future internet

MDPI

Article

Optimization of Wheelchair Control via Multi-Modal Integration: Combining Webcam and EEG

Lassaad Zaway [1,2,*], Nader Ben Amor [1], Jalel Ktari [1], Mohamed Jallouli [1], Larbi Chrifi Alaoui [3] and Laurent Delahoche [3]

[1] Laboratory of Computer Embedded System, National School of Engineering of Sfax, University of Sfax, Sfax 3029, Tunisia; nader.benamor@enis.tn (N.B.A.); jalel.ktari@enis.tn (J.K.); mohjallouli@gmail.com (M.J.)
[2] National School of Engineering of Gabes, University of Gabes, Gabes 6029, Tunisia
[3] Laboratory of Innovative Technologies, University of Picardie Jules Verne, 80000 Amiens, France; larbi.alaoui@u-picardie.fr (L.C.A.); laurent.delahoche@u-picardie.fr (L.D.)
* Correspondence: zawaylasaad@gmail.com

Abstract: Even though Electric Powered Wheelchairs (EPWs) are a useful tool for meeting the needs of people with disabilities, some disabled people find it difficult to use regular EPWs that are joystick-controlled. Smart wheelchairs that use Brain–Computer Interface (BCI) technology present an efficient solution to this problem. This article presents a cutting-edge intelligent control wheelchair that is intended to improve user involvement and security. The suggested method combines facial expression analysis via a camera with EEG signal processing using the EMOTIV Insight EEG dataset. The system generates control commands by identifying specific EEG patterns linked to facial expressions such as eye blinking, winking left and right, and smiling. Simultaneously, the system uses computer vision algorithms and inertial measurements to analyze gaze direction in order to establish the user's intended steering. The outcomes of the experiments prove that the proposed system is reliable and efficient in meeting the various requirements of people, presenting a positive development in the field of smart wheelchair technology.

Keywords: Electroencephalogram (EEG); facial expressions; Long Short-Term Memory (LSTM); fusion data; convolutional neural network (CNN); control wheelchairs

Citation: Zaway, L.; Ben Amor, N.; Ktari, J.; Jallouli, M.; Chrifi Alaoui, L.; Delahoche, L. Optimization of Wheelchair Control via Multi-Modal Integration: Combining Webcam and EEG. *Future Internet* **2024**, *16*, 158. https://doi.org/10.3390/fi16050158

Academic Editors: Christos Troussas, Akrivi Krouska, Cleo Sgouropoulou and Jaime Caro

Received: 7 March 2024
Revised: 23 March 2024
Accepted: 28 March 2024
Published: 3 May 2024

1. Introduction

The utilization of assistive technologies, particularly electric wheelchairs, has experienced a significant surge in recent decades, playing a pivotal role in supporting individuals facing mobility challenges and enhancing their overall quality of life. The World Health Organization (WHO) reports that over 1.3 billion people globally grapple with structural or functional impairments, with at least 80 million necessitating the use of wheelchairs [1]. While conventional control methods like keyboards, mice, joysticks, or touchscreens are effective for healthy users, they often prove impractical for those severely disabled due to conditions such as spinal cord injuries, paralysis, muscular dystrophy, multiple sclerosis, or stroke. Traditional control methods, such as joystick-based interfaces, exhibit limitations in usability, particularly for those with severe motor disabilities [2]. Consequently, the development of an efficient interface enabling users with physical disabilities to convey their intentions or commands to assistive devices becomes imperative.

In response, researchers have turned to innovative technologies, notably BCIs, to devise, design and propose alternative and more intuitive control strategies, as demonstrated by Rebsamen, B. et al. [3]. BCIs establish a direct communication link between the human brain and external devices, eliminating the reliance on traditional motor control mechanisms [4]. This technology holds immense potential for transforming the lives of individuals with limited mobility, granting them the ability to control assistive devices, such as wheelchairs, through their brain activity as discussed in [5].

Various techniques have been employed for wheelchair control, with low-level motion control systems converting EEG data into direct motion commands for wheelchair operation as demonstrated in [5]. Pioneering in this field, Tanaka et al. [6] introduced an electric wheelchair controlled by the user's brain, pioneering a groundbreaking device. By visualizing left or right limb movements during Motor Imagery (MI) tasks, users can dictate the direction of the subsequent movement.

