

Hybrid Deep Learning for EEG-Based Intelligence Prediction: Combining PSD, MSE, and CNN-GRU

Mohammed Alrowaili^{1*}, Mohammed J. Alhaddad^{2*}, Mutasem Jarrah³, Saim Rasheed^{4*}

Department of Information Technology-Faculty of Computing and Information Technology,

King Abdulaziz University, Jeddah, Saudi Arabia^{1,2,4}

Department of Computer Science-Faculty of Information Technology,

Applied Science Private University, Amman, Jordan³

Abstract—Electroencephalography (EEG) signals provide a non-invasive means of examining patterns of brain activity associated with cognitive processing. This study investigates whether spectral and complexity-based EEG characteristics can support the classification of intelligence quotient (IQ) levels using a hybrid deep-learning framework. EEG data were obtained from a publicly available dataset comprising 104 children aged 7–13 years who performed reading or mathematics tasks. Following data inspection, valid recordings from 90 participants were retained, yielding 212 task-level EEG samples for subsequent analysis. Because multiple task recordings could originate from the same participant, participant identity rather than individual task samples was considered when separating data for model evaluation. The EEG recordings were preprocessed and characterized using two complementary feature representations: Power Spectral Density (PSD), which describes frequency-domain activity, and Multiscale Entropy (MSE), which captures signal complexity across multiple scales. These representations were processed through a hybrid CNN-GRU architecture and used to classify participants into Low, Average, and High IQ categories. The proposed model achieved an overall classification accuracy of 95% in the reported evaluation and outperformed the conventional machine-learning and shallow neural-network models considered in the study. Class-specific precision, recall, and F1-score were also examined because of the unequal distribution of samples across the three IQ categories. The findings indicate that combining spectral and complexity-based EEG information may provide useful discriminative information for IQ-category classification within the investigated dataset. However, the relatively small number of independent participants, the pronounced class imbalance, the task-specific nature of the recordings, and the absence of external validation limit the extent to which these findings can be generalized. Further evaluation using larger and more diverse independent datasets is therefore needed to determine the robustness and broader applicability of the proposed framework.

Keywords—Electroencephalogram (EEG); intelligence prediction; Power Spectral Density (PSD); Multiscale Entropy (MSE); deep Learning; CNN-GRU; Brain-Computer Interface (BCI); signal processing

I. INTRODUCTION

Electroencephalography (EEG) is a non-invasive neurophysiological technique widely used to record brain activity with high temporal resolution. Because EEG directly reflects neural dynamics at the millisecond scale, it has become an important tool for investigating the neural mechanisms

underlying cognitive processes and brain function. In recent years, advances in signal processing and machine learning have enabled researchers to analyze complex EEG patterns and explore their relationship with higher-level cognitive abilities, including intelligence and cognitive performance [13], [3], [11].

Human intelligence is commonly assessed using psychometric instruments such as intelligence quotient (IQ) tests. However, behavioral assessments alone provide limited insight into the neurophysiological mechanisms that support cognitive performance. Neuroscientific research suggests that intelligence is associated with both structural and functional brain characteristics. One influential framework explaining this relationship is the Parieto-Frontal Integration Theory (P-FIT), which proposes that efficient communication between parietal and frontal brain regions plays a key role in complex reasoning and problem solving [2].

EEG provides a promising modality for investigating these neural mechanisms because it captures oscillatory activity across multiple frequency bands [12], [14]. Among the most widely used EEG features is Power Spectral Density (PSD), which quantifies the distribution of signal power across frequency bands such as delta, theta, alpha, beta, and gamma. Variations in spectral power have been associated with cognitive processes including attention, working memory, and information processing efficiency [15]. Consequently, PSD-based features have been extensively used in EEG-based cognitive assessment and classification tasks.

Beyond spectral features, recent research has emphasized nonlinear measures that capture the complexity of neural signals. Multiscale Entropy (MSE) is a widely used method for quantifying temporal irregularity across multiple time scales in physiological time series. Higher entropy values are often interpreted as reflecting richer and more flexible neural dynamics that may support efficient information processing. Several studies have demonstrated that EEG complexity measures can be associated with cognitive performance and intelligence-related metrics [5], [6].

In parallel with advances in feature extraction, deep learning techniques have increasingly been applied to EEG signal analysis. Convolutional Neural Networks (CNNs) have demonstrated strong capability in learning spatial-spectral representations from multi-channel EEG signals, while recurrent architectures such as Gated Recurrent Units (GRUs) are effective

*Corresponding author.

in modeling temporal dependencies in sequential data. Recent studies have shown that deep learning approaches can automatically extract meaningful EEG representations and improve classification performance in cognitive and brain-computer interface applications [11], [3], [7], [8].

Despite these developments, relatively few studies have investigated the integration of spectral and complexity-based EEG features within a unified deep learning framework for intelligence prediction. In particular, the combined use of PSD and Multiscale Entropy features for classifying intelligence levels remains largely unexplored. To address this gap, the present study proposes a hybrid CNN-GRU architecture that integrates PSD and MSE features extracted from task-based EEG recordings. By leveraging complementary spectral and temporal complexity information, the proposed framework aims to improve the classification of individuals into multiple IQ levels.

II. RELATED WORK

Research investigating the relationship between electroencephalography (EEG) and cognitive abilities has grown considerably over the past decade. Early studies mainly focused on spectral features derived from EEG signals, particularly power spectral density (PSD), to identify neural correlates of cognitive processes. Variations in spectral power across frequency bands such as theta and alpha have been associated with working memory, attention, and general cognitive performance [15], [12]. These findings suggest that oscillatory brain activity provides valuable information about individual differences in cognitive functioning.

