

Hybrid Temporal–Spatial Electroencephalographic Classification Using Principal Component Analysis and Ensemble Learning: Epilepsy, Attention-Deficit/Hyperactivity Disorder, and Schizophrenia

Dawood S. Hasan, Qais Al-Gayem, Hilal Al-Libawy

Department of Electrical Engineering, College of Engineering, University of Babylon, Babylon, Iraq

ARTICLE INFORMATION

Article History:

Received 18 May 2026

Revised 10 August 2026

Accepted 07 October 2026

Keywords:

EEG Signal Classification;
Hybrid Deep Learning;
Principal Component Analysis;
Ensemble Learning;
Biomedical Signal Processing

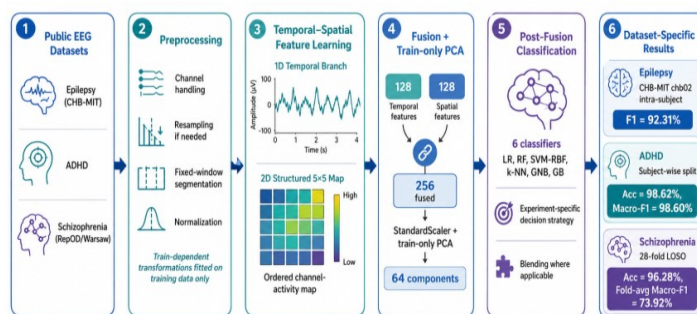
Corresponding Author:

Dawood S. Hasan,
Department of Electrical
Engineering, College of
Engineering, University of
Babylon, Babylon, Iraq.
Email:
eng357.dawood.salman@student.uobabylon.edu.iq

This work is licensed under a [Creative Commons Attribution-Share Alike 4.0](https://creativecommons.org/licenses/by-sa/4.0/)



ABSTRACT



Electroencephalography (EEG) allows noninvasive access to temporal brain activity; however, many EEG classifiers are tailored to a single dataset or disorder and are not easily compared across heterogeneous recording settings. A gap that this study fills is the evaluation of a fixed hybrid temporal–spatial EEG representation pipeline with an identical processing flow on public datasets for epilepsy, attention-deficit/hyperactivity disorder (ADHD), and schizophrenia. Each window passes through a one-dimensional temporal branch and a two-dimensional spatial branch. Their outputs are projected and concatenated into a 256-dimensional fused vector, then standardized and reduced to 64 components by principal component analysis (PCA) fitted only on the training data. The six post-fusion classifiers were tested under dataset-specific validation protocols, with experiment-specific blending applied where applicable, without using held-out test data for scaling, PCA fitting, model selection, threshold adjustment, or blending. Subject-wise ADHD classification achieved around 98.62% accuracy and approximately 98.60% macro-F1. Under 28-fold leave-one-subject-out evaluation, schizophrenia classification yielded an aggregated segment-level accuracy of 96.28% and a fold-averaged macro-F1 score of 73.92%. The selected intra-subject CHB-MIT chb02 experiment attained a seizure-class F1 of 92.31%. The internal four-class experiment was considered exploratory feature-space analysis rather than evidence of dataset-independent clinical diagnosis, as the classes were drawn from disparate datasets and acquisition conditions. The results provide a consistent engineering baseline across the investigated protocols and may inform future EEG decision-support studies, although broader clinical interpretation requires harmonized cohorts, broader multi-subject evaluation, and external validation.

Document Citation:

D. S. Hasan, Q. Al-Gayem, and H. Al-Libawy, “Hybrid Temporal–Spatial Electroencephalographic Classification Using Principal Component Analysis and Ensemble Learning: Epilepsy, Attention-Deficit/Hyperactivity Disorder, and Schizophrenia,” *Buletin Ilmiah Sarjana Teknik Elektro*, vol. 8, no. 5, pp. 1477-1492, 2026, DOI: [10.12928/biste.v8i5.16803](https://doi.org/10.12928/biste.v8i5.16803).

1. INTRODUCTION

EEG is a non-invasive way to measure electrical activity in the brain with high temporal resolution and has become an important source for biomedical signal processing, intelligent healthcare engineering, and automated analysis of brain disorders. Recent advances in deep learning have been particularly useful with regard to this problem, as they learn discriminative spatio-temporal patterns directly from multichannel EEG recordings [1]. However, EEG-based classification remains difficult due to the strong influence of intersubject variability, signal heterogeneity, and preprocessing choices on performance [1].

Three representative EEG-based classification problems with different signal characteristics and evaluation difficulties are considered: epilepsy, attention deficit hyperactivity disorder (ADHD), and schizophrenia. The CHB-MIT scalp EEG database has remained a popular public reference database in seizure-related EEG analysis [2]. However, experimental results also show that segment-level splitting can lead to an overestimation of performance because it may assign segments from the same subject to both the training and test partitions [3]. Recent ADHD and schizophrenia studies have shown strong performance using deep learning approaches leveraging multidomain features, structured EEG representations, and graph-based learning methods [4]–[6]. However, most of these methods are still developed and tested on a single disorder or dataset.

Characteristics of the assessed datasets are heterogeneous. CHB-MIT segments include 23 EEG channels sampled at 256 Hz, ADHD recordings include 19 channels, and RepOD/Warsaw recordings consist of 19 channels sampled at 250 Hz. The datasets were acquired under different acquisition settings and differ in electrode montage and channel configuration, recording duration, subject population, class balance, and preprocessing conditions. These differences can induce domain shift and may cause the learned representations to encode acquisition-related or dataset-origin features in addition to disorder-related EEG patterns. Hence, the internal four-class investigation is interpreted only as an exploratory feature-space assessment and not as proof of dataset-independent neurological separability.

Most pipelines introduced in the EEG-classification literature are limited to a single disorder or dataset, with preprocessing, representation design, model architecture, classifier selection, and validation procedures tailored to the corresponding benchmark [1],[3]–[6]. While such dataset-specific optimization may enhance performance within an individual benchmark, differences in methodological design and data-separation protocols limit direct comparisons across heterogeneous EEG tasks [1],[3],[6]. Thus, it is valuable to investigate whether a fixed core EEG-based representation and classification pipeline can be applied across multiple public datasets under separate dataset-specific protocols while explicitly accounting for acquisition heterogeneity and domain-related confounding.

An attempt is made in this study to introduce a hybrid temporal–spatial EEG representation pipeline with principal component analysis (PCA) and ensemble learning for evaluation across selected brain-disorder datasets. Moreover, the pipeline involves a 1D temporal branch to capture waveform-level dynamics of EEG, a 2D spatial branch to learn structured representations of EEG data, feature fusion, train-only PCA-based dimensionality reduction, and post-fusion ensemble classification. The same core pipeline is evaluated separately on public EEG datasets corresponding to epilepsy, ADHD, and schizophrenia according to dataset-specific validation protocols with leakage-aware processing. The epilepsy experiment is confined to a selected intra-subject CHB-MIT benchmark, whereas the ADHD and schizophrenia evaluations use subject-aware separation and LOSO-style evaluation, respectively.

This research primarily contributes an engineering-oriented evaluation of a fixed hybrid EEG pipeline across multiple disorder-related datasets. In particular, it provides: (1) a temporal–spatial EEG feature extraction strategy that combines raw-segment temporal learning with structured spatial representation; (2) a train-only PCA stage to reduce feature dimensionality while reducing leakage risk; (3) consistent post-fusion classification using the same candidate classifier pool, with experiment-specific blending applied where appropriate; and (4) an internal four-class feature-space experiment used only for exploratory representation analysis. This experiment should not be interpreted as evidence of dataset-independent clinical diagnosis or purely neurological separability because the four classes stem from distinct public datasets and acquisition conditions. By maintaining the same core methodology, the proposed study provides a coherent engineering framework in which to assess EEG representations under the considered protocols, as opposed to a wide-reaching cross-dataset clinical diagnostic model.

The rest of this paper is organized as follows. Section 2 reviews related studies on EEG-based classification of epilepsy, ADHD, and schizophrenia. Section 3 describes the proposed hybrid temporal–spatial EEG pipeline, datasets, validation settings, and reproducibility controls. The experimental results are reported and discussed in Section 4, including the disorder-specific evaluations, the internal four-class feature-space

analysis, and the dataset-aligned comparison with related studies. Section 5 concludes the paper and discusses future research directions.

2. RELATED WORK AND PROBLEM STATEMENT

2.1. Epilepsy EEG Classification

Seizure-related EEG signals are complicated and vary considerably [7][8], which makes EEG-based epilepsy classification an ongoing research topic. This includes the following recent work: a set of convolutional and recurrent deep-learning architectures reviewed for seizure detection and prediction [7][8], latent-space strange-attractor reconstruction for epileptic EEG classification [9], as well as a multi-attention model combining a Transformer encoder with spatiotemporal feature fusion [10]. Together, these studies demonstrate the potential of using deep representation learning for seizure-related EEG analysis; however, they also indicate that differences in models and evaluation criteria across studies can make direct comparison difficult.