Several approaches have been explored in the realm of BCI-based wheelchair control. MI tasks, wherein users mentally simulate specific movements, have been leveraged to detect intention and initiate corresponding wheelchair motions, as reported by Pires, C. P. et al. [7]. Additionally, Steady-State Visually Evoked Potentials (SSVEP) and P300-based BCIs have been employed to achieve higher accuracy in control through visual stimuli and attention-based paradigms, as proposed by Ortner, R. et al. [8].

Despite the promising results demonstrated by these BCI-based control methods, there remain challenges that require attention for practical implementation. Issues such as inter-subject variability in brain rhythms, the need for prolonged visual stimulation, and the computational complexity of hybrid control approaches pose limitations to achieving seamless and efficient wheelchair control [9].

In this paper, we present a human–machine interaction method for direct control of a robotic wheelchair based on hybrid control using facial expressions captured by a webcam and EEG signals. The main outcome of our research is the ability to recognize the user's control expression based on their gaze direction. This enables a more natural interaction style that could help with continuous control of the wheelchair.

Specifically, our method is dedicated to wheelchair users who are severely disabled but can perform basic tasks with the help of their eyes and heads to operate the wheelchair. To do so, we collected EEG data with an Emotiv Insight headset and used machine learning algorithms (e.g., CNN-LSTM) to recognize signal patterns triggered by various facial expressions, such as Smile (Backward), Eye Blink (Stop), Wink Left (Rotation Left), and Wink Right (Rotation Right). We used a forward-facing camera to track the user's head movements using computer vision methods.

This paper is structured as follows: Section 2 provides a review of related work, while Section 3 details the proposed system architecture. Section 4 delves into our methodology, offering insights into data acquisition, analysis, and classification algorithms. Tests and results are presented and discussed in Section 5. Finally, Section 6 is dedicated to the findings and conclusions of the article.

2. Related Work

On one hand, patients' data are private but on the other hand, remote applications, such as e-health ones, are becoming more and more common, and researchers' primary concern is improving user security. Furthermore, unauthorized users and even hackers pose a major risk to security and privacy. As a result, it becomes extremely difficult to provide effective e-health services while maintaining patient data availability, privacy, and validity. Without a doubt, the first prerequisite is access control, which facial recognition effectively ensures. This section examines related work from two perspectives: embedded solutions for some secured applications and access control mechanisms.

In [10,11], the technique employed for electrophysiological measurements was LORETA (Low-Resolution Electromagnetic Tomography of the Brain). LORETA is based on solving inverse problems to calculate the three-dimensional distribution of neuronal electrical activity. This method serves as a linear estimation technique for determining the sources of signals in the human brain without requiring additional data to be added to the EEG signal. In [10], the authors addressed the removal of biological artifacts, such as the subjects' anxious tics and short squinting of the eyelids. Furthermore, the authors of [12] studied the most common type of Electrooculographic artifact, namely eye blinking. In [13], authors combined an EEG with the power spectrum of eye blink artifacts to develop a

brain–computer interface. They have further conducted a spectrum analysis of EEG data for patients suffering from insomnia.

The authors of [14] focused on artifact reduction in the BCI hybrid. The researchers developed an approach that combines stationary wavelet transforms with adaptive thresholding to effectively remove artifacts from EEG signals. The study in [15] employed the IC MARC classifier to investigate the impact of various artifacts on a motor imagery-based Brain–Computer Interface (BCI) system. The findings demonstrated that when utilizing all 119 EEG channels, muscle artifacts had a detrimental effect on BCI performance. This was observed by comparing the results to a configuration with 48 centrally placed EEG channels. In [16], the authors introduced a new method for automatically eliminating eye-related EEG artifacts using independent component analyses and outlier identification techniques. The OD-ICA method demonstrated effectiveness in removing Ocular Artefacts (OA) while preserving significant EEG signals. Peak detection in online EEGs for BCIs, explored in [17], addressed the impact of filtering on BCI performance, emphasizing peak frequency detection. While peak detection improved with the filter, the BCI performance suffered from movement and increased artifact removal. The issue of noisy EEG data was resolved in [18] by employing the LombScargle periodogram for spectral power estimation and a denoising autoencoder (DAE) for training, successfully decoding insufficient EEG recordings.

Networks and highly classified expert systems, including AI solutions, are important. In [19], a convolutional neural network (CNN) was employed to perform skin cancer classification. Furthermore, [20] utilized a deep neural network (DNN) to classify histopathologic images of breast cancer. In contrast, ref. [21] employed a hybrid convolutional and recurrent deep neural network for classification purposes. Automation system control through artificial intelligence was addressed in [22], presenting a method for controlling a robot using an algorithm based on an artificial neural network.

