





Article

A Deep Learning Approach to Classifying User Performance in BCI Gaming

Aimilia Ntetska , Anastasia Mimou, Katerina D. Tzimourta , Pantelis Angelidis  and Markos G. Tsipouras 

Department of Electrical and Computer Engineering, University of Western Macedonia, 50100 Kozani, Greece; ece01515@uowm.gr (A.M.); ktzimourta@uowm.gr (K.D.T.); paggelidis@uowm.gr (P.A.)

* Correspondence: a.ntetska@uowm.gr (A.N.); mailto:mtsipouras@uowm.gr (M.G.T.)

Abstract

Brain–Computer Interface (BCI) systems are rapidly evolving and increasingly integrated into interactive environments such as gaming and Virtual/Augmented Reality. In such applications, user adaptability and engagement are critical. This study applies deep learning to predict user performance in a 3D BCI-controlled game using pre-game Motor Imagery (MI) electroencephalographic (EEG) recordings. A total of 72 EEG recordings were collected from 36 participants, 17 using the Muse 2 headset and 19 using the Emotiv Insight device, during left and right hand MI tasks. The signals were preprocessed and transformed into time–frequency spectrograms, which served as inputs to a custom convolutional neural network (CNN) designed to classify users into three performance levels: low, medium, and high. The model achieved classification accuracies of 83% and 95% on Muse 2 and Emotiv Insight data, respectively, at the epoch level, and 75% and 84% at the subject level, using LOSO-CV. These findings demonstrate the feasibility of using deep learning on MI EEG data to forecast user performance in BCI gaming, enabling adaptive systems that enhance both usability and user experience.

Keywords: Brain-Computer- Interface (BCI); Motor Imagery (MI); electroencephalography (EEG); deep learning; convolutional neural networks (CNN); user performance classification



Academic Editor: George A. Tsihrintzis

Received: 7 October 2025

Revised: 8 December 2025

Accepted: 13 December 2025

Published: 18 December 2025

Citation: Ntetska, A.; Mimou, A.; Tzimourta, K.D.; Angelidis, P.; Tsipouras, M.G. A Deep Learning Approach to Classifying User Performance in BCI Gaming. *Electronics* **2025**, *14*, 4974. <https://doi.org/10.3390/electronics14244974>

Copyright: © 2025 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

A Brain-Computer -Interface (BCI) is a direct neural communication pathway between the brain and an external electronic device [1]. BCIs are typically regarded as a specialized branch of human–computer interaction. Over the past decade, research in the field has primarily focused on supporting patients with motor impairments or disabilities [2]. However, with recent technological advances and the growing availability of commercial electroencephalography (EEG) devices, BCIs have increasingly expanded into areas such as gaming [3,4].

BCIs are commonly categorized into invasive and non-invasive systems [5]. Invasive BCIs employ surgically implanted electrodes to capture high-resolution neural activity and are suitable for applications such as restoring mobility or vision. Owing to risks including infection and tissue damage, they are typically reserved for severe clinical cases, such as paralysis or blindness. In contrast, non-invasive BCIs—particularly EEG-based systems—are safer and far more practical for everyday use, although they provide lower spatial resolution. Consequently, non-invasive BCIs have become increasingly popular for consumer-oriented applications such as gaming and neurofeedback [6].

A wide range of commercial EEG devices has been used in the development of BCI-controlled video games [7–12]. These devices, which vary from 1 to 64 electrodes, are affordable and widely accessible. The most commonly used EEG systems in this field are summarized in Table 1.

Table 1. Commercial non-invasive EEG devices.

Device	Manufacturer	Electrodes	Sampling Rate
Muse 2/Muse S	InteraXon Inc., Toronto, ON, Canada	4 dry	256 Hz
Neurosky MindWave	NeuroSky Inc., San Jose, CA, USA	1 dry	512 Hz
Emotiv EPOC+	Emotiv Inc., San Francisco, CA, USA	14 saline/dry	128/256 Hz
Emotiv Insight	Emotiv Inc., San Francisco, CA, USA	5 dry	128 Hz
OpenBCI Cyton	OpenBCI Inc., Brooklyn, NY, USA	8 (expandable to 16)	250 Hz (default)
Unicorn Hybrid Black	g.tec medical engineering GmbH, Schiedlberg, Austria	8 dry	250 Hz

All BCI systems generally follow the same processing pipeline [13]. First, electrical brain activity is measured and recorded using an EEG device. The raw signals are then preprocessed to remove noise and artifacts. Feature extraction follows, and the resulting features are fed into a classifier [14], which is typically trained offline. During real-time use, the classifier translates incoming EEG data into application-specific commands. Among the various paradigms used in BCIs, Motor Imagery (MI) has been widely adopted for voluntary control [2], as it enables users to generate distinct neural patterns without physical movement, making it suitable for hands-free gaming and assistive technologies.

Motor Imagery (MI) is defined as the mental simulation of movement without actual execution—is a well-established paradigm in EEG-based BCIs [15]. During MI tasks, individuals imagine specific limb movements, generating neural activity primarily in the sensorimotor cortex. These patterns can be captured using EEG and used to control external devices. However, decoding MI signals remains challenging due to inter-subject variability and low signal to noise ratios. Recent advances in deep learning, particularly convolutional neural networks (CNNs), have shown promise in improving MI classification accuracy [16,17]. CNNs can automatically extract spatial and temporal features from EEG signals, enhancing the robustness and generalization capabilities of BCI systems [18].

Research on human-machine interactions, carried out over a wider audience, shows that you can improve a user’s experience by being able to estimate differences between users. For example, researchers have recently developed methods for reducing cognitive load and improving task performance when there is a conflict between user goals and system goals [19]. Moreover, work in affective computing shows that robust modeling of human-generated signals requires representations that remain stable despite substantial variability across individuals, further emphasizing the need for systems that anticipate user-specific differences [20].

In this work, we investigate the prediction of user performance in a three-dimensional (3D) BCI-controlled game using pre-task MI EEG recordings. Participants used either the Muse 2 or the Emotiv Insight commercial EEG headset to perform MI tasks prior to gameplay. Based on their in-game performance, participants were categorized into three performance groups: low, medium, and high. Our goal is to predict a subject’s performance level solely from their MI recordings using a CNN-based model.

2. Related Work

Deep learning methods, and specifically CNNs, have significantly contributed to the development of BCI systems by performing well on the decoding of EEG signals in MI tasks. The majority of the existing literature, however, has focused on command-level classification, where the goal is to discriminate imagined movements (e.g., left vs. right

hand). More recently, an emerging line of research has attempted to move beyond command decoding and explore how deep learning can help identify individual variability in BCI performance. The subsequent section provides an overview of different studies that address command-level classification and subject-level prediction of performance using CNNs and other EEG-based approaches.

Fewer studies examine subject-level performance categorization, whereas the majority of BCI research employing CNNs concentrate on classifying motor commands, such as left or right hand motor images. Zhu et al. [21] performed a systematic comparison of five deep learning networks EEGNet, Deep ConvNet, Shallow ConvNet, ParaAtt, and MB3D—for MI EEG classification in BCI applications. On the basis of two large datasets (MBT-42 and Med-62), they optimized each network with hyperparameter tuning and evaluated subject-wise classification accuracy. The MB3D model with a 3D spatial representation of EEG channels worked well but with increased computational cost, whereas small models like EEGNet achieved comparable accuracy. Their results stress the significance of input representation and model structure on MI classification performance. These studies, however, focus exclusively on decoding MI commands, not on predicting user performance or aptitude.

Tibrewal et al. [22] evaluated the ability of deep learning models, specifically CNNs to improve classification efficiency in users of BCI motor imagery recordings who have previously been considered 'BCI inefficient'. A group of 54 participants underwent testing where the researchers compared a CNN model trained on EEG signals with a conventional Common Spatial Patterns (CSP) + Linear Discriminant Analysis (LDA) system and discovered the CNN produced superior results compared to the baseline approach with the most improvement seen with low-ability participants. The research team demonstrated that BCI underperformance cannot be attributed purely from human factors because CNN models could identify essential brain activity patterns which standard models usually overlook. The researchers implemented no manual feature extraction or preprocessing in their approach to prove that deep learning models maintain their performance across varied subjects while handling EEG signal imperfections. Nonetheless, their objective remains within task MI decoding enhancement, rather than forecasting overall user performance before interaction with the BCI system.

Another direction investigates pre-task prediction of MI performance based on neural traits. Cui et al. [23] proposed a method that uses microstate EEG features extracted from resting state recordings to predict whether a user will be a high or low performance MI BCI participant using microstate resting state features. The research findings showed that specific microstate parameters demonstrated a notable connection to MI task execution capabilities. Their algorithm trained on these microstate features showed excellent predictive performance (AUC = 0.83), demonstrating that the intrinsic dynamics of brain states relate to MI control capabilities. However, the framework uses resting-state EEG, takes a binary prediction approach, and uses classical ML machine learning models in its analyses of MI tasks or time-frequency distribution.