However, epilepsy EEG classification still suffers from class imbalance and from the complexity and variability of seizure-related EEG signals [7][8]. Moreover, if segments from the same subject appear in both the training and test sets, segment-level splitting can lead to an overestimation of performance when assessing generalization to previously unseen subjects [3]. Thus, results from an intra-subject seizure-window experiment should be viewed as within-subject performance rather than evidence of generalization to new subjects. In this paper, we use the epilepsy experiment as a benchmark-level seizure-window classification task to test whether the temporal-spatial representation pipeline retains discriminative seizure-related features within the selected intra-subject evaluation setting.

2.2. ADHD EEG Classification

The classification of ADHD from EEG data (EEG-based ADHD classification) is a topic that has recently attracted considerable attention, and several studies have presented analyses based on different EEG representations as well as classification strategies. The methodologies adopted range from EEG feature-map images combined with deep learning [11], multidimensional spectral, entropy, and functional-connectivity features evaluated using machine-learning models [12], Gabor-filter-based statistical features [13], to graph convolutional networks integrating time-domain, frequency-domain, and functional-connectivity information [5]. In aggregate, these studies demonstrate that ADHD-versus-control classification has been approached with a variety of feature representations and model families rather than following a single standardized pipeline.

The ADHD studies reviewed have largely been developed and validated within individual datasets, often involving disparate preprocessing procedures, feature representations, model architectures, and validation designs [5],[11]–[13]. Such methodological differences limit direct comparison of performance across studies, since reported performance often depends not only on the classifier used but also on windowing, electrode configuration, feature construction, and the data-separation protocol. Motivated by this, we evaluate a fixed EEG representation pipeline with a well-defined subject-wise protocol. Hence, the ADHD experiment is utilized in the present study to evaluate the performance of the proposed hybrid temporal-spatial pipeline with separate training, validation, and test subjects.

2.3. Schizophrenia EEG Classification

Recent EEG-classification studies use microstate semantic modeling [14] and deep-learning architectures that utilize multi-scale temporal feature extraction with adaptive weight fusion for schizophrenia classification [15]. Other related EEG studies have focused on spatial-spectral-temporal representation learning [16] and convolutional Transformer architectures for EEG decoding and brain-computer-interface tasks in various application settings [17]–[20]. These works offer methodological context for structured EEG representation learning; however, the studies in [16]–[20] were not conducted on schizophrenia datasets and thus are not interpreted herein as direct evidence of schizophrenia classification performance.

Additionally, the difficulty with schizophrenia EEG classification lies in the fact that available datasets are limited and heterogeneous in size, preprocessing and feature-extraction procedures vary between studies, and reported performance may be driven by preprocessing choices, classifier design, and class imbalance [6].

However, fair interpretation relies on reproducibility, which translates into explicit feature construction, dimensionality reduction, classifier design, and metric definitions. Here, schizophrenia classification is included to see whether the same core temporal-spatial EEG representation pipeline can remain effective beyond seizure and ADHD datasets.

2.4. Problem Statement

While there is increasing EEG-based machine-learning and deep-learning research on epilepsy, ADHD, and schizophrenia [7]–[15], the studies reviewed here are primarily disorder-specific and dataset-specific. Consequently, it is challenging to compare the effects of representation design, dimensionality reduction, classifier selection, and validation strategy across heterogeneous EEG tasks. Thus, this work responds to the need for a reproducible EEG engineering pipeline that allows consistent evaluation across selected public datasets without redesigning its central elements for every disorder under study.

Therefore, the present work examines a fixed hybrid temporal–spatial EEG pipeline with train-only PCA and post-fusion ensemble learning. This is intended to evaluate representation stability and classification behavior in a controlled manner under dataset-specific protocols, while cautiously avoiding overgeneralization beyond the available public datasets or the experimental validation scope.

3. METHODS

3.1. Proposed Temporal–Spatial EEG Pipeline

Our proposed pipeline is a fixed multistage temporal–spatial EEG representation and classification workflow. As illustrated in Figure 1, the processing chain consists of multichannel EEG preparation, window-based segmentation, temporal representation learning, spatial representation learning, feature fusion, train-only PCA-based dimensionality reduction, post-fusion classification, and an experiment-specific final decision stage. To ensure methodological consistency, we kept the same sequence of processing stages across the evaluated datasets, while dataset-specific input dimensions, recording conditions, encoder implementation details, and validation protocols were reported separately.

The first part of our pipeline is dataset preparation, which includes channel-wise organization, resampling, when necessary, fixed-window segmentation, and normalization. It is important to note that all fitted preprocessing and representation transformations were estimated using only the training partition or training fold. Validation data were used only for model selection and threshold determination, where appropriate. The test partition was processed using the fitted pipeline, but scaling and PCA fitting, classifier training, classifier selection, threshold determination, and meta-model training were not performed on it; only the fitted transformations and selected models were applied unchanged to the held-out test data during final evaluation.

Each EEG window was encoded through two complementary branches. Let $X_i \in \mathbb{R}^{C_i \times T_i}$ denote the i -th EEG window, where C_i is the number of channels and T_i is the number of temporal samples for the corresponding dataset. In the 19-channel ADHD and schizophrenia implementations, the temporal encoder comprised four one-dimensional convolutional layers with output channels of 32, 64, 128, and 256; kernel sizes of 7, 5, 5, and 3; unit stride; and padding values of 3, 2, 2, and 1, respectively. Each convolution was followed by batch normalization and ReLU activation, and a max-pooling operation with kernel size 2 and stride 2 followed the first convolution. Adaptive average pooling reduced the final temporal feature maps to 256 values, after which a fully connected projection mapped the temporal vector from 256 to 128 dimensions. The epilepsy benchmark used the corresponding one-dimensional residual temporal encoder for its CHB-MIT input, while retaining the same 256-dimensional post-fusion feature interface before PCA.

For the structured spatial branch, the activity assigned to channel c of window i was defined as the temporal mean-squared amplitude:

$$e_{i,c} = \frac{1}{T_i} \sum_{t=1}^{T_i} X_i(c,t)^2 \quad (1)$$

The resulting channel-activity vector was centered and standardized within each window as:

$$z_{i,c} = \frac{e_{i,c} - \mu_i}{\sigma_i + 10^{-6}} \quad (2)$$

where $X_i(c,t)$ in (1) denotes the EEG amplitude of channel c at temporal sample t in window i , and the quantity defined in (1) represents the temporal mean-squared activity of that channel. The standardized activity value is given by (2). For window i , T_i and C_i denote the numbers of temporal samples and channels, respectively, whereas μ_i and σ_i denote the mean and standard deviation of the C_i channel-activity values. The constant 10^{-6} was added to prevent division by zero. The standardized vector was zero-padded when $C_i < 25$ or truncated when $C_i > 25$, reshaped into a single-channel 5×5 map, and bilinearly resized to 32×32 . This fixed ordered

channel-activity layout was used in the implementation; it is not presented as an anatomical electrode-coordinate interpolation scheme.