Currently, several studies exist on deep learning theories, such as convolutional neural networks (CNNs). These techniques prove their effectiveness. The authors of [23,24] suggested a research strategy for producing 3D channel spectrograms that combines three different time–frequency representations (spectrograms, gamma-tone spectrograms, and continuous wavelet transform). Applications such as the automatic identification of phoneme classes for phone attribute extraction and the diagnosis of speech impairments in cochlear implant users have been successful.

In the context of wheelchair control, Ba-Viet et al. (2020) explored wheelchair navigation using EEG signals and 2D maps with a camera [25]. Other researchers investigated an intelligent wheelchair using eye detection and visual systems (Dorian 2021 [26], Agnes 2022 [27]). Notably, these experiments revealed a notable disparity in perceived convenience, referred to as preferred control, due to the inherent instability of photos influenced by light and the sensitivity of the EEG signal to an individual's state. Our study aims to enhance control system performance through the combination of facial expressions and neural signals.

3. Proposed System Architecture

3.1. System Architecture

Throughout this paper, we focus on the design and the implementation of a smart EPW with hybrid control technologies that will enhance/improve the safety of those with impairments. This EPW is composed of a webcam, an EEG headset, and two sensors. Artificial intelligence techniques will be employed to amalgamate the data obtained from the two sensors.

The architecture of our proposal/solution is illustrated in Figure 1. Data processing and fusion-utilizing Python-developed deep learning algorithms constitute the first phase. We employed the decision fusion technique, in which a directional choice (Forward, Rotation Left, Rotation Right, Stop) is provided by each sensor. The embedded system in charge of the wheelchair receives the final choice made by the fusion after that.

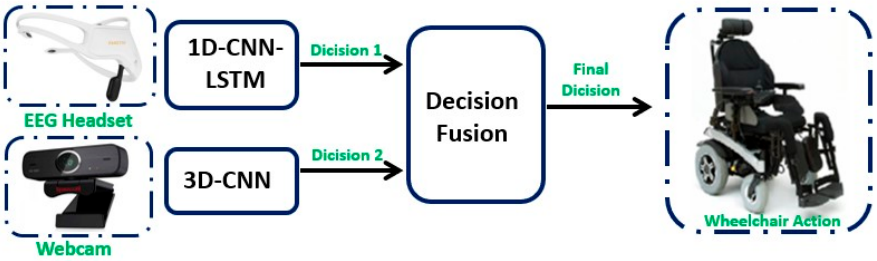


Figure 1. Architecture of the proposed system.

Our method includes using EEG waves to interpret facial expressions. We chose four emotions to be represented by four specific movements, as indicated by the corresponding weights in Table 1, as our objective is the EPW command.

Table 1. A table of feelings and the movements that go with them.

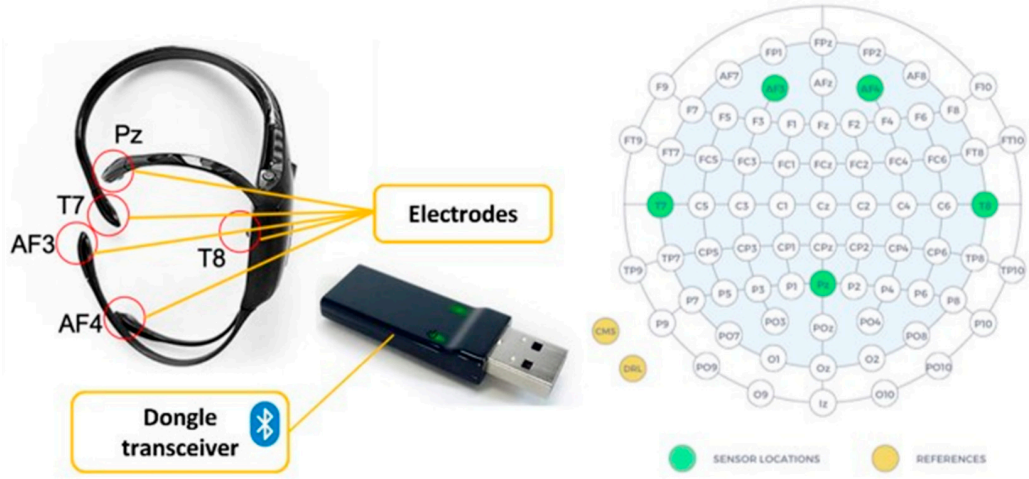
Number	Facial Expressions	Movements
1	Wink Right	Rotation Right
2	Wink Left	Rotation Left
3	Blink	Stop
4	Smile	Forward

The feelings sent by the headset are shown in the first column. The events linked to each emotion are listed in the second column. For instance, the Rotation Left movement and the Wink Left expression of number 1 match.

