Open Access Original Article

Exploring Electroencephalography-Based Visual Imagery Using Convolutional Neural Networks For 2D Shapes Classification

Received 06/30/2025
Review began 09/05/2025
Review ended 10/08/2025
Published 10/11/2025

DOI:
https://doi.org/10.7759/s44389-025-08315-6

Daniel Lizama [1] , David Serrano [1]

1. Mechanical Engineering, University of Puerto Rico, Mayaguez, USA

Corresponding author: Daniel Lizama, daniel.lizama@upr.edu

Abstract

Brain-Computer Interfaces (BCIs) using electroencephalography (EEG) enable intuitive, non-invasive neural communication with external systems. While many other BCI paradigms have been widely studied in EEG-BCI, visual imagery (VI) remains relatively unexplored, particularly for intra-class object differentiation. This study examines the viability of using state-of-the-art deep learning (DL) models for classifying EEG signals generated during imagination of 2D geometric shapes within the same visual category such as circles and triangles. Two leading convolutional neural network architectures, EEGNet and FBCNet, are evaluated for end-to-end learning from raw time-series data. The limitations of classical machine learning approaches such as feature extraction bottlenecks and low signal-to-noise ratio are addressed. Data augmentation with a crop window technique is used to mitigate restrictions on experimental data conditions while maintaining model performance. Results demonstrate that DL models can efficiently classify intra-category VI, predicting 216 testing samples in under 2 seconds. Challenges remain due to limited training data and subtle neural distinctions between simple geometric shapes. Topographic brain maps further illustrate class-specific activation patterns. This work presents, to our knowledge, the first implementation of DL classification on full waveform VI data for primitive 2D shapes in EEG-BCI systems.

Categories: Human-Machine Interface, Machine Learning (ML), Deep Learning
Keywords: machine learning, electroencephalogram signals classification, visual imagery, brain-computer interface (bci), convolutional neural networks (cnns)

Introduction

Brain-computer interfaces (BCI) aim to provide an alternative and intuitive modality for computer interaction to the traditional peripherals. BCIs achieve this by decoding the user's intention from observed neural activation using specialized computational methods. Electroencephalography (EEG) is a commonly used signal acquisition modality for BCI systems providing adequate temporal sampling with relatively less expensive equipment than systems requiring surgical implants or invasive procedures [1-3].

EEG-BCI has been proposed for motor rehabilitation [4-5], control of prosthesis and rehabilitation devices [6-7], and as computer software interface [8]. Nevertheless, EEG devices have two major disadvantages: low signal-to-noise ratio (SNR) and poor spatial resolution from the recording system, requiring specifically engineered processing methods for accurate interpretation and control. This makes autonomous and generic EEG classifiers complex or technically specific to design.

Training EEG-BCI classifiers must also overcome the hurdles of limited training data with high dimensionality and inter-trial variance, as well as inter-class variability associated with the required control paradigm of interest [8]. Machine learning (ML) has emerged as a promising tool for overcoming limitations inherent to EEG-BCI, thanks to its elaborate information extraction capabilities and adaptability. Classical ML techniques generally achieve this with the help of preprocessing methods that represent the raw low SNR signals onto a new characteristic physiological feature space with enhanced SNR and a reduced feature space, which optimizes ML performance. These techniques have proven to work well for a variety of BCI paradigms [9-10]. Motor imagery (MI) is one of the most used and extensively studied EEG-BCI control strategies. In MI, users perform a mental process, where they imagine voluntary limb motor movement [11]. It is known that different classes of MI can differ in their spectro-spatial distribution, making it the main feature to be extracted for accurate MI differentiation in successful ML models, such as Filter Bank Common Spatial Pattern [9]. Another much less studied BCI paradigm has emerged as a promising method for alternate system controls and computer-aided design, known as visual imagery (VI) [12-13].

In VI, the users perform a mental process, where they voluntarily visualize a distinct image of an object or scene from a third-person perspective [14]. However, there are far less studies performed for VI paradigms compared to MI. In many VI-BCI designs, the poor differentiation between tasks using classical ML is typically due to the EEG feature extraction capabilities as implemented by design. This can be attributed to limited knowledge of VI's physiological patterns for proper signal processing and feature selection [15]. Furthermore, most VI studies focus on differentiation of cross-category object visualization. where the image variables significantly change in many of their attributes (e.g. shape, colors, vividness, or familiarity).

Although VI classification has recently shown advancements in distinguishing static over dynamic image visualizations [15], differentiation of VI based on varying object shape remains a challenging task. Kosmyna et al. [16] studied the potentials of distinguishing EEG signals related to VI between two classes (hammers and flowers). Despite being two unrelated categories of objects with distinct features, they managed to achieve 52% classification accuracy using the Spectrally Weighted Common Spatial Patterns (SpecCSP) classification strategy. Other works, such as that by Esfahani [12], focused on intra-category visualization to classify VI signals of five basic shapes, as well as on the control of the feature variability of the objects themselves (including colors) to improve the analysis of crucial characteristics of objects that a BCI user can control. While the author reports a classification accuracy of only 44.6% using a linear discriminant analysis classifier, a 69.8% accuracy was achieved when distinguishing between round and sharp shapes. Although these results are well above chance probability, it is necessary to have much more consistent outcomes for a reliable BCI system. Furthermore, these methods tested for VI-BCI mainly rely on converting time-variant recordings into data-relevant features, which can be time consuming and may reduce the responsiveness of the BCI system in real-world applications.

