



Article Exploring the Ability to Classify Visual Perception and Visual Imagery EEG Data: Toward an Intuitive BCI System

Sunghan Lee D, Sehyeon Jang and Sung Chan Jun *D

School of Electrical Engineering and Computer Science, Gwangju Institute of Science and Technology, Gwangju 61005, Korea

* Correspondence: scjun@gist.ac.kr

Abstract: Providing an intuitive interface for the actual use of brain-computer interface (BCI) can increase BCI users' convenience greatly. We explored the possibility that visual imagery can be used as a paradigm that may constitute a more intuitive, active BCI. To do so, electroencephalography (EEG) data were collected during visual perception and imagery experiments. Three image categories (object, digit, shape) and three different images per category were used as visual stimuli. EEG data from seven subjects were used in this work. Three types of visual perception/imagery EEG data were preprocessed for classification: raw time series data; time-frequency maps; and common spatial pattern (CSP). Five types of classifiers (EEGNet, 1D convolutional neural network (CNN), Multi-Rocket, MobileNet, support vector machine (SVM)) were applied to each applicable data type among the three preprocessed types. Thus, we investigated the feasibility of classifying three-category or nine-class visual perception/imagery over various classifiers and preprocessed data types. We found that the MultiRocket network showed the best classification performance: yielding approximately 57.02% (max 63.62%) for three-category classification in visual perception and approximately 46.43% (max 71.38%) accuracy for three-category classification in visual imagery. However, no meaningfully improved performance was achieved in the nine-class classification in either visual perception or imagery, although visual perception yielded slightly higher accuracy than visual imagery. From our extensive investigation, we found that visual perception and visual imagery data may be classified; however, it is somewhat doubtful whether either may be applicable to an actual BCI system. It is believed that introducing better-designed advanced deep learning networks together with more informative feature extractions may improve the performance of EEG visual perception/imagery classifications. In addition, a more sophisticated experimental design paradigm may enhance the potential to achieve more intuitive visual imagery BCI.

Keywords: electroencephalography (EEG); visual imagery (VI); visual perception (VP); brain–computer interface (BCI); convolutional neural network (CNN); artificial neural network (ANN)

1. Introduction

Brain–computer interface (BCI) is a technology that allows users to control a machine or computer by decoding their intentions through brain activity without an input device such as a keyboard. BCI can be categorized into active, reactive, and passive according to the operating principle [1].

Among them, active BCI is a technique that allows users to communicate their intentions consciously, and motor imagery (MI) is a typical paradigm in active BCI. MI controls the device largely by distinguishing the corresponding brain signals from the user's imagination of moving the left or right arms, feet, or tongue. Generally, MI BCI classifies two classes of left and right arm imagery and often classifies up to four by adding the imagery of feet and tongue. Although this is an intuitive way to actually perform left–right movement, such as steering a wheelchair, the number of classes imagined is limited, thereby making it useful for simple communication only. In addition, users find



Citation: Lee, S.; Jang, S.; Jun, S.C. Exploring the Ability to Classify Visual Perception and Visual Imagery EEG Data: Toward an Intuitive BCI System. *Electronics* **2022**, *11*, 2706. https://doi.org/10.3390/ electronics11172706

Academic Editor: Sandra Gama

Received: 18 July 2022 Accepted: 25 August 2022 Published: 29 August 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2022 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/). that imagining movement without actually moving the limb is quite difficult. Therefore, various MI studies have reported that many users need to learn a specific imagery method or skill to imagine movement; neurofeedback training related to motion or alpha wave control [2,3], the use of virtual reality (VR) to enhance visual immersion and facilitate imagination [4–6], assistive motor-related stimulations, such as steady-state sensory-evoked potential (SSSEP), and electrical stimulation [7–10] have been proposed to enhance the ability to use a BCI system. However, approximately 15–30% of users find it difficult to perform MI BCI despite training [11]. Particularly, patients appear to demonstrate lower performance in MI BCI than healthy people.

On the other hand, the paradigms of steady-state visual-evoked potential (SSVEP) and P300, used primarily in reactive BCI, require external stimuli. Reactive BCI decodes the user's intention by discriminating the brain's reactive signal from the external stimuli and communicates a message through a speller technique that can create words or sentences. Furthermore, to convey intentions in the current BCI system, it is common to use SSVEP, while the P300 takes the indirect form of a speller to map a word or sentence one letter at a time, although such BCI paradigms are not very intuitive.

Visual imagery may have the potential to break through the limitations of the traditional aforementioned BCI method. If it were possible to classify visual imagery electroencephalography (EEG) signals, the user's intention could be conveyed as it is directly, rather than creating words and sentences by combining letters one by one, and the intention could be delivered more intuitively and quickly. As such, BCI through visual imagery has the potential to improve the degree of freedom of manipulation significantly compared to MI. In addition, the introduction of a paradigm that is more intuitive and simpler for BCI-illiterate or patient users who have difficulty using such an MI paradigm may improve the ability to use BCI. In addition, visual imagery-based BCI has the potential to improve the information transfer rate (ITR) significantly, not only for patients but also for healthy users, because it can deliver intentions more quickly and efficiently than the current BCI technique.

However, despite these many advantages, BCI that uses visual imagery is a relatively untested method and remains challenging. There are many obstacles, including an unestablished experimental paradigm, but one of the main causes may be that feature analysis, extraction, and classification techniques for visual imagery BCI have not yet been established well. The majority of BCI paradigms currently used primarily have distinct and well-known features. For example, in MI, a specific sensory motor rhythm (SMR) is observed in the sensory–motor area in the opposite hemispheres when imagining left and right arm movement. Furthermore, in SSVEP, the frequency at which the user is looking and the same frequency as its harmonics are observed in the occipital lobe. There are also effective classifiers that can extract features for a specific paradigm: common spatial pattern (CSP) in MI [12], and canonical correlation analysis (CCA) in SSVEP [13,14].

On the other hand, decoding brain signals through visual imagery BCI using noninvasive methods and its classifications are relatively recent approaches since the development of deep learning, and effective methods for extracting features are relatively unknown.

Therefore, in this work, our primary goal was to determine the potential of using visual imagery EEG as a more intuitive active BCI paradigm. To do so, we designed visual imagery experiments and collected EEG signals. The visual imagery EEG signals were preprocessed in time series, time–frequency maps, and CSP feature vectors. Furthermore, several deep neural network classifiers applicable to each preprocessed data type were compared to investigate whether visual imagery EEG can be classified. For comparison, visual perception EEG was processed and tested similarly.

We tried to explore the possibility of classifying visual imagery BCI through the EEG and various ML techniques, but unfortunately, the results we obtained were still insufficiently accurate. Moreover, visual imagery has many limitations and problems for it to be applied to a practical BCI system. However, the BCI system using visual imagery is still in the early stages of research, and our findings are expected to contribute to research teams seeking breakthroughs in noninvasive BCI, particularly those attempting

to discover more intuitive BCI systems. The visual imagery-based BCI can be an easier and more intuitive information delivery technique for BCI-illiterate or patient users who have difficulty using traditional BCI, and even for healthy users. We expect that visual imagery can be developed into a future BCI paradigm through sophisticated paradigm correction and feature extraction optimization.

In Section 3, the visual imagery experiment, EEG data processing and techniques used to extract features and classify visual imagery EEG signals are explored. Section 4 presents the classification results of visual perception and visual imagery through the techniques used in Section 3. Section 5 addresses similar studies on decoding visual information through brain signals, the limitations of this study and ways to improve them, and conclusions are presented in Section 6. In Section 2, which follows immediately, the latest efforts to decode visual information through noninvasive techniques, primarily of functional magnetic resonance imaging (fMRI) and EEG, are explored in depth.

2. Related Works

Efforts to decode human visual perception and imagery through brain signals have been long-standing challenges. Recently, with the development of deep learning approaches, more diverse attempts have been introduced to decode and reconstruct visual information through brain signals. Specifically, the use fMRI and EEG in such efforts are summarized in the following.

2.1. fMRI

Research on decoding visual brain signals with fMRI continues very actively, and after the great success of deep learning, there have been various attempts to do so in the field of neural engineering, as well as machine learning (ML). Kay et al. recorded the fMRI signal through an experiment that showed natural images and reconstructed the visual signal [15]. Using the same dataset as Kay et al.'s [16], Naselaris et al. attempted to reconstruct natural images through the Bayesian approach [17]. The reconstruction of visual information was attempted in the field of static images and even moving images, i.e., movies. In general, fMRI's blood oxygen level-dependent (BOLD) signals have a very low sampling rate; thus, it is inherently difficult to reconstruct rapidly changing dynamic images. Nishimoto et al. attempted to restore a dynamic image by proposing a motion-energy encoding model [18]. Furthermore, Horikawa et al. reported an attempt to read dreams by decoding the fMRI signal during sleep and showed a pattern similar to the image seen just before sleep [19]. In subsequent work, fMRI data were obtained while conducting experiments on visual perception and visual imagery for 150 classes of ImageNet fMRI [20,21]. Using two datasets—dream fMRI [19] and ImageNet fMRI [21]—Horikawa's team found that the value decoded from the dream fMRI data was correlated positively with the value related to the dream category in the upper-level deep neural network (DNN) layer [22], which supports the possibility of decoding a visual object during sleep. Since the release of an fMRI dataset using ImageNet images, many studies have been conducted to reconstruct visual perception and imagery. Shen et al. decoded fMRI information using DNN and generative models. ImageNet fMRI data were used for training, and artificial images or letters unseen in the training phase were used as test sets. They showed that the model generated can generalize visual information processing even for unseen data [23]. Recently, they also reconstructed visual fMRI signals with an end-to-end approach using deep learning techniques [24]. Beily and his team reported an unsupervised reconstruction approach to visual information using an encoder-decoder network architecture [25]. In addition, a study to reconstruct a facial image using the generative model was reported [26]. Thus far, various attempts to restore the visual experience have been made continuously.

