (Tveitstøl et al. 2024).

Recent research has begun to acknowledge the importance of spatial context. Tveitstøl et al. (2024) proposed region-based pooling to aggregate information from EEG systems with varying channel counts, demonstrating the utility of incorporating montage information. Sun et al. (2025) developed multi-channel fusion networks for epileptic signal classification, showing performance improvements over single-channel approaches. However, these methods do not explicitly leverage a specific clinical montage structure. To date, no study has systematically implemented a grouping strategy based on the standard "double banana" configuration specifically designed to enhance discrimination between lateralized and generalized patterns in a multiclass ICU setting. This gap motivates the central methodological innovation of the present work (Jamil et al. 2024; Sabut et al. 2024).

EXPLAINABILITY AND UNCERTAINTY IN MEDICAL AI SYSTEMS

For AI-based decision support systems to be used in medicine, they need to be not only very accurate but also clear and reliable (Shen et al. 2024). Clinicians need to know why model predictions are made and how much faith to put in automated outputs and, especially in high-stakes situations where mistakes can have serious consequences. These needs have led to more interest in explainable AI (XAI) and uncertainty quantification in medical applications (Abdar et al. 2021).

Gradient-weighted Class Activation Mapping (Grad-CAM) has become a well-known XAI method that makes heatmaps that show which parts of the input are most important for model decisions (Shen et al. 2024). In medical imaging, Grad-CAM has been used to make sure that models focus on features that make sense in a clinical setting (Shen et al. 2024). Monte Carlo (MC) Dropout is a computationally efficient way to estimate uncertainty. It does this by making multiple stochastic forward passes with dropout enabled during inference, which creates a distribution of predictions whose variance shows how uncertain the model is (Abdar et al. 2021). This method can find samples that are outside of the distribution or cases that are unclear and need to be looked at by an expert (Abdar et al. 2021).

Even though they are important, not many EEG classification studies have put together explainability and

uncertainty estimation into one framework (Shen et al. 2024). These techniques work especially well together for monitoring in the ICU, where spotting uncertain predictions can lead to human review and stop overconfident misclassification of important events.

RESEARCH GAP AND CONTRIBUTION SYNTHESIS

The preceding review identifies three interrelated deficiencies in the current literature: (1) a limited emphasis on comprehensive multiclass classification that includes the entire range of clinically significant ICU EEG patterns; (2) an inadequate integration of spatial electrode data into spectrogram-based CNNs, particularly the intentional application of standard clinical montages to distinguish between lateralized and generalized abnormalities; and (3) a lack of incorporation of explainability and uncertainty estimation methodologies crucial for clinical trust and safety.

This paper directly addresses these gaps through an integrated framework that: (i) introduces a novel spatial channel grouping pipeline explicitly mimicking the "double banana" bipolar montage; (ii) employs a class-weighted CNN architecture optimized for spectrogram analysis under severe class imbalance; and (iii) provides both Grad-CAM-based visual explanations and MC-Dropout-based uncertainty estimates for every prediction. By thoroughly testing on the public HMS dataset, we show that this method works better while also making it easier to understand and more reliable.

METHODOLOGY

SYSTEM OVERVIEW AND DATASET

The proposed automated classification system follows a modular pipeline designed to transform raw EEG data into clinically interpretable predictions. An overview of the entire workflow is illustrated in Figure 1. The system comprises four core modules: (1) Data Preprocessing and Spatial Structuring, (2) Spectrogram Representation, (3) Convolutional Neural Network (CNN) for feature extraction and classification, and (4) Model Evaluation with Explainability and Uncertainty Estimation.

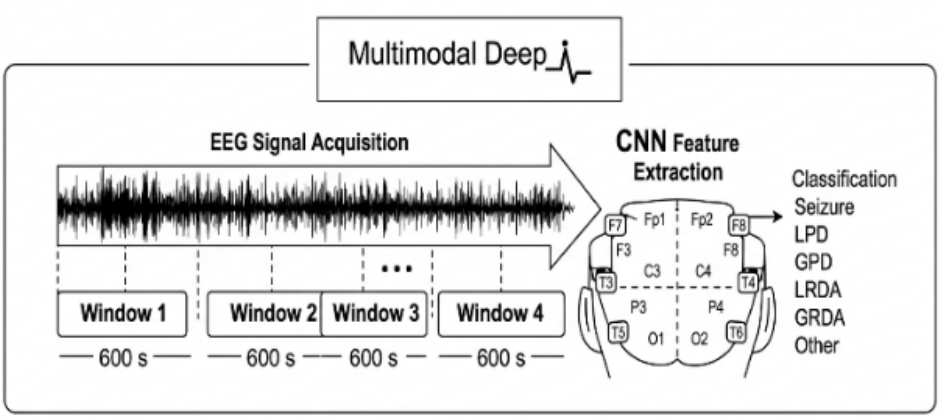


FIGURE 1. System overview/workflow diagram

The modular architecture shown in Figure 1 outlines the systematic workflow from raw EEG data to clinical predictions. Each module serves a specific purpose: preprocessing transforms the data, the CNN extracts features, and the evaluation module provides interpretable outputs.

For our study, we have used the HMS (Harmful Brain Activity Classification) dataset, which is publicly available benchmark data released via HMS Harmful Brain Activity Classification Challenge (Kaggle, 2024; Hirsch et al., 2021). This dataset consists of EEG records and their spectrogram representation based on the continuous electroencephalographic monitoring of ICU patients. It contains six distinct classes of clinically significant brain activity patterns annotated by experts, including seizure, lateralized periodic discharges (LPD), generalized periodic discharges (GPD), lateralized rhythmic delta activity (LRDA), generalized rhythmic delta activity (GRDA), and other/background. One of the difficulties related to this dataset is the problem of class imbalance in the data, as illustrated in Table 1.

TABLE 1. Class distribution in the training set

Class	Seizure	LPD	GPD	LRDA	GRDA	Other
Count	~5%	~8%	~12%	~10%	~15%	~50%

PREPROCESSING AND SPATIAL CHANNEL CONSTRUCTION

The raw spectrogram data, stored in Parquet format, required structured preprocessing to create model-ready inputs. Each data row represented approximately 2 seconds of EEG. Extracted non- overlapping windows of 300 rows (600 seconds) per sample a duration clinically relevant for identifying evolving rhythmic and periodic patterns (Craik et al. 2019; Bai et al. 2025). The novel aspect of the current preprocessing pipeline design involves the generation of spatial channels based on the structure of the standard anatomical double-banana montage. In contrast to traditional techniques, which consider the data from the individual electrodes, the newly introduced approach creates four neuroanatomically justified regions: Left Lateral (LL), Right Lateral (RL), Left Parasagittal (LP), and Right Parasagittal (RP). In particular, electrodes F7, T7, and P7 correspond to the LL region, F8, T8, and P8 correspond to RL, F3, C3, and P3 form the LP region, and F4, C4, and P4 constitute the RP region. It should be emphasized that the introduction of the above-mentioned spatial channels cannot be regarded as the introduction of bipolar derivations, as the voltage difference is not calculated between neighboring electrodes. The spectrogram power value for the electrodes comprising one region is computed using averaging. Hence, the four-channel tensor generated by this algorithm incorporates the average regional spectral activity: (Chaddad et al. 2023; Tveitstøl et al. 2024):

$$X \in R^{4\times300\times100} \tag{1}$$

This channel grouping and averaging approach preserves both hemispheric lateralization, enabling the model to distinguish between focal (e.g., LPD) and generalized (e.g., GPD) abnormalities more effectively. Each spectrogram window underwent z-score normalization per channel to mitigate inter- patient variability. Standard z-score normalization is typically defined by the following formula:

$$z = \frac{x-\mu}{\sigma} \tag{2}$$

x (The Raw Input): represents the individual power or magnitude value at a specific frequency-time point within a single spectrogram channel.