Recently, many advanced deep learning (DL) models have been proposed for BCI, and they have quickly outperformed classical approaches for EEG data classification [10,17-18]. Convolutional Neural Networks (CNN) stand out as a promising end-to-end solution due to their ability to learn local connectivity patterns from data without the need of 'hand-crafted' designed features. This is ideal for BCI development as it reduces the time and complexity of data processing, allowing the system to increase the interaction frequency. Therefore, a possible solution to improving VI physiological knowledge and VI-BCI reliability could come from emerging powerful deep DL models. For example, Llorella et al. [17] showed that a simple feature-based CNN model outperforms classical ML in the classification of diverse VI objects (i.e. tree, house, plane, and dog).

In this study, we investigate the performance of state-of-the-art (SOTA) DL models in classifying EEG-based VI recordings involving the imagination of 2D primitive geometric shapes (i.e. circles and triangles) with intra-class variations. This approach allows us to explore how the human mind mentally constructs basic shapes and assess the potential of modern DL techniques in decoding such abstract cognitive processes. We highlight the challenges associated with differentiating EEG signals corresponding to similar imagined shapes under conditions of limited training data. Two CNN variations are considered in this study as they have proven to be consistently the top performers for BCI applications: Filter-Bank Convolutional Network (FBCNet) [10] and EEGNet-8,2 [18]. Both models report to be effective in data-limited scenarios, which is the case for most BCI studies including the present one. These two CNNs are employed for end-to-end learning from raw EEG time-series data. We limit the data collection time to reduce participant fatigue and attempt to mitigate such data limitations with a data augmentation strategy using crop windows. Our results demonstrate that, despite the inherent difficulty of intra-class VI classification, DL models can achieve meaningful discrimination between imagined shapes. Additionally, we visualize class-specific cortical activity using topographic brain maps to identify distinct spatial patterns associated with each imagery class. To the best of our knowledge, this is the first study to apply DL methods to full EEG waveforms for classifying imagined 2D geometric shapes.

Materials And Methods

Model architectures

We use two of the most consistent SOTA DL models to decode various EEG-BCI paradigm datasets: FBCNet and EEGNet-8,2. These models are specifically designed to cater to the characteristics of EEG physiological variables in terms of spatio-temporal processing compared to other basic CNN models. Both models claim to be effective in data-limited scenarios, which is the case for most BCI studies including the present one. This is the first implementation of both architectures in a VI-related study. The main characteristics of the DL models implemented in this study are described as follows:

i. EEGNet-8,2 (which will be referred to as EEGNet in the rest of this manuscript) is a compact CNN model designed for raw EEG data inputs, featuring a unique combination of convolutional blocks, which allows for the interpretation of high-level features while minimizing model parameters. This is done by extracting information using a temporal kernel, followed by a depth-wise convolution and subsequent separable convolution. Finally, a classification block takes the resulting features extracted by the convolutional layers and determines the class with a softmax function. EEGNet has been proven to perform well over a variety of BCI paradigms within the event-related potentials and oscillatory component modalities, including sensory motor rhythms often studied in MI-BCI [18].

ii. FBCNet is a multi-input CNN designed to extract spectro-spatial discriminative information with a 'multi-view' representation of the EEG data. It uses multiple narrowband filters to produce a spatial convolution with various kernels to effectively extract discriminative features. In the convolutional kernel section, it also performs spatial convolution, followed by a temporal variance convolution feature extraction layer. Finally, it uses a log activation function and analyzes the extracted features with dense layers, consolidating them into the final classification layer for output. FBCNet has outperformed other SOTA DL models on a variety of MI datasets [10].

Data acquisition

In this study, we gathered EEG recordings from human subjects while they imagine or create a mental representation of primitive shapes in 2D. Throughout the experiment, participants are asked to imagine either a circle or a triangle after a given cue. The shapes or images selected for this experiment correspond to intra-categorical images with controlled figure variables (i.e. colors and size), thus minimizing the feature difference between classes to test DL model performance under difficult discriminative conditions. Figure *1a* shows the geometric shapes presented to the participants at the start of each session as examples of the type of images to mentally recreate after each cue for this experiment. The selected images are well-known primitive shapes with similar black background and white lines contrast, only varying in shape characteristics such as edges or roundness. However, the actual cues given throughout the experiment are the names of the corresponding shapes (also in black and white), as shown in Figure *1b*. This cue or stimulus modality is preferred to avoid the possibility of the mental task becoming the retention of the visual cue (which would be the case if shapes were used) instead of the task being the mental creation of the image. Previous research also supports that model performance is not affected by the stimulus modality [12,19].

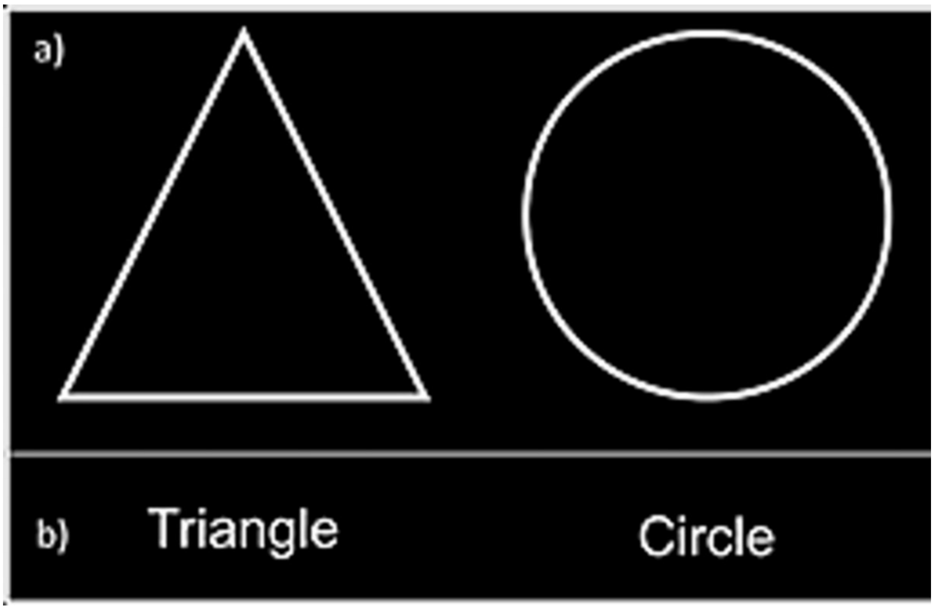

FIGURE 1: (a) Examples of figures shown to the participant before VI data capture, and (b) examples of visual cue words used during the VI session.