Beyond traditional spectral features, recent studies have explored nonlinear measures that capture the complexity of neural signals. Multiscale Entropy (MSE) has emerged as an important metric for quantifying temporal irregularity and information processing capacity in brain activity. Lewandowska et al. [5] applied multivariate multiscale entropy (mMSE) to resting-state EEG recordings and identified spatial-temporal patterns of neural complexity across functional brain networks. Similarly, Thiele et al. [6] demonstrated that multiple EEG complexity measures, including entropy-based features, can significantly predict individual differences in fluid intelligence in adult populations.

Recent research has also highlighted the importance of EEG signal complexity in understanding cognitive performance. For instance, Wan et al. [16] demonstrated that resting-state EEG complexity metrics can serve as predictors of cognitive performance, indicating that higher neural signal complexity may reflect more efficient information processing in the brain. These findings emphasize that cognitive ability may be reflected not only in local neural complexity but also in large-scale brain network interactions.

With the rapid advancement of artificial intelligence, deep learning methods have become increasingly popular for EEG analysis. Recent studies have also explored advanced deep learning architectures for EEG analysis, demonstrating improved capability in modeling complex neural dynamics and cognitive states [1].

Convolutional neural networks (CNNs) have demonstrated strong capability in extracting spatial and spectral represen-

tations from multi-channel EEG signals, while recurrent neural networks (RNNs) and gated recurrent units (GRUs) are effective in modeling temporal dependencies in neural time series [7], [8]. Recent studies have further explored deep learning frameworks for EEG-based cognitive state detection and mental workload estimation [11], [3].

More recent works have also investigated hybrid architectures and advanced representation learning approaches for EEG signals. For example, Zhang et al. [10] proposed a masked autoencoder-based pretraining framework to learn generalizable spatiotemporal representations from large-scale EEG data, while Song et al. [9] employed dynamical graph convolutional neural networks to model relationships between EEG channels for emotion recognition. Although these approaches demonstrate promising performance in EEG analysis tasks, relatively few studies have explored the integration of spectral and entropy-based EEG features within a unified deep learning framework for intelligence prediction.

Hybrid AI frameworks combine complementary techniques to improve system performance. For example, Aloqla et al. [22] integrated lexical retrieval, dense semantic scoring, and evidence-based verification to improve question answering in high-sensitivity domains. This approach highlights the potential of combining complementary methods, motivating the integration of spectral and nonlinear EEG features within a unified deep learning framework.

Therefore, despite the growing body of research linking EEG features to cognitive performance, the combined use of PSD and multiscale entropy features within hybrid deep learning models remains insufficiently explored. To address this gap, the present study proposes a hybrid CNN-GRU framework that integrates PSD and MSE features extracted from task-based EEG recordings to classify individuals into different intelligence levels (Table I).

III. DATASET

The dataset used in this study is publicly available and contains EEG recordings collected from 104 children aged 7–13 years who were evaluated for learning difficulties. The participants were divided into two groups according to the type of learning difficulty reported in the original dataset. The reading-difficulty (RD) group consisted of 54 participants, whereas the mathematics-difficulty (MD) group included 50 participants. Each participant belonged exclusively to one of these two groups.

The experimental protocol differed according to the participant group. Children in the RD group performed three reading-related tasks, denoted as R1, R2, and R3, while those in the MD group performed two mathematics-related tasks, denoted as M1 and M2. EEG signals were recorded during task performance using 32 scalp electrodes positioned according to the international 10–20 system. The recordings were sampled at 256 Hz. These task-based recordings provided the EEG signals used for the subsequent preprocessing and feature-extraction stages.

Although the original dataset included 104 participants, valid EEG recordings suitable for the present analysis were available for 90 participants after data inspection. Therefore,

TABLE I. COMPARATIVE SUMMARY OF RECENT EEG-BASED STUDIES RELATED TO COGNITIVE ABILITY AND INTELLIGENCE PREDICTION. THE FINAL ROW PRESENTS THE PROPOSED HYBRID CNN-GRU FRAMEWORK

Study	Features	Method	Data/Task	Key Findings
Jung and Haier, 2007 [2]	Neuroimaging connectivity	P-FIT theoretical framework	Multiple neuroimaging datasets	Proposed the parieto-frontal integration model explaining the neural basis of intelligence.
Dubois et al., 2018 [4]	Functional connectivity	Predictive modeling	Resting-state fMRI (Human Connectome Project, N=884)	Distributed resting-state brain networks significantly predicted general intelligence across individuals.
Lewandowska et al., 2023 [5]	Multivariate Multiscale Entropy (mMSE)	Entropy-based complexity analysis	Resting EEG (N=95)	Identified spatial-temporal complexity patterns across distributed brain networks.
Thiele et al., 2023 [6]	EEG complexity metrics	Multimodal entropy modeling	Resting EEG (N=144)	EEG signal complexity measures significantly predicted fluid intelligence.
Schirmmeister et al., 2017 [7]	Raw EEG signals	Deep CNN architecture	Multiple EEG datasets	CNN models effectively learned spatial-spectral EEG representations.
Lawhern et al., 2018 [8]	Raw EEG signals	Compact CNN architecture (EEGNet)	BCI EEG datasets	Efficient CNN architecture for EEG classification.
Proposed Method	PSD + Multiscale Entropy (MSE)	Hybrid CNN-GRU deep learning	Task-based EEG (Reading + Math)	Achieved high accuracy in IQ level classification.

only recordings from these 90 participants were retained for subsequent preprocessing, feature extraction, and classification. The present study did not treat the number of available task recordings as equivalent to the number of independent participants. This distinction is particularly important because individual participants could contribute EEG recordings from more than one task.