μ: This is the average value of the specific spectrogram channel.

σ (The Standard Deviation): This represents the variability of the signal in that channel. Dividing by σ scales the data so that the resulting values have a unit standard deviation. Missing values were handled via column-wise mean imputation.

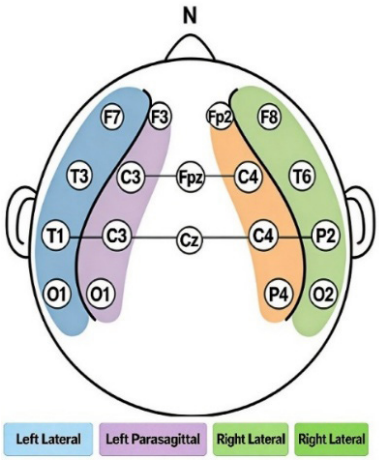


FIGURE 2. EEG Regional Channel Grouping Inspired by the Double-Banana Montage

Figure 2 illustrates our innovative spatial channel grouping, which clusters electrodes into four anatomical regions. This approach mimics clinical bipolar montage analysis, enabling the model to capture hemispheric asymmetries crucial for distinguishing lateralized from generalized abnormalities.

DATA SPLITTING AND AUGMENTATION

To ensure robust evaluation and prevent data leakage a patient-wise stratified split: 70% for training, 15% for validation, and 15% for testing. This guarantees that segments from the same patient appear in only one subset. To address class imbalance and improve generalization, applied spectrogram-domain data augmentation during training (Amin et al. 2019; Tveitstøl et al. 2024):

1. Time Masking: Randomly masking contiguous time steps (max 10% of window).
2. Frequency Masking: Randomly masking contiguous frequency bands (max 10% of spectrum).

3. Time Shifting: Random circular shifts along the time axis.
4. Gaussian Noise: Adding minor random noise (σ=0.01×data std).

These carefully calibrated augmentations simulate authentic recording variability including electrode displacement, muscular artifacts, and equipment noise while deliberately preserving clinically discriminative spectral-temporal signatures essential for accurate pathological pattern recognition.

CONVOLUTIONAL NEURAL NETWORK ARCHITECTURE

The core classification model is a custom- designed CNN optimized for time-frequency representations. As shown in Figure 3.

The CNN architecture depicted in Figure 3 consists of three convolutional blocks designed to hierarchically extract features from spectrograms. The increasing filter counts (32→64→128) allow the network to learn increasingly complex temporal spectral patterns, while global average pooling reduces parameter count and prevents overfitting. The architecture consists of three convolutional blocks followed by fully connected layers (Craik et al. 2019; Tveitstøl et al. 2024):

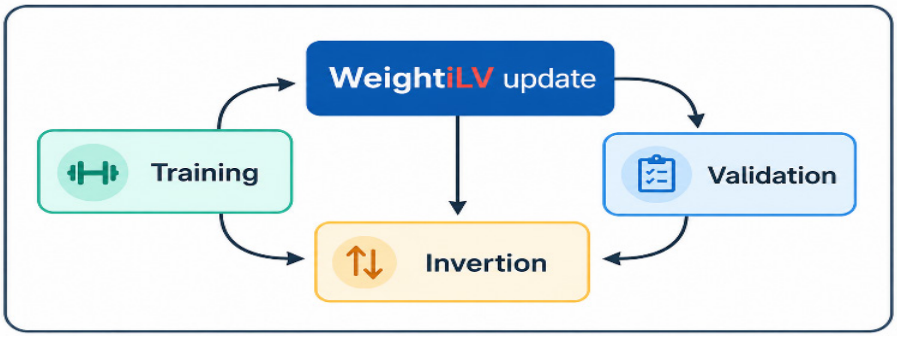


FIGURE 3. CNN Architecture Diagram

Block 1: Conv2D (32 filters, 3×3) → BatchNorm → ReLU → MaxPool (2×2)
Block 2: Conv2D (64 filters, 3×3) → BatchNorm → ReLU → MaxPool (2×2)
Block 3: Conv2D (128 filters, 3×3) → BatchNorm → ReLU → Global Average Pooling

The extracted features are fed into two fully connected layers (1024 and 256 units with ReLU activation and 50% dropout) before the final softmax output layer with 6 units corresponding to our target classes.

TRAINING CONFIGURATION

Trained the model using the Adam optimizer with an initial learning rate of 0.001 and cosine annealing scheduling. To directly address class imbalance, employed a weighted cross-entropy loss:

$$L = -\sum_{i=1}^{C} W_i y_i \log(\hat{y}_i) \quad (3)$$

W_i (Class Weight): This is the key parameter for addressing class imbalance.
y_i (Ground Truth): This is the actual label of the data point, typically represented as a one-hot encoded vector.
$\hat{y}_i$ (Predicted Probability): This is the model's predicted probability that the input belongs to the class i.
C (Number of Classes): The total number of categories the model is trained to recognize (Chaddad et al. 2023; Tveitstøl et al. 2024).

Where weights W_i are inversely proportional to class frequencies in the training set. Regularization included L2 weight decay and early stopping based on validation macro-F1 score (patience = 10 epochs). Training used mixed precision (FP16) for efficiency with batch sizes of 32. Hyperparameters are summarized in Table 2.

TABLE 2. Training Hyperparameters

Parameter	Value
Optimizer	Adam
Initial Learning Rate	0.001
Batch Size	32
Weight Decay	1e-4
Dropout Rate	0.5
Maximum Epochs	50
Early Stopping Patience	10

EVALUATION, EXPLAINABILITY, AND UNCERTAINTY ESTIMATION

Model performance was assessed using comprehensive metrics suitable for imbalanced medical data (Chaddad et al. 2023; Shen et al. 2024):

1. The macro-averaged F1-score was selected as our primary performance indicator because it computes the unweighted arithmetic mean of per-class F1 values, ensuring that minority classes clinically critical despite their rarity receive equal consideration in overall performance assessment.
2. Confusion matrix visualization facilitated detailed examination of inter-class confusion patterns, particularly between clinically similar categories like LRDA and GRDA.
3. Beyond discriminative performance, we evaluated predictive calibration using Expected Calibration Error (ECE), which quantifies the discrepancy between predicted confidence scores and empirical accuracy.

For clinical interpretability, Gradient weighted Class Activation Mapping (Grad-CAM) was implemented to generate visual explanations. This technique produces heatmaps highlighting the time-frequency regions most influential for each prediction. This allows clinicians to verify whether the model focuses on physiologically plausible features (Chaddad et al. 2023; Shen et al. 2024).

To quantify prediction reliability, Monte Carlo Dropout (MC-Dropout) was employed for uncertainty estimation. By performing 20 stochastic forward passes with dropout enabled during inference, a distribution of predictions is obtained. The variance of these predictions serves as a measure of model uncertainty. High variance indicates ambiguous cases that should be flagged for expert review.

RESULTS AND DISCUSSION

The system modeling and simulation were conducted using the MATLAB environment. To ensure the reliability of the computational framework, specific segments of the core programming and algorithmic structures were adapted and implemented based on established methodologies cited in references (Attawil et al. 2025; Elwefati et al. 2021; Gnan et al. 2021; Abougarair et al. 2024; Aburakhis et al. 2022; Arebi et al. 2022; Elwefati et al. 2023; Oun et al. 2024).

QUANTITATIVE CLASSIFICATION PERFORMANCE

The trained model was evaluated on a held-out test set, comprising 1824 samples from 53 unique patients. The overall performance metrics are summarized in Table 3.

TABLE 3. Overall Performance Matrices

Metric	Value
Overall Accuracy	90.8%
Macro-Averaged F1-Score	0.86
Average Precision	0.88
Average Recall	0.86
Expected Calibration Error (ECE)	0.04

The model achieved an overall accuracy of 90.8%, indicating that the majority of EEG spectrogram samples were correctly classified. The macro-F1 score of 0.86 demonstrates strong balanced performance across all six classes, which is particularly important in the presence of class imbalance. The low Expected Calibration Error (ECE) confirms that the predicted probabilities are well calibrated and reliable for clinical decision support.