VI, Visual Imagery

EEG data is recorded with the Emotiv EPOC Model 1.0 [20] neuroheadset developed by Emotiv. It is composed of 14 wet electrode channels oriented following the international 10-20 sensor location system. The Emotiv EPOCH is a relatively affordable and portable EEG interpreter with capabilities very similar to other medical grade counterparts [21]. The experimental sampling rate is set to 128 Hz, which is on the low-end capability of EEG devices typically used in most studies. The device also has an integrated 50/60 Hz notch filter to eliminate powerline noise artifacts, and a digital fifth-order Sinc filter for smoothing the signal. The Emotiv headset uses the EmotivPro software [22] to allow third-party softwares to interact with EEG data, as well as recording in real time. We use lab streaming layer (LSL) [23] to stream EEG recordings from the neural headset in real time and collect the data stream with OpenViBe [24] during the experiment. While collecting data, OpenViBe is scripted to present the experimental sequence cues and label the EEG recordings to mark the onset of each stimulus throughout the experiment. A schematic of the data acquisition system with the package communication streamline for stimulus synchronization in our VI experiment is shown in Figure 2.

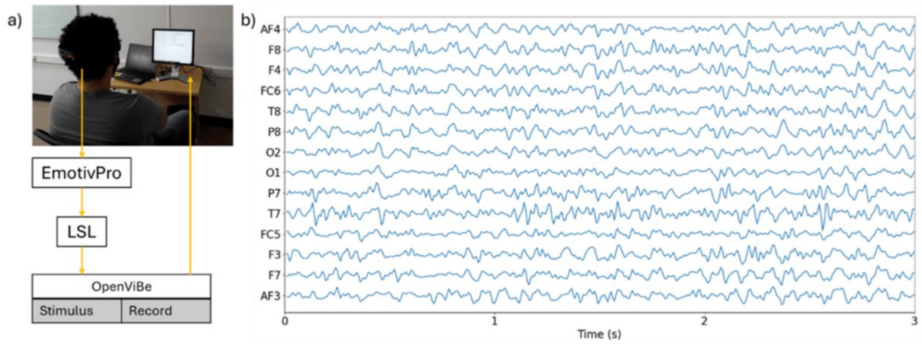

FIGURE 2: (a) Experimental station for data acquisition and stimulus presentation schematic. (b) Example of 3 seconds of EEG recordings.

EEG, electroencephalography; LSL, Lab Streaming Layer

Ten engineering students (five males and five females; 22-32 years old) participated in this study. Subjects had no perceptual or cognitive impairments, and they had normal or corrected vision. The data acquisition experiment was carried out in two sessions done in separate days, each consisting of six randomized sequences for each participant with 15 trials per sequence. This experiment was performed under the oversight and approval of the University of Puerto Rico Mayaguez Institutional Review Board (protocol number 2021020006). Safety information and task instructions for the experiment were given to each subject, and they signed an informed consent before the experiment. The participant was accommodated in the experimental station at least 3 feet away from the monitor and the neural headset was placed, making sure there was good contact on all electrode sensors. During the experiment, a randomized text cue would appear in the screen for 3 seconds and then disappear, leaving an empty black screen for 5 seconds. During those 5 seconds, participants imagined the figure corresponding to the given cue. This time arrangement constitutes 1 trial. Participants were specifically asked to think of the object appearing or projecting inside the screen. This helps constrain the given task to a finite space (also helping with consistency of the VI) and suggest a more objective task with real intent in creating the desired object on a specific place as one would perform with a BCI rather than by just free thinking. Afterwards, a new cue will appear to begin the next trial. Figure 3 shows a sequence schedule example. The process is repeated until all trials in the sequence have been completed. The sensor contact quality is verified in between sequences and corrected by adding more saline solution if necessary. Each cue class appears five times in each sequence.

Compared to most BCI studies, this work counts with less sample trials per class (60 trials per class). This was intentionally designed to evaluate the possibility of training a capable DL model with limited training samples, which would result in reduced user burden in data acquisition, as well as a reduction in training time. This is beneficial as it allows the user to quickly retrain or calibrate the DL model whenever necessary to mitigate external variables from the BCI system that could affect performance. These external variables include changes in the mental state, changes in dynamic activation of neural synapses, or deterioration of the EEG device sensors. To address this limitation, we employed the cropped training strategy [25], which creates small time windows within each trial and shifts the window over time to generate additional time-variant data samples, as permissible by our experimental design. Further details of this processing method are presented in the next section.

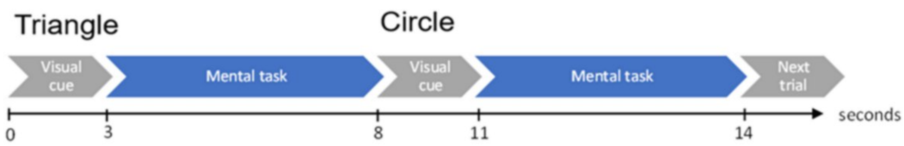

FIGURE 3: Example of EEG signal recording for each of the randomized sequences for experimental VI data acquisition.