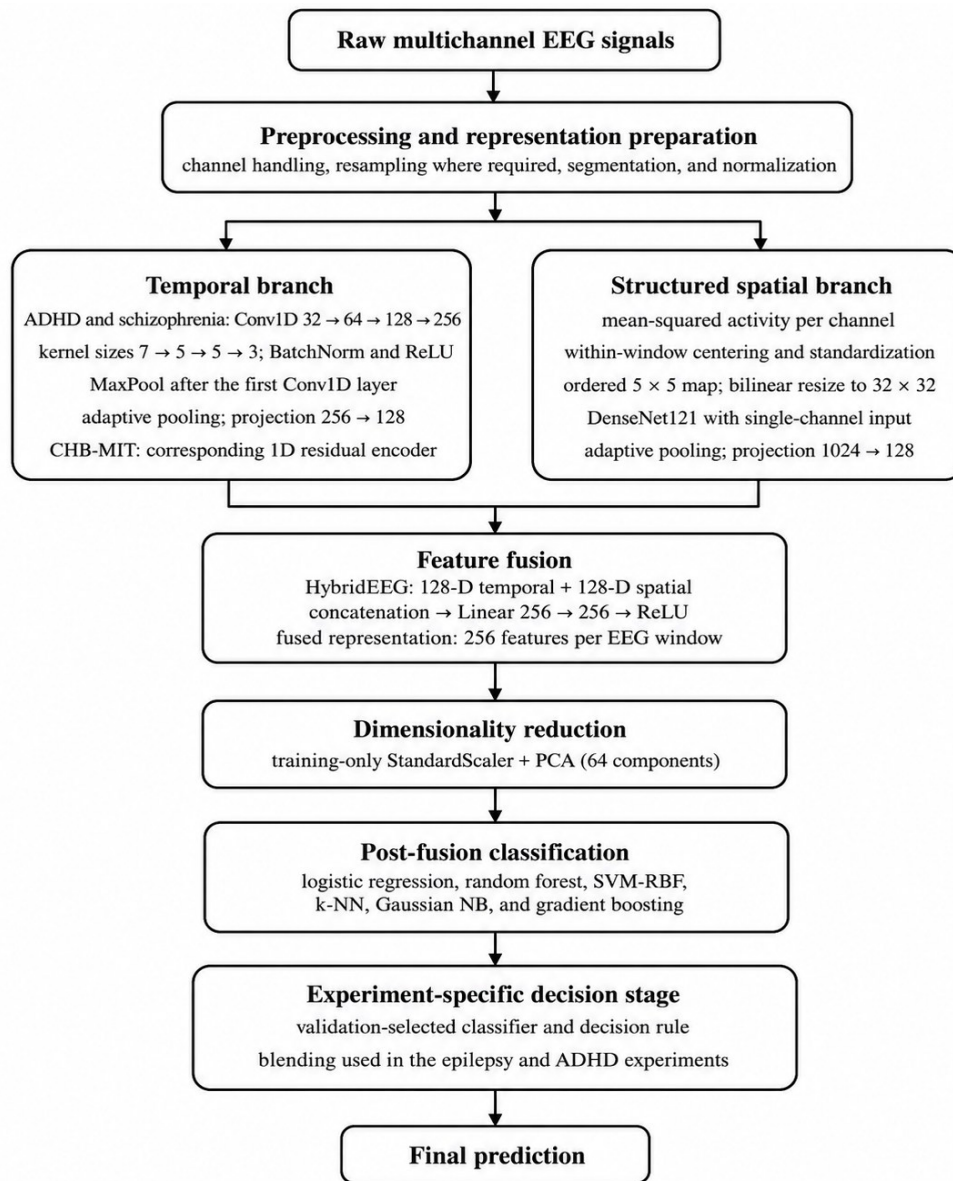


Figure 1. Block diagram and implementation details of the proposed temporal-spatial EEG classification pipeline

The HybridEEG spatial encoder used a DenseNet121 feature extractor initialized without pretrained weights. Its first convolution was modified to accept one input channel and used 64 filters, a 7×7 kernel, stride 2, and padding 3. Following the DenseNet121 feature blocks, ReLU activation and adaptive average pooling produced a 1024-dimensional vector, which was projected to 128 dimensions. The 128-dimensional temporal and spatial vectors were concatenated and passed through a fully connected $256 \rightarrow 256$ layer followed by ReLU activation. The resulting fused representation contained 256 features per EEG window, and the same 256-dimensional feature representation was used as the input to PCA in all disorder-specific experiments.

After feature fusion, the 256-dimensional hybrid vectors were standardized and reduced using PCA with 64 retained components. The StandardScaler and PCA transformations were fitted only on the training data within each split or LOSO fold and were then applied unchanged to the corresponding validation and test partitions. PCA used `n_components = 64`, `svd_solver = auto`, `whiten = False`, and `random_state = 42`. PCA

reduces redundancy by representing high-dimensional data using a smaller set of components [21]. In this study, the retained 64 components provided the same dimensional input to all post-fusion classifiers.

The reduced features were evaluated using six conventional classifiers: logistic regression, random forest, support vector machine with a radial-basis-function kernel, k-nearest neighbors, Gaussian naive Bayes, and gradient boosting. Their implementation settings are summarized in Table 1. Where blending was applied, validation-level or out-of-fold class-probability outputs from the base classifiers were used instead of test predictions. The retained decision model was selected only within the corresponding training/validation procedure. In the epilepsy experiment, the selected blend used meta-logistic regression and a validation-selected decision threshold. In the ADHD experiment, 10-fold out-of-fold base-classifier probabilities were generated, and a random-forest meta-model was selected through the corresponding validation procedure. Accordingly, blending is described as an experiment-specific post-fusion decision stage rather than as universally logistic-regression blending [22][23].

The pipeline was not designed to replace specialized clinical workflows. The study aims to test whether a consistent stage-level EEG representation and post-fusion classification strategy can provide engineering evidence within selected heterogeneous public datasets while maintaining leakage-aware processing. It was definitely not built as a dataset-invariant or universal clinical diagnostic system.

Table 1. Hyperparameter settings for post-fusion classification and blending

Component	Configuration
Logistic regression	StandardScaler; C=1.0; solver=lbfgs; max_iter=2000; class_weight=balanced; random_state=42
Random forest	n_estimators=300; max_depth=None; min_samples_split=2; min_samples_leaf=1; class_weight=balanced; random_state=42; n_jobs=-1
SVM-RBF	StandardScaler; C=1.0; kernel=rbf; gamma=scale; probability=True; class_weight=balanced; random_state=42
k-NN	StandardScaler; n_neighbors=7; weights=distance; metric=minkowski; p=2
Gaussian NB	StandardScaler; var_smoothing=1e-9
Gradient boosting	n_estimators=200; learning_rate=0.05; max_depth=3; random_state=42
Blending input	Validation-level or out-of-fold base-classifier probabilities; no test predictions used for meta-model fitting, classifier selection, or threshold selection
Epilepsy meta-model	Meta-logistic regression; decision threshold selected using validation data
ADHD meta-model	10-fold out-of-fold base-classifier probabilities; random-forest meta-model selected through the corresponding validation procedure
Schizophrenia decision model	Validation-selected post-fusion classifier within each LOSO fold; no test data used for classifier selection

3.2. Classifiers and Experiment-Specific Decision Strategy

After PCA-based dimensionality reduction, the 64-dimensional hybrid feature vectors were evaluated using the same candidate set of six conventional machine-learning classifiers: support vector machine with a radial-basis-function kernel, logistic regression, random forest, k-nearest neighbors, Gaussian Naive Bayes, and gradient boosting. Their implementation settings are summarized in Table 1. These classifiers represent complementary decision mechanisms, including margin-based classification, linear probabilistic modeling, tree-based ensemble learning, instance-based classification, probabilistic generative modeling, and boosting-based classification. The same candidate classifier pool was retained across the evaluated datasets to preserve a consistent post-fusion comparison, whereas the final classifier and decision rule were selected only within the corresponding training and validation procedure.

Where blending was used, the base classifiers generated class-probability outputs either at the validation level or through out-of-fold predictions, and these outputs were used as input features for the meta-model [22][23].

Test labels and test predictions were not used to train the meta-model, select the final classifier, or determine the decision threshold [23]. In the epilepsy experiment, the selected blend used meta-logistic regression with a decision threshold determined from the validation data. In the ADHD experiment, 10-fold out-of-fold base-classifier probabilities were combined using a random-forest meta-model selected through the corresponding validation procedure. For the schizophrenia experiment, the final post-fusion classifier was selected within the training and validation procedure of each LOSO fold. Accordingly, the same PCA-reduced

representation and candidate classifier pool were maintained across the experiments, while the final decision mechanism remained experiment-specific.

From a computational standpoint, the largest part of the processing cost is associated with deep feature extraction in the temporal and spatial branches. The temporal branch processes each multichannel EEG window through successive one-dimensional convolutional operations, while the DenseNet121-based spatial branch is expected to be the more computationally intensive component because of its deeper convolutional structure. After branch-level projection and fusion, PCA reduces the feature dimension from 256 to 64, thus lessening the computational overhead and memory requirements of the subsequent conventional classifiers. The post-fusion classifiers therefore operate on lower-dimensional compact feature vectors rather than on the original EEG windows or high-dimensional deep representations. Exact wall-clock runtime, memory usage, and floating-point operation counts were not benchmarked under a controlled hardware configuration and therefore are not reported here as comparative performance measures.

3.3. Datasets and Validation Settings

We then evaluated the proposed pipeline separately on three public EEG-based disorder datasets representing epilepsy, ADHD, and schizophrenia. Moreover, the three datasets differ widely in channel configuration, sampling conditions, recording duration, subject population, class balance, acquisition processes, and preprocessing requirements. We used 23 EEG channels sampled at 256 Hz for the implemented CHB-MIT inputs, 19 channels for the ADHD recordings, and 19 channels sampled at 250 Hz for the RepOD/Warsaw recordings. No cross-dataset domain-adaptation objective, acquisition harmonization procedure, or common electrode-coordinate interpolation was applied. Consequently, the three datasets were not viewed as a unified clinical cohort; rather, the same stage-level processing pipeline was assessed separately using dataset-specific protocols.