2.2. MEG

To measure brain signals using magnetoencephalography (MEG), a shielding room that can shield out the Earth's magnetic field [27] and a superconducting quantum inter-

ference device (SQUID) sensor [28] that can measure the extremely fine magnetic field of the human brain are required. As such, MEG requires a high level of technology, and the amount of equipment required is enormous and expensive. Therefore, in general, relatively few cases of BCI research have used MEG. Among them, as far as we know, Kim et al.'s research is the only attempt to classify brain signals according to visual stimuli [29]. A network referred to as CANet was proposed, and classification accuracy of 91.96% was obtained in ImageNet [20] for object and handwritten digit category binary classification.

2.3. EEG

EEG has a relatively higher temporal resolution than fMRI, although it has a weaker spatial resolution. Nevertheless, remarkable work to decode visual perception and visual imagery with EEG has been reported in recent years. Spampinato et al. used a 128-channel EEG system to collect brain activity from 6 subjects while 40 kinds of object (from ImageNet) image stimuli were presented. Their proposed RNN-based EEG classification approach yielded a mean accuracy of approximately 83% [30]. Furthermore, Palazzo et al. proposed EEG-ChannelNet [31], a model for learning the brain manifold for EEG classification, applied it to the same 40-class EEG datasets, and achieved 48.1% accuracy. Zheng et al. proposed an attention-based long short-term memory (LSTM) network and achieved a classification rate of 99.50% for 40 classes [32]. Ling et al. recorded EEG from 14 subjects during a word–image stimulation experiment and reported the ability to decode and reconstruct words based upon EEG [33].

Among the efforts in visual imagery, Kumar et al. recorded EEG data from 23 subjects using Emotiv EPOC + during a 10-class (consisting of 10 numbers, 10 letters, or 10 objects' images) visual imagery experiment, and reported a classification accuracy of 85.20% using the random forest technique [34]. Tirupattur et al. proposed a network referred to as ThoughtViz to reconstruct imagined images from EEG signals using the generative adversarial network (GAN); they used Kumar et al.'s dataset [34] and reported approximately 72% accuracy in 10-class classification [35]. Bang et al. recorded EEG from four subjects during an imagery experiment with six shapes. They used convolutional neural network (CNN) for classification and achieved a six-class classification performance of 32.56% in visual perception. In addition, the binary classification of visual perception and visual imagery showed a classification rate of 90.16% [36]. Recent work has reported remarkable performance with a massive EEG visual image dataset using the Choi–Williams time-frequency distribution (CWD) [37] technique. Sixteen subjects participated in the study and EEG was recorded during the visual imagery experiment with four categories (4 class objects, 10 class digits, 26 class letters, and 16 class arrows). They achieved classification accuracies of 96.67%, 93.64%, 88.95%, and 92.68%, respectively, for the four categories [38]. Furthermore, the same research team reported a classification accuracy of 95.47 for 36 classes using a CWD time–frequency map and deep learning techniques [39]. In addition, attempts to classify speech and visual imagery EEG for an intuitive BCI were also reported [40,41]. The authors performed 13 classes of visual/speech imagery experiments on 22 subjects and achieved a classification accuracy of 39.73% for speech imagery and 40.14% for visual imagery.

3. Materials and Methods

3.1. Experimental Settings

The experiment was designed to record both visual perception and visual imagery EEG data. The visual stimulus images had 3 categories and 3 classes for each; thus, they consisted of a total of 9 classes. The three categories consisted of an object, digit, and shape. The object category was composed of airplane, cup, and tree images selected from the ImageNet [20] dataset. For the digit category, images of 1, 3, and 5 (handwritten) numbers were used, and for the shape category, images of a heart, star, and triangle were used. The reasons for choosing these three categories of images stemmed from our previous studies. Our study using MEG [29] showed a high classification rate in the classification of objects

and handwritten digits, but it was relatively difficult to distinguish between classes within objects and those within digits. Therefore, it was predicted that adding a category would be advantageous for classification. All 9 images are illustrated in Figure 1.

	-	P		/	3	5		\star	
Category	Object		Digit		Shape				
Class	Airplane	Cup	Tree	One	Three	Five	Heart	Star	Triangle

^{. . . .}

A total of 21 healthy university students aged 22.9 ± 1.75 years, 13 of whom were female, participated in the experiment. To minimize their effects on the experiment, all participants were asked to sleep for at least 7 h the day before the experiment and to abstain from coffee, tobacco, and alcohol for 24 h before the experiment. All participants in the experiment were young adults, as it was difficult to recruit participants outside of the university because of the COVID-19 quarantine policy. All subjects reported that they had normal or normal corrected vision and had no history of cognitive or neurological impairment. All experimental instructions and guidance were given in their native language: Korean. The details of the experiment were explained before the experiment began, and signed informed consents were collected. The experiment was conducted with the approval of the Institutional Review Board of Gwangju Institute of Science and Technology (20180316-HR-34-03-02). Detailed participant information is tabulated in Table 1. Gray shaded areas in Table 1 indicate participants were not used for analysis due to insufficient data. Details are described in Section 3.2: Data Preprocessing.

Participants	Handedness	Age	Gender	Wearing Glasses	Eye Surgery	Past EEG Experiences	Participated Sessions
S01	Right	21	F	Т	F	0	10
S02	Left	26	М	F	T (LASEK ¹)	2 (MI)	10
S03	Right	23	F	Т	F	0	4
S04	Right	21	F	F	T (LASEK)	0	8
S05	Right	22	F	Т	F	0	10
S06	Right	22	М	F	T (LASEK)	0	7
S07	Right	21	F	Т	F	0	6
S08	Right	23	F	Т	F	0	2
S09	Right	22	М	Т	F	1	1
S10	Right	27	М	Т	F	0	2
S11	Right	22	М	Т	F	0	1
S12	Left	21	F	Т	F	0	2
S13	Right	22	F	F	F	0	1
S14	Right	22	F	Т	F	0	1
S15	Right	22	М	Т	F	3	1
S16	Right	24	F	F	T (LASEK)	1	2
S17	Right	21	М	Т	F	0	1
S18	Right	23	М	Т	F	0	2
S19	Right	25	F	Т	F	0	1
S20	Right	25	F	F	T (LASIK ²)	1	1
S21	Right	25	F	Т	F	1	1

Table 1. Subject information.

¹ Laser-assisted sub-epithelial keratomileusis, LASEK. ² Laser-assisted in situ keratomileusis, LASIK.

During the experiment, a trial was composed of fixation, visual perception, instruction, visual imagery, and feedback. Detailed information is depicted in Figure 2. A 24-inch

Figure 1. Images of 9 classes as visual stimuli.

monitor was used to present the experimental stimuli. The subject was asked to sit approximately 1.5 m from the monitor, attend to the image stimulus presented, and perform an experiment according to the instructions. First, after presenting a fixation cue for 1 s, one of the nine classes of visual stimuli was presented randomly for 2 s as a visual perception cue. For 1 s thereafter, the subjects were instructed to imagine the visual stimulus seen immediately before. Then, a blank screen was presented for 4 s as a visual imagery cue,

and they were asked to imagine the visual stimulus with their eyes open. Finally, they were asked to press the keyboard button using the index finger of the right hand within 1.5–2.0 s to report the imagery's vividness on a 3-point Likert scale, on which 1 indicated that imagery was poor/difficult, 2 that it was moderate, and 3 that it was vivid/easy. We note that, in this experiment, the first 2 s interval in which the visual stimulus was presented is referred to as visual perception (VP), and the next 4 s in which the image presented is imagined is defined as visual imagery (VI).



Figure 2. Experimental paradigm of visual perception (VP) and visual imagery (VI).

One session consisted of 3 runs, each with 60 trials. Thus, 1 session consisted of a total of 180 trials, with 60 trials per category, and 20 per class. One run required approximately 10 min. During the experiment, a rest period of approximately 5 min was given between each run to avoid artifacts attributable to eye fatigue and afterimages. This experiment allowed the same subject to participate in multiple sessions. However, to prevent the EEG signal quality from deteriorating because of cumulative fatigue, the subjects were allowed to participate in only one session per day. Finally, we recorded a total of 74 sessions from the 21 subjects, each of whom participated in a minimum of 1 session and a maximum of 10.

3.2. Data Preprocessing

EEG data were recorded using a BioSemi ActiveTwo system with Ag/AgCl wet electrodes, and 32 channels were used according to the international 10-20 system. In addition, the positions of the A1 and A2 mastoids of both ears were recorded together, and the vertical electrooculography (VEOG), horizontal electrooculography (HEOG), and electromyography (EMG) channels on the jaw were recorded together to remove some artifacts other than brain signals. EEG data were recorded at 16 kHz on a BioSemi and OpenViBE [42] platform and downsampled to 1 kHz. A MATLAB-based EEGLAB [43] toolbox was used to process the data. The EEG data were bandpass filtered at 1–50Hz after 60Hz power noise was removed with a notch filter. Then, artifact subspace reconstruction (ASR) was applied to interpolate the bad channels acquired during the experiment. Rereferencing was performed using EEG data at positions A1 and A2, and then artifacts were removed by applying independent component analysis (ICA) [44] using the EOG and EMG channels recorded together. Thereafter, some bad trials with an excessively large magnitude (over 200 μ V) or external noise were rejected through manual inspection. The EEG data used to analyze classification were 2 s of VP and 4 s of VI after preprocessing. The input data were baseline-corrected using a time window of 1000 ms before the onset of imagination. Lastly, the EEG data at 1 kHz were downsampled to 1/8 of 128 Hz. The data from those subjects who participated in only one or several sessions were not included in

the analysis because the machine learning approach could not be applied due to inadequacy. Finally, among the 21 subjects, data from 7 subjects with at least 4 sessions were analyzed in this work.

3.3. Classification

After common preprocessing and before classification, the EEG data were processed further into time series, time–frequency maps, and CSP formats. In addition, various classifiers that accepted each type of processed data were introduced. Thus, the EEGNet [45], 1D convolution network, and MultiRocket [46] were used to classify the time series data signals, and the MobileNet V2 [47] network pre-trained from ImageNet was used to classify the time–frequency map format data. Multi-class support vector machine (SVM) was used as the feature vector of the CSP technique. The subjects' performance was measured as the mean classification accuracy through 5-fold cross-validation. The overall analysis flow of this work is presented in Figure 3.