3.2. Used Sensors

3.2.1. The EEG Headset

We employed the EMOTIV Insight headset in this study, which can be seen in Figure 2a. This device has 5 electrode channels (AF3, T7, PZ, T8, and AF4). Figure 2b represents the location of all five channels in the brain. The data were sampled using a 16 bit analog-to-digital converter at a frequency of 128 Hz [28].



(**a**) (**b**)

Figure 2. (**a**) EMOTIV Insight headset; (**b**) location of channel.

Brain waves are analyzed by the EEG headset, which then wirelessly transmits the data to a computer via Bluetooth. EEG data gathering is made possible using the EMOTIV (BCI-OSC) V3.5 software interface. Figure 3 represents the architecture of the headset for the acquisition and processing of EEG signals from EMOTIV Insight.

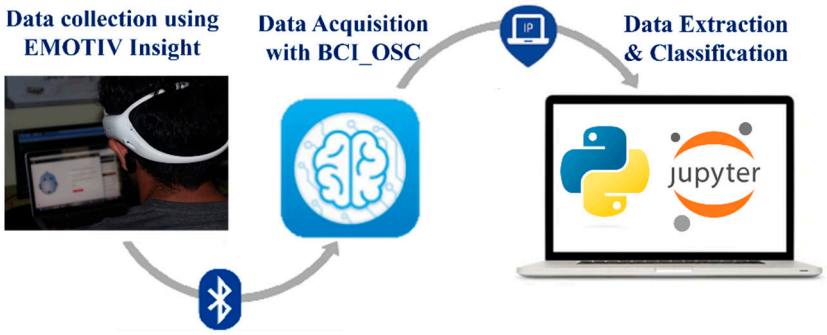


Figure 3. Interface of the EMOTIV BCI-OSC.

3.2.2. The Webcam

We used the integrated camera of our PC for quick validation purposes, which features an advanced RGB HD configuration, ensuring outstanding image quality. With a high resolution of 2 megapixels, it produces sharp and detailed still images and videos. The camera boasts a resolution of 1280 × 720 pixels (HD). For video capture, it operates seamlessly at 30 frames per second. Additionally, the camera's wide-angle lens offers a generous diagonal viewing angle of 74.9°.

The webcam was used to identify the same four facial expressions (Smile, Blink, Wink Left, and Wink Right) and equivalent commands (Forward, Stop, Rotation Left, and Rotation Right).

4. Methodology

In the "webcam-based command" section, we will present the fundamental structure of the 3D-CNN connected to image classification. Moving on to the "EEG headset-based command" section, we will delve into the essential components of the 1D-CNN-LSTM approach for processing EEG data. Lastly, in Section 3, we will introduce a feature fusion network that combines the two aforementioned methods and elaborate on our selection process for the network parameters.

4.1. Webcam Based Command

4.1.1. Detection of Face, Eye, and Mouth

The image acquisition and processing classification system, illustrated in Figure 4, consists of three primary steps. The first step employs the Open-CV library for face recognition. In the second step, learning takes place, followed by the third step, which involves image classification. Deep learning algorithms, specifically convolutional neural networks, are utilized for both learning and classification, enabling matching with the database. Ultimately, the classification results are displayed to the user.

Figure 4. The proposed system architecture.

4.1.2. Report of Closing and Opening Emotion

This method is used to separate the difference between blinking and winking left/right, naturally or artificially. Blinking is the rapid closing and reopening of the human eye. Each individual has a slightly different blinking pattern. The pattern differs in the speed of

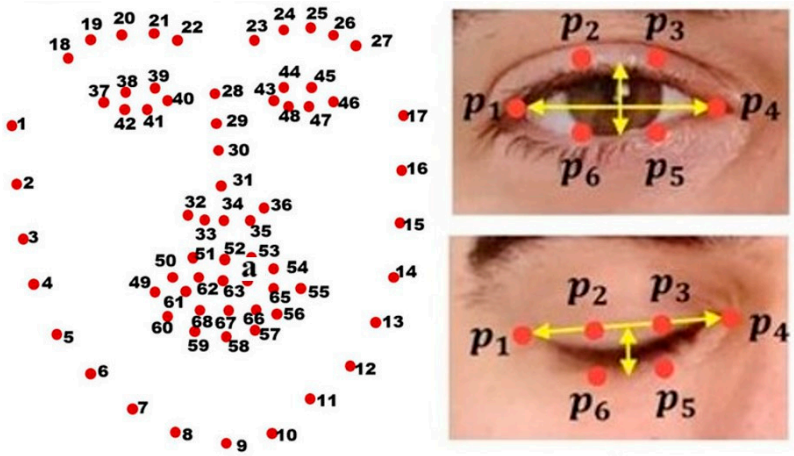
closing and opening, the degree of pressure on the eye, and the duration of blinking. The eye blink lasts approximately 100 to 400 ms [29].