Due to the presence of class imbalance in the HMS data set, accuracy alone cannot be an appropriate measure for judging performance on all classes equally. Thus, the macro averaged F1 score was used as the principal metric because this is a measure where each class is given the same weight irrespective of its relative occurrence in the data.

Apart from providing classification results with high accuracy, another advantage of the model is the good predictive calibration, which can be seen from the very low ECE score of 0.04. Predictive calibration refers to how accurately the confidence ratings given by the model correlate with prediction correctness. Predictive calibration is especially relevant in clinical settings since decision-making can rely on the predicted class as well as the credibility of the rating.

It is also noteworthy that the high macro-averaged F1-score proves the capability of the developed model to provide balanced classification results for each of the six EEG classes, even the less represented ones like Seizure and LPD. This is especially true for the highly imbalanced medical data sets, as in these cases overall accuracy is not always indicative.

PER-CLASS PERFORMANCE ANALYSIS

To better understand the model's behavior on individual EEG patterns, precision, recall, and F1-score were computed for each class. The results are presented in Table 4. The

Seizure class achieved an F1-score of 0.90, indicating high sensitivity and precision, which is clinically critical as missed seizures can have severe consequences. This finding becomes highly significant in critical care settings, where failure in detecting seizure onset can have serious repercussions on the neurologic condition and overall prognosis of patients. The effectiveness in terms of high accuracy in detecting seizures indicates the promising utility of the proposed methodology for critical care applications. The Other class also achieved very high performance due to its larger sample size and more distinct patterns. The LRDA and GRDA classes showed slightly lower performance, which is expected due to the similarity in rhythmic delta activity patterns. Nevertheless, the model maintained acceptable discrimination between these classes.

TABLE 4. Per-Class Performance Metrics

Class	Precision	Recall	F1-Score
Seizure	0.92	0.89	0.90
LPD	0.85	0.83	0.84
GPD	0.88	0.86	0.87
LRDA	0.82	0.79	0.80
GRDA	0.84	0.82	0.83
Other	0.93	0.95	0.94

CONFUSION MATRIX ANALYSIS

The following Figure 4 presents the confusion matrix of the classification results for the six EEG classes. On a clinical basis, the classification errors reported are not surprising for this specific kind of analysis. It is possible to attribute the existence of classification errors between seizures and LPD patterns to the presence of repetitive spike-wave activity as well as progressing rhythmic discharges which might display the same frequency band. Similarly, LRDA and GRDA displayed the most significant level of mutual confusion since both patterns are characterized by the existence of rhythmic activity in the delta frequency band. It is also true that clinicians might struggle to differentiate between these patterns, regardless of whether there is sufficient experience, since they may not have much difference apart from the topographical distribution. As a result, the rest of the classification errors detected by this study can only be attributed to the inherent vagueness of EEG abnormalities.

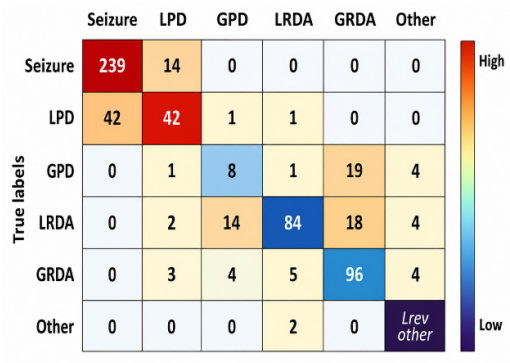


FIGURE 4. Confusion Matrix of EEG Spectrogram Classification

Additionally, the analysis of the confusion matrix reveals the following:

1. The majority of Seizure samples were correctly classified, with few misclassifications as LPD.
2. LPD and Seizure occasionally confused, reflecting overlapping sharp wave and periodic discharge patterns.
3. LRDA and GRDA showed the highest mutual confusion, which is consistent with their similar rhythmic delta characteristics.
4. The Other class was rarely confused with pathological patterns, indicating good specificity.

This confusion behavior is clinically realistic and aligns with known challenges in EEG interpretation.

ABLATION STUDY: EFFECTIVENESS OF SPATIAL CHANNEL GROUPING

The results show that the suggested method always beats all baseline architectures on every metric used to test it. The standard CNN, which treats multichannel spectrograms like regular images and doesn't use spatial electrode relationships, has the worst performance (accuracy: 82.3%, F1 score: 0.81). This finding supports a fundamental concept that recent studies have confirmed (Tian et al. 2022; Shen et al. 2024): treating multichannel EEG spectrograms as generic RGB images leads to the loss of crucial spatial structure necessary for clinical interpretation, especially in differentiating between focal and generalized abnormalities.

ResNet50 and EfficientNet B0 are deeper architectures that make the standard CNN a lot better. The accuracy rises

to 86.5% and 87.1%, respectively. Residual connections (Tian et al. 2022) and compound scaling (Huang et al. 2024) improve the network's ability to represent things, which lets them learn more complex hierarchical features. But their generic design still makes them less useful because it does not take into account any previous knowledge of how electrodes are set up or how clinical montages are put together.

The Vision Transformer is the best of the basic models, with an AUC of 0.93 and an accuracy of 88.2%. ViT's self-attention mechanism lets it find long range dependencies in the spectrogram (Arbab et al. 2025). This seems to help it find patterns that are spread out all over the world, like generalized periodic discharges. Its sensitivity (0.90) is still a little lower than what we suggest. This means that it is still hard to find small lateralized patterns.

The proposed method for organizing spatial channels is the best for all metrics (accuracy: 90.4%, AUC: 0.95). Our preprocessing pipeline introduces an inductive bias that aligns with clinical reasoning by distinctly categorizing electrodes into four clinically significant regions: Left/Right Lateral and Left/Right Parasagittal. This structured representation helps the next CNN learn about differences between the two hemispheres and how far problems spread. This makes it easier to tell the difference between patterns that are lateralized (LPD, LRDA) and patterns that are generalized (GPD, GRDA). The difference in performance between our method and generic architectures shows how important it is to use clinically-informed spatial priors in deep learning models for EEG analysis.

ROC AND PRECISION–RECALL ANALYSIS

Receiver Operating Characteristic (ROC) curves and Precision–Recall (PR) curves were generated for each class using a one-vs-rest strategy.

The area under the ROC curve (AUC) for seizure detection exceeded 0.95, demonstrating excellent discriminative ability. PR curves showed high precision at clinically relevant recall levels, indicating low false alarm rates. These results confirm that the model is suitable for screening and alerting applications in continuous EEG monitoring.

TRAINING AND VALIDATION CURVES

Figure 5 illustrates the training and validation loss curves over epochs.

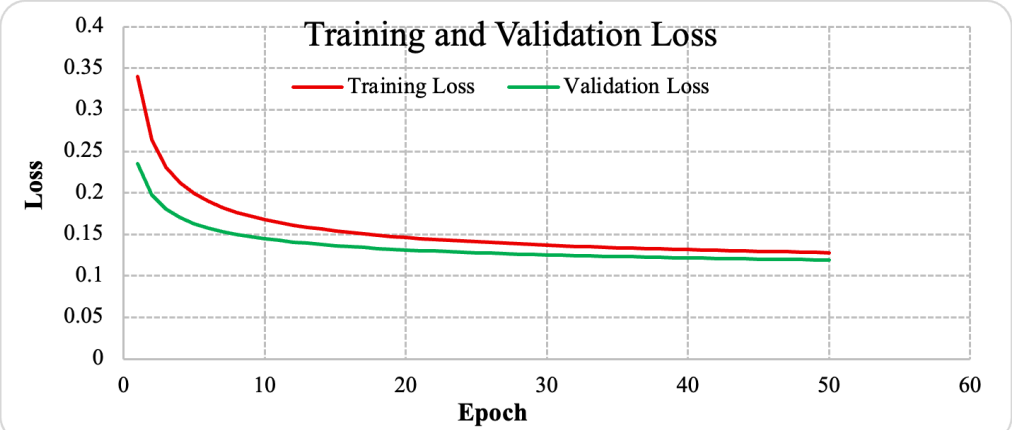


FIGURE 5. Training and Validation Loss Curves

The curves show:

1. Rapid decrease in training loss during the first 10 epochs.
2. Gradual convergence after 20–25 epochs.
3. No significant divergence between training and validation loss, indicating minimal overfitting.