Figure 3. Flowchart describing the detailed procedure of this work. EEG data in the visual perception/imagery experiments were preprocessed and transformed into three types—time series, a time—frequency map, and a CSP feature vector. Each processed data type was trained through artificial neural network (ANN) or the SVM classifier according to each type, and the classification accuracies were compared.

3.3.1. Data Type—Common Spatial Pattern (CSP)

The CSP algorithm is a feature extraction method that uses spatial filters to maximize two classes' discriminability. The CSP algorithm, as defined in Equation (1), identifies the CSP filter w that maximizes the variance ratio between two classes a and b:

$$w = \operatorname{argmax}_{w} \frac{w \Sigma_{a} w}{w \Sigma_{b} w},\tag{1}$$

Here, Σ_a is the covariance matrix of class 'a', which is defined in Equation (2) as follows:

$$\Sigma_{\rm a} = \frac{X_a X_a^{\ T}}{\operatorname{trace}(X_a X_a^{\ T})},\tag{2}$$

in which X_a represents the multivariate time signal of class 'a', X^T denotes the transpose of X, and trace(X) represents the sum of the diagonal element of X.

CSP is a traditional and popular approach in EEG analysis; it extracts spatial features very well, so it is used particularly to classify motor imagery (MI). Because CSP is designed commonly for binary classification, a special approach is required for it to classify multiple classes [48,49]. Among several methods, CSP feature vectors were extracted using the 1 vs. rest (OVR) method, which calculates CSP repeatedly for one class and all other

remaining classes. In general, all 32 CSP filters were used in this work. Therefore, a feature vector of $32 \times n$ was extracted for each trial, in which '*n*' refers to the number of classes. For example, 9-class classifications comprised 288 feature vectors while 3-category classifications comprised 96.

Classifier—Multi-class Support Vector Machine (SVM): SVM is a classifier specialized for binary classification, and a different approach is required to use it for multiple classes. In this work, a one vs. one (OVO) approach was introduced. For each binary classifier, all combinations of class pairs were divided into positive and negative and a total of $\binom{n}{2} = \frac{n(n-1)}{2}$ classifiers were generated. Here, 'n' refers to the total number of classes. In this work, the CSP feature vector was used as an input for the SVM classifier to classify categories and classes.

3.3.2. Data Type—Time Series Data

The simplest spatiotemporal EEG data format is time series, which is easy to use after simple preprocessing only. EEG data are two-dimensional after simple preprocessing and consist of channel \times time. In this work, this simple channel \times time data could be used as input in such neural networks as EEGNet, 1D Network, and the MultiRocket network specialized for time series classification.

Classifier—EEGNet: EEGNet is a compact CNN architecture for EEG-based BCIs that has been shown to classify EEG data well, such as MI, P300, and error-related negativity (ERN). EEGNet uses the depth-wise separable convolution technique for EEG features to lighten the network and extracts both temporal and spatial features of EEG. In this work, EEGNet was trained using the default settings without changing parameters.

Classifier—**1D CNN**: EEG time series has a two-dimensional form, channel \times time, so it has an unbalanced form in that the number of time samples is likely to be far larger than the number of channels. When a commonly used square-shaped 2D convolution filter is applied, the channel's dimension decreases rapidly before the features are learned sufficiently over time. To solve this problem, we used the 1D convolutional approach. This network performs convolution and pooling operations in the time direction first to extract the time features and reduce the dimension. Spatial features are extracted in the upper layer through the convolution operation in the channel direction when the features are extracted and the dimension is reduced sufficiently. After extracting temporal and spatial features, a convolution operation in the direction of the feature map is constructed so that the integrated spatial and temporal features can be extracted. In this work, we constructed a network consisting of 17 layers (including activation layers). The lower 6 layers can extract temporal information through convolution operation and pooling in the time direction. Thereafter, the 4 layers designed to perform the convolution and pooling operations were used in the channel direction to combine spatial information. Then, the next 4 layers extracted spatiotemporal features in the direction of the feature map, after which two fully connected (FC) layers and a softmax classifier were considered. Because the EEG data contained negative amplitudes, the leaky rectified linear unit (leaky ReLU) [50] was introduced as the activation function, the slope of the leaky ReLU was set to 0.3, and the Adam optimizer was used. Our proposed network is outlined in Figure 4.



Figure 4. Schematic diagram of the proposed 1D convolution network for EEG classification.

Classifier—MultiRocket: MultiRocket is a state-of-the-art network that demonstrates the fastest performance for both overall and classification performance in time series classification (TSC). The TSC benchmark measures its classification performance for the entire dataset of the UCR Time Series Classification Archive [51], which is a public dataset that includes 128 various time series data, such as sensor, motion, and electrocardiography (ECG). Although EEG are typical time series data, as far as we know, it is difficult to find EEG studies in which TSC-specialized networks, such as MultiRocket, InceptionTime [52], and HIVE-COTE 2.0 [53], are used. MultiRocket extracts the feature vector using both time series and its first derivatives. It has a convolutional kernel length of 9 and uses 4 time series-specific pooling operators that can extract features in the pooling operation without using simple max pooling or average pooling. The number of features used in this work was 300 k, rather than the default of 50 k, and the ridge regression classifier was introduced.

3.3.3. Data Type—Time–Frequency Maps

EEG data have spectral as well as temporal characteristics. Therefore, they can be processed in the form of a time–frequency map to extract the spectro-temporal feature. To do so, the preprocessed EEG data were transformed into 2D grayscale images through two kinds of time–frequency conversion, event-related spectral perturbation (ERSP) [54], and CWD [37].

In this work, to calculate the time–frequency map, the preprocessed time series data for 2 s of VP and 4 s of VI after onset were used. The time–frequency data were trained through a 2D CNN specialized in image processing. The size of the time–frequency map varied according to the techniques, and thus, the time–frequency map was finally resized to 224×224 to use as an input in the MobileNet network. Although the absolute size of each technique's two-dimensional matrix varied, the parameters were adjusted so that the ratio of time and frequency was similar when the time–frequency 2D map was generated. The detailed conversion process is illustrated in Figure 5.



Figure 5. Process of transforming EEG time series data into a time–frequency map. The EEG time series data acquired were converted into a time–frequency map according to the power spectrum function. Then, each channel's time–frequency map (TFMap) matrix was converted into two dimensions using valid frequency bands. Finally, the real-value time–frequency map matrix was converted into a grayscale image and used as an input to the classifiers.

Feature—Event-Related Spectral Perturbation (ERSP): ERSP is defined in Equation (3) as follows:

$$ERSP(t,f) = \frac{1}{n} \sum_{k=1}^{n} |F_k(f,t)|^2,$$
(3)

in which 'n' represents the number of trials, and $F_k(f, t)$ represents the power spectrum of the *k*-th trial at frequency 'f" and time 't'. In this experiment, we used the EEGLAB [43] toolbox to calculate the ERSP, which provides transform functions, such as the short-time Fourier transform (STFT) and Morlet wavelet for F_k in Equation (3).

ERSP is one of the traditional methods used in EEG analysis [54]. Because it can observe the change in the frequency components over time, it is used primarily to identify the event-related frequency change before and after the onset of stimulus presentation, such as event-related potential (ERP). The preprocessed EEG data were converted into an ERSP map with a 10 \times 200 (frequency \times time) matrix for a 1–50Hz frequency for 2 s after the visual stimulus and 4 s after visual imagery onset. The difference between the low- and high-frequency components in the converted ERSP had a very large amplitude, so the difference between the high-frequency components was nearly indistinguishable when converted to the image. Therefore, it was transformed into a log-scale value. The ERSP calculated for the entire channel was $10 \times 200 \times 32$ (frequency \times time \times channel) three-dimensional data. These must be converted into 2D data so that MobileNet can learn them. The ERSP matrix calculated for each channel was reshaped into a two-dimensional matrix of 320×200 by concatenating the time-frequency map in the frequency direction in the EEG channel order. Because this matrix has a real value, it could be converted into a 0–255 grayscale 2D image with an integer value so that it could be learned in MobileNet with a pretrained weight. Finally, the image was resized to a 224×224 grayscale image and used as an input in the network.

Feature—Choi–Williams time–frequency distribution (CWD): In 1966, Cohen proposed a general phase-space function referred to as Cohen's Class [55] that is expressed in Equation (4):

$$C(t,f) = \iiint e^{-j2\pi(\theta t + f\tau - \theta u)}\phi(\theta,\tau) \times x\left(u + \frac{\tau}{2}\right)x^*\left(u - \frac{\tau}{2}\right)du \ d\tau \ d\theta \tag{4}$$

in which x^* represents the complex conjugate of the signal x, and function ϕ is the kernel function. In CWD, the kernel function is defined as in Equation (5):

$$\phi(\theta,\tau) = e^{\frac{-\theta^2 \tau^2}{\sigma^2}}$$
(5)

in which σ is a parameter that changes the kernel function's distribution. Finding the optimal σ becomes an optimization problem; however, we determined σ to be 30 heuristically.

The MATLAB-based higher-order spectral analysis (HOSA) toolbox [56] was used to calculate CWD. When CWD is calculated with the parameters above, a 25×512 timefrequency map for one channel is created for VI. Because the CWD calculated had an imaginary value, the magnitude was extracted from the calculated value and then the real number value only was taken. The data calculated for one trial had a three-dimensional form of $25 \times 512 \times 32$ (frequency \times time \times channel). However, three-dimensional data must be reduced to two dimensions because 2D image data must be grayscale or have RGB channels to use the weight of the pretrained network. We attempted to make the time-frequency map as near as possible to 224×224 in size, which is the original input of MobileNet V2 because the extracted features are used as inputs in the pretrained MobileNet V2. Finally, an 800×512 time-frequency map was created by stacking the extracted features in the frequency direction using the method introduced in Figure 5. Finally, the real-valued matrix for a single trial was converted into a 2D grayscale image with integer values of 0–255 and used as input in the MobileNet V2 network.