In parallel with advances in MI decoding, recent work has shown that neurophysiological and behavioral signals reflect broader cognitive and performance-related states that are valuable for predictive modeling. Chen et al. [24] demonstrated that multi-band EEG functional connectivity is capable of reliably identifying impaired cognitive states in demanding operational scenarios, indicating that neural dynamics provide informative markers relevant to performance estimation and highlighting the value of connectivity-based EEG features. Similarly, Xu et al. [25] utilized gaze behavior and flight-control signals, in combination with multimodal deep learning to assess pilot situational awareness, reinforcing that predicting human performance in interactive environments has substantial

practical value. Collectively, these findings underscore that effective performance prediction relies on capturing the user's underlying cognitive state, a principle that remains central regardless of the specific modality or operational domain.

This work examines a different and understudied problem by predicting multilevel BCI gaming performance (low, medium, high) from multiclass pre-game MI EEG using spectrogram based CNNs that use low channel consumer devices. It is important to note that in our study, the performance labels do not indicate MI decoding accuracy—as seen in other BCI studies. The labels indicate actual user performance in a BCI-controlled game in three dimensions that required continuous control in multiple degrees of freedom.

3. Materials and Methods

3.1. Materials

The following sections describe the hardware and software components used. EEG data were acquired on a personal computer that interfaced with the recording devices via Bluetooth. Although multiple EEG headsets were employed, the same experimental protocol and data processing pipeline were followed in all implementations.

3.1.1. Muse 2 Headband

The Muse 2 (InteraXon Inc., Toronto, ON, Canada) headset [26] was used to record EEG data. Muse 2 is a commercial EEG device that offers good flexibility due to its portability and ease of use. It has been widely adopted in both research and consumer applications, including BCI systems. The headset transmits data to a computer via Bluetooth.

Muse 2 uses four dry electrodes, which are more sensitive to environmental noise and movement artifacts compared to gel-based systems. Nevertheless, its affordability and portability made it suitable for this experiment. The electrodes are positioned according to the 10–20 system: two over the frontal lobe (AF7 and AF8) and one over each temporal lobe (TP9 and TP10). Figure 1 shows the device and its electrode locations. Although the device also includes sensors such as a gyroscope, accelerometer, and pulse oximeter, only the EEG channels were used in this study.

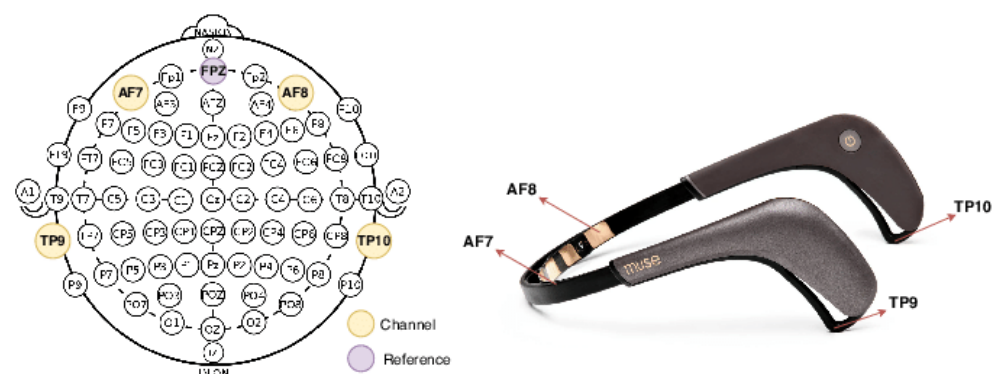


Figure 1. The Muse 2 EEG headband (right) and its electrode sites (left) mapped onto the 10–20 international system. The Muse 2 has four active electrodes: AF7, AF8, TP9, and TP10.

3.1.2. Emotiv Insight

The second recording device used was the Emotiv Insight (Emotiv Inc., San Francisco, CA, USA) [27], a lightweight, wireless EEG headset designed for quick setup and ease of use. It consists of five active electrodes, AF3, AF4, T7, T8, and Pz along with two reference electrodes (CMS/DRL). These sensors are arranged according to the international 10–20 placement system, as shown in Figure 2.

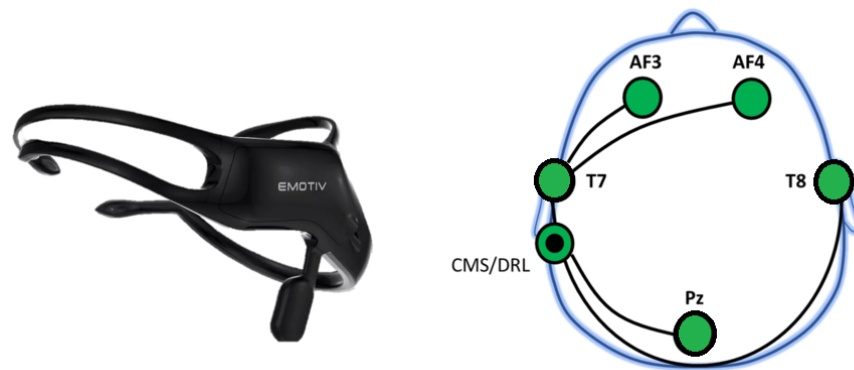


Figure 2. The Emotiv Insight EEG headset (**left**) and the corresponding electrode placement (**right**) projected onto a standard head model. The Insight uses five channels: AF3, AF4, T7, T8, and Pz, with CMS/DRL as reference.

The frontal electrodes (AF3, AF4) are associated with attentional and decision making processes, while T7 and T8 are located over the temporal lobes and are linked to memory related activity. The Pz electrode, positioned over the parietal region, is associated with visual processing. The Emotiv Insight uses semi-dry, gel-coated polymer sensors to improve signal quality. It records data at a 128 Hz sampling rate with 16-bit resolution, supports a bandwidth of 0.5–43 Hz, and includes built-in 50/60 Hz notch filters. EEG signals are transmitted wirelessly to a PC or smartphone via Bluetooth.

3.1.3. Game Design

The BCI video game developed in this study was a 3D interactive environment created using the Unity game engine. Players controlled an avatar within the game using mental commands based on MI tasks, specifically imagined left or right hand movements and eye blinking. The objective was to move the avatar along a track and collect virtual coins, creating a functional and motivating environment for evaluating real-time BCI performance. The game environment developed in this research is depicted in Figure 3.



Figure 3. Screenshot of the 3D BCI game environment created using the Unity engine. The player controls an avatar with mental commands generated from EEG signals, guiding it along the track to collect virtual coins. The setting provides an engaging and interactive platform for evaluating real-time BCI performance.

Two commercial EEG headsets were used to record data: the Muse 2 headband and the Emotiv Insight headset. EEG signals from the Muse 2 were transmitted wirelessly via Bluetooth using the BlueMuse application, while signals from the Emotiv Insight were streamed using the EmotivPro software [28]. Both applications supported the Lab Streaming Layer (LSL) protocol, enabling real-time streaming, synchronization, and visualization of EEG signals. These signals were sent to the OpenViBE platform, which served as the acquisition and signal processing environment.

The experimental procedure was divided into two main stages: offline training and online gameplay. In the offline phase, subjects performed three types of mental tasks: left hand MI, right hand MI, and eye blinking under supervised conditions. The EEG signals were preprocessed to retain frequencies in the 8–40 Hz range, then decomposed into sub-bands (Alpha, Beta1, Beta2, Gamma1, Gamma2) to compute energy features. These features were used to train two classifiers: a Multilayer Perceptron and a LDA model, using a one-vs-all strategy. In the subsequent online phase, the output of these classifiers was used to translate the user's real-time EEG signals into commands to control the game.

3.1.4. Dataset

This research utilized data from 36 subjects. Seventeen participants (7 females, 10 males), aged between 21 and 24 years (mean age: 22.06), participated in the experiment using the Muse 2 headband. The remaining 19 participants (7 females, 12 males), aged between 19 and 26 years (mean age: 22.53), took part in this study using the Emotiv Insight headset. All participants had normal or corrected to normal vision and signed written informed consent forms in accordance with ethical standards. Each participant completed three distinct EEG recording sessions under the supervision of a researcher, seated comfortably in a quiet, controlled environment to minimize movement and external interference.

Both the first and second recordings lasted five minutes, during which participants were asked to look left or right and imagine moving the corresponding hand. The third recording, lasting two and a half minutes, required forceful blinking every second. In total, 108 recordings were collected. However, in this study, only the left and right hand MI recordings were used.

Following ten gameplay attempts, each participant received an average performance score based on the number of virtual coins collected per attempt. Based on these scores, each subject was assigned to one of three performance categories: low, medium, or high. This grouping enabled further analysis of how individual differences in MI performance and classifier accuracy correlated with in-game control quality, offering insights into the usability of BCI systems in gaming contexts.

In both device cohorts, performance-based participant grouping was determined exclusively on the basis of each subject's average gameplay score across the ten coin-collection attempts. This average score served as the primary and objective criterion for categorization of the participant into one of performance classes. Specifically, subjects with an average score below 20 were assigned to the low-performance group, those with scores between 20 and 27 to the medium-performance group, and those who achieved average scores from 27 to 35 to the high-performance group. Brief qualitative observations made during gameplay by the two present researchers were used solely to confirm that these score-based thresholds were consistent with the participant's observable control ability during the BCI sessions, without influencing the numerical criteria employed for class assignment.