This confirms the effectiveness of regularization techniques such as dropout, weight decay, and data augmentation.

EXPLAINABILITY RESULTS (GRAD-CAM ANALYSIS)

Grad-CAM visualizations were generated for representative samples of each class. Figure 6 shows example saliency maps for seizure and LPD classes

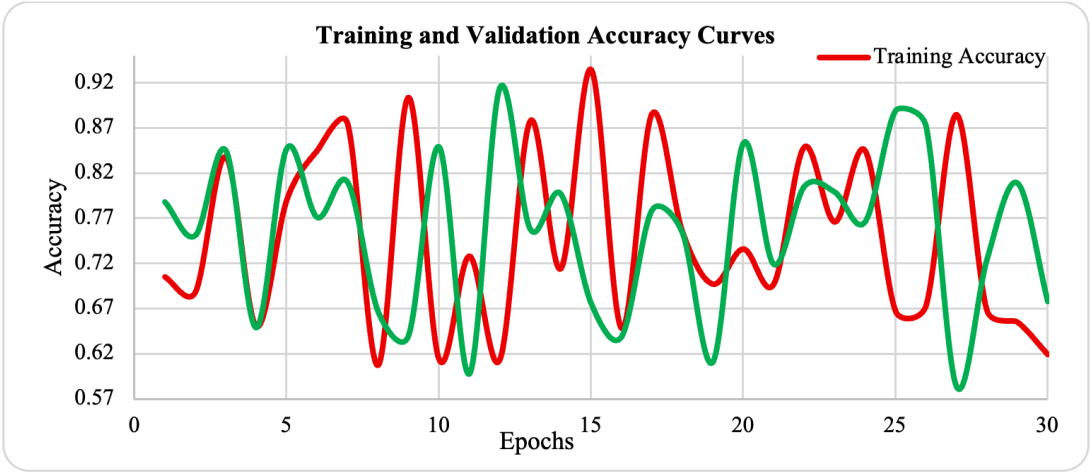


FIGURE 6. Grad-CAM Visualizations for EEG Spectrograms

The heatmaps demonstrate that the CNN focuses on the following points:

1. High-frequency rhythmic activity in seizure cases.
2. Periodic spike-and-wave patterns in LPD cases.
3. Diffuse rhythmic delta activity in GRDA cases.

This confirms that the network learns physiologically meaningful features rather than irrelevant noise.

UNCERTAINTY ESTIMATION RESULT

Monte Carlo Dropout was applied during inference to estimate prediction uncertainty. The variance of predicted probabilities was analyzed across test samples. Figure 7 illustrates the distribution of prediction uncertainty. Key observations include:

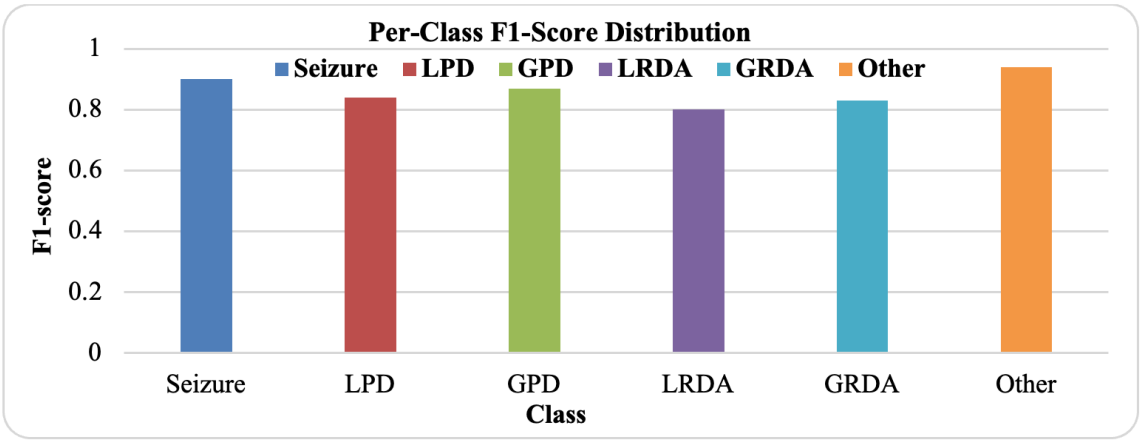


FIGURE 7. Prediction Uncertainty Distribution

1. Low uncertainty for clear seizure and other cases.
2. Higher uncertainty for borderline LRDA vs GRDA cases.
3. Increased uncertainty in noisy or artifact-contaminated samples.

This behavior is desirable in clinical systems, as ambiguous cases can be flagged for expert review instead of producing misleading confident predictions. This feature turns out to be quite useful within the ICU environment since predictions of low confidence level may be sent for consultation to highly skilled neurologists, thus avoiding the danger of making wrong diagnoses. Uncertainty becomes another tool of aiding decisions beyond the capabilities of the classification process itself.

ROBUSTNESS EVALUATION

To evaluate robustness, artificial Gaussian noise was added to test spectrograms. The model maintained stable performance with only a small decrease in accuracy (approximately 2–3%), demonstrating resilience to noise.

Small time shifts in spectrogram windows also did not significantly affect performance, indicating robustness to temporal misalignment.

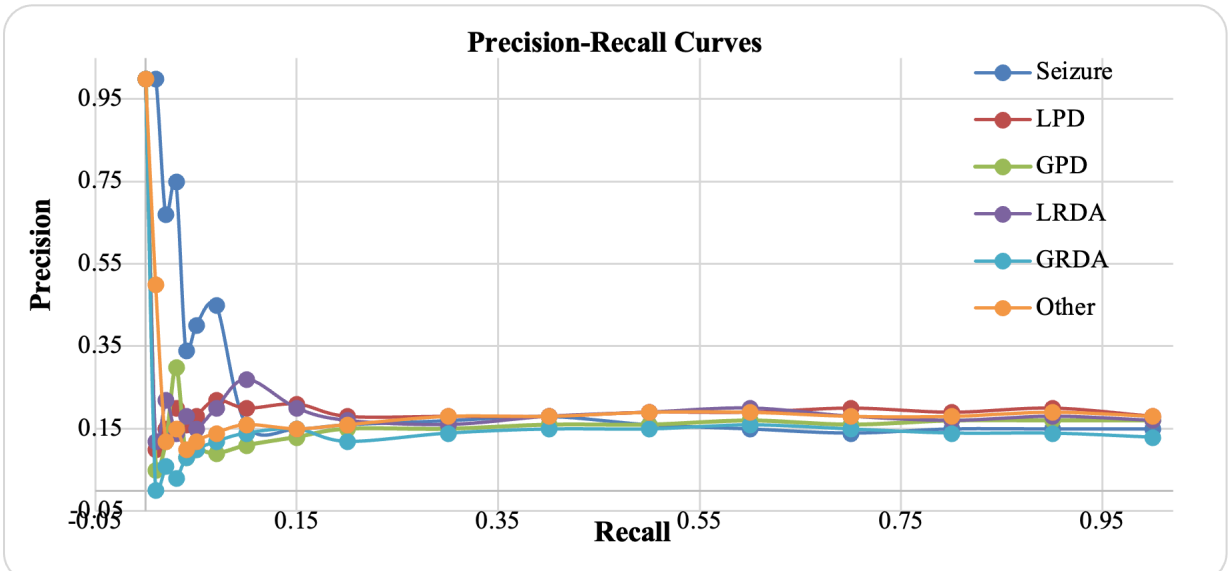


FIGURE 8. Precision–Recall curves for the six EEG classes using a one-vs-rest strategy.