Classifier—**MobileNet V2**: Transfer learning or fine-tuning using weights of a wellpretrained network has the potential to increase classification performance and reduce learning time effectively rather than training from the outset [57]. To do so, we introduced the MobileNet V2 network pretrained for ImageNet provided by Tensorflow [58]. The input image should have a size of 224×224 , and also be in grayscale or RGB 3-channel format to use MobileNet's pretrained weights. Therefore, the ERSP and CWD results calculated were resized to a 224×224 grayscale image. The original MobileNet V2 model Tensorflow supports has 154 layers. In this work, we added 5 layers (one serialization layer, one FC layer of size 1000, one 30% dropout layer, one FC layer of size 512, and one softmax layer for EEG data class classification) in place of the last prediction layer of the basic MobileNet. Therefore, a total of 159 layers of networks were constructed. In the custom MobileNet network, the lower 100 layers were fixed as the pretrained weight of ImageNet. Then, the classifier was trained through a fine-tuning process in which the upper 59 layers could update their weights only during the learning procedure, and the parameters were adjusted while the learning trend was observed. Adam was used as the activation function, the learning rate was set to 1×10^{-4} , and up to 1000 epochs were trained for the classifier.

4. Results

4.1. VP Classification

EEG visual perception was explored with various data types and corresponding classifiers. Our investigation focused particularly on three-category and nine-class classifications. Table 2 shows the nine-class classification results (mean accuracies with 5-fold cross-validation) in the VP condition for six types of approaches. The highest mean classification accuracy among all subjects was 24.02%. Of the various approaches, the combination of time series data and MultiRocket or 1D CNN classifiers yielded better performance than others. While 1D CNN showed the highest classification accuracy for three subjects (S01, S04, and S07), MultiRocket showed the highest performance for four subjects. We found that most classification approaches using EEGNet and MobileNet V2 did not appear to learn effectively, and thereby yielded results near the random chance level (0.1111). Furthermore, most classifiers did not yield reasonably high performance in nine-class classification, although S02 showed an accuracy of 34.01%, which is still not significantly high. We note that the combination of CSP features and SVM classifier yielded relatively greater accuracy in S02 alone.

Data				Time–Freq	COD	
Format	1	ime Series EE	EG –	ERSP	CWD	CSr
Classifier	EEGNet	1D CNN	MultiRocket	Mobile	Net V2	SVM
S01	0.1201	0.2661	0.2607	0.1165	0.1099	0.1822
S02	0.1212	0.2633	0.3401	0.1232	0.1239	0.2954
S03	0.1185	0.1511	0.1719	0.1081	0.1067	0.1681
S04	0.1148	0.2178	0.1985	0.1126	0.1022	0.1208
S05	0.1171	0.2258	0.2595	0.1117	0.1123	0.1039
S06	0.1070	0.1383	0.2658	0.1062	0.1078	0.0706
S07	0.1263	0.2182	0.1848	0.1020	0.1242	0.0991
Mean	0.1179	0.2115	0.2402	0.1115	0.1124	0.1486

Table 2. Each subject's classification accuracy of 9-class VP according to the data type and classifiers.

The three-category classification results in VP are presented in Table 3 for comparison with the nine-class classification. In contrast to nine-class classification, EEGNet yielded the highest classification accuracy, 57.59%, for three-category classification, while MultiRocket still yielded stable classification performance as good as EEGNet. Similar to nine-class classification, time series data yielded better classification performance in three-category classification than other data types.

Data	T			Time–Freq	COD	
Format	1	ime Series EE		ERSP	CWD	CSP
Classifier	EEGNet 1D CNN MultiRocket		MobileNet V2		SVM	
S01	0.6018	0.6216	0.5754	0.3517	0.3661	0.4793
S02	0.5785	0.5535	0.6067	0.5431	0.6011	0.4734
S03	0.5511	0.4519	0.4904	0.3243	0.3709	0.3615
S04	0.5696	0.5563	0.5719	0.3665	0.4018	0.3993
S05	0.5724	0.5724	0.5838	0.3547	0.3511	0.4294
S06	0.6066	0.4700	0.6362	0.3645	0.3600	0.4346
S07	0.5515	0.6212	0.5273	0.3762	0.3543	0.4202
Mean	0.5759	0.5496	0.5702	0.3830	0.4008	0.4282

Table 3. Each subject's classification accuracy of 3-category VP according to the data type and classifiers.

4.2. VI Classification

In the same way as VP classification, EEG VI classification was explored with various combinations of data types and classifiers. For nine-class classification in EEG VI data, it was observed that most methods did not achieve significantly higher than random-chance-level performance (0.111). In addition, we note that EEGNet did not converge. However, similar to VP classification, the MultiRocket classifier yielded relatively better classification performance compared to other approaches, although it still did not yield very high performance. The classification results of nine-class VI overall are presented in Table 4. Expectedly, VI's classification performance was poor overall and near the random chance level. It is noted that MultiRocket for S02 yielded a significantly high classification accuracy, which interestingly was higher than that in the VP. It was expected that the even VI data would be classifiable in nine classes if reasonable features and classifiers were found.

Data				Time–Freq		
Format	1	ime Series EE	EG –	ERSP	CWD	CSP
Classifier	EEGNet 1D CNN MultiRo		MultiRocket	MobileNet V2		SVM
S01	-	0.1279	0.1333	0.1051	0.1027	0.1135
S02	-	0.1037	0.3663	0.1293	0.1515	0.1556
S03	-	0.1141	0.1407	0.1200	0.1067	0.1200
S04	-	0.1148	0.1311	0.0881	0.1059	0.1119
S05	-	0.1249	0.1273	0.1093	0.1051	0.1159
S06	-	0.0938	0.1309	0.0996	0.1119	0.0922
S07	-	0.1455	0.1566	0.1131	0.1323	0.1111
Mean	-	0.1178	0.1695	0.1092	0.1166	0.1172

Table 4. Each subject's classification accuracy of 9-category VI according to the data type and classifiers.

As in the VP case, the three-category classification of VI is presented in Table 5. Similar to VP, EEGNet and MultiRocket showed relatively higher performance than other approaches, while MultiRocket showed the highest classification accuracy of 46.79%. Notably, MultiRocket for S02 yielded a classification accuracy of 73.71%, which may be a performance as good as applicable in practice. The classification performance overall appeared to be higher in the VP than the VI, although S02 showed better classification accuracies in VI.

Data	T			Time–Freq	CCD	
Format	1	ime Series EE		ERSP	CWD	CSP
Classifier	EEGNet 1D CNN MultiRocket			Mobile	SVM	
S01	0.4192	0.3523	0.3962	0.3483	0.3387	0.3692
S02	0.4503	0.5712	0.7371	0.3865	0.4391	0.6261
S03	0.4471	0.3771	0.3413	0.3422	0.3467	0.3933
S04	0.4107	0.3508	0.4111	0.3556	0.3378	0.3471
S05	0.4353	0.3724	0.3692	0.3351	0.3339	0.3797
S06	0.4365	0.3835	0.4903	0.3424	0.3424	0.3797
S07	0.4797	0.3753	0.5301	0.3414	0.3515	0.4692
Mean	0.4398	0.3975	0.4679	0.3502	0.3557	0.4235

Table 5. Each subject's classification accuracy of 3-category VI according to the data type and classifiers.

Overall, our findings from this extensive classification analysis are summarized as follows:

- Classification performance varied depending upon classification approaches and subjects. In particular, S02 demonstrated significantly better performance than the other subjects.
- The VP task yielded relatively higher performance than the VI task, although both VP and VI are difficult to compare in the nine-class classification. Presumably, this is attributable to the fact that VP may be time-locked to stimulation onset more accurately than VI because the time that imagery initiated could not be fixed over trials.
- The overall classification using time series data showed higher performance than the other data types, such as time–frequency transformation or CSP feature vector extraction.
- Among all possible combinations of data types and classifiers, the combination of EEG time series and the MultiRocket classifier showed the best performance overall in most classification problems. In particular, EEGNet showed quite poor performance in nine-class classification; however, it showed relatively better performance in threecategory classification.

4.3. Classification Accuracy over Sessions

We investigated the way that classification accuracy proceeded over the sessions (accumulated sessions) using the MultiRocket classifier. Figure 6 shows VP and VI's nine-class and three-category classification results for each subject's sessions. For session learning, the total dataset from the previous sessions was used as the training set and the current session data were used for the test. For example, session 4 was trained with all data from sessions 1, 2, and 3, and tested with session 4. Naturally, the first session had no previous data, so performance was measured using 80% of the training data and 20% of the test data. As a result, despite EEG's nonstationary characteristics, it was observed that the classification accuracy increased slightly over sessions in the case of VP. However, in VI, the classification accuracy did not appear to increase, except for that of S02 and S07. Interestingly, S02's classification accuracy in the second session was very high compared to other sessions; this phenomenon was observed in both VP and VI. We note that S02's second session was collected the day after the first. Perhaps this inter-session transfer learning was attributable to the short period (1 day) between sessions. Furthermore, it is expected that VP may enhance performance with repeated training, just as in the traditional BCI paradigms.



Figure 6. Classification accuracy over accumulated sessions.

5. Discussion

5.1. Individual Differences

A total of 21 subjects participated in the experiment. However, only those who participated in more than four sessions were used for the comparative analysis. As addressed in Section 3.2 above, the subjects who participated in only a few sessions (fewer than four) did not have sufficiently good data for use with machine learning. There was a large difference in the classification accuracy even with a very small bias in the training or test set, so we elected to discard such cases. However, we note that even when insufficient EEG data are collected, it is necessary to apply a transfer learning method that can derive an acceptable performance estimation after the network is trained. Thus, there is potential to improve the performance when fine-tuning or when a transfer learning technique is applied to the data of subjects who participated in only a small number of sessions after the network was trained, so that their data may be combined with that of other subjects.

S02 showed a very high classification performance compared to the other subjects. Before the experiment, the participants were asked to fill out a questionnaire about handedness, age, gender, disease status, and EEG experience as described in Table 1. In this pre-experimental questionnaire, most of the participants said that our experiment was their first experience measuring EEG signals, but S02 had already experienced MI experiments twice before. It is unclear how much influence S02's previous MI experience had on these VI classification results. However, the human brain tends to learn from previous experiences, and having repeated training session experiences can enhance BCI performance. Therefore, experience in MI may be an important factor that influenced S02's results. Furthermore, we noted that S02 scarcely moved during the experiment, was focused highly, and had a positive attitude. This observation appears to be relevant to existing studies, in that the user's motivation is correlated strongly with BCI classification performance [59–62]. In this work, VI classification performances were not particularly high; however, it may be possible to increase the performance of the VI task by sensory/electrical stimulation and neurofeedback training approaches that motivate users or help them engage in visual imagery. In that case, more intuitive BCI using VI may be practical.