Following the selection of valid recordings, a total of 212 task-level EEG samples were available for analysis. A task-level sample represents an EEG recording associated with a completed reading or mathematics task rather than an independent participant. Consequently, multiple task-level samples could originate from the same participant. The 212 samples should therefore be interpreted as repeated task observations obtained from 90 independent participants rather than as 212 independent subjects.

IQ information was provided at the participant level in the original dataset. Accordingly, all task-level EEG samples originating from the same participant were assigned the same IQ category. The IQ categories used in this study were Low, Average, and High, following the labeling information available in the source dataset. Because the IQ label is participant-specific while EEG recordings are task-specific, the resulting data have a hierarchical structure in which several task-level observations may correspond to the same individual.

This structure was taken into consideration when preparing the data for model evaluation. Participant identity, rather than the individual task-level sample, was used as the basis for data separation. Thus, EEG samples originating from the same participant were kept within the same data partition and were not distributed across both training and testing sets. This subject-wise separation was adopted to reduce the possibility of information leakage arising from participant-specific EEG characteristics and to provide a more appropriate assessment of model performance on unseen participants.

The demographic information available in the dataset was limited. Age was available for the participants, whereas other potentially relevant demographic characteristics were not consistently available for all records. Therefore, the analysis was primarily based on the available EEG recordings, task information, learning-difficulty group, and IQ labels.

The restricted demographic information, together with the specific age range and learning-difficulty characteristics of the participants, should be considered when interpreting the generalizability of the findings.

The dataset used in this study is publicly available through the Mendeley Data repository [17].

A. Samples and Labeling

Following the selection of participants with valid EEG recordings, the available task recordings were organized into task-level samples for subsequent analysis. A total of 212 samples were obtained from the 90 retained participants. Each sample corresponds to an EEG recording collected during a completed reading or mathematics task. Because participants performed more than one task, multiple samples could originate from the same individual. Therefore, the 212 samples represent repeated task-level observations rather than 212 independent participants.

IQ information was available at the participant level in the original dataset. Accordingly, the IQ label assigned to a participant was applied consistently to all task-level EEG samples belonging to that individual. Participants were grouped into three IQ categories: Low, Average, and High. The thresholds used to define these categories followed the labeling scheme adopted for the dataset and were applied consistently throughout the analysis.

After organizing the recordings and assigning the corresponding IQ labels, the resulting dataset contained 131 samples in the Low-IQ category, 72 samples in the Average-IQ category, and 9 samples in the High-IQ category. This distribution indicates a clear imbalance among the three classes, particularly because the High-IQ category is represented by relatively few task-level samples. For this reason, overall classification accuracy should not be considered in isolation when evaluating model performance. Class-specific precision, recall, and F1-score were also considered to provide a more informative assessment of performance across the three IQ categories.

An important consideration in organizing the dataset was the presence of multiple task-level samples from the same participant. Treating these samples as fully independent observations during data partitioning could introduce participant-

level information leakage and produce overly optimistic performance estimates. To reduce this risk, participant identity was used as the unit of data separation. All task-level samples associated with a given participant were therefore kept within the same data partition, ensuring that recordings from the same individual were not simultaneously included in both the training and testing data.

This subject-wise organization preserves the distinction between the number of independent participants and the number of available task-level observations. It also provides a more appropriate basis for evaluating whether the proposed model can generalize to EEG recordings from participants who were not represented during model training (Fig. 1). Further details regarding the training, validation, and testing procedure are provided in the experimental evaluation section.

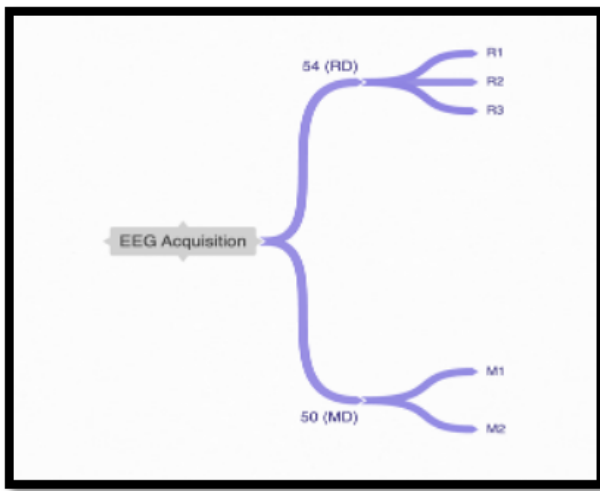


Fig. 1. Distribution of participants in the EEG dataset: 54 children with reading difficulties (RD) across three reading tasks (R1–R3) and 50 children with math difficulties (MD) across two math tasks (M1–M2).

IV. METHODOLOGY

A. EEG Data Acquisition

EEG signals were acquired from 32 scalp electrodes following the international 10–20 system at 256 Hz [17]. Participants were split into two non-overlapping groups: 54 children with reading difficulties (R1–R3 tasks) and 50 with math difficulties (M1–M2 tasks); each participant belonged to a single group only. All recordings included a resting baseline. Ethical approvals and written consent were obtained according to institutional guidelines.