Building on this verification, these qualitative observations were not used to modify any labels but served only to verify that the score-based grouping reflected genuine differences in BCI control. Low-performance users generally struggled to generate stable MI commands and often faced difficulty switching lanes consistently, medium performers showed intermittent but still imperfect control with delayed or occasionally incorrect activations, and high performers demonstrated reliable, timely commands enabling smooth lane switching. Two researchers monitored all gameplay sessions and confirmed that the numerical scores accurately captured these behavioral distinctions. In the Muse 2 dataset, the three groups showed clearly distinct control patterns. On the other hand, the Emotiv

Insight dataset has higher signal quality which led to subtler, yet still consistent with the defined score differences between medium and high performers, ensuring that group assignments reflected genuine variation in BCI control rather than subjective judgment.

In the Muse 2 cohort, participant distribution across the three performance groups is shown in Figure 4. Specifically, five participants were categorized into the low-performance group, with their average scores ranging from 15.6 to 19.3 out of 50. Six participants belonged to the medium-performance group (scores: 21.0–26.4), and six participants were placed in the high-performance group, with scores between 27.1 and 31.5.

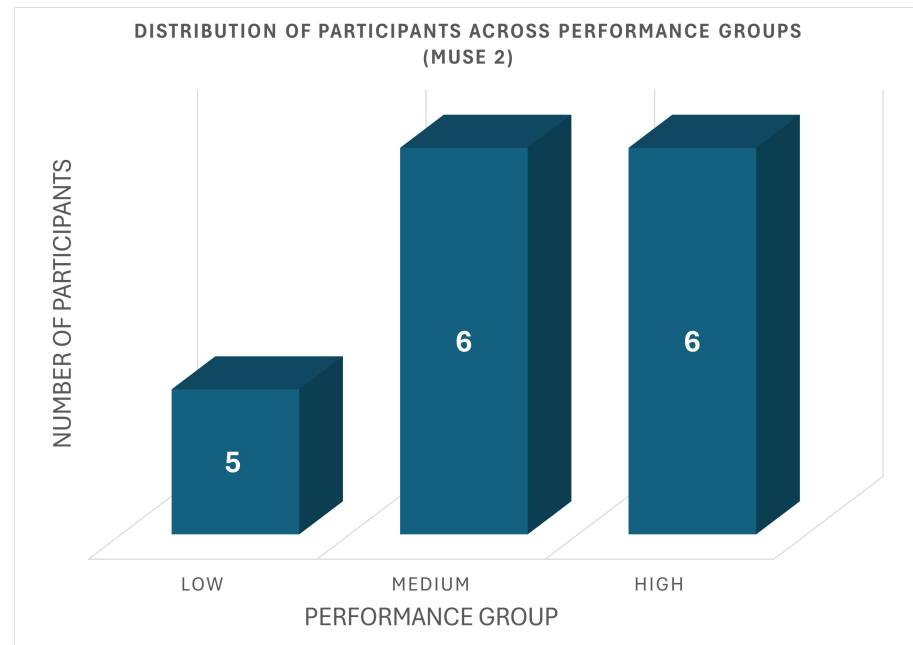


Figure 4. Distribution of Muse 2 Participants Across Performance Groups.

Similarly, the Emotiv Insight cohort showed the distribution presented in Figure 5. Six participants were in the low-performance group, with scores ranging from 14.2 to 17.6. Nine participants were classified as medium performers (scores: 20.0–25.9), and the remaining four participants formed the high-performance group, with scores from 27.6 to 35.0.

3.2. Methods

The following section provides a step by step explanation of the methodology adopted in this study that encompasses all the major stages of the experimental pipeline. This includes preprocessing operations applied to the raw EEG signals, the generation of time frequency spectrograms used as input representations to the CNN classifier, and the model training, validation, and performance metrics adopted. This methodological approach ensures consistency between the recordings from the two EEG devices and allows for reproducibility and reliability of the results.

3.2.1. Preprocessing

All EEG recordings followed the same standardized preprocessing steps to ensure the quality and comparability of the signal between subjects. The raw signals were band-pass filtered between 0.5–40 Hz to retain activity related to MI and notch-filtered at 50 Hz to remove the powerline interference. Artifact attenuation was then performed using ICA, where components reflecting ocular, muscular, or motion artifacts were identified and removed based on visual inspection. Finally, the signals were separated into 3 s epochs with a second overlap.

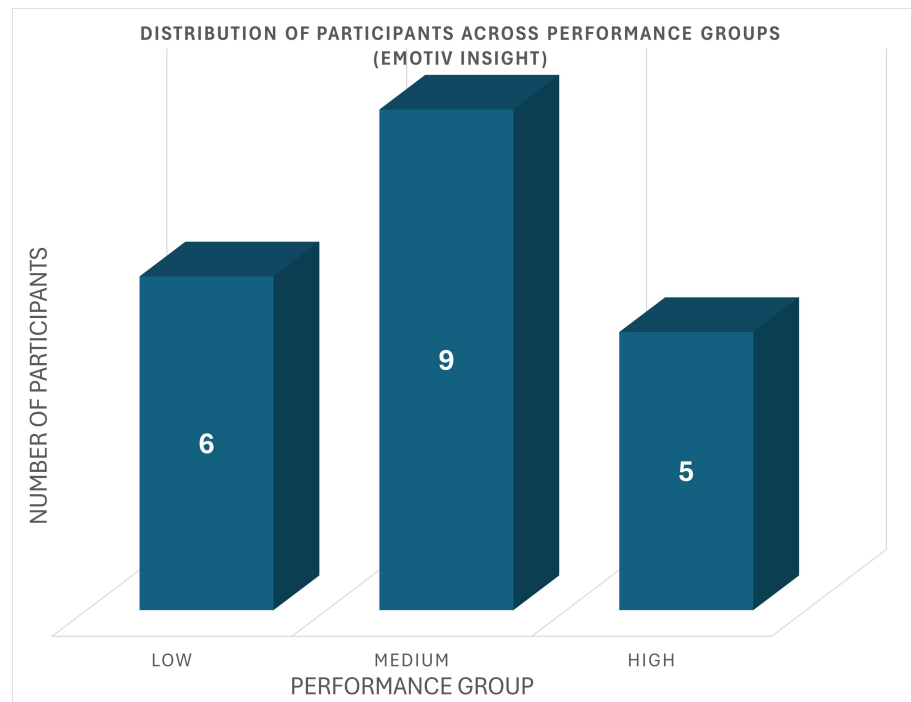


Figure 5. Distribution of Emotiv Participants Across Performance Groups.

The cleaned recordings were then divided into fixed length epochs, representing the basic units for subsequent time-frequency analysis. On every epoch and channel, a Short Time Fourier Transform was computed by means of a Hamming window ($n_{\text{perseg}} = 128$, $n_{\text{overlap}} = 64$); the resulting power spectra were then log scaled and converted to spectrograms. The spectrograms were stacked across channels to build 3D tensors ($\text{channels} \times \text{frequencies} \times \text{time}$) that could be used as input for the CNN.

In this study, time–frequency spectrograms are employed as the primary input representation. Spectrograms highlight frequency-specific changes associated with MI, such as event-related desynchronization and synchronization, events that are not easily captured in raw waveforms—especially when using low-density consumer headsets (Emotiv Insight, Muse 2) [16]. Additionally, converting EEG data into 2D time–frequency maps enables the application of spatial 2D convolutions, which have been shown to learn more stable and interpretable representations. [29] This choice is supported by prior work demonstrating improved separability of MI classes when spectral features are emphasized.

3.2.2. Data Balancing and Augmentation

The original datasets showed unequal representation across the three performance classes. Within the Emotiv Insight recordings, the distribution of samples across the three performance labels was: 3592 low performance samples (Label 0); 5379 medium performance samples (Label 1); and 2390 high performance samples (Label 2), resulting in an original distribution of 11,361 total samples. The Muse 2 recordings had total counts of 4187 for low (Label 0), 2955 for medium (Label 1), and 3595 for high (Label 2), resulting in the original distribution of samples totaling 10,737 samples.

To address this imbalance, a targeted augmentation procedure was applied to the minority classes in the training partition directly on the time-frequency representations. For each synthetically generated sample, one of three transformations was selected at random: Gaussian noise injection, implemented by adding zero-mean Gaussian noise with standard deviation $\sigma = 0.05$ to the log-scaled spectrogram; temporal shifting, performed as a circular shift along the time axis by an offset sampled uniformly from $[-5, 4]$ spectrogram time bins, corresponding to a maximum temporal displacement of approximately ± 1.25 s

for each device's sampling frequency; or spectral-temporal smoothing using Gaussian filtering with $\sigma = 1$. Each minority class was oversampled until it matched the size of the majority class. After augmentation, all classes were balanced to 5379 samples for Emotiv Insight (16,137 total) and 4187 samples for Muse 2 (12,561 total), resulting in more stable model training and improved generalization across performance categories. Importantly, no augmented samples were used in the validation or test sets, ensuring that the evaluation remained unbiased and reflective of real unseen data without any data leakage.