Before feature extraction, dataset-specific window segmentation was conducted. The CHB-MIT recordings were divided into 4096-sample windows with a stride of 2048 samples, representing 16-s windows, an 8-s stride, and 50% overlap. Whenever a CHB-MIT window intersected a seizure annotation, it was assigned to the seizure class. For the ADHD recordings, 256-sample windows with a 128-sample stride were created, corresponding to 50% overlap. The RepOD/Warsaw recordings were cut into 4096-sample windows with a 2048-sample stride, corresponding to 16.384-s windows, an 8.192-s stride, and 50% overlap.

The CHB-MIT scalp EEG database serves as a seizure-window benchmark for epilepsy in accordance with recent studies based upon the CHB-MIT [24]–[26]. The epilepsy experiment reported was performed in the selected intra-subject chb02 setting, with training, validation, and test windows coming from the same subject. Consequently, this experiment assesses within-subject seizure discrimination and does not serve as evidence of cross-subject or full-cohort CHB-MIT generalizability. Recent studies for ADHD included work using the public EEG dataset of 121 participants [27], separate diagnostic and attention-state datasets [28], and a further study on the same 121-participant dataset [29]. In the current study, this 121-participant dataset was partitioned at the subject level into training subjects ($n=84$), validation subjects ($n=18$), and test subjects ($n=19$), with no subject present in more than one of the three partitions. For schizophrenia, the RepOD/Warsaw dataset with 28 participants was used [30]–[32]. This was assessed with 28-fold leave-one-subject-out validation, where one participant was held out for testing during each fold. All training-dependent transformations were fitted using only the corresponding training data, while classifier selection was performed within the corresponding training and validation procedure for each experimental setting.

In addition to these disorder-specific evaluations, an exploratory four-class feature-space analysis was run based on previously extracted 256-dimensional representations. The operational labels were healthy control, epilepsy seizure windows, ADHD, and schizophrenia. We included only seizure-window representations under the epilepsy label, but this choice does not eliminate the differences in acquisition equipment, recording protocols, electrode arrangements, subject populations, and preprocessing conditions among the source datasets. Because the four classes were constructed from heterogeneous public EEG datasets, the analysis might capture dataset-origin or acquisition-related signatures in addition to disorder-related information. Thus, it is regarded only as an exploratory representation-level analysis, rather than evidence of dataset-independent neurological separability or a cross-dataset clinical diagnostic model.

This exploratory analysis uses the class distribution presented in Table 2. The disorder-specific experiments are still the main evidence for classification performance, while the four-class analysis provides only supplementary information about how the extracted representations behave when combined in a shared feature space.

As shown in Figure 2, the internal four-class feature-space experiment starts from the extracted 256-dimensional disorder-specific features, applies training-only standardization and PCA, and then performs class-weighted four-class classification using a model selected from the validation data.

Table 2. Class distribution used in the exploratory internal four-class feature-space analysis.

Class	Total	Train	Validation	Test
Healthy control	11584	8108	1738	1738
Epilepsy seizure windows	46	32	7	7
ADHD	6735	4714	1010	1011
Schizophrenia	1912	1339	287	286

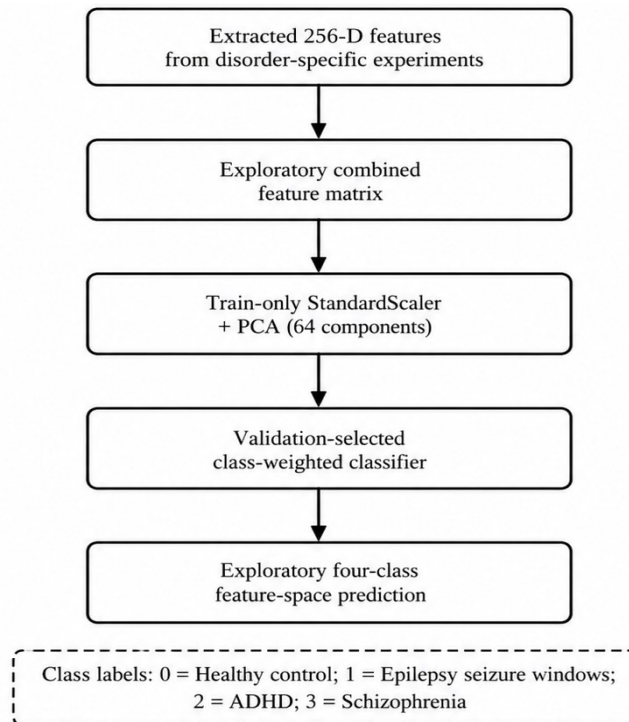


Figure 2. Workflow of the exploratory internal four-class feature-space analysis

3.4. Evaluation Metrics and Reproducibility Controls

Based on the binary or multiclass setting of each experiment, performance was computed using accuracy, macro-F1, weighted-F1, balanced accuracy, sensitivity, specificity, and confusion matrices. Since the evaluated datasets differed in class balance, number of subjects, and validation protocol, we report metrics within each disorder-specific experiment and avoid a single global measure. In the case of the epilepsy experiment, we report seizure-class F1 and macro-F1 separately since they embody different aspects of performance under severe class imbalance. For schizophrenia, accuracy from the aggregated confusion matrix was presented together with fold-averaged macro-F1, allowing overall segment-level performance to be differentiated from performance variability across the 28 LOSO folds. Results from the exploratory four-class feature-space analysis were interpreted solely in the context of that experiment—our data were not presented as evidence either for (i) dataset-independent neurological separability or (ii) clinical generalization.

All transformations needing parameter estimation, such as StandardScaler and PCA, were fitted only on the training data within the corresponding split or LOSO fold [3],[21]. Neither base classifiers nor any meta-models were trained using the final test partition. If a separate validation procedure was performed, validation data were used only for classifier selection and decision-threshold selection. The transformations fitted on the training data and the chosen models were applied unchanged to the test data during final evaluation. The random states were fixed, the split definitions and subject identifiers were saved, and fitted preprocessing objects, saved feature representations, available model checkpoints, prediction outputs, and confusion matrices were retained to support verification of the reported experiments.

4. RESULTS AND DISCUSSION

Notably, the experimental results were reported separately for epilepsy, ADHD, and schizophrenia because the datasets differed in subject population, acquisition conditions, class distribution, and validation protocol. Despite retaining the same stage-level temporal-spatial EEG pipeline across the experiments, the results were interpreted relative to each dataset and evaluation setting. No performance metric was aggregated across the datasets into a single global score. All train-dependent transformations, model selection, blending, and decision-threshold selection were performed without access to the final test data, as detailed in Section 3.4.

4.1. Epilepsy Results

In the selected intra-subject CHB-MIT chb02 experiment, the validation-selected blending model with a logistic regression meta-classifier produced the test confusion matrix $\begin{bmatrix} 9443 & 0 \\ 1 & 6 \end{bmatrix}$. This corresponds to an accuracy of 99.99%, a seizure-class F1 of 92.31%, a sensitivity of 85.71%, and a specificity of 100%. Six of the seven seizure windows were detected correctly, one seizure window was missed, and no non-seizure window was falsely classified as a seizure.

The extremely high accuracy must be interpreted with caution because the test partition was highly imbalanced, with seven seizure windows compared with 9443 non-seizure windows. Consequently, seizure-class F1 and sensitivity offer more informative indicators of seizure-window discrimination than accuracy alone. The macro-F1 derived from the same confusion matrix was approximately 96.15%, which is the unweighted average of the F1 scores of both classes, seizure and non-seizure.

These results provide evidence of intra-subject seizure discrimination only for the selected chb02 experimental setting. They do not demonstrate cross-subject generalization or full-cohort CHB-MIT performance because the training, validation, and test windows all originated from the same subject, and the number of seizure windows in the test partition was very small. Broader multi-subject and external validation remain necessary before drawing more general conclusions.

4.2. ADHD Results

For the subject-wise ADHD evaluation, the validation-selected random-forest meta-model provided the following test confusion matrix: $\begin{bmatrix} 1335 & 5 \\ 29 & 1090 \end{bmatrix}$. Out of 1340 control samples, 1335 were classified correctly and five were misclassified as ADHD samples. Out of the 1119 ADHD samples, 1090 were correctly classified and 29 were misclassified as control. These outcomes correspond to an accuracy of 98.62%, a macro-F1 of approximately 98.60%, an ADHD sensitivity of 97.41%, and a specificity of 99.63%.

The low error rates in both classes reflect the adequate performance of the temporal-spatial representation, PCA-based dimensionality reduction, and experiment-specific post-fusion decision strategy within the held-out subject-separated test partition. As shown in Figure 3, ADHD samples were misclassified as control slightly more often than control samples were classified as ADHD. Thus, the result provides substantial evidence of subject-separated classification performance within this ADHD dataset. As a caveat, the result is interpreted within the established dataset and validation protocol rather than as evidence of independent external or cross-dataset clinical validation.