B. Psychometric Assessment (IQ Measurement)

The Intelligence Quotient (IQ) scores used in this study were obtained from the original dataset, where standardized psychometric evaluations were conducted. Specifically, IQ was estimated using the Wechsler Intelligence Scale for Children – Fourth Edition (WISC-IV), based on selected subtests including Vocabulary and Block Design, which are considered sufficient for reliable IQ estimation.

TABLE II. SUMMARY OF EEG PREPROCESSING STEPS APPLIED TO ENHANCE SIGNAL QUALITY AND REDUCE ARTIFACTS

Step	Specification
Band-pass filtering	FIR filter (1–40 Hz), order $\approx 2 \times$ sampling rate, with zero-phase filtering.
Notch filtering	60 Hz (50 Hz where applicable) to suppress power-line interference.
Artifact inspection	Independent Component Analysis (ICA) using the Infomax algorithm; no components were removed.
Re-referencing	Common Average Reference (CAR).

The resulting IQ scores were then categorized into three groups (Low, Average, and High) according to the classification provided in the dataset annotations. Participant demographic information was partially available in the dataset. All participants were children aged between 7 and 13 years, representing a developmental stage relevant to cognitive and learning processes.

Gender information was provided in categorical form (Male/Female), although the exact distribution of male and female participants was not explicitly specified in the dataset.

C. Preprocessing

EEG preprocessing was performed using standard procedures to enhance signal quality and reduce noise. A finite impulse response (FIR) band-pass filter (1–40 Hz, order $\approx 2 \times$ sampling rate) was applied to retain physiologically relevant frequency components. A 60 Hz notch filter was used to suppress power-line interference. Independent Component Analysis (ICA) using the Infomax algorithm was then applied to decompose the EEG signals into statistically independent components for artifact inspection. No ICA components were removed, and all components were retained to preserve the original signal structure. Finally, the EEG signals were re-referenced using the Common Average Reference (CAR) method (Table II).

D. Epoching (Segmentation)

Continuous EEG recordings were segmented into epochs time-locked to stimulus onset using windows of -1 to $+2$ s. Baseline correction was applied by subtracting the mean pre-stimulus activity. No automated threshold-based rejection was performed at this stage, as artifact removal had already been addressed during preprocessing (filtering and ICA).

E. Feature Extraction

1) *Feature relevance for cognitive EEG analysis:* The selection of appropriate EEG features plays a critical role in modeling cognitive processes and brain dynamics. Spectral features such as Power Spectral Density (PSD) have been widely used in cognitive neuroscience because they reflect oscillatory brain activity associated with attention, working memory, and information processing efficiency. Previous studies have demonstrated that variations in EEG spectral power, particularly within the theta and alpha frequency bands, are strongly linked to cognitive performance and neural efficiency [12], [14].

In addition to spectral measures, nonlinear features have been increasingly employed to capture the complexity of neural signals. Multiscale Entropy (MSE) provides a quantitative measure of temporal irregularity across multiple time scales and has been shown to reflect the richness of neural information processing. Higher entropy values are often associated with greater neural adaptability and cognitive flexibility [5].

By combining PSD and MSE features, the present study aims to capture complementary aspects of EEG dynamics, including both oscillatory activity and signal complexity. This combination enables the proposed model to represent multiple dimensions of brain activity, which may contribute to improved classification of intelligence levels.

2) *Power Spectral Density (PSD)*: PSD was computed with Welch's method (Hamming windows, 50% overlap, FFT length matched to the sampling rate). Band powers were integrated over Delta (1–4 Hz), Theta (4–8 Hz), Alpha (8–13 Hz), Beta (13–30 Hz), and Gamma (30–40 Hz) for each of the 32 channels. Features included absolute and relative band power (band/total 1–40 Hz), yielding $5 \times 32 = 160$ PSD features per epoch.

3) *Multiscale Entropy (MSE)*: Multiscale Entropy (MSE) quantifies the complexity of a time series across multiple temporal scales. For a given time series $\{x_1, x_2, \dots, x_N\}$, coarse-grained time series at scale factor τ are constructed by averaging non-overlapping segments:

$$y_j^{(\tau)} = \frac{1}{\tau} \sum_{i=(j-1)\tau+1}^{j\tau} x_i, \quad 1 \leq j \leq \left\lfloor \frac{N}{\tau} \right\rfloor. \quad (1)$$

Sample Entropy (SampEn) is then computed for each coarse-grained time series:

$$\text{SampEn}(m, r, N) = -\ln \left(\frac{A}{B} \right), \quad (2)$$

where, m is the embedding dimension, r is the tolerance threshold, A is the number of template vector pairs of length $m + 1$ within tolerance r , and B is the number of template vector pairs of length m within tolerance r .

The MSE value at scale τ corresponds to the Sample Entropy computed from $y^{(\tau)}$. In this study, entropy features were computed based on Sample Entropy. The embedding dimension (m) was set to 2, and the tolerance (r) was defined as 0.5 times the standard deviation of the signal.

F. Feature Integration

PSD and MSE features were z-score normalised (per feature across the training set) and merged into a unified feature space. Each sample was labelled with the participant's psychometric IQ score. IQ scores were categorized into three levels based on standardized thresholds: Low (< 90), Average ($90 - 109$), and High (≥ 110). Since IQ is measured at the subject level, all samples originating from the same participant shared the same label. Subject identifiers were retained to enforce strict subject-wise data separation in all evaluations.