3.2.3. Convolutional Neural Network

The generated spectrograms were imported into a custom designed CNN to classify subjects into three BCI performance categories: low, medium, or high. The architecture of the network is illustrated in Figure 6. The input to the model consists of multi-channel EEG spectrograms formatted as 4D tensors with dimensions (epochs, frequency_bins, time_bins, channels).

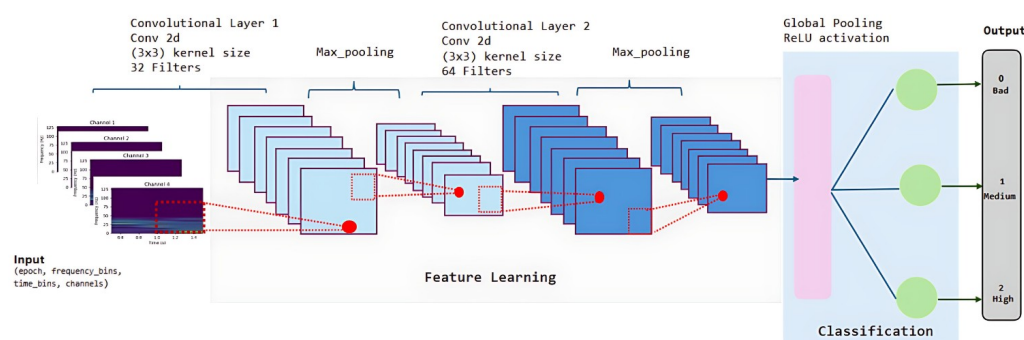


Figure 6. Architecture of the CNN for BCI performance classification.

The CNN architecture begins with two consecutive convolutional layers. The first convolutional layer uses 32 filters of size 3×3 followed by a ReLU activation and a MaxPooling2D layer with a 2×2 window, resulting in an output size of (65, 3, 32). The second convolutional layer uses 64 filters with ReLU activation, followed again by a 2×2 MaxPooling2D operation, yielding an output shape of (33, 3, 64).

Next, a GlobalAveragePooling2D layer is employed to flatten the spatial features into a one-dimensional feature vector of size 64. This is followed by a fully connected dense layer with 128 ReLU neurons and a Dropout layer with a 50% rate to mitigate overfitting. The final dense layer uses softmax activation to output a probability distribution over the three performance categories: low (0), medium (1), and high (2). In total, the model includes approximately 40,000 trainable parameters.

The model was trained with a batch size of 32, a learning rate of 0.001, and a maximum of 50 epochs, with early stopping after 5 epochs of no improvement in validation loss. To further interpret the learned representations of the CNN, we included a post-hoc visualization analysis. This involved inspecting first layer feature maps for representative samples and generating Grad-CAM heatmaps to identify the time-frequency regions that contributed most to the model's decisions. The corresponding visual results are presented in Section 3.2.4.

To estimate the model's efficiency in epoch level, the CNN was trained using the Adam optimizer and categorical cross-entropy loss, with early stopping based on validation loss to avoid overfitting. Performance evaluation was conducted using an 80/20 train-test split and classification results were assessed using accuracy, confusion matrices, and precision-recall metrics.

To assess subject-specific variability and model generalization, Leave-One-Subject-Out Cross Validation (LOSO-CV) was performed. In each fold, data from one subject were kept as the test set, while data from the remaining subjects formed the training set. During this LOSO-CV experiment, the same CNN architecture was used but trained with focal loss to better handle potential class imbalance and to emphasize harder to classify examples. Subject-level accuracy was calculated using majority voting over the test subject's evaluation samples, providing a more reliable metric of generalization across individuals.

3.2.4. Visualization of Learned Spectrogram Features

To better understand how the network encodes information from EEG spectrograms, we examined the internal representations at multiple stages of the CNN. Figure 7 shows the class averaged spectrograms, which reveal visible differences in the overall time–frequency structure across all three performance groups. These group level patterns suggest that the network has access to distinctive spectral–temporal signatures associated with different user profiles.

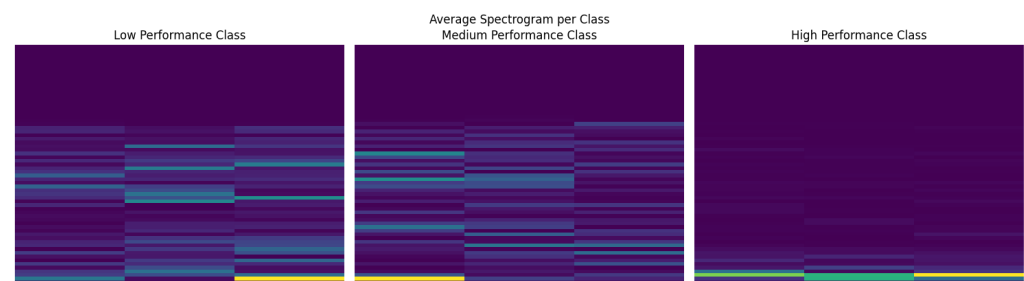


Figure 7. Class-averaged spectrograms for low, medium, and high performers. Color intensity represents the normalized spectral power at each time–frequency bin.

In order to explore how the CNN interprets these inputs, Figure 8 presents activation maps from the first convolutional layer of a correctly classified example from each class. Although only a subset of eight filters is displayed for clarity, they already illustrate the diversity of feature extraction strategies in the network. Some filters act as detectors of localized spectral variations, while others respond to broader temporal patterns or transitions. The heterogeneity and structure of these activations indicate that the CNN develops a rich, multiscale representation of the input spectrograms rather than relying on a single dominant feature type.

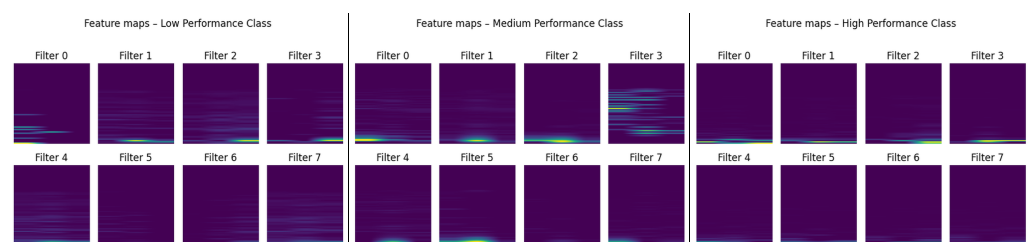


Figure 8. First-layer feature maps for one representative sample of each performance class. The first row corresponds to the low-performance class, the second to the medium-performance class, and the third to the high-performance class. Color intensity represents the normalized activation magnitude of the feature maps.

To further interpret the model's decision process, Grad-CAM heatmaps were generated for one sample per class in Figure 9. These visualizations highlight the regions of the spectrogram that contribute the most to each prediction. The highlighted areas differ systematically between classes, suggesting that the CNN focuses on distinct spectral-temporal patterns depending on performance level. This consistent class-specific saliency supports

the view that the model learns meaningful and discriminative representations rather than relying on noise or isolated outliers.

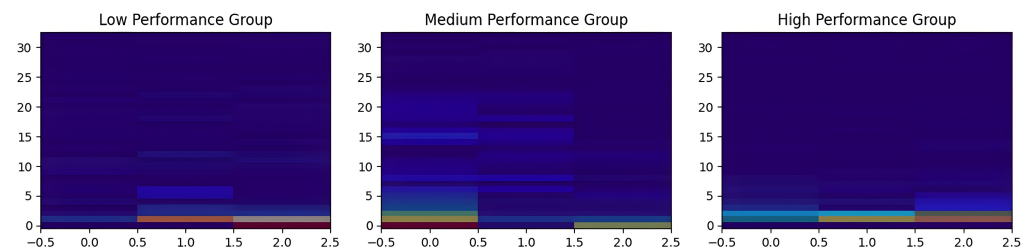


Figure 9. Grad-CAM heatmaps showing class-discriminative time-frequency regions for each performance class. The first image corresponds to the low-performance class, the second to the medium-performance class, and the third to the high-performance class. Color intensity indicates the relative contribution of each time-frequency region to the class prediction.

4. Results

This section presents the classification results of the CNN on spectrogram representations of EEG signals from two different datasets: one acquired using the Muse 2 headset and the other using the Emotiv Insight device. The CNN was trained to discriminate between three categories of users representing low, medium, and high BCI game performance. Performance was evaluated in terms of precision, recall, F1-score, and overall accuracy, along with the corresponding confusion matrices.

4.1. Model Performance

4.1.1. Performance Classification Using Muse 2 EEG Data

The CNN model applied to the Muse 2 data achieved a mean classification accuracy of 83%, as shown in Table 2. Overall, the model exhibited balanced classification performance across the three user classes. As observed in Table 2, the model successfully classified high- and low-performing users, achieving F1-scores of 85% and 84%, respectively, while medium-performing subjects had an F1-score of 77%. Precision and recall values were consistently high across all classes, with the model reaching a recall of 88% for both low and high performers and a precision of 88% for medium performers, indicating reliable sensitivity and specificity.

The confusion matrix in Figure 10 illustrates very good class separation, with few misclassifications between low and high performance classes. However, the network demonstrated slightly more difficulty distinguishing medium performers, with some overlap in predictions with the other two classes.