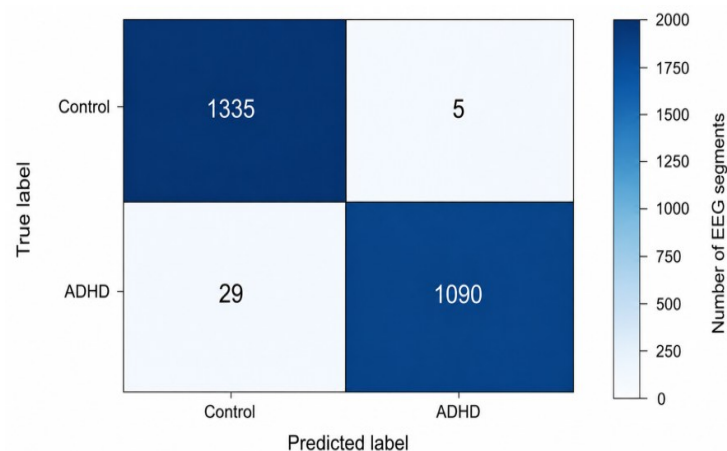


Figure 3. Confusion matrix for the subject-wise ADHD test partition

4.3. Schizophrenia Results

As shown in Fig. 4, the aggregated confusion matrix for the 28-fold LOSO evaluation on the RepOD/Warsaw dataset was $[[1644, 42], [91, 1799]]$, where class 0 indicates healthy control and class 1 indicates schizophrenia. It contains 3443 correct predictions among 3576 EEG segments, yielding an aggregated accuracy of 96.28%. This also corresponds to a schizophrenia sensitivity of 95.19%, a specificity of 97.51%, and an aggregated macro-F1 of around 96.27%.

In contrast, the fold-averaged macro-F1 across the 28 LOSO folds was 73.92%. The difference between the metrics aggregated across all test segments and those averaged over folds suggests that subject-to-subject variability can be obscured when pooling all test segments, especially when the held-out subjects differ substantially in segment count and classification difficulty. Hence, the aggregated confusion matrix characterizes overall segment-level performance, while the fold-averaged macro-F1 provides a more conservative summary of generalization across held-out subjects.

These results imply strong aggregate performance within the analysed RepOD/Warsaw LOSO protocol. However, they do not establish external or cross-dataset clinical generalization. To determine whether this performance is consistent across recording systems and subject populations, additional independent cohorts recorded under distinct acquisition conditions are needed.

The main dataset-specific results are presented in Table 3, where the exploratory four-class feature-space analysis is treated as a separate entry. The exploratory analysis is not considered evidence of dataset-independent neurological separability or cross-dataset clinical diagnosis.

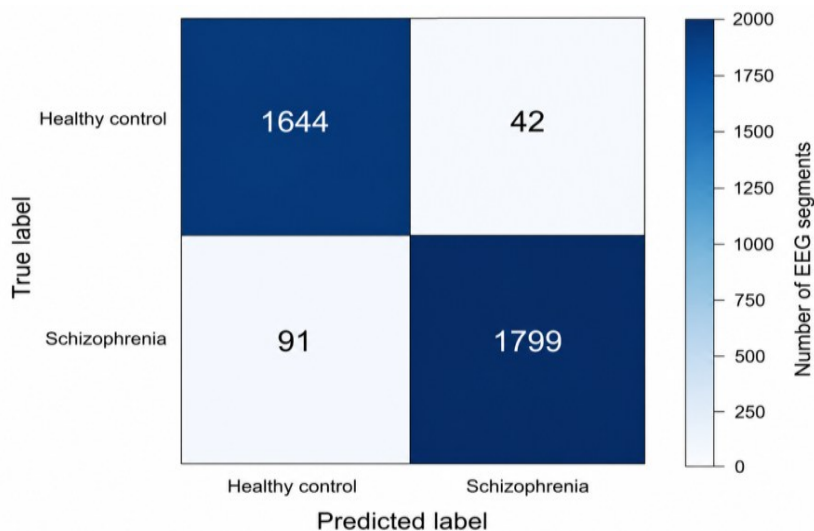


Figure 4. Aggregated confusion matrix across the 28 LOSO folds for the schizophrenia experiment

Table 3. Summary of dataset-specific results and the exploratory four-class analysis

Experiment	Protocol	Main result
Epilepsy	CHB-MIT chb02 intra-subject evaluation	Acc. 99.99%; seizure-class F1 92.31%; macro-F1 96.15%
ADHD	Subject-wise train/validation/test split	Acc. 98.62%; macro-F1 98.60%
Schizophrenia	RepOD/Warsaw 28-fold LOSO evaluation	Aggregated Acc. 96.28%; aggregated macro-F1 96.27%; fold-averaged macro-F1 73.92%
Four-class	Exploratory combined feature-space analysis	Acc. 96.61%; macro-F1 95.47% — exploratory feature-space analysis

4.4. Exploratory Internal Four-Class Feature-Space Results

First, we conducted an exploratory four-class analysis to understand how the previously extracted 256-dimensional representations behaved when combined in a shared feature space. The operational classes were healthy control, epilepsy seizure windows, ADHD, and schizophrenia. After training-only standardization and PCA, a validation accuracy of 96.98% and a validation macro-F1 of 95.71% were obtained using a class-weighted random-forest classifier selected through the validation partition. The performance summary of the classifier on the held-out test partition is presented in Table 4, where accuracy was 96.61%, macro-F1 was 95.47%, and balanced accuracy and weighted-F1 were reported as 95.10% and 96.59%, respectively.

These scores suggest high predictive performance among the four operationally defined labels within the constructed feature-space experiment. Nevertheless, the labels were assembled from several heterogeneous public EEG datasets. For the healthy-control class, we combined 10,014 control-window representations from the ADHD dataset with 1,570 healthy-subject representations from the RepOD/Warsaw dataset, yielding a total of 11,584 samples. The ADHD class originated from the ADHD dataset, while the epilepsy seizure-window and schizophrenia classes were derived from the CHB-MIT and RepOD/Warsaw datasets, respectively. The four-class analysis excluded CHB-MIT non-seizure windows and did not assign them to the healthy-control class. These sources differ in acquisition conditions, electrode configurations, subject populations, class distributions, and preprocessing procedures. Therefore, the observed performance may also exhibit dataset-origin or acquisition-related signatures alongside disorder-related EEG information. Thus, this analysis is interpreted only as an exploratory representation-level experiment and not as evidence of dataset-independent neurological separability or validation of a cross-dataset clinical diagnostic model.

The test confusion matrix is presented in Figure 5. All seven epilepsy seizure windows were classified correctly; however, the very small size of this class prevents this result from being interpreted as strong evidence of generalization. Among the errors, 45 schizophrenia segments were labelled as healthy control, followed by 33 healthy-control segments labelled as schizophrenia. Moreover, 12 healthy-control segments were misclassified as ADHD, and 13 ADHD segments were misclassified as healthy control. Thus, the matrix reflects generally high classification performance within the exploratory setting while revealing class-specific error patterns and limitations associated with the highly unequal class sizes.

Thus, under their respective validation protocols, the disorder-specific epilepsy, ADHD, and schizophrenia evaluations provide the primary evidence. While the four-class analysis offers additional representation-level evidence within the constructed experimental setting, its interpretation must account for the heterogeneity of the contributing datasets.

Table 4. Validation and test metrics for the exploratory four-class analysis.

Metric	Validation	Test
Accuracy	96.98%	96.61%
Macro-F1	95.71%	95.47%
Balanced accuracy	—	95.10%
Weighted-F1	—	96.59%

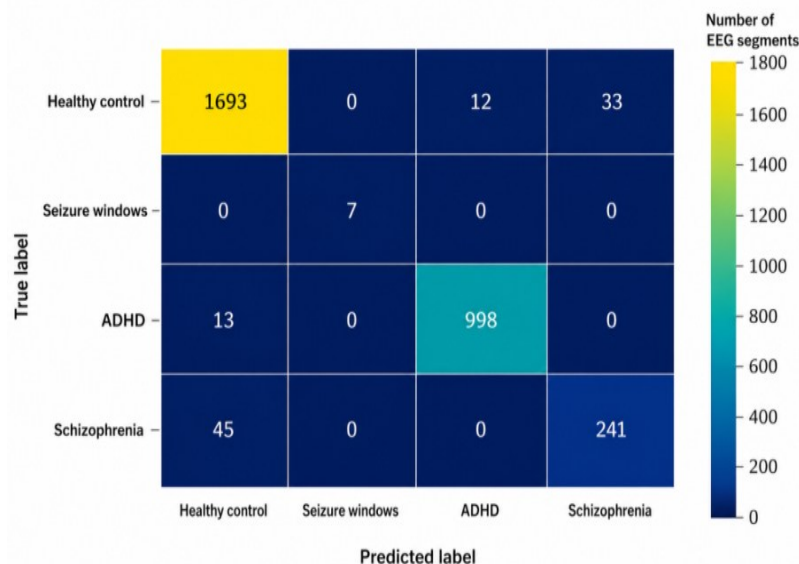


Figure 5. Test confusion matrix for the exploratory internal four-class feature-space analysis

4.5. Dataset-Aligned Comparison and Methodological Considerations

Table 5 provides a dataset-aligned comparison between the proposed experiments and existing EEG-based studies conducted on the same datasets or under closely related evaluation settings [24][25],[27][28],[30],[32]. This type of comparison is meant to provide a descriptive contextual assessment rather than a controlled head-to-head ranking. However, direct numerical comparisons should be treated with

caution because the studies differ in subject composition, segmentation strategy, preprocessing procedure, subject-separation protocol, model configuration, class balance, and metric definition.