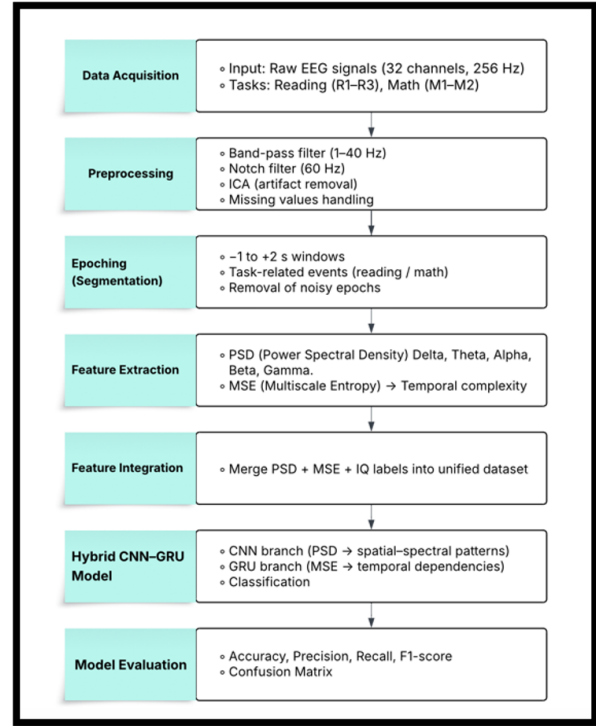


Fig. 2. Pipeline of the proposed EEG-based IQ prediction framework, summarizing the steps from EEG acquisition, preprocessing, and epoching, through feature extraction (PSD and MSE), feature integration, hybrid CNN-GRU classification, and final evaluation.

G. Hybrid CNN-GRU Architecture

To leverage complementary spectral and temporal information, we designed a dual-branch architecture (Fig. 3). The PSD branch receives a 1-D vector of 160 features (5 bands $\times$ 32 channels) reshaped as a sequence and is processed by two Conv1D blocks: Conv1D (kernel = 3, filters = 32, 'same') $\rightarrow$ BatchNorm $\rightarrow$ ReLU; followed by Conv1D (kernel = 3, filters = 64, 'same') $\rightarrow$ BatchNorm $\rightarrow$ ReLU. Global average pooling yields a compact PSD embedding. The MSE branch receives a sequence of 40 time steps with 32 features per step (scales $\times$ channels) and is processed by a GRU layer (64 units, OutputMode = 'last') to capture temporal dependencies. The branch embeddings are concatenated and passed to a fully connected layer (64 units) with dropout ($p = 0.3$), followed by a softmax layer for three-class prediction.

- Optimizer: Adam, learning rate = 10^{-3} (schedule: constant).
- Epochs: 50. Mini-batch size: 64.
- Regularization: Dropout $p = 0.3$, L2 weight decay.

H. Evaluation Metrics

Let TP_c , FP_c , and FN_c denote the true positives, false positives, and false negatives for class c (Fig. 2).

The overall accuracy is defined as:

$$Accuracy = \frac{\text{Number of correctly classified samples}}{\text{Total number of samples}}$$

For each class c , precision, recall, and F1-score are computed as:

$$Precision_c = \frac{TP_c}{TP_c + FP_c} \quad (3)$$

$$Recall_c = \frac{TP_c}{TP_c + FN_c} \quad (4)$$

$$F1_c = \frac{2 \cdot Precision_c \cdot Recall_c}{Precision_c + Recall_c} \quad (5)$$

Macro-F1 computes the unweighted mean of class-wise F1 scores:

$$Macro-F1 = \frac{1}{C} \sum_{c=1}^C F1_c \quad (6)$$

Weighted-F1 accounts for class imbalance by weighting each class by its support n_c :

$$Weighted-F1 = \frac{\sum_{c=1}^C n_c \cdot F1_c}{\sum_{c=1}^C n_c} \quad (7)$$

Given the class imbalance (131 Low vs. 9 High samples), Macro-F1 prevents dominance by the majority class, whereas Weighted-F1 reflects overall performance while accounting for class support.

I. Baselines

We implemented four baselines using the same PSD+MSE feature set and subject-wise splits: SVM (RBF), Decision Tree, k-NN, and a shallow MLP. Hyperparameters were tuned by inner 3-fold cross-validation on the training set.

V. RESULTS

After completing preprocessing and segmentation, the cleaned EEG data were successfully matched with the participants' psychometric IQ scores. The extracted PSD and MSE features were then used as input to the proposed hybrid CNN-GRU model. The CNN-GRU framework for three-level IQ classification. The CNN-GRU model demonstrated strong performance, achieving a test accuracy of 95%. Fig. 4 shows the training and validation accuracy and loss curves across 50 epochs, indicating stable convergence and minimal overfitting.

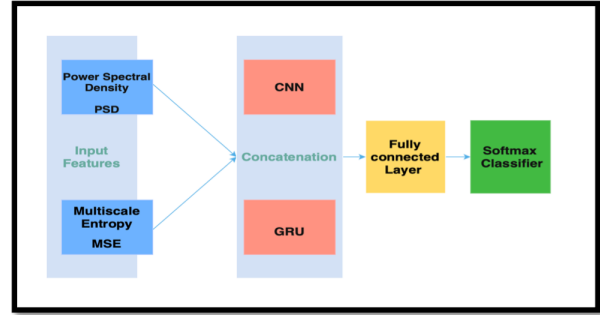


Fig. 3. Architecture of the proposed hybrid CNN-GRU model. PSD features are processed through convolutional layers to capture spatial-spectral representations, while MSE features are modeled by GRU layers to capture temporal dependencies. The outputs of both branches are concatenated and passed to a fully connected layer followed by a softmax classifier for final IQ level prediction.