We used four expressions of Wink Right, Wink Left, Blink, and Smile with measurements of the Eye Aspect Ratio (EAR). There are numerous algorithms for face recognition, but we will only focus on Dlib's method in this paper. Dlib uses the HOG (Histogram of Oriented Gradients).

To do this, we must first localize the human face in the overall image. Face detection is a technique that identifies a human face in an image and returns the value of the bounding box or rectangle associated with the face in x, y, w, and h coordinates [29].

We must first determine the position of the face in the image before determining the its smallest features, such as the lips and eyebrows. By using points within this rectangle, the face recognition software can identify all the necessary features of a human face.

The 68-point model of Dlib is shown in Figure 5, where points from 1 to 68 are visible. We will discuss how to recognize these emotions (Blink, Wink Left, Wink Right, and Smile) and how to recognize the EAR.



(**a**) (**b**)

Figure 5. (**a**) Dlib landmark: 68 points for face; (**b**) 6 points for eye [30].

We used state-of-the-art facial feature recognition systems to localize the contours of the eyes and eyelids. From the 68 points in the image, we derived the EAR, which was used as a parameter to estimate the opening state of the eyes, as shown.

$$\text{EAR} = \frac{||\text{p2} - \text{p6}|| + ||\text{p3} - \text{p5}||}{2||\text{p1} - \text{p4}||} \quad (1)$$

where p1, p2, p3, p4, p5, and p6, shown in Figure 5b, are the same points found on the circumference of the left eye, respectively 37, 38, 39, 40, 41, and 42 in Figure 5a. Using this metric, a classifier was used to recognize eye blinks and left and right winks.

The distance between the corners of the mouth increased. However, since different people have different mouth sizes, we normalized this metric by dividing it by the jaw distance to obtain a general ratio that can be used on different people.

In our base detector, we will use the x-coordinates of points 49, 55, 3, and 15 to calculate the EAR_Smile defined in Equation (2). p49, p51, p53, p55, p57, and p59 are the same points found on the circumference of the mouth in Figure 6.

$$\text{EAR\_Smile} = \frac{||\text{p51} - \text{p59}|| + ||\text{p53} - \text{p57}||}{2||\text{p49} - \text{p55}||} \quad (2)$$

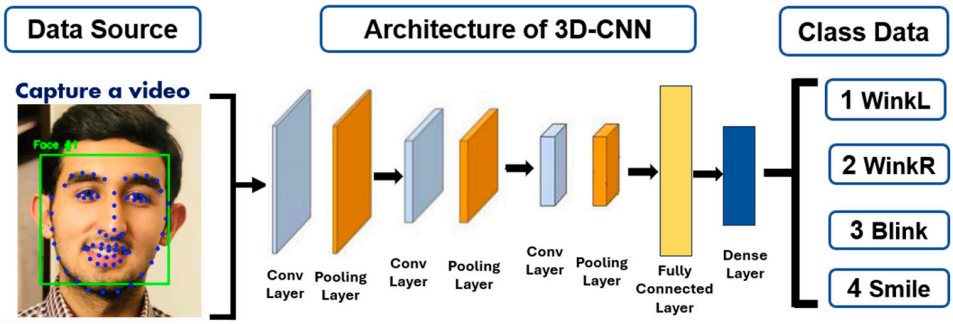


Figure 6. The architecture of algorithm 3D-CNN [30].

4.1.3. Algorithm of Classification

The task of emotion estimation is to determine whether an eye blinks, is winking left, is winking right, or is smiling. In cases where faces are not fully frontal, the proposed CNN was used to extract more features that are resilient and effectively classify four emotions. To improve the performance of these methods, we tracked a sequence of images instead of inputting a single image [31,32].

The neural network topology (3D-CNN) is depicted in Figure 6. CNN has extensive applications in various domains, such as natural language processing, recommendation systems, and image and video recognition. These networks were utilized in our situation to process images for smiling, winking to the left or right, and blinking of the eyes.