EEG, electroencephalography; VI, Visual Imagery

Metrics

In this section, we describe the data processing and the DL experimental setup used in this study. With 5 seconds of data acquisition per trial, we slice 3-second data windows from each trial, detrend, and standardize each signal to be used as inputs by our DL models. This data window size was preferred due to our EGG device sampling rate, allowing sufficient datapoints in the window for adequate feature extraction. The first sample window extracted at each trial starts when the first text queue disappears from the screen. Then, we shift the window forward by 250 milliseconds, resulting in a total of nine training samples per trial. All channel signals *X* of each trial are normalized individually using Z-score defined as $Z = \frac{X - \mu}{\sigma}$. This entails detrending the mean value μ and dividing the signal by the standard deviation σ of each channel, allowing to transform the data so that it has a mean value of 0 and a standard deviation of 1.

Finally, all samples are temporally filtered with a sixth-order Butterworth bandpass filter (4-50 Hz). This frequency range is suitable as it removes signals below 4 Hz corresponding to noise and relaxed mental states and removes signals above 50 Hz related to muscle movement or electromyography. Therefore, the main artefactual concern can come from signals related to eye movement or electro-oculography since their frequency range tends to overlap those from EEG. However, with our intent to minimally process the data for fast real-time applications, we do not condition the data any further and allow models to overcome such artifacts themselves through the training data observations.

We evaluate the performance of our DL models based on the classification accuracy over the testing dataset. We define the accuracy as:

$$[\text{Accuracy} = \frac{TP + TN}{P + N}]$$

where *TP* is the true positive (correctly classified sample belonging to the predicted class), *TN* is true negative (correctly classified samples not belonging to the target class), *P* number of positives or total correct predictions, and *N* number of negatives or total false predictions.

Training parameters

We compared the performance of two DL models (FBCNet and EEGNet) over our experimental VI dataset. All models were generated, trained, and tested using the TorchEEG [26] library from PyTorch [27]. We trained our models using k-fold cross-validation (CV) with k = 5. Our data samples are randomly split into 80% of the data for training and 20% for testing, where the training set is used for the CV optimization, and the testing set is used to evaluate the best trained model. The best trained model saved during training is determined by the lowest validation loss obtained during the CV. Each k-fold comprises 500 optimization steps or epochs.

All models are optimized with an Adam optimizer with a learning rate of 1e-4 and have cross-entropy loss implemented. No further data processing is done for the EEGNet inputs. On the other hand, FBCNet utilizes the multi-view representation of the input streams using narrow band frequency filters. Here, we use four common EEG frequency ranges associated with active or conscious mental states, namely, theta (4-8 Hz), alpha (8-13 Hz), beta (13-30 Hz), and gamma (higher than 30 Hz). For the gamma rhythm, we use a 30-50 Hz range filter. These rhythms are known to be associated with a specific mental state as follows:

i. Theta: A state of daydreaming or drowsiness; it may also be present during periods of emotional stress [28].

ii. Alpha: Cognitive processes of attention and memory [29].

iii. Beta: Intense mental concentration or intellectual activity [28].

iv. Gamma: A state of consciousness regarding attention, perception, and cognition [28].

We perform an intra-session, subject-specific analysis to assess different DL model performance over our VI dataset. In the next section, we compare the progression in terms of optimization for each model, the performance over unseen data for the average optimized models, and the brain activations associated with each class.

Artifact removal

For better understanding of the spatial patterns of brain activity, EEG data are often summarized using topographic head maps. To generate the topographic map features, signal samples were first processed to remove as many artifacts signals possible and enhance the visualization of underlying activations or patterns associated with each class. This is achieved by implementing independent component analysis

(ICA) [30] over each sample's mixed signal X to separate the signal into independent sources S over a discrete time t_n. Assuming there are equal number of signals as there are sources, determined by the number of channels C from our EEG sensor, then X and S are dimensions $C \times t_n$. If we consider a mixing matrix W with dimensions $C \times C$, we can describe the mixture of source signals received by each recording channels as:

$$[X(t) = W \cdot S(t)]$$

Thus, we can estimate the independent sources of the data y, solving:

$$[S(t) = W^{-1} \cdot X(t)]$$

Where W^{-1} is the inverse of W, otherwise known as the unmixing matrix. Following the method outlined by Lindsen and Bhattacharya [31], artefactual components were removed. We apply empirical mode decomposition to each channel of S to break down non-linear and non-stationary signals into a set of simpler components called Intrinsic Mode Functions (IMFs). Therefore, our sources can be described as a sum of N number of IMF's u and a residue r in the following way:

$$[S(t) = \sum_{i=1}^{n} u_i(t) + r(t)]$$

The standard deviation of artifact related components is generally much larger than those of EEG-related components. Therefore, we calculate the standard deviation of the first component u and use it as reference to compare with other components from that independent source. If the standard deviation of the component u exceeded 2 times the value of the reference, then that component u is eliminated. Then, the remaining components and r are added together to form an artifact free source S^*. This process is repeated with all the independent sources. Finally, we mix the sources once more to obtain an artifact free signal X^* for any given trial.

For each trial, we extract the mean power spectral density (PSD) of the four frequency bands of interest on the mixed signal X^*. Finally, we average the PSD features across the trial in each class and generate the topological head map.