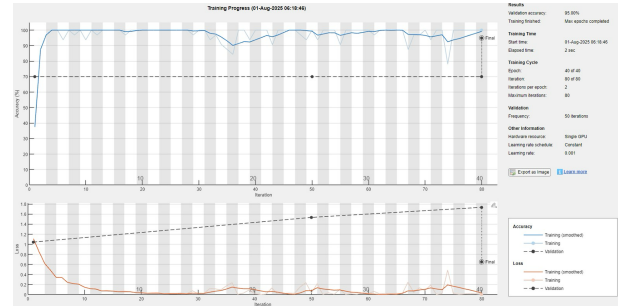


Fig. 4. Training and validation accuracy and loss curves of the proposed CNN-GRU model, showing stable convergence and high classification performance

A. Confusion Matrix Analysis

To further examine the classification performance, a confusion matrix was constructed. The model showed high recall for the Low IQ group, highlighting its sensitivity in detecting this class. Some overlap was observed between the Average and High IQ groups, which is expected given their closer cognitive performance ranges. The detailed precision, recall, and F1-scores are summarized in Table III.

TABLE III. CLASSIFICATION PERFORMANCE OF THE CNN-GRU MODEL ACROSS THREE IQ GROUPS, REPORTING PRECISION, RECALL, F1-SCORE, AND CLASS SUPPORT

Class	Precision	Recall	F1-score	Support
Low IQ	0.96	0.94	0.95	$n = 131$
Average IQ	0.92	0.90	0.91	$n = 72$
High IQ	0.89	0.93	0.91	$n = 9$
Macro F1	—	—	0.92	—
Weighted F1	—	—	0.94	—

B. Comparison with Baseline Models

For benchmarking, the proposed model was compared against several traditional machine learning methods applied on the same dataset. As presented in Table IV, the CNN-GRU framework clearly outperformed all baselines, achieving 95% accuracy compared to a maximum of 82% with shallow MLPs.

TABLE IV. ACCURACY COMPARISON BETWEEN THE PROPOSED CNN-GRU MODEL AND BASELINE MACHINE LEARNING METHODS APPLIED TO THE SAME DATASET

Model	Accuracy (%)
Support Vector Machine (SVM)	74
Decision Tree	68
K-Nearest Neighbors (KNN)	71
MLP (shallow)	82
CNN-GRU (proposed)	95

Overall, these results confirm the effectiveness of combining PSD and MSE features within a hybrid CNN-GRU framework. The integration of spectral and temporal-complexity information enabled the model to capture complementary aspects of brain dynamics, leading to superior classification of IQ levels compared to traditional machine learning approaches.

VI. DISCUSSION

A. Interpretation of Findings

The results obtained in this study demonstrate that the proposed CNN-GRU model can effectively learn discriminative patterns from EEG signals for intelligence-level classification. The achieved performance suggests that combining spectral and nonlinear EEG features provides complementary information about neural dynamics associated with cognitive ability.

Recent studies have increasingly emphasized the effectiveness of deep learning models for analyzing EEG signals in cognitive and neurophysiological research. For example, Bashivan et al. demonstrated that deep recurrent-convolutional neural networks can learn meaningful spatial-spectral representations from multi-channel EEG signals, enabling improved decoding of brain activity patterns [18]. Similarly, Roy et al. presented a comprehensive review showing that deep learning approaches significantly improve EEG-based classification performance compared with traditional machine learning models [3].

More recent work has also explored hierarchical feature learning for EEG analysis. Zhang et al. reported that deep neural networks can effectively extract multilevel representations from EEG signals, leading to improved performance in cognitive-state recognition tasks [1]. Recent reviews have emphasized that integrating multiple EEG feature representations, including spectral and nonlinear measures, can enhance the robustness of machine learning models for cognitive-state prediction [19].

Recent advances have also explored transformer-based and hybrid architectures for EEG analysis, further demonstrating the effectiveness of deep neural networks in modeling complex brain signals and cognitive states [11].

For example, Zhang et al. [10] proposed a masked autoencoder-based pretraining framework for learning generalizable spatiotemporal EEG representations, while Song et al. [9] used dynamical graph convolutional neural networks to capture spatial relationships between EEG channels.

These findings are consistent with the results obtained in the present study, where the integration of PSD and multiscale entropy features allowed the model to capture complementary aspects of neural dynamics related to cognitive ability.

B. Relation to Previous Work

Prior studies using traditional machine learning classifiers, such as Support Vector Machines (SVM) and k-Nearest Neighbors (KNN), have reported moderate classification accuracies ranging between 65% and 80% in EEG-based cognitive assessment and brain-computer interface tasks [20], [21].

More recently, deep learning approaches have been introduced for EEG analysis, including convolutional neural networks (CNNs) and recurrent neural networks (RNNs). These models have demonstrated improved performance in learning spatial-spectral and temporal representations of EEG signals, although reported accuracies in many experimental settings often remain below 90% [11], [3].

By combining CNN and GRU architectures, our study demonstrates that a hybrid temporal-spatial framework can achieve strong performance within the evaluated experimental setting by jointly modeling spectral features and temporal signal complexity.

This study presents novelty in two main aspects:

- **Feature Integration:** Few works have systematically combined PSD and MSE features within a deep learning model for intelligence prediction. This dual representation captures both frequency-specific oscillatory dynamics and temporal complexity of brain signals.
- **Hybrid Architecture:** The CNN-GRU framework effectively bridges spatial filtering and temporal modeling, allowing the network to learn more robust representations compared to single-branch models.