5.2. VP and VI Classifications

We investigated the way each trial was classified using the MultiRocket classifier. Figure 7 shows one exemplary classifier result (confusion matrix), which is the first-fold among S02's five-fold cross-validations in VP. We confirmed that S02's relatively high classification accuracy was attributable to the fact that the number of classes classified correctly overall was higher than the misclassified results. Looking at the square section marked in red (in Figure 7), the classes in the object category (airplane, cup, tree) tended to be classified within the same object category regardless of whether they were misclassified or classified correctly. Furthermore, the classes belonging to the digit and shape categories tended to be classified within the two categories. It may be inferred that the object category and digit+shape category could be classified easily, while the digit and shape categories were difficult to classify separately.



Figure 7. Exemplary confusion matrix of MultiRocket classifier for S02 in VP.

Initially, we conducted a nine-class classification; however, based upon the results above, we expected that a three-category classification of visual images might be more effective. This is even more pronounced in Figure 8, which classified the VP into three categories. In S02's VP confusion matrix in the upper right of Figure 8, the recall of the object category was 73%, which is very high compared to the other two categories. Further, the digit and shape categories tended to be misclassified, rather than the object category, in both S01 and S02. We note that this phenomenon was observed in most VP cases, regardless of classifiers and subjects. It is difficult to determine the reason for this phenomenon. However, we speculated that the images in the digit and shape categories were relatively simple and monochromatic, while those in the object category had multiple colors and complex images, as illustrated in Figure 1. However, it was not possible to determine accurately whether the images' color or complexity had a greater effect on the classification performance, which should be investigated in future work. Therefore, when the VP paradigm is applied to actual BCI, it is likely that it will be important to configure the visual complexity (including colors) among classes.



Figure 8. Exemplary confusion matrix of S01 and S02's 3-category classification using MultiRocket classifier in VP and VI.

We assumed that the VI results would show a trend similar to those of VP. Interestingly, as Figure 8 shows, the complexity of the proposed visual image did not affect the classification in VI strongly. As discussed above, the digit and shape categories tended to be classified within two categories in VP, regardless of subjects and classifiers; however, a phenomenon similar to VP was not observed in VI, and thus the classification results varied between subjects. This may be attributable to the difference in information processing between VP and VI. The neurophysiological processing of visual stimuli is exogenous because the stimuli are external, while visual imagery is endogenous, in that the user generates his/her intention internally.

5.3. Classification Accuracies over Methods

As reported in Section 4, VP and VI's classification accuracies were insufficiently high and, therefore, are impractical for an actual BCI system. However, we believe that there is considerable room to improve performance in the classification results. For example, it may be necessary to downsample the original EEG data more carefully to do so. In this work, it was necessary to maintain the consistency of input data to compare the approaches accurately. The 4 s long VI EEG data were recorded at a 1 kHz sampling rate and had 4028 samples. We downsampled the EEG data for rapid computation in processing. In addition, the EEG data were downsampled to form a 2D image similar to a square to use as input in the neural networks when they were converted into the time–frequency map data. This was because a 224×224 square image was used as the input in the MobileNet V2 network in this work, and the image must be resized to 224×224 if the input size does not match. To do so, EEG data with a sampling rate of 1 kHz were downsampled by 1/8. However, if all EEG data are used without downsampling, it may be possible to improve classification accuracy at the expense of longer data processing time.

Furthermore, classification accuracy can be improved by determining the optimal parameters, i.e., solving the optimization problem. For example, in the CWD technique, the results of the time–frequency map differed greatly depending upon the σ value (Equation (5)), and thus, changing the σ value may have a great influence on the classification result. Similarly, when calculating ERSP or estimating CSP features, various parameters may be optimized according to the domain knowledge, such as window spectrum, the number of CSP filters, and the selection of effective channels. In addition, there are many tunable parameters, such as the number of layers in the neural networks, optimizers, pretraining status, dropout ratio, learning rate, and so on. To achieve the best performance in each approach, it is necessary to fine-tune the parameters for each. However, because the goal of this work was to investigate the feasibility of VI and VP as an actual BCI paradigm and to compare various techniques, a thorough exploration of each combination of classifier and feature extraction technique was not considered. Thus, future work needs to conduct an intensive investigation of a specific approach such as MultiRocket.

5.4. Performance Variation in Related Studies

The classification accuracies in reported studies differ significantly, even if they use EEG data for visual imagery. Kumar's research team showed a classification accuracy of 85.2% for 10-class visual imagery within the categories of objects, digits, and letters using the random forest technique [34]. Tirupattur et al. [35] obtained a classification accuracy of over 71% by applying the CNN method to the same data as those of Kumar et al. Furthermore, another study reported a very high classification accuracy of 76.39% for 56-class classification using CWD [38]. On the other hand, some studies have reported relatively poor performance, unlike the studies above. Lee et al. [40] obtained a classification accuracy of only 26.7% in the 13-class classification of VI using CNN. The following year, they achieved a classification accuracy of 40.14% using spectral analysis and SVM for the same visual imagery dataset [41]. In this work, we obtained a relatively low classification accuracy of 16.95% in the nine-class classification of visual imagery. It is quite surprising to observe such significant differences in visual imagery classification in the literature. Based upon our experience in EEG research, our expectation is as follows:

First, in various visual imagery experiments, the EEG acquisition equipment, experimental paradigm, image imagination time, and the number of EEG trials acquired have differed. In Kumar et al.'s [34] study, visual imagery was performed for 10 s using 14channel Emotiv EPOC+ equipment. Lee et al. used the 64-channel Brain Vision/Recorder from Brain Products, and 2 s of visual imagery was repeated five times after the imagery onset. Alazrai et al. [39] used 16-channel ActiveTwo EEG equipment from BioSemi in their study, and visual imagery was performed for 8 s. We used 32-channel BioSemi ActiveTwo equipment, and visual imagery was performed for 4 s after onset. As such, the EEG acquisition equipment, experimental paradigm, and the number of trials executed differed significantly, and such experimental parameters may have great effects on the classification results. In future work, an in-depth review of visual imagery classification may be highly beneficial in designing the most optimal experimental paradigm possible.

Second, the difference may have been attributable to the preprocessing of the raw EEG data and external noise during the experiment. Extreme signal-to-noise attenuation occurs when EEG data pass through the skull. Therefore, the EEG data acquired are contaminated by many artifacts, such as EOG, EMG, ECG, and electrical power noise. When external noise remains because of insufficient preprocessing, the subsequent analysis may yield quite different results. In addition, the magnitude of EEG data may vary depending upon the way the baseline is aligned before or after stimulus presentation. Furthermore, the use of the gamma band may have an effect because gamma signals above 30 Hz are generally known to be involved in attention, perception, and memory [63–66]. However, because most of the artifacts aforementioned have high-frequency components, the gamma

band's high-frequency dynamics may be contaminated easily, and thus, the high-frequency component of EEG should be managed more carefully [67]. In general, many techniques are applied to remove the potential noise in raw EEG data and the gamma band. For example, ICA is applied to remove external artifacts attributable to movement, a notch filter is applied to remove 50 Hz or 60 Hz harmonic power noise, and trials that exceed the threshold are rejected to remove the noise that remains. Kumar et al. [34] did not address such details and reported simply that artifacts were removed by applying a moving average (MA) filter. Lee et al. [41] used a frequency range of 0.2–145 Hz, which contains a high gamma band. However, their other studies used different frequency ranges of 0.5–40 Hz, even with the same EEG dataset [40]. Alazrai et al. [38], who used a frequency range of 1–45 Hz, and our team, who used 1–50 Hz, did not consider the high-gamma band. Whether to use high gamma or not and the way to preprocess EEG to keep high gamma clean should be investigated thoroughly.

Third, data augmentation may create distortion. EEG data are acquired during human experiments, so the amount of data may be insufficient for machine learning approaches. Therefore, to overcome this problem, the amount of data can be increased through data augmentation. In the image processing field, such augmentation methods as shuffling, flipping, and distorting are used commonly; however, these techniques are not applied directly to time series EEG data [68] because the temporal characteristics associated with a specific event may be broken if time samples are mixed or the sampling rate is changed. Generating virtual EEG using generative models such as autoencoder and GAN [69–71], decomposing signals using EEG's spatiotemporal properties [72,73], adding noise, and windowing without breaking the temporal characteristics [34,35,74,75] are methods used commonly to augment EEG data. Among these, great attention is required to manage overlapping when the method chosen is augmenting data with a sliding window. This is because the data included in the training set may be included in the test set in the overlapping process. For example, suppose that 1 s long data are windowed with a size of 100 ms and an overlap of 50%, that 200–300 ms of data are allocated to the training set, and 250–350 ms of data are allocated to the test set after onset in the process of dividing the training set and the test set. Notably, 50 ms of the window is already held in the training set, although the test set does not contain the exact same data as the training set, so the test set becomes contaminated. Therefore, when augmenting data with the sliding window method, including overlapping, it is necessary to be very careful so that a part of the test set is not included in the training set. In Tirupattur et al.'s [35] study, one trial had a very long duration of 10 s, so they chose the sliding window with overlapping to augment the data. In this work, classification was performed without data augmentation. We performed comparative analysis according to the EEG data type and classifier, and thus, the input data needed to have the same parameters to the extent possible in each classification method. Because of this, it was difficult to apply the sliding window augmentation technique, which demonstrates high performance, as in Tirupattur et al.'s study [35].

We note that our EEG visual imagery was not a well-designed paradigm; to the best of our knowledge, there is no typical standard paradigm that many researchers have verified, such as MI and P300. There are many challenges in the experimental setting, such as the equipment and paradigm, the preprocessing and analysis process, the data augmentation method, and the research purpose. Such issues should be investigated, which we plan to do in our future work.