Results And Discussion

Results

Experimental Setup of VI-BCI Models

Figure *4* presents the workflow for evaluating our VI-BCI experiment using two DL models, as well as processing and interpretation of our data. EEG data is recorded using the Emotiv EPOC Model 1.0 neuroheadset, offering a conveniently portable option for BCI. A combination of specialized software packages (EmotivPro, LSL and OpenVibe) manage the acquisition of data while simultaneously presenting the participants with the stimuli in the corresponding sequences. Thereafter, the data is prepared or processed by cleaning signals outside of the 4-50 Hz range and extracting the windows of each trial corresponding to VI. These samples are then introduced to the DL models for proper classification of the object design intent.

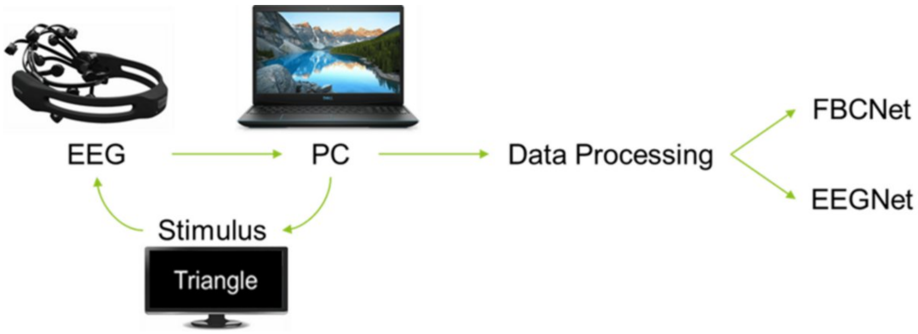

FIGURE 4: Experimental setup for the VI-BCI.

BCI, Brain-Computer Interface; EEG, Electroencephalography; FBCNet, Filter-Bank Convolutional Network; VI, Visual Imagery

Evaluation of VI-BCI Models

We evaluate each model's accuracy over samples unseen by our models during the training phase to assess the reliability of each model with future or online case samples. Table *1* presents the average performance of our optimized models over the testing dataset for each participant. The average performance across all participants for FBCNet is 51.3± 2.1%, while EEGNet accuracy is 50.3± 0.7%. With a two-way ANOVA analysis for prediction accuracy between the two models, we get a p-value of 0.17008, indicating no significant statistical difference between classification accuracy for both models. The training time for FBCNet (4.67 hours) was slightly longer than EEGNet (4.5 hours) for one participant's dataset on a GTX 1050 Ti GPU. Once trained, both models are capable of processing and predicting 216 testing samples in under 2 seconds.

	FBCNet						EEGNet					
Participant	k=1	k=2	k=3	k=4	k=5	Ave. Accuracy	k=1	k=2	k=3	k=4	k=5	Ave. Accuracy
p1	0.523	0.528	0.560	0.556	0.468	0.527	0.495	0.500	0.519	0.500	0.500	0.503
p2	0.528	0.537	0.514	0.519	0.514	0.522	0.519	0.500	0.500	0.505	0.514	0.507
p3	0.477	0.546	0.481	0.519	0.532	0.511	0.505	0.509	0.500	0.523	0.500	0.507
p4	0.505	0.454	0.514	0.458	0.491	0.484	0.560	0.509	0.532	0.458	0.532	0.519
p5	0.449	0.472	0.505	0.454	0.463	0.469	0.495	0.514	0.468	0.500	0.491	0.494
p6	0.560	0.505	0.477	0.500	0.537	0.516	0.500	0.491	0.500	0.514	0.500	0.501
p7	0.491	0.523	0.616	0.514	0.523	0.533	0.491	0.500	0.500	0.500	0.486	0.495
p8	0.523	0.551	0.514	0.495	0.500	0.517	0.500	0.491	0.500	0.500	0.519	0.502
p9	0.509	0.463	0.486	0.579	0.574	0.522	0.500	0.500	0.500	0.509	0.500	0.502
p10	0.546	0.528	0.616	0.481	0.468	0.528	0.500	0.500	0.500	0.500	0.505	0.501
Average:	0.511	0.511	0.528	0.507	0.507	0.513	0.506	0.501	0.502	0.501	0.505	0.503
STD:	0.033	0.036	0.052	0.039	0.036	0.021	0.020	0.008	0.016	0.017	0.014	0.007

TABLE 1: Comparison of FBCNet and EEGNet over the VI testing dataset for subject-specific optimized models.

FBCNet, Filter-Bank Convolutional Network; STD, Standard Deviation; VI, Visual Imagery

To examine class-specific neural patterns elicited during VI, we visualized the average brain activity across four standard EEG frequency bands. Figure *5* presents topographic head plots depicting PSD distributions for each frequency band averaged over 60 trials per class for a single participant. These plots reflect cortical activation during the imagery-maintenance phase, providing spatial insight into the neural correlates of imagining different 2D shapes. By comparing frequency-specific brain activity, we aim to identify distinctive topographical signatures associated with each VI class.

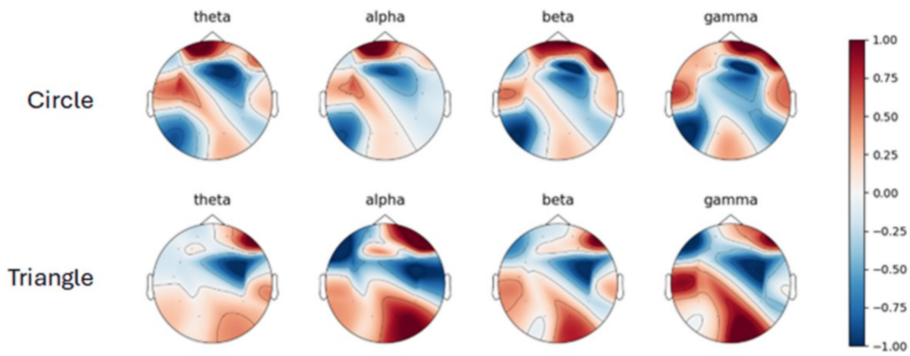

FIGURE 5: Participant 3 brain activation for the independent frequency bands during each VI task.