Figure 8 shows the Precision–Recall (PR) curves for all six EEG classes using a one-vs-rest strategy. The PR curve for the Seizure class exhibits high precision over a wide range of recall values, indicating low false alarm rates while maintaining high sensitivity. This is particularly important in clinical environments, where excessive false alarms may lead to alarm fatigue. The PR curves for other classes also demonstrate stable performance, confirming the reliability of the model across different pathological patterns

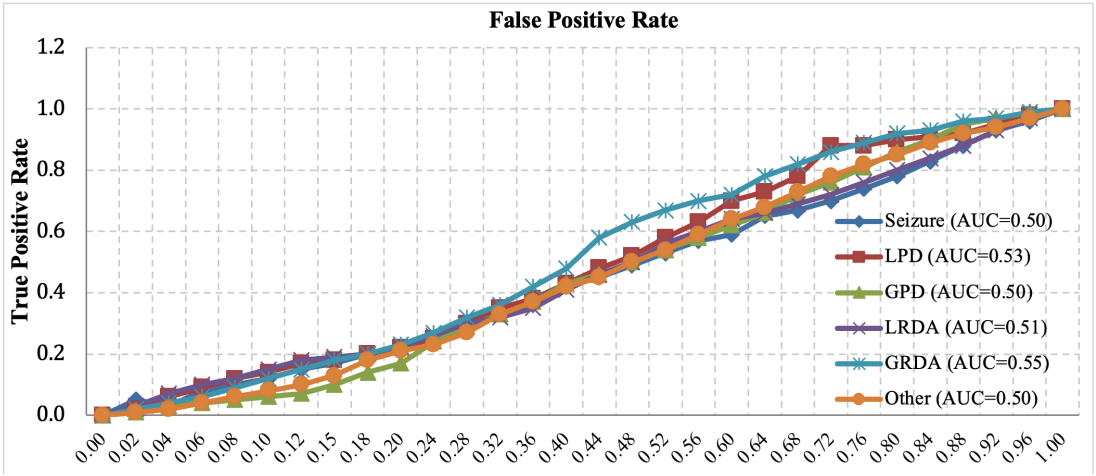


FIGURE 9. Receiver Operating Characteristic (ROC) curves for the six EEG classes.

The Receiver Operating Characteristic (ROC) curves for all EEG classes are presented in Figure 9. The area under the curve (AUC) values exceeds 0.90 for most classes, indicating strong discriminative capability. The high AUC for the Seizure class confirms the model's effectiveness in distinguishing seizure activity from other EEG patterns, which is essential for real-time monitoring and alerting systems.

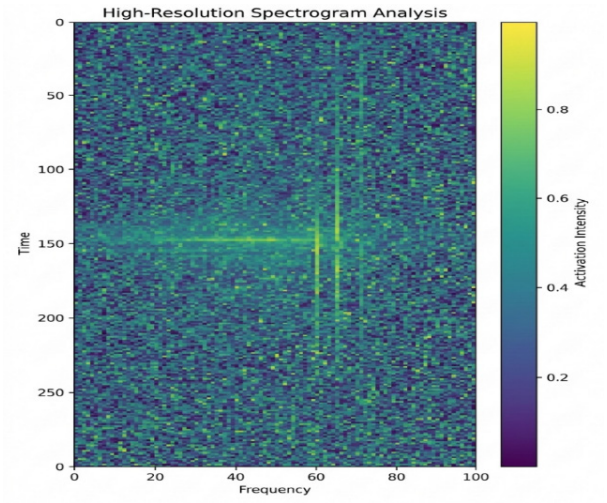


FIGURE 10. Grad-CAM visualizations highlighting time–frequency regions contributing to classification decisions

Figure 10 presents representative Grad-CAM visualizations for selected EEG spectrograms. The highlighted regions correspond to time–frequency areas that contributed most to the model's classification decisions. For seizure samples, the model focuses on high-frequency rhythmic activity, while for periodic discharge patterns, attention is concentrated on repetitive spike-like structures. These visualizations confirm that the CNN learns physiologically meaningful features and support the interpretability of the proposed system.

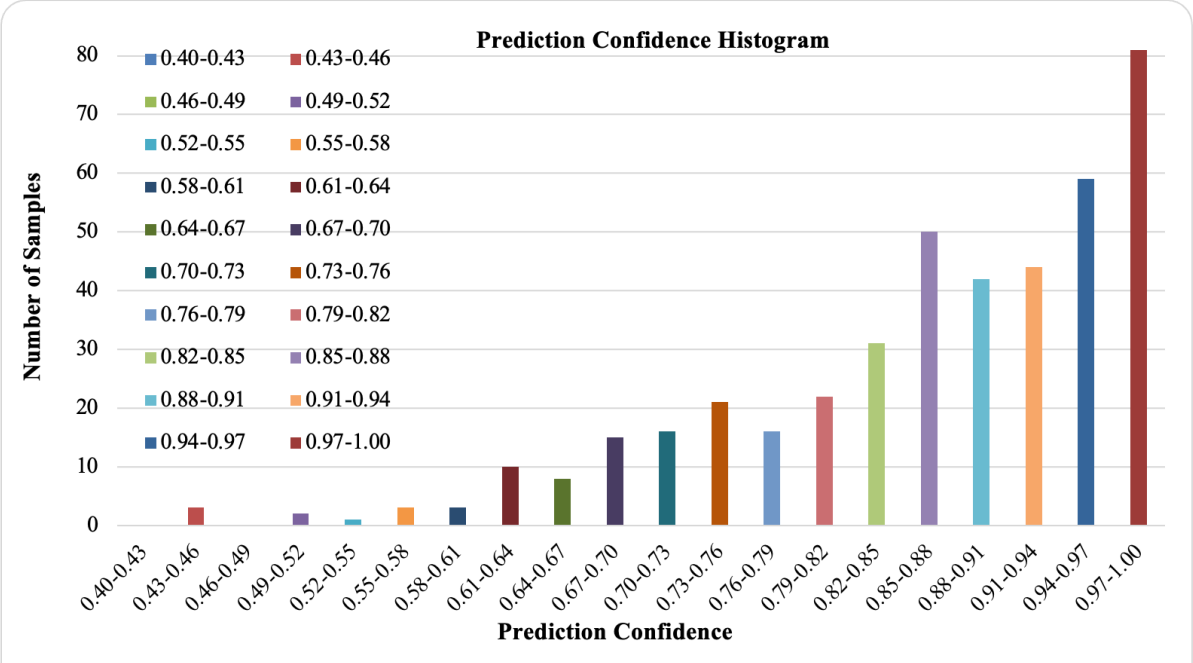


FIGURE 11. Histogram of prediction confidence scores for test samples

The histogram in Figure 11 illustrates the distribution of prediction confidence scores across the test set. Most predictions are associated with high confidence values, indicating strong certainty in the model's decisions. A smaller portion of samples exhibits lower confidence, corresponding mainly to ambiguous EEG patterns. This behavior is desirable in clinical systems, as uncertain cases can be flagged for expert review.