Table 2. Classification performance of the CNN model on Muse 2 EEG data across the three BCI performance classes.

Class	Precision	Recall	F1-Score
0 (Low)	0.81	0.88	0.84
1 (Medium)	0.88	0.69	0.77
2 (High)	0.82	0.88	0.85
Accuracy		0.83	

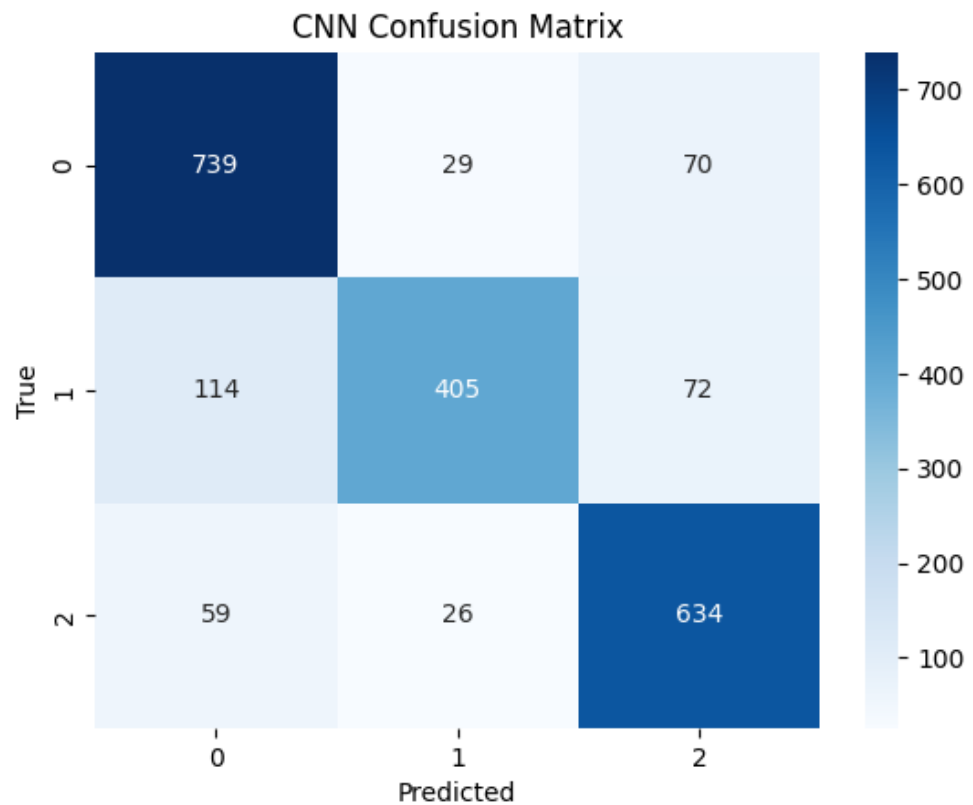


Figure 10. Confusion matrix for the CNN model predictions on Muse 2 EEG recordings.

4.1.2. Performance Classification Using Emotiv Insight EEG Data

The same CNN was also trained and tested on the Emotiv Insight dataset. The model achieved a higher accuracy of 95%, outperforming the Muse 2 configuration across all evaluated metrics. Table 3 highlights high precision and recall scores for all three classes, with recall ranging from 92% in the high performance class to 93% for the low performance class, and 96% in the medium class. Moreover, the model's precision ranged from 92% in the low performance class to 96% in the medium class. Regarding F1-scores, the values were 93% for the low and high classes, and 96% for the medium-performing class. The confusion matrix in Figure 11 reveals minimal misclassification and consistent model confidence across performance categories.

Table 3. Classification performance of the CNN model on Emotiv Insight EEG data across the three BCI performance classes.

Class	Precision	Recall	F1-Score
0 (Low)	0.93	0.92	0.93
1 (Medium)	0.96	0.96	0.96
2 (High)	0.92	0.93	0.93
Accuracy	0.95		

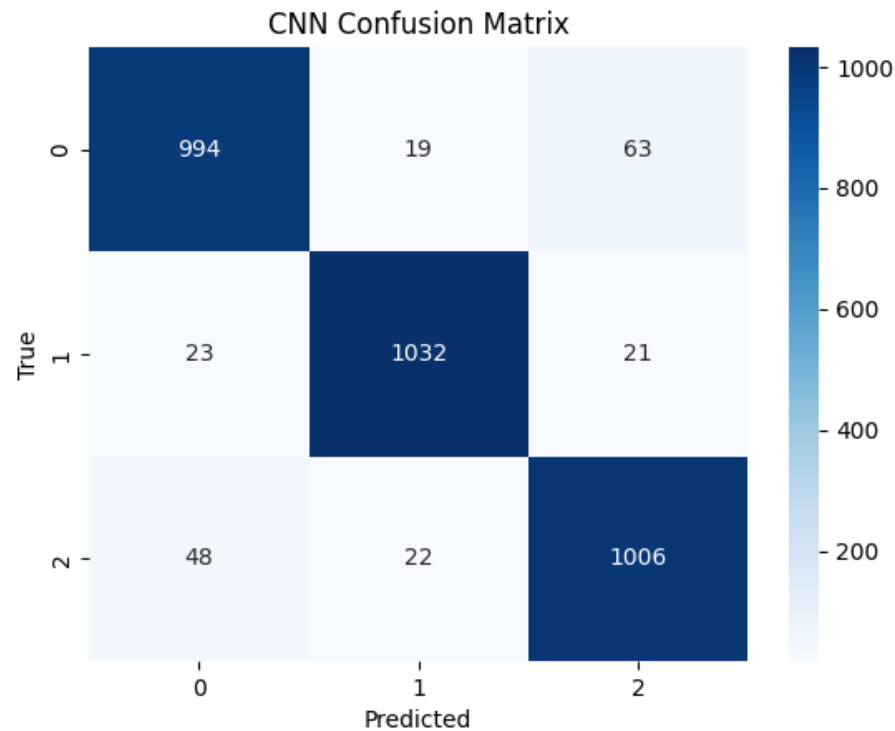


Figure 11. Confusion matrix for the CNN model predictions on Emotiv Insight EEG recordings.

4.2. Leave-One-Subject-Out Cross Validation

This section presents the results from the LOSO-CV experiments. In this setting, data from one subject were excluded during training and served as the test set in each fold. Final class predictions were obtained via majority voting over the predicted class labels of the test samples belonging to each subject.

4.2.1. Subject Classification Using Muse 2 Dataset

The LOSO-CV applied to the Muse 2 dataset yielded an average accuracy of 75% across all 17 subjects. The model maintained competitive performance across all three classes. A mild reduction in accuracy was observed compared to the standard train-test split, indicating increased difficulty in cross-subject generalization. The true and predicted labels for each subject are presented in Table 4. As shown, the majority of subjects were assigned to the correct category. Only four subjects (Subject 4, Subject 6, Subject 10, and Subject 12) were misclassified. Overall, the class-wise mean accuracies were 67% for Low (0), 80% for Medium (1), and 83% for High (2) performance categories. Specifically, 4 out of 6 low performance subjects were correctly classified, along with 4 out of 5 medium and 5 out of 6 high-performing participants.

4.2.2. Subject Classification Using Emotiv Insight Dataset

In the Emotiv Insight dataset, the LOSO-CV evaluation achieved an overall accuracy of 84% across all 19 participants. The model demonstrated strong classification performance for each of the three target classes. As expected from the previous results, a slight decline in accuracy was noted compared to the conventional train-test split, highlighting the challenge of cross-subject generalization. As detailed in Table 5, most subjects were correctly classified, with only 3 individuals (Subject 1, Subject 5, and Subject 13) being misassigned. The class-wise analysis revealed consistent performance, with mean accuracies of 83%, 89%, and 75% for the Low (0), Medium (1), and High (2) performance categories, respectively.

Table 4. LOSO-CV subject-level classification results in the Muse 2 dataset. True and predicted labels are compared for each subject. The final prediction was derived through majority voting.

Subject ID	True Label	Predicted Label
S1	1	1
S2	1	1
S3	2	1
S4	0	0
S5	2	2
S6	2	1
S7	0	0
S8	1	1
S9	2	2
S10	1	0
S11	1	1
S12	0	2
S13	2	2
S14	0	0
S15	2	2
S16	0	0
S17	0	0
Subject-Level Accuracy: 0.75		

Table 5. LOSO-CV subject-level classification results in the Emotiv Insight dataset. True and predicted labels are compared for each subject. Final prediction is obtained via majority voting.

Subject ID	True Label	Predicted Label
S1	2	0
S2	1	1
S3	1	1
S4	1	1
S5	1	0
S6	2	2
S7	0	0
S8	0	0
S9	1	1
S10	1	1
S11	2	1
S12	1	1
S13	0	2
S14	0	0
S15	0	0
S16	1	1
S17	2	2
S18	0	0
S19	1	1
Subject-Level Accuracy: 0.84		

5. Discussion

The results demonstrate that CNNs can effectively predict BCI user performance in a BCI gaming environment based solely on MI (MI) EEG recordings. The findings indicate strong potential for robust classification between different performance levels using spectrogram-based deep learning, with the CNN achieving 83% accuracy on Muse 2 data and 94% on Emotiv Insight data.