VI, Visual Imagery

Comparison with previous studies

In this section, our work is compared with similar VI studies to assess its contributions and results. As demonstrated by Table *2*, there are not many studies available for VI, which limits our base knowledge about mental behavior for VI as a BCI paradigm. We test the limitations of SOTA DL models for classifying intra-class 2D geometric shapes under very limited conditions. Compared to other studies, here we use fewer training samples (60 samples per class) to achieve reasonable data acquisition times. We also rely on a relatively affordable commercial-grade EEG device with a maximum of 14 channels, only used by one other study [12] for a similar task. In terms of classification accuracy, our experimentation yielded results closest to chance probability performance compared to other studies.

Publication	VI Task	EEG Device	Accuracy	Features	Best Classification Model
This work	Circle, triangle	Emotiv EPOC	51.30	Raw	CNN
[12]	Cube, sphere, cone, pyramid, cylinder	Emotiv EPOC	44.60	HHT	LDA
[16]	Flower, hammer	g.USBAmp	52.00	Power spectrum	SpecCSP
[17]	Tree, house, plane, dog	g.Nautilus	60.50	PSD	CNN+GA
[15]	Static star and star moving right	NT9200	78.40	HHT	SVM
[32]	Drone swarm: spread-out, fall in, and hovering	BrainAmp	72.00	Raw	CNN

TABLE 2: Comparison of previous EEG studies focused on VI of objects.

Raw refers to no transformation.

CNN, Convolutional Neural Networks; EEG, electroencephalography; GA, Genetic Algorithm; HHT, Hilbert–Huang Transform; LDA, Linear Discriminant Analysis; PSD, Power Spectral Density; SpecCSP, Spectrally Weighted Common Spatial Patterns; SVM, Support Vector Machine; VI, Visual Imagery

Discussion

The ability to differentiate user intents though brain activity is imperative in signal processing models for the implementation of BCI. This is particularly challenging when performing BCI tasks with a scarce number of studies, such as VI. Thus, experimenting with different aspects for this particular task can provide insight into making this a more reliable BCI paradigm. We demonstrate the performance of DL models in classifying a VI-related EEG dataset with limited data samples. Our experimental data consists of EEG data acquisition from participants during voluntary exercise of imagination or mental creation of primitive figures (circles and triangles). This is different to other VI studies as the mental images considered here are intentionally very similar in many aspects except in the geometric shape. This is important as we intended to minimize the variability of the classes to evaluate the extent to which SOTA EEG-BCI models can map the separability of classes under such conditions. Furthermore, our dataset is considerably smaller in sample volume and number of sensor channels than other open-source datasets and other VI-BCI studies, adding to the challenges to be overcome by the DL models.

We evaluated two SOTA DL models (FBCNet and EEGNet) in an offline VI-BCI study. With a p-value of 0.17008 between classification results over all participants with both models, it is clear that there is no significant statistical difference between the methods used in this particular experimental setup. However, we can acknowledge some particularities in our results. FBCNet achieved higher classification accuracy for most participants compared to EEGNet. That was reflected in the prediction accuracies shown in Table *1*, where FBCNet achieves an average classification accuracy across all participants of 51.3%, compared to 50.3% from EEGNet. Nonetheless, this would suggest that neither model is fully capable of extracting VI-related physiological patterns in our dataset to distinguish between our intra-class tasks. We believe our results may have been greatly affected by our data acquisition system, and better classification for this study could be achieved using high-quality medical-grade EEG systems. Other studies describe having similar acquisition quality with the Emotiv EPOC [33]. Nonetheless, since our device has proven to be effective in other BCI paradigms [34] and tends to have advantages (in terms of affordability and non-technical usability), it is worth performing further experimentations with it. Further VI experimentation may include evaluating the model performance using additional classes in hope to improve the class separability by analyzing more feature spaces or even incorporating complex 2D images to induce higher visualization focus on the given task.

The training duration for FBCNet proved to be only a few minutes longer than EEGNet in processing all five folds on a single participant. This is impressive considering FBCNet is a much larger model with 3,331 trainable parameters, while EEGNet has 1,904 trainable parameters optimized during each epoch. In other words, despite being a more memory-consuming model, there is not much sacrifice in the required training time to achieve better results under the same study conditions. Furthermore, the trained models are capable of efficient classification prediction at a reasonable rate, making their response time ideal for BCI implementations.

After performing the artifact removal processing, we generated topological head plots of average PSD activation for multiple frequency bands. Despite the efforts, the high energies being depicted in the frontal channels suggest that remnants of eyeblink artefacts may still be present in the signals. We believe our ICA implementation was not able to optimize the separability of the independent components due to the quality of our acquired data. Therefore, our artifact removal method would only be capable of removing as many artifacts as permissible by the best estimated unmixed ICs. Nonetheless, we can recognize certain patterns from our results.

Strong similarities between classes in the topological maps were expected since both mental tasks probably recruit neural synapses originating from similar sources. Brain activation for both classes shows consistent energy drop around the central cortex of the brain and significant localized energy increase in the right occipital channels. This last observation is partly consistent with previous findings [35] reporting the right hemisphere as being more active during VI. Other reports [36] also suggest VI induces delta and activation in the prefrontal lobe, similar to our results. This suggests that a portion of the high energies developed in the frontal lobe could come from the VI task, but this would merit further investigation to confirm.