5.5. Limitations

We investigated the feasibility of nine-class and three-class (category) classification for VP and VI EEG data, in the hope that either may be used as a new BCI paradigm and accordingly, a more intuitive BCI system can be developed. In this work, nine-class classifications for VP and VI were far less effective, while three-category classifications were more accurate. Thus, at this moment, we believe that VI (and even VP) is unlikely to be practically applicable in a BCI system. However, it is expected that VI may have some potential, in that developing an imagery training method for users, developing classifiers, and optimizing feature extraction techniques may improve classification performance to a much greater extent. In this work, S02 showed evidence that a better-designed experimental paradigm that elicits users' strong motivation may improve classification performance.

One of the potential limitations in this work is our doubt that the classification results are related directly to the category (objects, digits, shapes) the users intend to imagine visually or are related more closely to neurophysiological EEG features that depend upon the image's complexity. The difficulty of classifying images within the categories aforementioned may be one clue to resolving the problem. In any case, it is necessary to investigate and verify these possibilities in subsequent experiments.

Another limitation is that the EEG data's functional connectivity characteristics were not considered in this work. The functional connectivity of time series data such as EEG is a unique characteristic and has received more attention in neuroscience research. In subsequent work, we plan to explore classification using functional connectivity features; to do so, a graph neural network is one of the potential classifiers that uses such features.

In addition, the participants' emotional status, motivation, and empathy were not considered in depth in this study. In various BCI paradigms, such as P300 and SMR, many studies have been reported in which a positive attitude, mood, and strong motivation affected users' performance [59–62]. However, these factors were not recorded sufficiently in the questionnaires in this experiment. If these factors were tracked properly, it would have been possible to infer more precisely the reason that S02's patterns differed from those of the other participants. Therefore, it is necessary to record the users' motivation and mood in future experiments.

Another factor that may be a limitation of the study is the bias in the participants' age. In this study, we analyzed the signals of young adults in their early twenties. However, the brain activity of young adults and older groups differs [76,77]. Furthermore, it is known that some emotional attention circuits continue to mature when people are in their twenties [76,78]. Therefore, it is possible that brain signals' characteristics appear differently depending upon age, and studies have reported that there is a difference in SSVEP BCI performance between young adults and older groups [79,80]. Accordingly, the maturational imbalance could be a potential disturbance factor that affected our results and it should be considered when designing and conducting future experiments.

Furthermore, we focused on studying the feasibility of VI and VP classifications. However, we did not explore their neurophysiological features from the perspective of information processing in the human brain. Although classification performance is likely to be important for a practical BCI system, it is necessary to understand what kinds of classifiable features can be extracted from EEG and the way such features are associated with brain information processing or brain functionality. In this context, we have an investigation underway to analyze the differences and commonalities between VI and VP.

In general, the ability to explain a classifier and its performance are often inversely proportional, and it is particularly difficult to interpret the results of the deep learning technique [81,82]. Thus, the research field of explainable artificial intelligence (XAI) has emerged to seek interpretable results from deep learning. It is expected that various XAI techniques, such as Grad-CAM [83] and layer-wise relevance propagation (LRP) [84], may help interpret classification results and neurophysiological analysis.

6. Conclusions

In this work, we explored the possibility of identifying a visual imagery paradigm that has the potential to increase BCI systems' intuitiveness and degree of freedom. In this context, we searched for effective methods to classify visual imagery (and visual perception) by converting EEG time series data into various other data formats (CSP feature vector, time series, and time–frequency maps) and applying various classifiers (EEGNet, MultiRocket, 1D CNN, and MobileNet V2) that are suitable for each format. As a result, nine-class and three-category classification in VP yielded 24.02% and 57.59% mean accuracy, respectively,

over the subjects, while for VI, nine-class and three-category classification yielded 16.95% (max 36.63%) and 46.50% (max 73.20%) in mean accuracy, respectively.

Notably, we observed that the MultiRocket classifier showed the best performance overall for the end-to-end learning of time series EEG data. To the best of our knowledge, this work is the first to introduce MultiRocket for EEG classification, particularly of VI and VP data. Achieving high performance by introducing such an advanced network suggests that introducing well-established networks verified in other fields could be used to improve BCI performance. However, although creating a new network and adjusting parameters each time according to the new EEG dataset and experimental paradigm may be possible, it is quite time-consuming.

We found that the nine-class classification was quite difficult in VI (and even VP) in this work. However, we confirmed that there was some visual commonality within the categories, and the three-category classification may be reasonably better. The careful selection of visual images may be quite important and have the ability to improve classification performance. In any case, although the experimental results are preliminary, we believe that the VI task has the potential to be applied as a new BCI paradigm when the experimental paradigm is designed carefully and the classifier is optimized. In this context, various future works should be conducted to develop effective intuitive BCI paradigms.

Author Contributions: Conceptualization, S.L., S.J. and S.C.J.; methodology, S.L. and S.J.; software, S.L. and S.J.; validation, S.L., S.J. and S.C.J.; formal analysis, S.L.; investigation, S.L. and S.J.; resources, S.C.J.; data curation, S.L. and S.J.; writing—original draft preparation, S.L.; writing—review and editing, S.C.J.; visualization, S.L.; supervision, S.C.J.; project administration, S.C.J.; funding acquisition, S.C.J. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported by the Republic of Korea's MSIT (Ministry of Science and ICT), under the High-Potential Individuals Global Training Program (No. 2021-0-01537) supervised by the IITP (Institute of Information and Communications Technology Planning & Evaluation). It was also supported by the IITP grants (No. 2017-0-00451, No. 2019-0-01842) and the National Research Foundation grant (No. 2018R1A2B2005687) funded by the Korea government.

Institutional Review Board Statement: The study was conducted in accordance with the Declaration of Helsinki, and approved by the Institutional Review Board of Gwangju Institute of Science and Technology (protocol code 20180316-HR-34-03-02).

Informed Consent Statement: Informed consent was obtained from all subjects involved in the study.

Data Availability Statement: The data presented in this study are available on reasonable request from the corresponding author. The data are not publicly available due to the anonymity of the participants.

Acknowledgments: This research is the result of a study of the "HPC Support" Project, supported by the 'Ministry of Science and ICT' and NIPA. It was also supported by 'Artificial intelligence industrial convergence cluster development project' funded by the Ministry of Science and ICT (MSIT, Korea) and Gwangju Metropolitan City.

Conflicts of Interest: The authors declare no conflict of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript; or in the decision to publish the results.

References

- Zander, T.O.; Kothe, C.; Jatzev, S.; Gaertner, M. Enhancing Human-Computer Interaction with Input from Active and Passive Brain-Computer Interfaces. In *Brain-Computer Interfaces: Applying Our Minds to Human-Computer Interaction*; Tan, D.S., Nijholt, A., Eds.; Human-Computer Interaction Series; Springer: London, UK, 2010; pp. 181–199. ISBN 978-1-84996-272-8.
- Wang, Z.; Zhou, Y.; Chen, L.; Gu, B.; Liu, S.; Xu, M.; Qi, H.; He, F.; Ming, D. A BCI Based Visual-Haptic Neurofeedback Training Improves Cortical Activations and Classification Performance during Motor Imagery. J. Neural Eng. 2019, 16, 066012. [CrossRef]
- 3. Gomez-Pilar, J.; Corralejo, R.; Nicolas-Alonso, L.F.; Álvarez, D.; Hornero, R. Neurofeedback Training with a Motor Imagery-Based BCI: Neurocognitive Improvements and EEG Changes in the Elderly. *Med. Biol. Eng. Comput.* **2016**, *54*, 1655–1666. [CrossRef]
- 4. Choi, J.W.; Kim, B.H.; Huh, S.; Jo, S. Observing Actions Through Immersive Virtual Reality Enhances Motor Imagery Training. *IEEE Trans. Neural Syst. Rehabil. Eng.* **2020**, *28*, 1614–1622. [CrossRef]