The proposed chb02 result is therefore not directly comparable to cross-subject or multi-subject CHB-MIT evaluations reported elsewhere [24][25], since it was obtained using a specific intra-subject protocol for epilepsy. For ADHD, the proposed outcome was obtained using a subject-wise train/validation/test split, making comparison most applicable to studies performed on the same dataset, although differences in validation design must still be considered [27],[29]. The 28-fold LOSO evaluation on the RepOD/Warsaw dataset [30],[32] was used for the proposed schizophrenia experiment. Thus, the aggregated accuracy of 96.28% and the fold-averaged macro-F1 of 73.92% represent different aspects of performance and should not be compared interchangeably with results obtained using conventional cross-validation or a single train/test split.

Recent seizure-based EEG investigations have pursued deep-learning and hybrid architectures for seizure detection and classification [33]–[35]. Attention-based learning, deep models, dataset diversity, and the influence of validation strategy on reported performance have been considered in ADHD-oriented and broader EEG-classification studies [36]–[39]. Other studies on structured EEG representation learning have investigated hybrid temporal–functional features and adaptive graph-convolution approaches [40],[41]. Visibility-graph-based representations, neural-network models, and convolutional frameworks with integrated attention further illustrate alternative approaches to structured feature learning [42]–[44]. Studies of motor-imagery EEG classification have addressed transformer-based and multi-scale methods [45]–[48], whereas subject-independent studies have evaluated representations on participants not used during model development [49][50]. Further recent scholarship on heterogeneous domain adaptation has indicated that device-related and cross-domain heterogeneity in EEG data may require explicit adaptation strategies rather than direct comparison across datasets [51]. Although these studies help to contextualise the methods, their results do not constitute a controlled ranking because their datasets, preprocessing procedures, and evaluation protocols differ.

Table 5. Dataset-aligned comparison with selected EEG studies.

Disorder	Study and protocol	Acc./F1 (%)
Epilepsy	Abdallah <i>et al.</i> [24]; CHB-MIT cross-subject	F1 97.90
Epilepsy	Ravi and Radhakrishnan [25]; CHB-MIT 5-fold CV	Acc. 95.90; F1 95.95
Epilepsy	Proposed; CHB-MIT chb02 intra-subject evaluation	Acc. 99.99; seizure-class F1 92.31
ADHD	Atila <i>et al.</i> [27]; same dataset	Acc. 97.46
ADHD	Alhussen <i>et al.</i> [28]; separate diagnostic and attention-state datasets	Acc. 98.52; F1 98.26
ADHD	Proposed; subject-wise train/validation/test split	Acc. 98.62; macro-F1 98.60
Schizophrenia	Alazzawi <i>et al.</i> [30]; RepOD/Warsaw	Acc. 93.00
Schizophrenia	Latreche <i>et al.</i> [32]; RepOD/Warsaw 10-fold + unseen test	Acc. 99.18
Schizophrenia	Proposed; RepOD/Warsaw 28-fold LOSO evaluation	Aggregated Acc. 96.28; fold-averaged macro-F1 73.92

For each of the evaluated experiments, the same stage-level processing sequence was preserved: temporal and structured spatial representation learning, 256-dimensional feature fusion, training-only PCA reduction to 64 components, and post-fusion classification. Nevertheless, the final classifier and decision rule remained experiment-specific, as described in Section 3.2. The main contribution is therefore the consistent implementation and evaluation of this processing pipeline across selected public EEG datasets under dataset-specific protocols, rather than evidence of a dataset-invariant or universally generalizable clinical model.

Nonetheless, the comparisons should be interpreted with caution given some methodological limitations. The evaluated datasets are heterogeneous in terms of subject populations, channel configurations, acquisition conditions, class distributions, and validation designs. The epilepsy result is limited to the selected intra-subject chb02 experiment, while the exploratory four-class analysis may be influenced by dataset-origin or acquisition-related signatures in addition to disorder-related EEG information. Because the CHB-MIT experiment used overlapping windows within a selected intra-subject setting, its result should be regarded only as an internal seizure-window benchmark and not as evidence of cross-subject generalization. Due to the nature of the present experiments, a systematic ablation study was not conducted; therefore, the contributions of the temporal branch, spatial branch, PCA stage, and experiment-specific decision strategy were not quantified separately. Formal statistical significance testing and confidence interval estimation were not included in the present experiments. Future controlled comparisons should use paired predictions or repeated subject-level resampling within each dataset-specific protocol to determine whether observed differences among classifiers are

statistically reliable. The present experiments did not include model-specific explainability analyses. Future studies should utilize methods such as Grad-CAM on the spatial branch, along with SHAP-based attribution at the post-fusion classification stage, to investigate which spatial patterns and fused-feature inputs contribute most strongly to dataset-specific predictions. Future studies should encompass broader multi-subject epilepsy evaluation, harmonized acquisition and preprocessing settings, independent external cohorts, and prospective validation across recording systems and subject populations.

5. CONCLUSIONS

This paper proposed a hybrid temporal–spatial EEG representation and classification pipeline involving temporal feature learning, structured spatial representation, 256-dimensional feature fusion, training-only PCA reduction to 64 components, and post-fusion machine-learning classification. The same stage-level processing sequence was consistently applied across selected public EEG datasets, while dataset-specific input dimensions, validation protocols, classifiers, and decision rules were retained where required. This design provides a reproducible engineering framework for evaluating temporal–spatial EEG representations under different disorder-specific experimental settings.

The results showed strong classification performance under the corresponding validation protocols. For the subject-wise ADHD assessment, the accuracy and macro-F1 of the validation-selected random-forest meta-model were 98.62% and 98.60%, respectively. Using the aggregated confusion matrix, the accuracy was 96.28% and the aggregated macro-F1 was approximately 96.27%, while the fold-averaged macro-F1 across the 28 LOSO folds was 73.92%. These complementary metrics indicate strong overall segment-level performance while also revealing variation across subjects. In the selected intra-subject CHB-MIT chb02 experiment, the seizure-class F1 was 92.31%, providing encouraging evidence of seizure-window discrimination within the evaluated subject-specific setting.

This exploratory four-class feature-space analysis further showed that the extracted representations retained considerable predictive information when combined within the constructed shared feature framework. Since the operational labels were assembled from heterogeneous public EEG sources, the observed performance may include dataset-origin or acquisition-related signatures in addition to disorder-related EEG information. Consequently, this analysis is presented as supplementary representation-level evidence and as a useful foundation for future harmonized cross-dataset studies, rather than as definitive evidence of dataset-independent clinical discrimination.

The main contribution of this work is the consistent implementation and evaluation of a fixed temporal–spatial EEG processing sequence across selected dataset-specific experiments. The strong results obtained under the corresponding validation protocols support the engineering feasibility and methodological value of the proposed framework. Before practical deployment, the framework should also be evaluated for inference latency, memory requirements, computational cost, robustness to variations in EEG signal quality, and compatibility with heterogeneous EEG acquisition systems. Future studies may extend this foundation through broader multi-subject epilepsy evaluation, harmonized acquisition and preprocessing settings, independent external cohorts, and prospective validation across recording systems and subject populations.

DECLARATION

Supplementary Materials

Not applicable.

Sustainable Development Goals

This work is related to Industry, Innovation and Infrastructure (SDG 9) through the development of reproducible biomedical signal processing methods.

Author Contribution

Conceptualization, Dawood S. Hasan and Qais Al-Gayem; methodology, Dawood S. Hasan; software, Dawood S. Hasan; validation, Dawood S. Hasan, Qais Al-Gayem, and Hilal Al-Libawy; formal analysis, Dawood S. Hasan; writing—original draft preparation, Dawood S. Hasan; writing—review and editing, Dawood S. Hasan, Qais Al-Gayem, and Hilal Al-Libawy; supervision, Qais Al-Gayem and Hilal Al-Libawy. All authors read and approved the final paper.