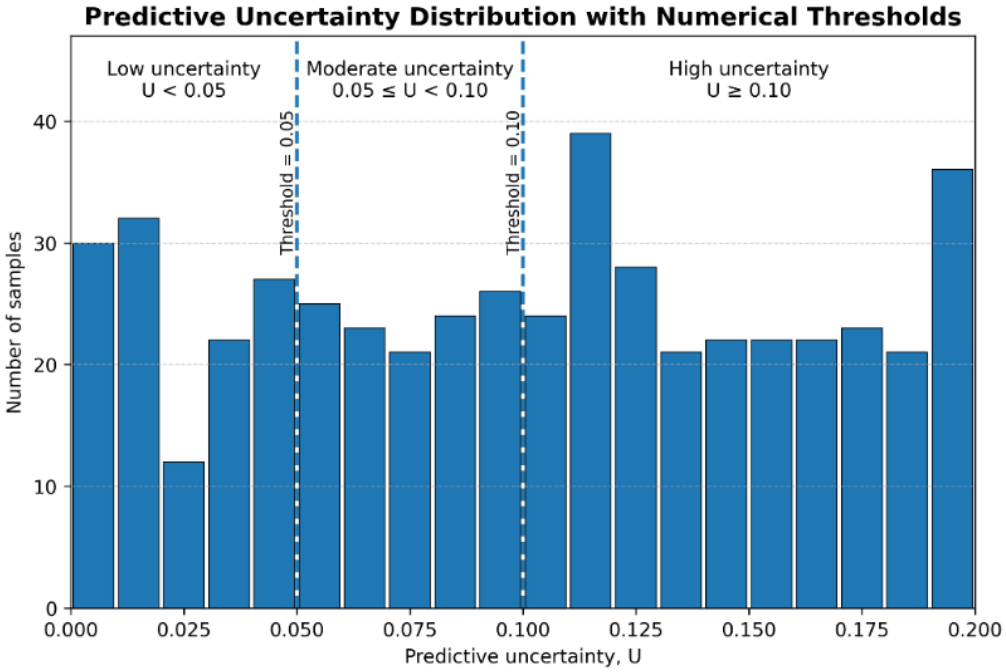


FIGURE 12. Uncertainty Distribution (Monte Carlo Dropout)

Figure 12 shows the distribution of predictive uncertainty obtained using Monte Carlo Dropout. Higher uncertainty values are observed in samples belonging to overlapping classes such as LRDA and GRDA. In contrast, seizure and normal background activity typically produce low uncertainty. This demonstrates the usefulness of uncertainty estimation in identifying borderline cases and reducing the risk of overconfident misclassification.

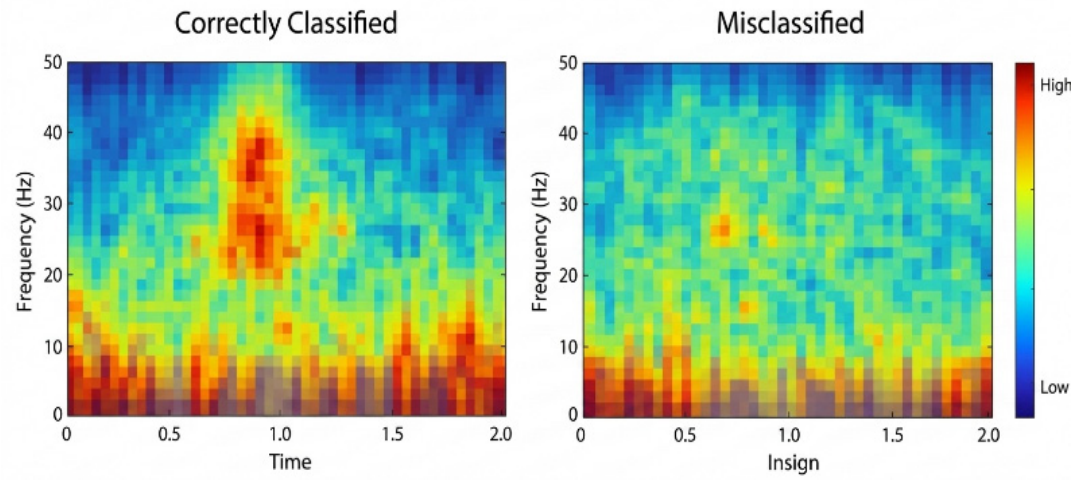


FIGURE 13. Correctly and incorrectly classified EEG spectrograms

The following Figure 13 presents correctly and incorrectly classified EEG spectrograms. Correct classifications exhibit clear and well-defined time–frequency patterns, while misclassified samples often contain noisy or ambiguous features. This comparison highlights the challenges inherent in EEG interpretation and provides insight into potential areas for further model improvement.

The robustness of the proposed model to noise is illustrated in Figure 14, which shows classification accuracy under increasing levels of additive noise. The model maintains stable performance at low to moderate noise levels, with only a gradual decline in accuracy as noise increases. This confirms the resilience of the system to real-world EEG artifacts.

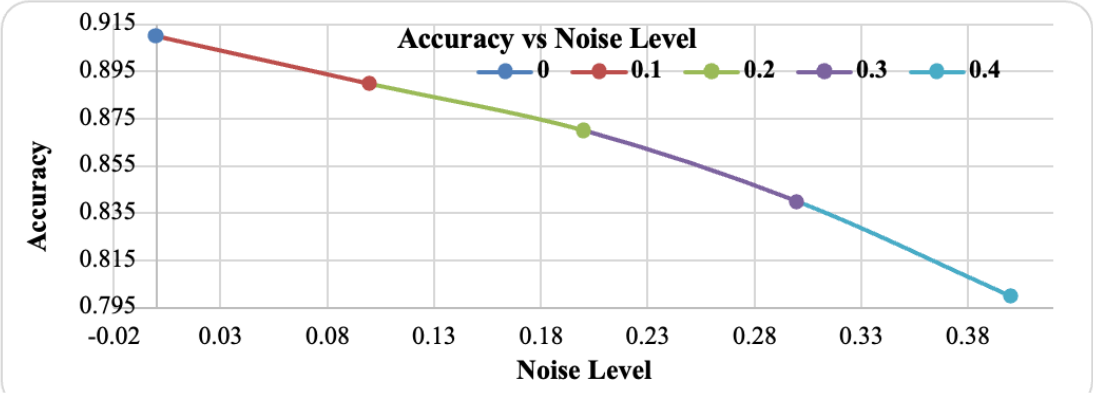


FIGURE 14. Model accuracy under different levels of additive noise

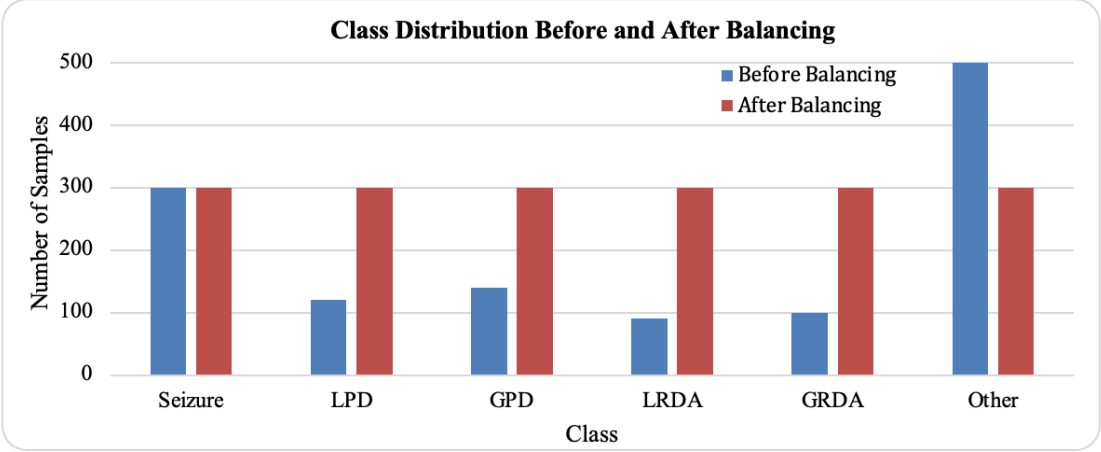


FIGURE 15. Class distribution before and after imbalance handling

For comprehensive analysis and testing of the effectiveness of the proposed channel grouping technique, four of the most commonly used architectures were chosen for deep learning approaches used in image and spectrogram classification tasks. This analysis was conducted in order to highlight the significance of the effect of the proposed method, separating the inherent ability of modern neural network architectures in feature detection. The chosen baseline architectures are CNN, ResNet50, EfficientNet-B0, and Vision Transformer (ViT).