Conclusions

In this study, we perform a VI-BCI-related experiment, where we record EEG signals while participants imagine 2D geometric shapes and evaluate multiple SOTA DL model performances in decoding these recordings for accurate classification. We demonstrate that despite challenging conditions present in our dataset, such as limited data size and subtle differences between VI classes (circle and triangle), FBCNet consistently outperforms EEGNet, achieving better classification accuracy across participants. Further analysis shows that while artifact removal via ICA improved signal quality, residual eye-blink artifacts in the frontal regions suggest limitations in component separation, likely due to suboptimal data quality. As a result, the effectiveness of artifact removal was constrained by the degree to which the independent components could be unmixed.

Overall, our findings advocate for further exploration of advanced DL models like FBCNet in EEG-BCI systems to offer promising avenues for enhancing classification accuracy and reducing user training burden in VI applications. In future studies, a medical-grade EEG device that can provide higher quality and spatio-temporal data would be preferable to replicate this experiment. It would provide better insight into the subtle underlaying brain activity differences for this VI task and should gain improved classification results. Exploring classification results with different or additional objects, or other imagery related controlled features, would also be of interest to gain insight into VI characteristics and establish it as a reliable BCI paradigms.

Additional Information

Author Contributions

All authors have reviewed the final version to be published and agreed to be accountable for all aspects of the work.

Concept and design: Daniel Lizama, David Serrano

Acquisition, analysis, or interpretation of data: Daniel Lizama

Drafting of the manuscript: Daniel Lizama, David Serrano

Critical review of the manuscript for important intellectual content: Daniel Lizama, David Serrano

Disclosures

Human subjects: Consent was obtained or waived by all participants in this study. University of Puerto Rico Mayaguez Institutional Review Board issued approval 2021020006. Approval: Expedited. **Animal subjects:** All authors have confirmed that this study did not involve animal subjects or tissue. **Conflicts of interest:** In compliance with the ICMJE uniform disclosure form, all authors declare the following: **Payment/services info:** All authors have declared that no financial support was received from any organization for the submitted work. **Financial relationships:** All authors have declared that they have no financial relationships at present or within the previous three years with any organizations that might have an interest in the submitted work. **Other relationships:** All authors have declared that there are no other relationships or activities that could appear to have influenced the submitted work.

Acknowledgements

Data are available on reasonable request. The data are stored as de-identified participant data, which are available on request to daniel.lizama@upr.edu

References

1. Jerbi K, Vidal JR, Mattout J, et al.: Inferring hand movement kinematics from MEG, EEG and intracranial EEG: From brain-machine interfaces to motor rehabilitation. IRBM. 2011, 32:8-18. 10.1016/j.irbm.2010.12.004
2. Zhang J, Li K: A multi-view CNN encoding for motor imagery EEG signals. Biomedical Signal Processing and Control. 2023, 85:105063. 10.1016/j.bspc.2023.105063
3. Ma X, Qiu S, Wei W, Wang S, He H: Deep channel-correlation network for motor imagery decoding from the same limb. IEEE Transactions on Neural Systems and Rehabilitation Engineering. 2020, 28:297-306. 10.1109/tnsre.2019.2953121
4. Schwartz AB, Cui XT, Weber DJ, Moran DW: Brain-controlled interfaces: movement restoration with neural prosthetics. Neuron. 2006, 52:205-220. 10.1016/j.neuron.2006.09.019
5. Padfield N, Zabalza J, Zhao H, Masero V, Ren J: EEG-based brain-computer interfaces using motor-imagery: Techniques and challenges. Sensors. 2019, 19:1423. 10.3390/s19061423
6. Elstob D, Secco EL: A low cost EEG based BCI prosthetic using motor imagery. International Journal of Information Technology Convergence and Services. 2016, 6:23-36. 10.5121/ijitcs.2016.6103
7. Khan MA, Das R, Iversen HK, Puthusserypady S: Review on motor imagery based BCI systems for upper limb post-stroke neurorehabilitation: From designing to application. Computers in Biology and Medicine. 2020, 123:103843. 10.1016/j.compbiomed.2020.103843
8. Rao RPN: Brain-Computer Interfacing: An Introduction. Cambridge University Press, Cambridge, UK; 2013. 10.1017/CBO9781139032803
9. Ang KK, Chin ZY, Zhang H, Guan C: Filter bank common spatial pattern (FBCSP) in brain-computer interface. 2008 IEEE International Joint Conference on Neural Networks (IEEE World Congress on Computational Intelligence), Hong Kong, China. 2008, 2390-2397. 10.1109/IJCNN.2008.4634130
10. Mane R, Robinson N, Vinod AP, Lee S-W, Guan C: A multi-view CNN with novel variance layer for motor imagery brain computer interface. 2020 42nd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC), Montreal, QC, Canada. 2020, 2950-2953. 10.1109/embc44109.2020.9175874
11. Kwon O-Y, Lee M-H, Guan C, Lee S-W: Subject-independent brain-computer interfaces based on deep convolutional neural networks. IEEE Transactions on Neural Networks and Learning Systems. 2020, 31:3839-3852. 10.1109/TNNLS.2019.2946869
12. Esfahani ET: Investigation of Brain Computer Interface as a New Modality in Computer Aided Design/Engineering Systems. University of California, Riverside, CA; 2012.
13. Bobrov P, Frolov A, Cantor C, Fedulova I, Bakhnyan M, Zhavoronkov A: Brain-computer interface based on generation of visual images. PLoS ONE. 2011, 6:e20674. 10.1371/journal.pone.0020674
14. Neuper C, Scherer R, Reiner M, Pfurtscheller G: Imagery of motor actions: Differential effects of kinesthetic and visual-motor mode of imagery in single-trial EEG. Cognitive Brain Research. 2005, 25:668-677. 10.1016/j.cogbrainres.2005.08.014
15. Fu Y, Li Z, Gong A, Qian Q, Su L, Zhao L: Identification of visual imagery by electroencephalography based on empirical mode decomposition and an autoregressive model. Computational Intelligence and Neuroscience. 2022, 2022:1-10. 10.1155/2022/1038901
16. Kosmyna N, Lindgren JT, Lécuyer A: Attending to visual stimuli versus performing visual imagery as a control