- Bermudez i Badia, S.; Garcia Morgade, A.; Samaha, H.; Verschure, P.F.M.J. Using a Hybrid Brain Computer Interface and Virtual Reality System to Monitor and Promote Cortical Reorganization through Motor Activity and Motor Imagery Training. *IEEE Trans. Neural Syst. Rehabil. Eng.* 2013, 21, 174–181. [CrossRef] [PubMed]
- Mirelman, A.; Maidan, I.; Deutsch, J.E. Virtual Reality and Motor Imagery: Promising Tools for Assessment and Therapy in Parkinson's Disease. *Mov. Disord.* 2013, 28, 1597–1608. [CrossRef] [PubMed]
- Corbet, T.; Iturrate, I.; Pereira, M.; Perdikis, S.; del Millán, J.R. Sensory Threshold Neuromuscular Electrical Stimulation Fosters Motor Imagery Performance. *NeuroImage* 2018, 176, 268–276. [CrossRef] [PubMed]
- Yao, L.; Meng, J.; Zhang, D.; Sheng, X.; Zhu, X. Combining Motor Imagery With Selective Sensation Toward a Hybrid-Modality BCI. *IEEE Trans. Biomed. Eng.* 2014, 61, 2304–2312. [CrossRef] [PubMed]
- 9. Ahn, S.; Ahn, M.; Cho, H.; Jun, S.C. Achieving a Hybrid Brain–Computer Interface with Tactile Selective Attention and Motor Imagery. J. Neural Eng. 2014, 11, 066004. [CrossRef]
- 10. Yi, W.; Qiu, S.; Wang, K.; Qi, H.; Zhao, X.; He, F.; Zhou, P.; Yang, J.; Ming, D. Enhancing Performance of a Motor Imagery Based Brain–Computer Interface by Incorporating Electrical Stimulation-Induced SSSEP. J. Neural Eng. 2017, 14, 026002. [CrossRef]
- Blankertz, B.; Sannelli, C.; Halder, S.; Hammer, E.M.; Kübler, A.; Müller, K.-R.; Curio, G.; Dickhaus, T. Neurophysiological Predictor of SMR-Based BCI Performance. *NeuroImage* 2010, *51*, 1303–1309. [CrossRef]
- 12. Koles, Z.J.; Lazar, M.S.; Zhou, S.Z. Spatial Patterns Underlying Population Differences in the Background EEG. *Brain Topogr.* **1990**, 2, 275–284. [CrossRef] [PubMed]
- Bin, G.; Gao, X.; Yan, Z.; Hong, B.; Gao, S. An Online Multi-Channel SSVEP-Based Brain–Computer Interface Using a Canonical Correlation Analysis Method. J. Neural Eng. 2009, 6, 046002. [CrossRef] [PubMed]
- 14. Lin, Z.; Zhang, C.; Wu, W.; Gao, X. Frequency Recognition Based on Canonical Correlation Analysis for SSVEP-Based BCIs. *IEEE Trans. Biomed. Eng.* **2006**, *53*, 2610–2614. [CrossRef] [PubMed]
- Kay, K.N.; Naselaris, T.; Prenger, R.J.; Gallant, J.L. Identifying Natural Images from Human Brain Activity. *Nature* 2008, 452, 352–355. [CrossRef]
- 16. Kay, K.N.; Naselaris, T.; Gallant, J.L. FMRI of Human Visual Areas in Response to Natural Images. CRCNS 2011. [CrossRef]
- 17. Naselaris, T.; Prenger, R.J.; Kay, K.N.; Oliver, M.; Gallant, J.L. Bayesian Reconstruction of Natural Images from Human Brain Activity. *Neuron* **2009**, *63*, 902–915. [CrossRef]
- 18. Nishimoto, S.; Vu, A.T.; Naselaris, T.; Benjamini, Y.; Yu, B.; Gallant, J.L. Reconstructing Visual Experiences from Brain Activity Evoked by Natural Movies. *Curr. Biol.* **2011**, *21*, 1641–1646. [CrossRef] [PubMed]
- Horikawa, T.; Tamaki, M.; Miyawaki, Y.; Kamitani, Y. Neural Decoding of Visual Imagery During Sleep. Science 2013, 340, 639–642. [CrossRef] [PubMed]
- Deng, J.; Dong, W.; Socher, R.; Li, L.-J.; Li, K.; Fei-Fei, L. ImageNet: A Large-Scale Hierarchical Image Database. In Proceedings of the 2009 IEEE Conference on Computer Vision and Pattern Recognition, Miami, FL, USA, 20–25 June 2009; pp. 248–255.
- Horikawa, T.; Kamitani, Y. Generic Decoding of Seen and Imagined Objects Using Hierarchical Visual Features. *Nat. Commun.* 2017, *8*, 15037. [CrossRef] [PubMed]
- Horikawa, T.; Kamitani, Y. Hierarchical Neural Representation of Dreamed Objects Revealed by Brain Decoding with Deep Neural Network Features. *Front. Comput. Neurosci.* 2017, 11, 4. [CrossRef] [PubMed]
- Shen, G.; Horikawa, T.; Majima, K.; Kamitani, Y. Deep Image Reconstruction from Human Brain Activity. *PLOS Comput. Biol.* 2019, 15, e1006633. [CrossRef]
- 24. Shen, G.; Dwivedi, K.; Majima, K.; Horikawa, T.; Kamitani, Y. End-to-End Deep Image Reconstruction From Human Brain Activity. *Front. Comput. Neurosci.* 2019, 13, 21. [CrossRef] [PubMed]
- Beliy, R.; Gaziv, G.; Hoogi, A.; Strappini, F.; Golan, T.; Irani, M. From Voxels to Pixels and Back: Self-Supervision in Natural-Image Reconstruction from FMRI. In Proceedings of the Advances in Neural Information Processing Systems; Curran Associates, Inc.: Red Hook, NY, USA, 2019; Volume 32.
- VanRullen, R.; Reddy, L. Reconstructing Faces from FMRI Patterns Using Deep Generative Neural Networks. *Commun. Biol.* 2019, 2, 193. [CrossRef] [PubMed]
- Cohen, D. Magnetoencephalography: Detection of the Brain's Electrical Activity with a Superconducting Magnetometer. *Science* 1972, 175, 664–666. [CrossRef] [PubMed]
- Jaklevic, R.C.; Lambe, J.; Silver, A.H.; Mercereau, J.E. Quantum Interference Effects in Josephson Tunneling. *Phys. Rev. Lett.* 1964, 12, 159–160. [CrossRef]
- Kim, Y.; Jang, S.; Won, K.; Jun, S.C. CANet: A Channel Attention Network to Determine Informative Multi-Channel for Image Classification from Brain Signals. In Proceedings of the 2019 41st Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), Berlin, Germany, 23–27 July 2019; pp. 680–683.
- Spampinato, C.; Palazzo, S.; Kavasidis, I.; Giordano, D.; Souly, N.; Shah, M. Deep Learning Human Mind for Automated Visual Classification. In Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, HI, USA, 21–26 July 2017; pp. 4503–4511.
- 31. Palazzo, S.; Spampinato, C.; Kavasidis, I.; Giordano, D.; Schmidt, J.; Shah, M. Decoding Brain Representations by Multimodal Learning of Neural Activity and Visual Features. *IEEE Trans. Pattern Anal. Mach. Intell.* **2021**, *43*, 3833–3849. [CrossRef] [PubMed]
- Zheng, X.; Chen, W. An Attention-Based Bi-LSTM Method for Visual Object Classification via EEG. *Biomed. Signal Process. Control* 2021, 63, 102174. [CrossRef]

- 33. Ling, S.; Lee, A.C.H.; Armstrong, B.C.; Nestor, A. How Are Visual Words Represented? Insights from EEG-Based Visual Word Decoding, Feature Derivation and Image Reconstruction. *Hum. Brain Mapp.* **2019**, *40*, 5056–5068. [CrossRef] [PubMed]
- Kumar, P.; Saini, R.; Roy, P.P.; Sahu, P.K.; Dogra, D.P. Envisioned Speech Recognition Using EEG Sensors. Pers. Ubiquitous Comput. 2018, 22, 185–199. [CrossRef]
- 35. Tirupattur, P.; Rawat, Y.S.; Spampinato, C.; Shah, M. ThoughtViz: Visualizing Human Thoughts Using Generative Adversarial Network. In Proceedings of the 26th ACM International Conference on Multimedia, New York, NY, USA, 15 October 2018; Association for Computing Machinery: New York, NY, USA; pp. 950–958.
- Bang, J.-S.; Jeong, J.-H.; Won, D.-O. Classification of Visual Perception and Imagery Based EEG Signals Using Convolutional Neural Networks. In Proceedings of the 2021 9th International Winter Conference on Brain-Computer Interface (BCI), Gangwon, Korea, 22–24 February 2021; pp. 1–6.
- Choi, H.-I.; Williams, W.J. Improved Time-Frequency Representation of Multicomponent Signals Using Exponential Kernels. IEEE Trans. Acoust. Speech Signal Process. 1989, 37, 862–871. [CrossRef]
- Alazrai, R.; Al-Saqqaf, A.; Al-Hawari, F.; Alwanni, H.; Daoud, M.I. A Time-Frequency Distribution-Based Approach for Decoding Visually Imagined Objects Using EEG Signals. *IEEE Access* 2020, *8*, 138955–138972. [CrossRef]
- Alazrai, R.; Abuhijleh, M.; Ali, M.Z.; Daoud, M.I. A Deep Learning Approach for Decoding Visually Imagined Digits and Letters Using Time–Frequency–Spatial Representation of EEG Signals. *Expert Syst. Appl.* 2022, 203, 117417. [CrossRef]
- Lee, S.-H.; Lee, M.; Jeong, J.-H.; Lee, S.-W. Towards an EEG-Based Intuitive BCI Communication System Using Imagined Speech and Visual Imagery. In Proceedings of the 2019 IEEE International Conference on Systems, Man and Cybernetics (SMC), Bari, Italy, 6–9 October 2019; pp. 4409–4414.
- Lee, S.-H.; Lee, M.; Lee, S.-W. Neural Decoding of Imagined Speech and Visual Imagery as Intuitive Paradigms for BCI Communication. *IEEE Trans. Neural Syst. Rehabil. Eng.* 2020, 28, 2647–2659. [CrossRef] [PubMed]
- Renard, Y.; Lotte, F.; Gibert, G.; Congedo, M.; Maby, E.; Delannoy, V.; Bertrand, O.; Lécuyer, A. OpenViBE: An Open-Source Software Platform to Design, Test, and Use Brain–Computer Interfaces in Real and Virtual Environments. *Presence Teleoperators Virtual Environ.* 2010, 19, 35–53. [CrossRef]
- Delorme, A.; Makeig, S. EEGLAB: An Open Source Toolbox for Analysis of Single-Trial EEG Dynamics Including Independent Component Analysis. J. Neurosci. Methods 2004, 134, 9–21. [CrossRef]
- 44. Comon, P. Independent Component Analysis, A New Concept? Signal Process. 1994, 36, 287–314. [CrossRef]
- 45. 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] [PubMed]
- 46. Tan, C.W.; Dempster, A.; Bergmeir, C.; Webb, G.I. MultiRocket: Multiple Pooling Operators and Transformations for Fast and Effective Time Series Classification. *Data Min. Knowl. Discov.* **2022**, 1–24. [CrossRef]
- Sandler, M.; Howard, A.; Zhu, M.; Zhmoginov, A.; Chen, L.-C. MobileNetV2: Inverted Residuals and Linear Bottlenecks. In Proceedings of the 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–23 June 2018; pp. 4510–4520.
- 48. Dornhege, G.; Blankertz, B.; Curio, G.; Muller, K.-R. Boosting Bit Rates in Noninvasive EEG Single-Trial Classifications by Feature Combination and Multiclass Paradigms. *IEEE Trans. Biomed. Eng.* **2004**, *51*, 993–1002. [CrossRef] [PubMed]
- Wu, W.; Gao, X.; Gao, S. One-Versus-the-Rest(OVR) Algorithm: An Extension of Common Spatial Patterns(CSP) Algorithm to Multi-Class Case. In Proceedings of the 2005 IEEE Engineering in Medicine and Biology 27th Annual Conference, Shanghai, China, 17–18 January 2005; pp. 2387–2390.
- 50. Maas, A.L.; Hannun, A.Y.; Ng, A.Y. Rectifier Nonlinearities Improve Neural Network Acoustic Models. In Proceedings of the 30th International Conference on Machine Learning, Atlanta, GA, USA, 16–21 June 2013; p. 3.
- 51. Dau, H.A.; Bagnall, A.; Kamgar, K.; Yeh, C.-C.M.; Zhu, Y.; Gharghabi, S.; Ratanamahatana, C.A.; Keogh, E. The UCR Time Series Archive. *IEEECAA J. Autom. Sin.* 2019, *6*, 1293–1305. [CrossRef]
- Ismail Fawaz, H.; Lucas, B.; Forestier, G.; Pelletier, C.; Schmidt, D.F.; Weber, J.; Webb, G.I.; Idoumghar, L.; Muller, P.-A.; Petitjean, F. InceptionTime: Finding AlexNet for Time Series Classification. *Data Min. Knowl. Discov.* 2020, 34, 1936–1962. [CrossRef]
- 53. Middlehurst, M.; Large, J.; Flynn, M.; Lines, J.; Bostrom, A.; Bagnall, A. HIVE-COTE 2.0: A New Meta Ensemble for Time Series Classification. *Mach. Learn.* 2021, *110*, 3211–3243. [CrossRef]
- Makeig, S. Auditory Event-Related Dynamics of the EEG Spectrum and Effects of Exposure to Tones. *Electroencephalogr. Clin.* Neurophysiol. 1993, 86, 283–293. [CrossRef]
- 55. Cohen, L. Generalized Phase-Space Distribution Functions. J. Math. Phys. 1966, 7, 781–786. [CrossRef]
- SAYIN, F.S.; AKGÜN, Ö. Higher Order Spectral Analysis of Ventricular Arrhythmic ECG Signals with MATLAB HOSA Toolbox. In Proceedings of the 2018 6th International Conference on Control Engineering & Information Technology (CEIT), Istanbul, Turkey, 25–27 October 2018; pp. 1–4.
- 57. Yosinski, J.; Clune, J.; Bengio, Y.; Lipson, H. How Transferable Are Features in Deep Neural Networks? In Proceedings of the Advances in Neural Information Processing Systems; Curran Associates, Inc.: Red Hook, NY, USA, 2014; Volume 27.
- Abadi, M.; Barham, P.; Chen, J.; Chen, Z.; Davis, A.; Dean, J.; Devin, M.; Ghemawat, S.; Irving, G.; Isard, M.; et al. TensorFlow: A System for Large-Scale Machine Learning. In Proceedings of the 12th USENIX Conference on Operating Systems Design and Implementation, Savannah, GA, USA, 2–4 November 2016; pp. 265–283.
- 59. Kleih, S.C.; Kübler, A. Empathy, Motivation, and P300 BCI Performance. Front. Hum. Neurosci. 2013, 7, 642. [CrossRef] [PubMed]