Funding

This research received no external funding.

Acknowledgement

The authors acknowledge the providers of the public EEG datasets used in this study.

Conflicts of Interest

The authors declare no conflict of interest.

ABBREVIATIONS

The following abbreviations are used in this manuscript.

EEG	:	Electroencephalography
PCA	:	Principal Component Analysis
ADHD	:	Attention-Deficit/Hyperactivity Disorder
LOSO	:	Leave-One-Subject-Out
CNN	:	Convolutional Neural Network
SVM	:	Support Vector Machine
RBF	:	Radial Basis Function
LR	:	Logistic Regression
RF	:	Random Forest
k-NN	:	k-Nearest Neighbors
GNB	:	Gaussian Naive Bayes
GB	:	Gradient Boosting
F1	:	F1-score
CHB-MIT	:	Children's Hospital Boston–Massachusetts Institute of Technology
RepOD	:	Repository for Open Data
SDG	:	Sustainable Development Goal
ML	:	Machine Learning
DL	:	Deep Learning
ReLU	:	Rectified Linear Unit
CV	:	Cross-Validation

REFERENCES

- [1] R. Lyu, "Deep learning approaches for EEG-based healthcare applications: a comprehensive review," *Frontiers in Human Neuroscience*, vol. 19, p. 1689073, 2026, <https://doi.org/10.3389/fnhum.2025.1689073>.
- [2] J. Guttag, "CHB-MIT Scalp EEG Database," *PhysioNet*, Version 1.0.0, Jun. 2010, <https://doi.org/10.13026/C2K01R>.
- [3] G. Brookshire, J. Kasper, N. M. Blaich, Y. C. Wu, R. Glatt, D. A. Merrill, S. Gerrol, K. J. Yoder, C. Quirk, and C. Lucero, "Data leakage in deep learning studies of translational EEG," *Frontiers in Neuroscience*, vol. 18, p. 1373515, 2024, <https://doi.org/10.3389/fnins.2024.1373515>.
- [4] M. Feng and J. Xu, "Electroencephalogram-based ConvMixer architecture for recognizing attention deficit hyperactivity disorder in children," *Brain Sciences*, vol. 14, no. 5, p. 469, 2024, <https://doi.org/10.3390/brainsci14050469>.
- [5] L. Li, X. Guo, Z. Yang, Y. Zhao, X. Liu, J. Yang, Y. Chen, X. Peng, and L. Han, "ADHD detection from EEG signals using GCN based on multi-domain features," *Frontiers in Neuroscience*, vol. 19, p. 1561994, 2025, <https://doi.org/10.3389/fnins.2025.1561994>.
- [6] J. Rahul, D. Sharma, L. D. Sharma, U. Nanda, and A. K. Sarkar, "A systematic review of EEG based automated schizophrenia classification through machine learning and deep learning," *Frontiers in Human Neuroscience*, vol. 18, p. 1347082, 2024, <https://doi.org/10.3389/fnhum.2024.1347082>.
- [7] R. Mourad, A. Diab, Z. Merhi, M. Khalil, and R. Le Bouquin Jeannes, "Machine and deep learning methods for epileptic seizure recognition using EEG data: A systematic review," *Brain Research*, vol. 1864, p. 149797, 2025, <https://doi.org/10.1016/j.brainres.2025.149797>.
- [8] K. Slama, A. Yahyaouy, J. Riffi, M. A. Mahraz, and H. Tairi, "Comprehensive review of machine learning and deep learning techniques for epileptic seizure detection and prediction based on neuroimaging modalities," *Visual Computing for Industry, Biomedicine, and Art*, vol. 8, p. 27, 2025, <https://doi.org/10.1186/s42492-025-00208-8>.
- [9] Y. Lin, L. Dong, Y. Jiang, and J. Lian, "Epileptic EEG classification via deep learning-based strange attractor," *Biomedical Signal Processing and Control*, vol. 100, p. 106965, 2025, <https://doi.org/10.1016/j.bspc.2024.106965>.
- [10] J. Su, Z. Huang, Y. Ma, H. Shi, Y. Yang, M. Xi, and B. Li, "A model for epileptic EEG detection and recognition based on Multi-Attention mechanism and Spatiotemporal," *Scientific Reports*, vol. 15, p. 31993, 2025, <https://doi.org/10.1038/s41598-025-17256-y>.

- [11] O. K. Cura, A. Akan, and S. K. Atli, "Detection of Attention Deficit Hyperactivity Disorder based on EEG feature maps and deep learning," *Biocybernetics and Biomedical Engineering*, vol. 44, no. 3, pp. 450-460, 2024, <https://doi.org/10.1016/j.bbe.2024.07.003>.
- [12] Y. Mao, X. Qi, L. He, S. Wang, Z. Wang, and F. Wang, "Advanced machine learning techniques reveal multidimensional EEG abnormalities in children with ADHD: a framework for automatic diagnosis," *Frontiers in Psychiatry*, vol. 16, p. 1475936, 2025, <https://doi.org/10.3389/fpsyt.2025.1475936>.
- [13] E. Sathiyar, T. D. Rao, and T. S. Kumar, "Gabor filter-based statistical features for ADHD detection," *Frontiers in Human Neuroscience*, vol. 18, p. 1369862, 2024, <https://doi.org/10.3389/fnhum.2024.1369862>.
- [14] H. Li, C. Wang, L. Ma, C. Xu, and H. Li, "EEG analysis in patients with schizophrenia based on microstate semantic modeling method," *Frontiers in Human Neuroscience*, vol. 18, p. 1372985, 2024, <https://doi.org/10.3389/fnhum.2024.1372985>.
- [15] A. Al Mazroa, M. M. Eltahir, S. A. Ebad, F. A. Alotaibi, K. Venkatachalam, and J. Cho, "EEG-based schizophrenia diagnosis using deep learning with multi-scale and adaptive feature selection," *PeerJ Computer Science*, vol. 11, p. e2811, 2025, <https://doi.org/10.7717/peerj-cs.2811>.
- [16] W. Mai, F. Wu, and X. Mai, "Learning spatial-spectral-temporal EEG representations with dual-stream neural networks for motor imagery," *Biomedical Signal Processing and Control*, vol. 92, p. 106003, 2024, <https://doi.org/10.1016/j.bspc.2024.106003>.
- [17] Y. Song, Q. Zheng, B. Liu, and X. Gao, "EEG Conformer: convolutional transformer for EEG decoding and visualization," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 31, pp. 710-719, 2023, <https://doi.org/10.1109/TNSRE.2022.3230250>.
- [18] Y. Ding, Y. Li, H. Sun, R. Liu, C. Tong, C. Liu, X. Zhou, and C. Guan, "EEG-Deformer: a dense convolutional transformer for brain-computer interfaces," *IEEE Journal of Biomedical and Health Informatics*, vol. 29, no. 3, pp. 1909-1918, 2025, <https://doi.org/10.1109/JBHI.2024.3504604>.
- [19] A. Keutayeva, N. Fakhrutdinov, and B. Abibullaev, "Compact convolutional transformer for subject-independent motor imagery EEG-based BCIs," *Scientific Reports*, vol. 14, p. 25775, 2024, <https://doi.org/10.1038/s41598-024-73755-4>.
- [20] J. Zhang, K. Li, B. Yang, and X. Han, "Local and global convolutional transformer-based motor imagery EEG classification," *Frontiers in Neuroscience*, vol. 17, p. 1219988, 2023, <https://doi.org/10.3389/fnins.2023.1219988>.
- [21] W. Jia, M. Sun, J. Lian, and S. Hou, "Feature dimensionality reduction: a review," *Complex & Intelligent Systems*, vol. 8, pp. 2663-2693, 2022, <https://doi.org/10.1007/s40747-021-00637-x>.
- [22] A. Mohammed and R. Kora, "A comprehensive review on ensemble deep learning: Opportunities and challenges," *Journal of King Saud University - Computer and Information Sciences*, vol. 35, no. 2, pp. 757-774, 2023, <https://doi.org/10.1016/j.jksuci.2023.01.014>.
- [23] A. Fanizzi *et al.*, "Developing an ensemble machine learning study: insights from a multi-center proof-of-concept study," *PLOS ONE*, vol. 19, no. 9, p. e0303217, 2024, <https://doi.org/10.1371/journal.pone.0303217>.
- [24] T. Abdallah, N. Jrad, F. Abdallah, A. Humeau-Heurtier, and P. Van Bogaert, "A self-attention model for cross-subject seizure detection," *Computers in Biology and Medicine*, vol. 165, p. 107427, 2023, <https://doi.org/10.1016/j.combiomed.2023.107427>.
- [25] S. Ravi and A. Radhakrishnan, "A hybrid 1D CNN-BiLSTM model for epileptic seizure detection using multichannel EEG feature fusion," *Biomedical Physics & Engineering Express*, vol. 10, no. 3, p. 035040, 2024, <https://doi.org/10.1088/2057-1976/ad3afd>.
- [26] J. Zhang, S. Zheng, W. Chen, G. Du, Q. Fu, and H. Jiang, "A scheme combining feature fusion and hybrid deep learning models for epileptic seizure detection and prediction," *Scientific Reports*, vol. 14, p. 16916, 2024, <https://doi.org/10.1038/s41598-024-67855-4>.
- [27] O. Atila, E. Deniz, A. Ari, A. Sengur, S. Chakraborty, P. D. Barua, and U. R. Acharya, "LSGP-USFNet: Automated attention deficit hyperactivity disorder detection using locations of Sophie Germain's primes on Ulam's spiral-based features with electroencephalogram signals," *Sensors*, vol. 23, no. 16, p. 7032, 2023, <https://doi.org/10.3390/s23167032>.
- [28] A. Alhussen, A. I. Alotaibi, S. K. Sharma, A. R. Khan, F. Ahmad, and G. G. Tejani, "Early attention-deficit/hyperactivity disorder (ADHD) with NeuroDCT-ICA and rhinofish optimization (RFO) algorithm based optimized ADHD-AttentionNet," *Scientific Reports*, vol. 15, p. 6967, 2025, <https://doi.org/10.1038/s41598-025-90649-1>.
- [29] J. Bansal, G. Gangwar, M. Aljaidi, A. Alkoradees, and G. Singh, "EEG-based ADHD classification using autoencoder feature extraction and ResNet with double augmented attention mechanism," *Brain Sciences*, vol. 15, no. 1, p. 95, 2025, <https://doi.org/10.3390/brainsci15010095>.
- [30] A. Alazzawi, S. Aljumaili, A. D. Duru, O. N. Ucan, O. Bayat, P. J. Coelho, and I. M. Pires, "Schizophrenia diagnosis based on diverse epoch size resting-state EEG using machine learning," *PeerJ Computer Science*, vol. 10, p. e2170, 2024, <https://doi.org/10.7717/peerj-cs.2170>.
- [31] A. Al-Azzawi and O. N. Ucan, "Investigate schizophrenia classification based on EEG electrode reduction using machine learning techniques," *Traitement du Signal*, vol. 42, no. 3, pp. 1707-1720, 2025, <https://doi.org/10.18280/ts.420340>.