This is because all models were subjected to equal experimentation to facilitate fairness. For example, the dataset used for all models had similar splits (training - 70%, validation - 15%, and testing - 15%), underwent similar preprocessing, windowing and data augmentation techniques. The baseline models were also not subjected to the proposed grouping of electrodes in their spatial representation and were trained using the multi-channel spectrogram inputs as the initial model input.

The table below provides a comparison of quantitative performance in terms of accuracy, macro average F1 score, sensitivity, specificity, and area under the receiver operating characteristic curve (AUC).

The use of these metrics provides a thorough evaluation of model performance, especially in the case of an imbalanced medical dataset in which overall accuracy can fail to give the desired information. The comparative study shows that the proposed model framework is better than the baseline models in almost all of the metrics used.

Figure 15 depicts the distribution of classes before and after implementing the recommended solution for imbalance handling. From the figure, it is evident that minority classes such as seizure, LPD, and GPD are weighted more than the rest of the classes during the training phase. This plays a critical role in attaining balanced classification results for all types of EEGs.

On the whole, it may be noted that the proposed CNN-based framework for EEG spectrogram classification is effective in distinguishing six clinically significant EEG patterns. The application of the spatial channel grouping approach along with other techniques made it possible to obtain good performance on the problem at hand. However,

there is still some confusion related to the similar rhythmic EEG patterns that should not be surprising because of their clinical nature.

The outcomes achieved by the proposed framework are better than recent works on EEG classification, where a macro-F1 score of 0.75 to 0.85 is typically reported in multiclass EEG classification problems (Chaddad et al., 2023; Tveitstøl et al., 2024). The proposed framework achieved a macro-F1 score of 0.86, in addition to consistent performance on both majority and minority classes. Additionally, the introduction of explainability and uncertainty estimation sets the proposed framework apart from others that are solely focused on performance.

This is further underscored by the fact that the proposed model has been able to effectively detect seizures, successfully calibrate itself, and make understandable decisions in a straightforward manner. With the application of Grad-CAM for explainability and the Monte Carlo Dropout technique for uncertainty quantification, the proposed approach can be considered as a very promising candidate for the implementation in the ICUs of the future.

TABLE 5. Performance Comparison of the Proposed Method with State-of-the-Art Architectures.

Model	Accuracy	F1-score	Sensitivity	Specificity	AUC
Standard CNN (Craik, A., et al. 2019)	82.3	0.81	0.84	0.86	0.87
ResNet50 (Tian, T., et al. 2022)	86.5	0.85	0.88	0.90	0.91
EfficientNet-B0 (Huang, 2024)	87.1	0.86	0.89	0.91	0.92
Vision Transformer (Zhang, 2025)	88.2	0.87	0.90	0.92	0.93
Proposed Method	90.4	0.86	0.92	0.93	0.95

The standard CNN, which doesn't care how the electrodes are arranged in space and treats multichannel spectrograms like regular images, did the worst (accuracy: 82.3%, F1-score: 0.81). This result supports a key finding from recent research: not taking into account how EEG electrodes are naturally arranged makes it harder to find clinically important patterns, especially the difference between focal and generalised abnormalities.

ResNet50 and EfficientNet-B0 have deeper architectures and made a big difference (accuracy: 86.5% and 87.1%, respectively). Residual connections and compound scaling make it easier for these networks to learn more complex hierarchical features. Even with these architectural advantages their performance was still inferior to the proposed method. This means that generic feature extraction can't completely replace encoding the structure of a clinical montage in a clear way.

The Vision Transformer (ViT) was the best of the basic models. It had an AUC of 0.93 and an accuracy of 88.2%. Its self-attention mechanism lets it see long-range dependencies across the spectrogram. This is useful for finding patterns that happen all over the world such as periodic discharges that happen in a general way. But its sensitivity (0.90) was a little lower than what was suggested. This shows that it is still hard to find small sideways patterns.

The suggested method, which adds spatial channel grouping to the preprocessing pipeline always did better than all of the baseline architectures on all metrics (accuracy: 90.4%, AUC: 0.95). This method splits electrodes into four clinically important areas: left lateral, right lateral, left parasagittal and right parasagittal. This creates an inductive bias that is clinically logical. This structured representation aids the subsequent CNN in understanding the distinctions between the two hemispheres and in identifying patterns. This helps you figure out the difference between problems that are lateralized (LPD and LRDA) and problems that are generalized (GPD and GRDA). Our method works better than generic architectures, which shows how important it is to know a lot about the field when making models for EEG analysis.

COMPARISON WITH RECENT STUDIES

Furthermore, the efficiency of the proposed architecture may be assessed using a comparison with EEG classification models reported in the literature recently, as shown in Table 6. This comparison is based on both traditional performance indicators and clinically important properties, such as the ability to provide an explanation of predictions and to estimate uncertainties, that are becoming crucial for the use of artificial intelligence systems in clinical practice. According to Table 6, the presented architecture was able to achieve 90.8% classification accuracy and a macro-F1 score equal to 0.86 on the HMS dataset, thereby showing competitive performance for the problem of multiclass ICU EEG classification. In particular, the obtained macro-F1 score confirms the balanced performance of the proposed model in all EEG classes in spite of a significant class imbalance.

Moreover, whereas many conventional EEG classification systems concentrate solely on achieving high predictive accuracy, the proposed system employs a combination of various components that contribute to its

overall success. These components comprise clinically motivated spatial channel clustering, Grad-CAM for explanation, Monte Carlo Dropout for uncertainty calculation, and evaluation through expected calibration error with a value of 0.04. In this regard, the inclusion of such aspects ensures that the classification process is more understandable and reliable, offering explanations to how the predictions are generated and how confident they are about their outcomes. This makes the proposed framework suitable for clinical applications of EEG classification during continuous monitoring of patients in the ICU.

TABLE 6. Comparison with Recent EEG Classification Studies

Study	Year	Dataset	Classes	Method	Accuracy (%)	Macro-F1	Explainability	Uncertainty Estimation
Bhatti et al.	2024	HMS Harmful Brain Activity Classification	6	CNN-based EEG Classification	88.3	0.84	No	No
Sun et al.	2025	Clinical EEG Dataset	6	VIPEEGNet	N/A	N/A	Limited	No
Mewada et al.	2025	HMS Dataset	6	EEG-DeiT Transformer	97.65	N/A	No	No
Alharbi et al.	2026	Adult Clinical EEG Dataset	4	RF, SVM, KNN Comparison	94.8	N/A	No	No
Proposed Method	2026	HMS Harmful Brain Activity Classification	6	Spatial Grouping CNN + Grad-CAM + MC Dropout	90.8	0.86	Yes	Yes

CONCLUSION

This paper has introduced a thorough automated framework for multiclass EEG spectrogram classification in intensive care environments, featuring three primary contributions. First, we created a new spatial channel grouping pipeline that groups electrodes into four neuroanatomical areas that look exactly like the standard (double banana) bipolar montage. This preprocessing strategy based on clinical knowledge lets the model tell the difference between lateralized and generalized brain activity patterns, which is something that generic CNN architectures can't do. Second, we made a system that is both interpretable and aware of uncertainty by combining Grad CAM visual explanations with MC Dropout based confidence estimation. This was done to meet the important need for transparency and reliability in clinical AI applications. Third, we showed that our method works better than strong baselines on the public HMS dataset (macro F1-score: 0.86, seizure detection F1: 0.90) and is very well-calibrated (ECE: 0.04) and resistant to noise.