A key observation is the clear difference in performance between the two datasets. The model performed better on the Emotiv Insight dataset across all classification metrics—precision, recall, and F1-score—within all three performance categories. This improvement is attributed to the device’s electrode coverage and optimized sensor placement,

which provide better spatial resolution. In contrast, although Muse 2 is more accessible and affordable, it exhibited slightly lower performance, particularly for the medium-performing group, which showed greater overlap in misclassifications. These findings are consistent with prior studies and are also reflected in the class distributions observed for each device.

Regarding real-world applicability, the LOSO-CV technique approximates real-time performance evaluation. The results, although slightly lower than those obtained from the standard train–test split, remain promising. The Emotiv Insight dataset achieved 84% subject-level accuracy, with only 3 out of 19 subjects misclassified. Similarly, the Muse 2 dataset achieved 75% subject-level accuracy, misclassifying only 4 users. This reduction in performance compared to the epoch-level results (as shown in Sections 4.1.2) was expected due to the more demanding nature of cross-subject generalization.

Overall, the superior performance exhibited in epoch level is consistent with the structured temporal dependencies inherent in EEG data. Recent work has shown that exploiting within-trial dependencies enhances affective BCI performance, as adjacent temporal segments carry structured information rather than independent noise [30]. This provides a compelling explanation for the increased discriminability of epoch-level spectrograms in our experiments, where local temporal frequency patterns appear to encode informative markers of subsequent BCI control performance.

Building on this observation, the results demonstrate the CNN’s ability to exploit temporal and frequency-based information to automatically derive correlations between EEG and motor skill units. The ability of the CNN to extract informative features through an end-to-end learning process supports the increased movement toward fully automated BCI systems, and ultimately, the development of BCI systems which rely on user-specific data for predicting user capability without the need for manual feature construction.

5.1. Cross-Device Robustness and Statistical Significance Analysis

To further examine the robustness of the proposed CNN across different EEG hardware, we compared the subject-level performance obtained through the LOSO-CV procedure for the Muse 2 and Emotiv Insight datasets. Figure 12 presents a violin plot illustrating the distribution of continuous per-subject accuracies for each device. Despite substantial differences in electrode placement, sampling rate, and signal characteristics, both datasets exhibit consistently high accuracies, indicating that the spectrogram-based CNN remains reliably effective across distinct commercial EEG systems.

To statistically validate these findings, we performed exact binomial significance testing on the LOSO subject-level outcomes. For the Muse 2 dataset, 13 out of 17 subjects were correctly classified (exact binomial p -value = 0.00034), while for the Emotiv Insight dataset, 16 out of 19 subjects were correctly classified (p = 0.000007). In both cases, performance was significantly above the three-class chance level of 33.3%.

We further quantified uncertainty by computing 95% Wilson confidence intervals for the subject-level accuracy proportions. As shown in Figure 13, the confidence interval for the Muse 2 dataset ranged from 0.501 to 0.932, whereas that for the Emotiv Insight dataset ranged from 0.604 to 0.966, confirming that the true accuracy lies well above chance for both devices. Finally, Fisher’s exact test revealed no statistically significant difference in classification reliability between the two devices (p = 0.684), suggesting that the proposed model generalizes comparably across hardware platforms.

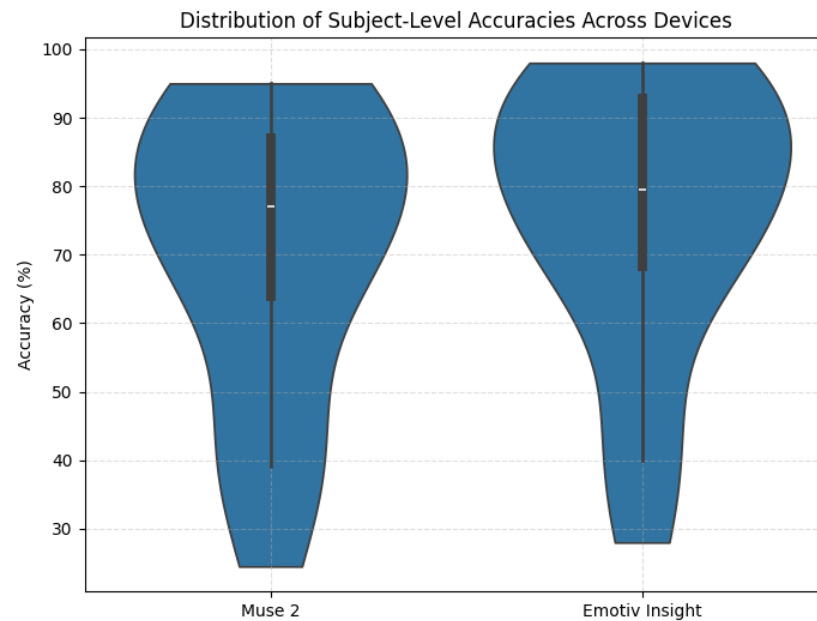


Figure 12. Violin plot showing the distribution of per-subject LOSO-CV accuracies for the Muse 2 and Emotiv Insight datasets.

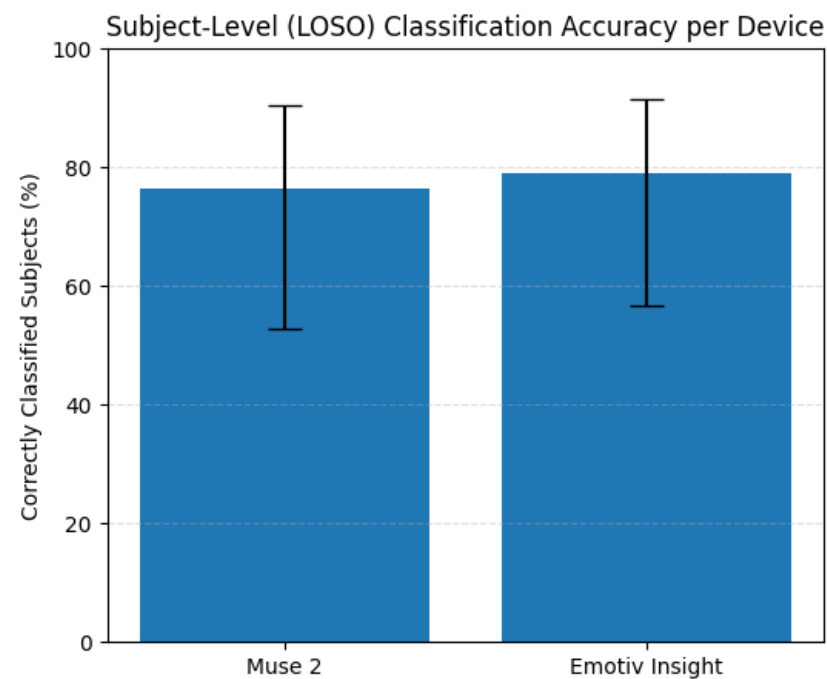


Figure 13. Comparison of Muse 2 and Emotiv Insight datasets using Wilson binomial confidence intervals under the LOSO-CV evaluation.

As per the epoch-level classification, for the Muse 2 dataset, the epoch-level accuracy of 83% corresponded to a 95% Wilson confidence interval of [80.7%, 83.4%] based on 2148 test epochs. On the other hand for the Emotiv Insight dataset, the epoch-level accuracy of 95% corresponded to a 95% Wilson confidence interval of [93.48%, 95.08%] computed over 3228 test epochs. Figure 14 depicts the confidence intervals for both devices under the epoch-level classification. These intervals reflect the precision of the accuracy estimates, with narrower intervals indicating greater statistical reliability.

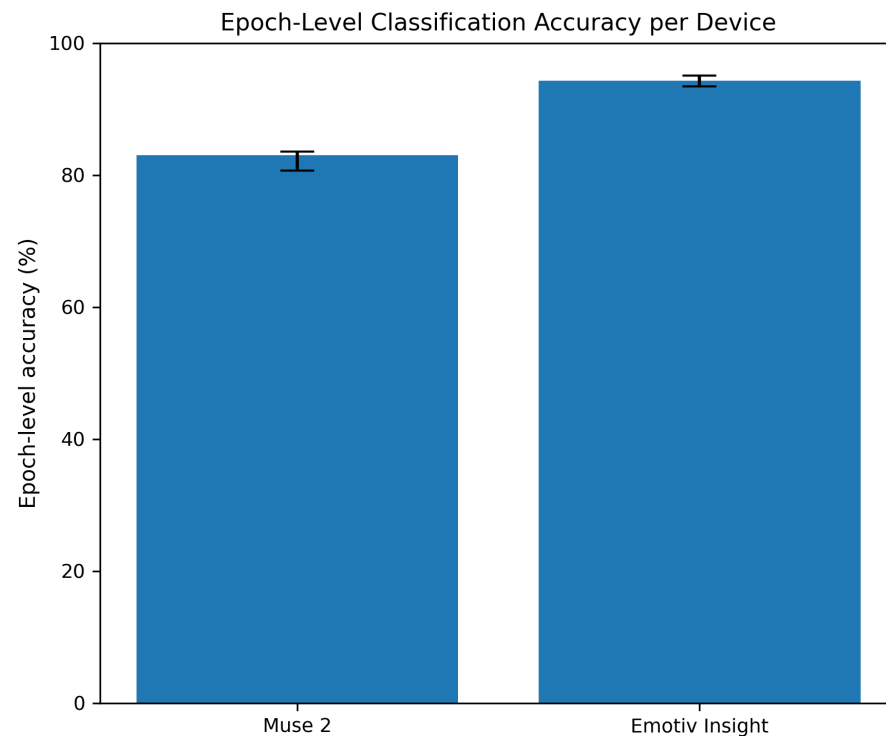


Figure 14. Epoch-level classification accuracy for the Muse 2 and Emotiv Insight datasets with Wilson binomial confidence intervals.