- [32] I. Latreche, S. Slatnia, O. Kazar, S. Harous, and M. A. Khelili, "Identification and diagnosis of schizophrenia based on multichannel EEG and CNN deep learning model," *Schizophrenia Research*, vol. 271, pp. 28-35, 2024, <https://doi.org/10.1016/j.schres.2024.07.015>.
- [33] M. Alarfaj, M. A. Zeb, M. H. Al-Adhaileh, A. Alhamadi, and N. Ebrahim, "Deep learning approaches for diagnosing seizure based on EEG signal analysis," *Frontiers in Human Neuroscience*, vol. 19, p. 1669919, 2025, <https://doi.org/10.3389/fnhum.2025.1669919>.
- [34] H. Kashefi Amiri, M. Zarei, and M. R. Daliri, "Epileptic seizure detection from electroencephalogram signals based on 1D CNN-LSTM deep learning model using discrete wavelet transform," *Scientific Reports*, vol. 15, p. 32820, 2025, <https://doi.org/10.1038/s41598-025-18479-9>.
- [35] L. Bai, G. Litscher, and X. Li, "Epileptic seizure detection using machine learning: a systematic review and meta-analysis," *Brain Sciences*, vol. 15, no. 6, p. 634, 2025, <https://doi.org/10.3390/brainsci15060634>.
- [36] N. Chugh, S. Aggarwal, and A. Balyan, "The hybrid deep learning model for identification of attention-deficit/hyperactivity disorder using EEG," *Clinical EEG and Neuroscience*, vol. 55, no. 1, pp. 22-33, 2024, <https://doi.org/10.1177/15500594231193511>.
- [37] S. K. Khare and U. R. Acharya, "An explainable and interpretable model for attention deficit hyperactivity disorder in children using EEG signals," *Computers in Biology and Medicine*, vol. 155, p. 106676, 2023, <https://doi.org/10.1016/j.combiomed.2023.106676>.
- [38] A. Belhadi, A. Yazidi, P. G. Lind, and Y. Djenouri, "EEG data classification: review and taxonomy," *ACM Transactions on Computing for Healthcare*, vol. 6, no. 4, p. 45, pp. 1-29, 2025, <https://doi.org/10.1145/3742795>.
- [39] S. Khalfallah, W. Puech, M. Tlija, and K. Bouallegue, "Exploring the effectiveness of machine learning and deep learning techniques for EEG signal classification in neurological disorders," *IEEE Access*, vol. 13, pp. 17002-17015, 2025, <https://doi.org/10.1109/ACCESS.2025.3532515>.
- [40] M. Soltani-Nejad, F. Salar-pour, S. A. Rakhshan, and H. Nezamabadi-pour, "Enhanced hybrid deep neural network for EEG-based schizophrenia diagnosis using functional and temporal features," *Scientific Reports*, vol. 15, p. 42592, 2025, <https://doi.org/10.1038/s41598-025-26627-4>.
- [41] G. Yin, J. Yuan, Y. Chen, G. Guo, D. Shi, L. Wang, Z. Zhao, Y. Zhao, M. Zhang, Y. Dong, B. Wang, and S. Tan, "Schizophrenia recognition based on three-dimensional adaptive graph convolutional neural network," *Scientific Reports*, vol. 15, p. 4067, 2025, <https://doi.org/10.1038/s41598-024-84497-8>.
- [42] A. Das, S. Singh, J. Kim, T. A. Ahanger, and A. A. Pise, "Enhanced EEG signal classification in brain computer interfaces using hybrid deep learning models," *Scientific Reports*, vol. 15, p. 27161, 2025, <https://doi.org/10.1038/s41598-025-07427-2>.
- [43] A. Belhadi, P. G. Lind, Y. Djenouri, and A. Yazidi, "Enhanced visibility graph for EEG classification," *Frontiers in Neuroscience*, vol. 19, p. 1541062, 2025, <https://doi.org/10.3389/fnins.2025.1541062>.
- [44] A. S. Jangde and G. K. Verma, "EEG-based schizophrenia classification using attention-integrated deep convolutional networks," *Psychiatry Research: Neuroimaging*, vol. 357, p. 112138, 2026, <https://doi.org/10.1016/j.psychresns.2026.112138>.
- [45] M. Liu, Y. Liu, W. Shi, Y. Lou, Y. Sun, Q. Meng, D. Wang, F. Xu, Y. Zhang, L. Zhang, and J. Leng, "EMPT: a sparsity Transformer for EEG-based motor imagery recognition," *Frontiers in Neuroscience*, vol. 18, p. 1366294, 2024, <https://doi.org/10.3389/fnins.2024.1366294>.
- [46] W. Zhao, X. Jiang, B. Zhang, S. Xiao, and S. Weng, "CTNet: a convolutional transformer network for EEG-based motor imagery classification," *Scientific Reports*, vol. 14, p. 20237, 2024, <https://doi.org/10.1038/s41598-024-71118-7>.
- [47] H. Altaheri, F. Karray, and A.-H. Karimi, "Temporal convolutional transformer for EEG based motor imagery decoding," *Scientific Reports*, vol. 15, p. 32959, 2025, <https://doi.org/10.1038/s41598-025-16219-7>.
- [48] W. Zhao, B. Zhang, H. Zhou, D. Wei, C. Huang, and Q. Lan, "Multi-scale convolutional transformer network for motor imagery brain-computer interface," *Scientific Reports*, vol. 15, p. 12935, 2025, <https://doi.org/10.1038/s41598-025-96611-5>.
- [49] J. Luo, Y. Wang, S. Xia, N. Lu, X. Ren, Z. Shi, and X. Hei, "A shallow mirror transformer for subject-independent motor imagery BCI," *Computers in Biology and Medicine*, vol. 164, p. 107254, 2023, <https://doi.org/10.1016/j.combiomed.2023.107254>.
- [50] S. Sartipi and M. Cetin, "Subject-independent deep architecture for EEG-based motor imagery classification," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 32, pp. 718-727, 2024, <https://doi.org/10.1109/TNSRE.2024.3360194>.
- [51] H. Wu, Q. Xie, Z. Yu, J. Zhang, S. Liu, and J. Long, "Unsupervised heterogeneous domain adaptation for EEG classification," *Journal of Neural Engineering*, vol. 21, no. 4, p. 046018, 2024, <https://doi.org/10.1088/1741-2552/ad5fbd>.