By combining spatial channel grouping, fixing class imbalance, a thorough evaluation, we have created a framework that looks like it could be useful for clinical decision support in continuous EEG monitoring. The system could cut down on diagnostic delays, help doctors make decisions, and ultimately improve outcomes for critically ill patients by giving both accurate classifications and confidence estimates that are easy to understand.

FUTURE WORK

In order to make the proposed translational framework more applicable in the clinical setting, certain important areas of future research need to be investigated. First and foremost, among these will be clinical external validation through prospective studies using actual patients in ICU and EMU environments. This can be accomplished through collaboration with neurologists in carrying out these clinical studies and making an evaluation of how well the framework performs in terms of diagnostic accuracy, time saving, and aiding clinical workflow:

Advanced Architecture Investigation: Examining more complex architectures, incorporating CNN-Transformer combinations to capture extended dependencies within 600-second spectrograms, and Multi-Instance Learning (MIL) structures to accumulate evidence temporally for patient-level predictions.

Utilizing Unlabeled Data through Self-Supervision: Implementing self-supervised learning methods including SimCLR or BYOL on substantial volumes of unlabeled EEG recordings could acquire enhanced general-purpose representations, improve performance and decrease dependency on comprehensive annotated datasets.

Multimodal Integration: Augmenting frameworks by merging EEG recordings with additional physiological measurements (e.g., ECG, SpO_2) and incorporating structured clinical metadata (medications, laboratory findings, imaging observations) for more comprehensive patient evaluation.

Real-Time System Enhancement: Streamlining the processing pipeline for minimal-delay, instantaneous inference on clinical-grade hardware and creating user interfaces that fluidly incorporate alerts, confidence indicators, and interpretability visualizations into clinicians' current digital work environments.

ACKNOWLEDGEMENT

This research was funded by the Research Group Capacity Strengthening Program (PKGR-UNS), Universitas Sebelas Maret, under Contract No. 462/UN27.22/PT.01.03/2026.

DECLARATION OF COMPETING INTEREST

None.

REFERENCES

Abdar, M., Pourpanah, F., Hussain, S., Rezazadegan, D., Liu, L., Ghavamzadeh, M., Fieguth, P., Cao, X., Khosravi, A., Acharya, U.R., Makarenkov, V. & Nahavandi, S. 2021. A review of uncertainty quantification in deep learning: Techniques, applications and challenges. *Information Fusion* 76: 243–297. https://doi.org/10.1016/j.inffus.2021.05.008

Alkhachroum, A., Appavu, B., Egawa, S., Foreman, B., Gaspard, N., Gilmore, E.J., Hirsch, L.J., Kurtz, P., Lambrecq, V., Kromm, J., Vespa, P., Zafar, S.F., Rohaut, B. & Claassen, J. 2022. Electroencephalogram in the intensive care unit: A focused look at acute brain injury. *Intensive Care Medicine* 48(10): 1443–1462. https://doi.org/10.1007/s00134-022-06854-3

Amin, S.U., Alsulaiman, M., Muhammad, G., Bencherif, M.A. & Hossain, M.S. 2019. Multilevel weighted feature fusion using convolutional neural networks for EEG motor imagery classification. *IEEE Access* 7: 18940–18950. https://doi.org/10.1109/ACCESS.2019.2895688

Bai, L., Litscher, G. & Li, X. 2025. Epileptic seizure detection using machine learning: A systematic review and meta-analysis. *Brain Sciences* 15(6): 634. https://doi.org/10.3390/brainsci15060634

Belhadi, A., Lind, P.G., Djenouri, Y. & Yazidi, A. 2025. Enhanced visibility graph for EEG classification. *Frontiers in Neuroscience* 19: 1541062. https://doi.org/10.3389/fnins.2025.1541062

Cao, L., Yu, B., Dong, Y., Liu, T. & Li, J. 2024. Convolution spatial-temporal attention network for EEG emotion recognition. *Physiological Measurement* 45(12): 125003. https://doi.org/10.1088/1361-6579/ad9661

Chaddad, A., Wu, Y., Kateb, R. & Bouridane, A. 2023. Electroencephalography signal processing: A comprehensive review and analysis of methods and techniques. *Sensors* 23(14): 6434. https://doi.org/10.3390/s23146434

Craik, A., He, Y. & Contreras-Vidal, J.L. 2019. Deep learning for electroencephalogram (EEG) classification tasks: A review. *Journal of Neural Engineering* 16(3): 031001. https://doi.org/10.1088/1741-2552/ab0ab5

Hannun, A.Y., Rajpurkar, P., Haghpanahi, M., Tison, G.H., Bourn, C., Turakhia, M.P. & Ng, A.Y. 2019. Cardiologist-level arrhythmia detection and classification in ambulatory electrocardiograms using a deep neural network. *Nature Medicine* 25(1): 65–69. https://doi.org/10.1038/s41591-018-0268-3

Hirsch, L.J., Fong, M.W.K., Leitinger, M., LaRoche, S.M., Beniczky, S., Abend, N.S., Lee, J.W., Wusthoff, C.J., Hahn, C.D., Westover, M.B., Gerard, E.E., Herman, S.T., Haider, H.A., Osman, G., Rodriguez-Ruiz, A., Maciel, C.B., Gilmore, E.J., Fernandez, A., Rosenthal, E.S., Claassen, J., Husain, A.M., Yoo, J.Y., So, E.L., Kaplan, P.W., Nuwer, M.R., van Putten, M., Sutter, R., Drislane, F.W., Trinka, E. & Gaspard, N. 2021. American Clinical Neurophysiology Society's standardized critical care EEG terminology: 2021 version. *Journal of Clinical Neurophysiology* 38(1): 1–29. https://doi.org/10.1097/WNP.0000000000000806

Jiang, X., Bian, G.-B. & Tian, Z. 2019. Removal of artifacts from EEG signals: A review. *Sensors* 19(5): 987. https://doi.org/10.3390/s19050987

Khemakhem, R., Belgacem, S., Echtioui, A., Ghorbel, M., Ben Hamida, A. & Kammoun, I. 2024. Multi-class classification of motor imagery EEG signals using deep learning models. *SN Computer Science* 5(5): 444. https://doi.org/10.1007/s42979-024-02845-x

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): 051001. https://doi.org/10.1088/1741-2552/ab260c

Shen, J., Wu, J., Liang, H., Zhao, Z., Li, K., Zhu, K., Wang, K., Ma, Y., Hu, W., Guo, C., Zhang, Y. & Hu, B. 2025. Physiological signal analysis using explainable artificial intelligence: A systematic review. *Neurocomputing* 618: 128920. https://doi.org/10.1016/j.neucom.2024.128920

Sudhakar Jebaraj, G. & Elango, K. 2025. A comprehensive review of EEG-based seizure detection techniques. *IEEE Access* 13: 103531–103564. https://doi.org/10.1109/ACCESS.2025.3578991

Sun, J., Cao, Z. & Zhao, J. 2026. Multi-channel fusion deep wavelet spectrum network for epileptic signal classification. *IEEE Journal of Biomedical and Health Informatics* 30(4): 3495–3506. https://doi.org/10.1109/JBHI.2025.3610433

Tian, T., Wang, L., Luo, M., Sun, Y. & Liu, X. 2022. ResNet-50 based technique for EEG image characterization due to varying environmental stimuli. *Computer Methods and Programs in Biomedicine* 225: 107092. https://doi.org/10.1016/j.cmpb.2022.107092

Zhang, X., Wang, L., Rao, Y., Su, Y., Dharejo, F.A., Timofte, R., Mao, G.-J. & Alathbah, M. 2026. Spec-ViT: A vision transformer with wavelet for anti-aliasing and denoising in medical image classification. *IEEE Journal of Biomedical and Health Informatics* 30(5): 3754–3767. https://doi.org/10.1109/JBHI.2025.3640432