C. Limitations

Several limitations should be considered when interpreting the findings of this study. First, although 212 task-level EEG samples were available for analysis, these observations originated from only 90 independent participants. Therefore, the effective sample size for assessing model generalizability is determined primarily by the number of participants rather than by the number of task-level observations. This relatively limited participant sample may increase model variance and the risk of overfitting, particularly when using deep-learning models. Consequently, the reported performance should be interpreted cautiously and requires confirmation in larger independent cohorts.

Second, the distribution of IQ categories was substantially imbalanced, with considerably fewer samples in the High-IQ category than in the Low- and Average-IQ categories. This imbalance increases the uncertainty associated with performance estimates for the minority class and may limit the reliability of conclusions drawn for High-IQ participants. Accordingly, overall classification accuracy should be interpreted together with class-specific measures such as precision, recall, and F1-score, as well as imbalance-sensitive evaluation metrics.

Third, the dataset consisted of children aged 7–13 years who were evaluated for learning difficulties and belonged to either the reading-difficulty or mathematics-difficulty group. Participants in these groups performed different task paradigms;

therefore, some of the EEG patterns learned by the model may reflect task- or group-related differences in addition to differences associated with IQ categories. Because the present study is a secondary analysis of an existing dataset, these experimental factors could not be fully controlled. The findings should therefore not be directly generalized to other age groups, typically developing populations, resting-state EEG, or substantially different cognitive tasks.

Fourth, the IQ labels used in this study were derived from the psychometric information provided with the original dataset and were based on selected WISC-IV subtests. Therefore, these labels should be interpreted within the context of the original dataset rather than as an independently administered comprehensive full-scale IQ assessment conducted as part of the present study. This limitation should be considered when interpreting associations between the extracted EEG characteristics and the reported IQ categories.

D. Future Work

Future research should address the following directions: Apply cross-validation and external validation datasets to ensure generalizability. Explore data augmentation methods (e.g., sliding windows, synthetic EEG generation via GANs) to overcome limited sample size. Investigate additional EEG biomarkers such as functional connectivity (PLV, coherence) and nonlinear dynamics (dPLI, HFD) to enrich feature representation. Extend classification to regression-based IQ prediction, where the model estimates continuous IQ scores rather than categorical levels.

VII. CONCLUSION

This study presented a hybrid CNN-GRU model for EEG-based IQ classification, integrating Power Spectral Density (PSD) and Multiscale Entropy (MSE) features. Using a dataset of 90 participants and 212 cleaned epochs from reading and math tasks, the model achieved an accuracy of 95%, outperforming classical machine learning baselines. The findings highlight the effectiveness of combining frequency-domain and complexity-based EEG features with a temporal-spatial deep learning framework.

Importantly, the results suggest that EEG signals can serve as objective neurophysiological markers of intelligence, complementing traditional psychometric assessments. While the dataset size and class imbalance remain limitations, the proposed approach provides a promising step toward novel, non-invasive methods for cognitive assessment. Future work should focus on validating these results across larger and more diverse datasets, as well as extending the model to continuous IQ prediction.

FUNDING

The project was funded by KAU Endowment (WAQF) at King Abdulaziz University, Jeddah, Saudi Arabia. The authors, therefore, acknowledge with thanks WAQF and the Deanship of Scientific Research (DSR) for technical and financial support.

INSTITUTIONAL REVIEW BOARD STATEMENT

The dataset analyzed in this study was obtained from a publicly available repository (EBRAINS / Mendeley Data; <https://doi.org/10.17632/bnfgm5jbb.4>). According to the original dataset publication, the experimental protocol was approved by the Neuroscience Institute of University of Guadalajara Ethics Committee (Approval Code: ET122022-356, Approval Date: 28 April 2023).

INFORMED CONSENT STATEMENT

The data used in this study were obtained from a publicly available dataset. The original data collection, including informed consent from participants, was conducted by the dataset authors. Therefore, additional informed consent was not required for this study.

DATA AVAILABILITY STATEMENT

The dataset used in this study is publicly available from the EBRAINS / Mendeley Data repository: <https://doi.org/10.17632/bnfgm5jbb.4>

CONFLICTS OF INTEREST

The authors declare no conflict of interest.