5.2. Novelty and Contribution

This work investigates a relatively unexplored direction in BCI research, namely, the prediction of user performance levels in a game-based BCI environment using only MI EEG recordings prior to actual task engagement. Unlike previous studies that focus either on classifying MI commands or on improving decoding accuracy during task execution, our objective is to determine whether early task-evoked EEG activity can reliably forecast how well a user will subsequently control a BCI system.

Compared to the existing literature, as presented in Table 6 this study differs in several key aspects. Zhu et al. [21] focused on command level MI classification from high density clinical EEG systems, without addressing performance prediction or pre-task aptitude assessment. On the other hand Tiberwal et al. [22] investigate BCI inefficiency and use CNNs to enhance MI decoding for low-performing users, yet their analysis remains strictly within task and does not attempt to predict future user performance before interaction with the BCI. The work conducted by Ciu et al. [23] use resting state microstate features to perform binary prediction of MI aptitude in a smaller cohort of 28 subjects; however, their work does not incorporate MI recordings, time–frequency representations, or real world BCI performance metrics.

This study provides an initial proof of concept for three-level performance prediction using low channel consumer devices, with class labels derived from interaction in a 3D BCI game involving multiple degrees of freedom rather than from offline MI accuracy metrics. The inclusion of task based spectrograms and a cross-subject LOSO evaluation further differentiates this work, as only the second experiment in Zhu et al. [21] employed LOSO for MI classification. While preliminary, the findings suggest that early MI responses may contain informative patterns related to subsequent BCI control performance, motivating future studies with larger and more diverse datasets.

Table 6. Comparative results with other studies.

Author	Dataset	Model/Validation Method	Results
Zhu et al. [21]	MBT-42 dataset (42 subjects, 64-channel Neuroscan EEG).	Five DL models: EEGNet, Shallow ConvNet, DeepConvNet, MB3D, ParaAtt. Validation: within-subject (Sessions 1–2 for training, Session 3 for testing).	All DL models achieved comparable accuracy and consistently outperformed the online baseline decoder.
	Med-62 dataset (62 subjects, 64-channel Neuroscan EEG).	Validation: LOSO-CV (62 folds).	EEGNet achieved the highest cross-subject accuracy. Preprocessing substantially improved stability. MB3D exhibited the highest computational cost.
Tibrewal et al. [22]	MI EEG (Left/Right), 13 electrodes, 120 trials per subject (54 subjects).	Custom CNN vs. CSP + LDA. Validation: 70/30 per-subject split with additional 10-fold CV.	CNN significantly outperformed CSP + LDA (69.42% vs. 52.56%), with the largest gains observed in low-performing users (+18.46%).
Cui et al. [23]	Resting-state EEG (64-channel), 28 subjects, 6 sessions \times 5 runs.	CSP + SVM for MI accuracy prediction; LDA regression using microstate features (MS1–MS4). Validation: 10-fold CV.	AUC = 0.83; microstate parameters robustly predicted MI skill (high vs. low performers).
Ntetska et al. (This Study)	MI EEG from Muse 2 (17 subjects) and Emotiv Insight (19 subjects).	Custom CNN using spectrogram inputs. Validation: Two distinct validation methodologies were employed: 80/20 epoch-level classification and LOSO-CV with majority voting.	Percentage Split (80/20) Muse 2: 83% accuracy, Emotiv Insight: 95%. LOSO-CV: Muse 2: 75%, Emotiv Insight: 84%. Higher performance with Emotiv attributed to improved spatial sampling.

5.3. Baseline Comparison

To evaluate the effectiveness of the proposed spectrogram-based CNN, an additional baseline experiment was conducted using two widely adopted deep learning architectures for EEG analysis: EEGNet and ShallowConvNet and CSP + LDA, a classical BCI classification method. All models were re-implemented within the same training and evaluation framework used in our study with identical data splits, augmentation strategy, and LOSO protocol, thus ensuring a consistent and fair comparison at the classification level. This setup allowed us to assess whether more established EEG-specific architectures could outperform our spectrogram oriented approach under identical experimental conditions.

The results of the epoch-level evaluation are presented in Table 7. The proposed spectrogram-based CNN outperformed all models across both datasets. Although CSP + LDA is a classical method in BCI classification, it does not perform adequately in this three-class setting, achieving only 56% and 62% accuracy in the respective datasets. EEGNet achieved higher accuracy but still reached only 76% and 78%. ShallowConvNet performed better than both, achieving 82% accuracy in the Muse 2 dataset and 92% in the Emotiv dataset. Notably, while ShallowConvNet reached performance levels close to those of our model in the Emotiv dataset (94%), this outcome is expected due to its shallow architecture, which is known to fit smaller datasets more easily and to benefit from reduced inter-subject variability.

A more pronounced contrast between the models emerges under the LOSO-CV evaluation, as shown in Table 8. In this setting, all models, CSP + LDA, EEGNet and ShallowConvNet exhibited a substantial drop in performance, with the CSP + LDA model achieving accuracy of 39% in the Muse 2 cohort and 49% in the Emotiv cohort. On the other hand, EEGNet and ShallowConvNet achieved only 52% and 53% accuracy on the Muse 2 dataset

and 53% and 55% accuracy in the Emotiv Dataset respectively. This steep decline indicates poor generalization to unseen individuals, an expected limitation of these architectures, as they rely heavily on subject-specific temporal patterns and are primarily tuned for raw EEG inputs. This degradation is particularly evident for the CSP + LDA pipeline, which relies on subject-specific spatial filters that do not generalize well to unseen individuals. Furthermore, CSP + LDA is traditionally tailored for binary motor-imagery tasks (e.g., left vs. right hand), making it inherently less suitable for the multi-class performance prediction problem addressed here. In contrast, by employing spectrograms as inputs, the proposed CNN achieved significantly higher subject-level accuracies of 75% for Muse 2 and 83% for Emotiv, demonstrating strong robustness against inter-subject variability. These results suggest that the proposed model not only performs competitively at the epoch level but also generalizes markedly better across subjects, a crucial characteristic for practical deployment in real world BCI applications.

Table 7. Epoch-level accuracy comparison between the proposed Methodology EEGNet, and ShallowConvNet for both datasets.

Model	Muse 2 Accuracy	Emotiv Accuracy
Proposed Methodology	0.83	0.95
EEGNet	0.76	0.78
ShallowConvNet	0.82	0.94
CSP + LDA	0.56	0.67

Table 8. Subject-level LOSO accuracy comparison across the three models.

Model	Muse 2 LOSO Accuracy	Emotiv LOSO Accuracy
Proposed Methodology	0.75	0.84
EEGNet	0.52	0.53
ShallowConvNet	0.53	0.55
CSP + LDA	0.39	0.49

5.4. Limitations and Future Work

While the results are promising, this study has several limitations that future work should address. First, although the sample size is sufficient for a proof of concept, it remains relatively small. This limits the generalizability of the findings to larger populations with more diverse EEG characteristics and gaming abilities. Second, the approach was tested using pre-task EEG recordings only. While such recordings enable early prediction, they do not incorporate real-time adaptation and feedback mechanisms that are essential in dynamic gaming environments. Integrating adaptive in-game feedback could enhance both prediction accuracy and user engagement.

Furthermore, although spectrograms capture informative time–frequency information, they are less sensitive to spatial and topological aspects of brain activity. Future work could explore hybrid representations that combine spectrogram inputs with graph-based EEG features to better model inter-channel connectivity and spatial dynamics. Another limitation concerns cross-device generalization. Although two independent consumer grade EEG devices were analyzed, a model trained on one device does not necessarily transfer to another due to differences in electrode positioning and signal characteristics. Future work should therefore explore domain adaptation and transfer learning strategies to improve hardware-agnostic robustness. Moreover, incorporating a brief user-specific calibration trial prior to gameplay could further improve classification accuracy by accounting for inter-subject variability. Finally, evaluating the proposed model in real-time BCI scenarios, where decisions are made continuously from live EEG input, would help determine its

suitability for deployment in practical neuroadaptive gaming systems. Taken together, these directions open promising avenues toward predictive personalization in BCI games, enabling customized user experiences based on early neurophysiological profiling.