- Kleih, S.C.; Kübler, A. Psychological Factors Influencing Brain-Computer Interface (BCI) Performance. In Proceedings of the 2015 IEEE International Conference on Systems, Man, and Cybernetics, Hong Kong, China, 9–12 October 2015; pp. 3192–3196.
- Baykara, E.; Ruf, C.A.; Fioravanti, C.; K\u00e4thner, I.; Simon, N.; Kleih, S.C.; K\u00fcbler, A.; Halder, S. Effects of Training and Motivation on Auditory P300 Brain–Computer Interface Performance. *Clin. Neurophysiol.* 2016, 127, 379–387. [CrossRef]
- Kleih-Dahms, S.C.; Botrel, L.; Kübler, A. The Influence of Motivation and Emotion on Sensorimotor Rhythm-Based Brain– Computer Interface Performance. *Psychophysiology* 2021, 58, e13832. [CrossRef]
- 63. Kahlbrock, N.; Butz, M.; May, E.S.; Schnitzler, A. Sustained Gamma Band Synchronization in Early Visual Areas Reflects the Level of Selective Attention. *NeuroImage* **2012**, *59*, 673–681. [CrossRef]
- 64. Knoblich, U.; Siegle, J.; Pritchett, D.; Moore, C. What Do We Gain from Gamma? Local Dynamic Gain Modulation Drives Enhanced Efficacy and Efficiency of Signal Transmission. *Front. Hum. Neurosci.* **2010**, *4*, 185. [CrossRef] [PubMed]
- Gaona, C.M.; Sharma, M.; Freudenburg, Z.V.; Breshears, J.D.; Bundy, D.T.; Roland, J.; Barbour, D.L.; Schalk, G.; Leuthardt, E.C. Nonuniform High-Gamma (60–500 Hz) Power Changes Dissociate Cognitive Task and Anatomy in Human Cortex. *J. Neurosci.* 2011, 31, 2091–2100. [CrossRef]
- Jerbi, K.; Ossandón, T.; Hamamé, C.M.; Senova, S.; Dalal, S.S.; Jung, J.; Minotti, L.; Bertrand, O.; Berthoz, A.; Kahane, P.; et al. Task-Related Gamma-Band Dynamics from an Intracerebral Perspective: Review and Implications for Surface EEG and MEG. *Hum. Brain Mapp.* 2009, 30, 1758–1771. [CrossRef]
- 67. Nottage, J.F.; Horder, J. State-of-the-Art Analysis of High-Frequency (Gamma Range) Electroencephalography in Humans. *Neuropsychobiology* **2015**, *72*, 219–228. [CrossRef] [PubMed]
- 68. He, C.; Liu, J.; Zhu, Y.; Du, W. Data Augmentation for Deep Neural Networks Model in EEG Classification Task: A Review. *Front. Hum. Neurosci.* **2021**, *15*, 765525. [CrossRef] [PubMed]
- 69. Zhang, Q.; Liu, Y. Improving Brain Computer Interface Performance by Data Augmentation with Conditional Deep Convolutional Generative Adversarial Networks. *arXiv* 2018, arXiv:1806.07108.
- Abdelfattah, S.M.; Abdelrahman, G.M.; Wang, M. Augmenting The Size of EEG Datasets Using Generative Adversarial Networks. In Proceedings of the 2018 International Joint Conference on Neural Networks (IJCNN), Rio de Janeiro, Brazil, 8–13 July 2018; pp. 1–6.
- Aznan, N.K.N.; Connolly, J.D.; Moubayed, N.A.; Breckon, T.P. Using Variable Natural Environment Brain-Computer Interface Stimuli for Real-Time Humanoid Robot Navigation. In Proceedings of the 2019 International Conference on Robotics and Automation (ICRA), Montreal, QC, Canada, 20–24 May 2019; pp. 4889–4895.
- 72. Kalaganis, F.P.; Laskaris, N.A.; Chatzilari, E.; Nikolopoulos, S.; Kompatsiaris, I. A Data Augmentation Scheme for Geometric Deep Learning in Personalized Brain–Computer Interfaces. *IEEE Access* 2020, *8*, 162218–162229. [CrossRef]
- 73. Zhang, Z.; Duan, F.; Solé-Casals, J.; Dinarès-Ferran, J.; Cichocki, A.; Yang, Z.; Sun, Z. A Novel Deep Learning Approach With Data Augmentation to Classify Motor Imagery Signals. *IEEE Access* 2019, 7, 15945–15954. [CrossRef]
- Zhang, P.; Wang, X.; Zhang, W.; Chen, J. Learning Spatial–Spectral–Temporal EEG Features With Recurrent 3D Convolutional Neural Networks for Cross-Task Mental Workload Assessment. *IEEE Trans. Neural Syst. Rehabil. Eng.* 2019, 27, 31–42. [CrossRef]
- 75. Jiao, Z.; Gao, X.; Wang, Y.; Li, J.; Xu, H. Deep Convolutional Neural Networks for Mental Load Classification Based on EEG Data. *Pattern Recognit.* **2018**, *76*, 582–595. [CrossRef]
- Vetter, N.C.; Fröhner, J.H.; Hoffmann, K.; Backhausen, L.L.; Smolka, M.N. Adolescent to Young Adult Longitudinal Development across 8 Years for Matching Emotional Stimuli during Functional Magnetic Resonance Imaging. *Dev. Cogn. Neurosci.* 2022, 57, 101131. [CrossRef]
- 77. Togo, T.; Sanjo, Y.; Sakai, K.; Nomura, T. Brain Activity in Healthy Elderly Persons When Presented with Swallowing-Related Videos: A Functional Magnetic Resonance Imaging Study. *J. Oral Maxillofac. Surg. Med. Pathol.* **2022**, in press. [CrossRef]
- Casey, B.; Galván, A.; Somerville, L.H. Beyond Simple Models of Adolescence to an Integrated Circuit-Based Account: A Commentary. Dev. Cogn. Neurosci. 2016, 17, 128–130. [CrossRef] [PubMed]
- Volosyak, I.; Gembler, F.; Stawicki, P. Age-Related Differences in SSVEP-Based BCI Performance. *Neurocomputing* 2017, 250, 57–64. [CrossRef]
- Hsu, H.-T.; Lee, I.-H.; Tsai, H.-T.; Chang, H.-C.; Shyu, K.-K.; Hsu, C.-C.; Chang, H.-H.; Yeh, T.-K.; Chang, C.-Y.; Lee, P.-L. Evaluate the Feasibility of Using Frontal SSVEP to Implement an SSVEP-Based BCI in Young, Elderly and ALS Groups. *IEEE Trans. Neural* Syst. Rehabil. Eng. 2016, 24, 603–615. [CrossRef]
- 81. Angelov, P.P.; Soares, E.A.; Jiang, R.; Arnold, N.I.; Atkinson, P.M. Explainable Artificial Intelligence: An Analytical Review. *WIREs Data Min. Knowl. Discov.* 2021, *11*, e1424. [CrossRef]
- 82. Nesvijevskaia, A.; Ouillade, S.; Guilmin, P.; Zucker, J.-D. The Accuracy versus Interpretability Trade-off in Fraud Detection Model. *Data Policy* **2021**, *3*, e12. [CrossRef]
- 83. Selvaraju, R.R.; Cogswell, M.; Das, A.; Vedantam, R.; Parikh, D.; Batra, D. Grad-CAM: Visual Explanations from Deep Networks via Gradient-Based Localization. *Int. J. Comput. Vis.* **2020**, *128*, 336–359. [CrossRef]
- Bach, S.; Binder, A.; Montavon, G.; Klauschen, F.; Müller, K.-R.; Samek, W. On Pixel-Wise Explanations for Non-Linear Classifier Decisions by Layer-Wise Relevance Propagation. *PLoS ONE* 2015, 10, e0130140. [CrossRef]