strategy for EEG-based brain-computer interfaces. Scientific Reports. 2018, 8:13222. 10.1038/s41598-018-31472-9

17. Llorella FR, Patow G, Azorín JM: Convolutional neural networks and genetic algorithm for visual imagery classification. Physical and Engineering Sciences in Medicine. 2020, 43:973-983. 10.1007/s13246-020-00894-z
18. Lawhern VJ, Solon AJ, Waytowich NR, Gordon SM, Hung CP, Lance BJ: EEGNet: a compact convolutional neural network for EEG-based brain-computer interfaces. Journal of Neural Engineering. 2018, 15:056013. 10.1088/1741-2552/aace8c
19. Suppes P, Han B, Epelboim J, Lu Z: Invariance of brain-wave representations of simple visual images and their names. Proceedings of the National Academy of Sciences. 1999, 96:14658-14663. 10.1073/pnas.96.25.14658
20. Emotiv. Accessed: June 28, 2021: https://www.emotiv.com./.
21. Duvinage M, Castermans T, Petieau M, Hoellinger T, Cheron G, Dutoit T: Performance of the Emotiv Epoc headset for P300-based applications. BioMedical Engineering OnLine. 2013, 12:56. 10.1186/1475-925x-12-56
22. Emotiv: EmotivPro Software, Version 3.2.3.420. https://www.emotiv.com/emotivpro/.
23. Kothe C: Lab Streaming Layer (LSL). Accessed: May 27, 2022: https://github.com/sccn/labstreaminglayer.
24. Renard Y, Lotte F, Gibert G, et al.: OpenViBE: An open-source software platform to design, test, and use brain-computer interfaces in real and virtual environments. Presence: Teleoperators and Virtual Environments. 2010, 19:35-53. 10.1162/pres.19.1.35
25. Schirrmeister RT, Springenberg JT, Fiederer LDJ, et al.: Deep learning with convolutional neural networks for EEG decoding and visualization. Human Brain Mapping. 2017, 38:5391-5420. 10.1002/hbm.23730
26. Zhang Z, Zhong S, Liu Y: TorchEEGEMO: A deep learning toolbox towards EEG-based emotion recognition. Expert Systems with Applications. 2024, 249:123550. 10.1016/j.eswa.2024.123550
27. Paszke A, Gross S, Massa F, et al.: PyTorch: An imperative style, high-performance deep learning library. arXiv. 2019, 10.48550/arXiv.1912.01703
28. Schomer DL: The normal EEG in an adult. The Clinical Neurophysiology Primer. Blum AS, Rutkove SB (ed): Humana Press, Totowa, NJ; 2007. 57-71. 10.1007/978-1-59745-271-7_5
29. Klimesch W, Schimke H, Pfurtscheller G: Alpha frequency, cognitive load and memory performance. Brain Topography. 1993, 5:241-251. 10.1007/bf01128991
30. Najmi AH, Moon TK: Advanced Signal Processing: A Concise Guide. McGraw-Hill Education, New York, NY, USA; 2020.
31. Lindsen JP, Bhattacharya J: Correction of blink artifacts using independent component analysis and empirical mode decomposition. Psychophysiology. 2010, 47:955-960. 10.1111/j.1469-8986.2010.00995.x
32. Lee D-H, Han D-K, Kim S-J, Jeong J-H, Lee S-W: Subject-independent brain-computer interface for decoding high-level visual imagery tasks. 2021 IEEE International Conference on Systems, Man, and Cybernetics (SMC), Melbourne, Australia. 2021, 3396-3401. 10.1109/SMC52423.2021.9659151
33. Fakhruzzaman MN, Riksakomara E, Suryotrisongko H: EEG wave identification in human brain with Emotiv EPOC for motor imagery. Procedia Computer Science. 2015, 72:269-276. 10.1016/j.procs.2015.12.140
34. Sumardi T, Afriyanti A, Suhendra MA, Ajidarmo S, Basri MI, Amri MF: Classifying EEG-based motor imagery for drone control using artificial neural network. 2024 IEEE International Conference on Smart Mechatronics (ICSMech), Yogyakarta, Indonesia. 2024, 141-145. 10.1109/ICSMech62936.2024.10812269
35. Mellet E, Tzourio N, Denis M, Mazoyer B: A positron emission tomography study of visual and mental spatial exploration. Journal of Cognitive Neuroscience. 1995, 7:433-445. 10.1162/jocn.1995.7.4.433
36. Sousa T, Amaral C, Andrade J, Pires G, Nunes UJ, Castelo-Branco M: Pure visual imagery as a potential approach to achieve three classes of control for implementation of BCI in non-motor disorders. Journal of Neural Engineering. 2017, 14:046026. 10.1088/1741-2552/aa70ac