In this study, we introduced a deep learning framework capable of predicting user performance in BCI-based gaming using pre-game MI EEG recordings. The proposed CNN achieved strong and consistent accuracy across both datasets, demonstrating that time–frequency representations of pre-task EEG activity can reliably predict BCI control performance during gameplay. These findings provide an initial demonstration of the feasibility of early performance prediction in BCI systems and motivate further refinement and investigation. Finally, this work serves as an initial exploration of pre-task user profiling, which can be further enhanced in future studies. Larger datasets, additional performance metrics, and the inclusion of different EEG hardware could support broader validation of this approach and establish its applicability across diverse users and devices. Expanding these areas is not a limitation of our method but rather a natural progression toward consolidating and generalizing the promising results observed here.

Author Contributions: Formal analysis, M.G.T. and K.D.T.; Resources, M.G.T. and P.A.; Data curation, A.N. and A.M.; Writing—original draft, A.N.; Project administration, M.G.T. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: The study was conducted following approval by the Research Ethics and Deontology Committee of the University of Western Macedonia, under Application No. 172/20-03-2023 and Protocol No. 179/2023.

Informed Consent Statement: Informed consent for publication was obtained from the participants.

Data Availability Statement: Data not available for this study.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Saha, S.; Mamun, K.A.; Ahmed, K.; Mostafa, R.; Naik, G.R.; Darvishi, S.; Khandoker, A.H.; Baumert, M. Progress in brain computer interface: Challenges and opportunities. *Front. Syst. Neurosci.* **2021**, *15*, 578875. [[CrossRef](#)] [[PubMed](#)]
2. Liu, X.; Wang, W.; Liu, M.; Chen, M.; Pereira, T.; Doda, D.Y.; Ke, Y.; Wang, S.; Wen, D.; Tong, X.; et al. Recent applications of EEG-based brain-computer-interface in the medical field. *Mil. Med Res.* **2025**, *12*, 14. [[CrossRef](#)] [[PubMed](#)]
3. Patiño, J.; Vega, I.; Becerra, M.A.; Duque-Grisales, E.; Jimenez, L. Integration Between Serious Games and EEG Signals: A Systematic Review. *Appl. Sci.* **2025**, *15*, 1946. [[CrossRef](#)]
4. Vărbu, K.; Muhammad, N.; Muhammad, Y. Past, present, and future of EEG-based BCI applications. *Sensors* **2022**, *22*, 3331. [[CrossRef](#)]
5. Birbaumer, N. Breaking the silence: Brain–computer interfaces (BCI) for communication and motor control. *Psychophysiology* **2006**, *43*, 517–532. [[CrossRef](#)]
6. Guger, C.; Grünwald, J.; Xu, R. Noninvasive and invasive BCIs and hardware and software components for BCIs. In *Handbook of Neuroengineering*; Springer: Berlin/Heidelberg, Germany, 2023; pp. 1193–1224.
7. Hjørungdal, R.M.; Sanfilippo, F.; Osen, O.L.; Rutle, A.; Bye, R.T. A Game-Based Learning Framework For Controlling Brain-Actuated Wheelchairs. In Proceedings of the ECMS 2016 Proceedings, Regensburg, Germany, 31 May–3 June 2016; pp. 554–563. [[CrossRef](#)]
8. Rosca, S.D.; Leba, M. Design of a Brain-Controlled Video Game based on a BCI System. *MATEC Web Conf.* **2019**, *290*, 01019. [[CrossRef](#)]
9. Joselli, M.; Binder, F.; Clua, E.; Soluri, E. Mindninja: Concept, Development and Evaluation of a Mind Action Game Based on EEGs. In Proceedings of the 2014 Brazilian Symposium on Computer Games and Digital Entertainment, Porto Alegre, Brazil, 12–14 November 2014; pp. 123–132. [[CrossRef](#)]
10. Vasiljevic, G.; Miranda, L.; Menezes, B. Mental War: An Attention-Based Single/Multiplayer Brain-Computer Interface Game. In *Lecture Notes in Computer Science*; Springer: Cham, Switzerland, 2018; pp. 450–465. [[CrossRef](#)]

11. Glavas, K.; Prapas, G.; Tzimourta, K.D.; Giannakeas, N.; Tsipouras, M.G. Evaluation of the User Adaptation in a BCI Game Environment. *Appl. Sci.* **2022**, *12*, 12722. [\[CrossRef\]](#)
12. Prapas, G.; Glavas, K.; Tzallas, A.T.; Tzimourta, K.D.; Giannakeas, N.; Tsipouras, M.G. Motor imagery approach for bci game development. In Proceedings of the 2022 7th South-East Europe Design Automation, Computer Engineering, Computer Networks and Social Media Conference (SEEDA-CECNSM), Ioannina, Greece, 23–25 September 2022; IEEE: Piscataway, NJ, USA, 2022; pp. 1–5.
13. Mousa, F.A.; El-Khoribi, R.A.; Shoman, M.E. A novel brain computer interface based on principle component analysis. *Procedia Comput. Sci.* **2016**, *82*, 49–56. [\[CrossRef\]](#)
14. Singh, A.K.; Krishnan, S. Trends in EEG signal feature extraction applications. *Front. Artif. Intell.* **2023**, *5*, 1072801. [\[CrossRef\]](#)
15. Moran, A.; O'Shea, H. Motor imagery practice and cognitive processes. *Front. Psychol.* **2020**, *11*, 394. [\[CrossRef\]](#)
16. Al-Saegh, A.; Dawwd, S.A.; Abdul-Jabbar, J.M. Deep learning for motor imagery EEG-based classification: A review. *Biomed. Signal Process. Control* **2021**, *63*, 102172. [\[CrossRef\]](#)
17. Lionakis, E.; Karampidis, K.; Papadourakis, G. Current trends, challenges, and future research directions of hybrid and deep learning techniques for motor imagery brain–computer interface. *Multimodal Technol. Interact.* **2023**, *7*, 95. [\[CrossRef\]](#)
18. 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. *J. Neural Eng.* **2018**, *15*, 056013. [\[CrossRef\]](#)
19. Pan, Y.; Xu, J. Human-machine plan conflict and conflict resolution in a visual search task. *Int. J. Hum.-Comput. Stud.* **2025**, *193*, 103377. [\[CrossRef\]](#)
20. Lin, Z.; Wang, Y.; Zhou, Y.; Du, F.; Yang, Y. MLM-EOE: Automatic Depression Detection via Sentimental Annotation and Multi-Expert Ensemble. *IEEE Trans. Affect. Comput.* **2025**, *16*, 2842–2858. [\[CrossRef\]](#)
21. Zhu, H.; Forenzo, D.; He, B. On the Deep Learning Models for EEG-Based Brain-Computer Interface Using Motor Imagery. *IEEE Trans. Neural Syst. Rehabil. Eng.* **2022**, *30*, 2283–2291. [\[CrossRef\]](#)
22. Tibrewal, N.; Leeuwis, N.; Alimardani, M. Classification of motor imagery EEG using deep learning increases performance in inefficient BCI users. *PLoS ONE* **2022**, *17*, e0268880. [\[CrossRef\]](#)
23. Cui, Y.; Xie, S.; Fu, Y.; Xie, X. Predicting Motor Imagery BCI Performance Based on EEG Microstate Analysis. *Brain Sci.* **2023**, *13*, 1288. [\[CrossRef\]](#)
24. Chen, J.; Wang, Y.; Cui, Y.; Wang, H.; Polat, K.; Alenezi, F. EEG-based multi-band functional connectivity using corrected amplitude envelope correlation for identifying unfavorable driving states. *Comput. Methods Biomech. Biomed. Eng.* **2025**, *9*, 1–13. [\[CrossRef\]](#) [\[PubMed\]](#)
25. Xu, G.J.W.; Pan, S.; Sun, P.Z.H.; Guo, K.; Park, S.H.; Yan, F.; Gao, M.; Wanyan, X.; Cheng, H.; Wu, E.Q. Human-Factors-in-Aviation-Loop: Multimodal Deep Learning for Pilot Situation Awareness Analysis Using Gaze Position and Flight Control Data. *IEEE Trans. Intell. Transp. Syst.* **2025**, *26*, 8065–8077. [\[CrossRef\]](#)
26. Muse. Muse 2—EEG-Powered Meditation & Sleep Headband. Available online: <https://choosemuse.com/products/muse-2> (accessed on 12 December 2025).
27. EMOTIV. Insight—5 Channel Wireless EEG Headset. Available online: <https://www.emotiv.com/products/insight> (accessed on 12 December 2025).
28. EMOTIV. EmotivPro Software. Available online: <https://www.emotiv.com/products/emotivpro> (accessed on 12 December 2025).
29. Tabar, Y.R.; Halici, U. A novel deep learning approach for classification of EEG motor imagery signals. *J. Neural Eng.* **2016**, *14*, 016003. [\[CrossRef\]](#)
30. Pei, Y.; Zhao, S.; Xie, L.; Ji, B.; Luo, Z.; Ma, C.; Gao, K.; Wang, X.; Sheng, T.; Yan, Y.; et al. Toward the enhancement of affective brain–computer interfaces using dependence within EEG series. *J. Neural Eng.* **2025**, *22*, 026038. [\[CrossRef\]](#)

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.