REFERENCES

- [1] Zhang, X.; Yao, L.; Wang, X.; Monaghan, J.; McAlpine, D.; Zhang, Y. A survey on deep learning-based non-invasive brain signals: recent advances and new frontiers. *Journal of Neural Engineering* **2021**, *18*, 031002. <https://doi.org/10.1088/1741-2552/abc902>.
- [2] Jung, R.E.; Haier, R.J. The Parieto-Frontal Integration Theory (P-FIT) of Intelligence: Converging Neuroimaging Evidence. *Behavioral and Brain Sciences* **2007**, *30*, 135–154. <https://doi.org/10.1017/S0140525X07001185>.
- [3] Roy, Y.; Banville, H.; Albuquerque, I.; Gramfort, A.; Falk, T.; Faubert, J. Deep learning-based electroencephalography analysis: A systematic review. *Journal of Neural Engineering* **2019**, *16*, 051001. <https://doi.org/10.1088/1741-2552/ab260c>.
- [4] Dubois, J.; Galdi, P.; Han, Y.; Paul, L.; Adolphs, R. A distributed brain network predicts general intelligence from resting-state human neuroimaging data. *Philosophical Transactions of the Royal Society B* **2018**, *373*, 20170284. <https://doi.org/10.1098/rstb.2017.0284>.
- [5] Lewandowska, M.; Tolpa, K.; Rogala, J.; et al. Multivariate multiscale entropy (mMSE) as a tool for understanding the resting-state EEG signal dynamics. *Behavioral and Brain Functions* **2023**, *19*, 18. <https://doi.org/10.1186/s12993-023-00218-7>.
- [6] Thiele, J.; Richter, A.; Hilger, K. Multimodal Brain Signal Complexity Predicts Human Intelligence. *eNeuro* **2023**, *10*, ENEURO.0345-22.2022. <https://doi.org/10.1523/ENEURO.0345-22.2022>.
- [7] Schirrmester, R.T.; Springenberg, J.T.; Fiederer, L.; Glasstetter, M.; Eggensperger, K.; Tangermann, M.; Hutter, F.; Burgard, W.; Ball, T. Deep learning with convolutional neural networks for EEG decoding and visualization. *Human Brain Mapping* **2017**, *38*, 5391–5420. <https://doi.org/10.1002/hbm.23730>.
- [8] Lawhern, V.J.; Solon, A.J.; Waytowich, N.R.; Gordon, S.M.; Hung, C.P.; Lance, B.J. EEGNet: A compact convolutional neural network for EEG-based brain-computer interfaces. *Journal of Neural Engineering* **2018**, *15*, 056013. <https://doi.org/10.1088/1741-2552/aace8c>.
- [9] Song, T.; Zheng, W.; Song, P.; Cui, Z. EEG emotion recognition using dynamical graph convolutional neural networks. *IEEE Transactions on Affective Computing* **2020**, *11*, 532–541. <https://doi.org/10.1109/TAFFC.2018.2817622>.

- [10] Zhang, Y.; Liu, X.; Wang, Z.; Chen, L.; Li, H. DMAE-EEG: A pretraining framework for EEG spatiotemporal representation learning. *IEEE Transactions on Neural Networks and Learning Systems* **2025**, *36*, 17664–17678. <https://doi.org/10.1109/TNNLS.2025.3581991>.
- [11] Craik, A.; He, Y.; Contreras-Vidal, J.L. Deep learning for electroencephalogram classification tasks: A review. *Journal of Neural Engineering* **2019**, *16*, 031001. <https://doi.org/10.1088/1741-2552/ab0ab5>.
- [12] Klimesch, W. EEG alpha and theta oscillations reflect cognitive and memory performance. *Brain Research Reviews* **1999**, *29*, 169–195. [https://doi.org/10.1016/S0165-0173\(98\)00056-3](https://doi.org/10.1016/S0165-0173(98)00056-3).
- [13] Deary, I.J. *Intelligence: A Very Short Introduction*; Oxford University Press: Oxford, UK, 2020.
- [14] Basar, E.; Basar-Eroglu, C.; Karakas, S.; Schürmann, M. Brain oscillations in perception and memory. *International Journal of Psychophysiology* **2000**, *35*, 95–124. [https://doi.org/10.1016/S0167-8760\(99\)00047-1](https://doi.org/10.1016/S0167-8760(99)00047-1).
- [15] Santarnecchi, E.; Emmendorfer, A.; Tadayon, S.; Rossi, S.; Rossi, A.; Pascual-Leone, A. Network connectivity correlates of variability in fluid intelligence performance. *Intelligence* **2017**, *65*, 35–47. <https://doi.org/10.1016/j.intell.2017.10.002>.
- [16] Wan, W.; Gao, Z.; Zhang, Q.; Gu, Z.; Chang, C.; Peng, C.K.; Cui, X. Resting state EEG complexity as a predictor of cognitive performance. *Physica A* **2023**, *624*, 128952. <https://doi.org/10.1016/j.physa.2023.128952>.
- [17] Corona-González, C.; De Stefano-Ramos, C.; Ramos-Flores, M.; Alonso-Valerdi, L.; Ibarra-Zárate, D.; Gómez-Velázquez, F. EEG data and psychometric results from children with learning difficulties. *Mendeley Data* **2024**. <https://doi.org/10.17632/bnfgm5jbb.4>.
- [18] Bashivan, P.; Rish, I.; Yeasin, M.; Codella, N. Learning representations from EEG with deep recurrent-convolutional neural networks. *arXiv* **2016**. <https://doi.org/10.48550/arXiv.1511.06448>.
- [19] Hossain, K.M.; Islam, M.A.; Hossain, S.; Nijholt, A.; Ahad, M.A.R. Status of deep learning for EEG-based brain–computer interface applications. *Frontiers in Computational Neuroscience* **2022**, *16*, 1006763. <https://doi.org/10.3389/fncom.2022.1006763>.
- [20] Lotte, F.; Congedo, M.; Lecuyer, A.; Lamarche, F.; Arnaldi, B. A review of classification algorithms for EEG-based brain–computer interfaces. *Journal of Neural Engineering* **2007**, *4*, R1–R13. <https://doi.org/10.1088/1741-2560/4/2/R01>.
- [21] Blankertz, B.; Lemm, S.; Treder, M.; Haufe, S.; Müller, K.R. Single-trial analysis and classification of ERP components—A tutorial. *NeuroImage* **2011**, *56*, 814–825. <https://doi.org/10.1016/j.neuroimage.2010.06.048>.
- [22] Aloqla, A.M.; Salama, R.; Alghamdi, W.; Abi Sen, A.A. Enhanced Hybrid Retrieval-Augmented Model for Question Answering in High-Sensitivity Domains. *Future Internet* **2026**, *18*, 434. <https://doi.org/10.3390/fi18080434>.