Table 2 displays the architecture of our neural network. Filtering the input signals is the responsibility of the convolutional layer. Its objective is to minimize training time and minimize data volume while preserving quality. It has a max-pooling layer connected to it. This process was performed three times on Layers 1, 2, and 3 as indicated in Figure 6. By connecting to a fully connected class, this layer establishes the relationship between a class and the positions of features in an image. Essentially, the thick layer modifies the size of the tensor, which is the basic data structure that forms the basis of all machine and deep learning methods.

Table 2. Characteristics of the 3D-CNN structure [30].

Layer Type	Filter Shape	Input Size
Convolution (ReLu)	$3 \times 3 \times 3 \times 128$	$227 \times 227 \times 128$
Max-Pooling	2×2	$225 \times 225 \times 128$
Convolution (ReLu)	$3 \times 3 \times 3 \times 64$	$112 \times 112 \times 64$
Max-Pooling	2×2	$110 \times 110 \times 64$
Convolution (ReLu)	$3 \times 3 \times 3 \times 32$	$55 \times 55 \times 32$
Max-Pooling	2×2	$27 \times 27 \times 32$
Fully Connected	32×4	$27 \times 27 \times 32$
Dense (Sigmoid)	4	4×1

4.2. EEG Headset-Based Command

4.2.1. Data Acquisition

To obtain EEG data with the EMOTIV Insight headset, subjects were exposed to 4 different emotions (Smile, Blink, Wink Left, and Wink Right) for 5 min, so that EEG data were available for each subject for a total of 20 min. A twenty-minute data collection protocol was created, with the different emotions first recorded in a separate file. Subsequently, all emotions of each person were consolidated into one file to avoid interference between different expressions and persons. This consolidation also ensured a balanced distribution of each emotion and facilitated a more accurate analysis without compromising the integrity of the dataset.

The data collected in the different situations mentioned above need to be preprocessed for later use in the machine learning component. To better understand the preprocessing, the data source as well as the format and features are discussed below.

The 5-channel EEG device provided 128 samples for each channel in a second. After transforming the collected data for the frequency domain, each collected datum is represented by the weighted and arithmetic mean for each of the 5 device channels and the 4 wave classifications.

4.2.2. Preprocessing

The process of modifying, resolving, and organizing data inside a dataset to make it generally consistent and ready for analysis is known as data cleaning. To ensure the best possible analysis, this entails purging any corrupted or unnecessary data and formatting these in a computer-readable manner [33].

It is therefore important to carry out proper data cleaning to ensure that the best possible results are obtained.

Data cleaning is composed of six steps:

- Remove irrelevant data;
- Duplicate data;
- Fix structural errors;
- Handle missing data;
- Filter outliers;
- Validate the data.

The process of making altered copies of a data set is known as data augmentation, and it is a mean of artificially expanding the training set. This entails either creating new data points via deep learning or making small adjustments to the dataset. We used the EEG data to use this data augmentation method. To expand the breadth and diversity of the training set, some small changes were made to the original data before generating these new data. This included removing duplicate data and handling missing data by calculating the average of the data before and after these data [33].

These data were derived from the original data, with some minor modifications to increase the size and variety of the training set. These are examples of straightforward data augmentation techniques which include random swapping, insertion, and synonym replacement.

4.2.3. Algorithm of Classification

Figure 7 represents the architecture of our system. In this work, we processed EEG signals for the estimation of facial expressions using a CNN and LSTM combination [34]. The ability to extract robust features from CNNs and go beyond the drawbacks of conventional techniques is a true advantage of CNN networks. Three layers make up the network: LSTM, max-pooling, and convolutional. The convolutional layer filters the input signals. The max-pooling layer reduces the size of the data while preserving their features, which helps to shorten the training period. Using the EEG database, the LSTM layer trains the model and performs classification [34].

Table 3 provides the specifics of our neural network's construction. The EEG modalities' 1D-CNN-LSTM architecture is displayed. Sequence data include EEG recordings that last one second. Two one-dimensional convolutional layers, Conv Layers 1 and 2, were then applied to this sequence as is indicated in Figure 7. Max-pooling and ReLU activation layers came next for each, allowing for the direct extraction of temporal properties from the time series data. The collected features were then flattened for the LSTM layer. This LSTM layer determines the order link between the gathered temporal features to categorize time series data. Next, by connecting the layer to a fully linked class, the relationship between the locations of features in the EEG data and a class was established. In essence, the dimensions of the tensor are altered by the dense layer.

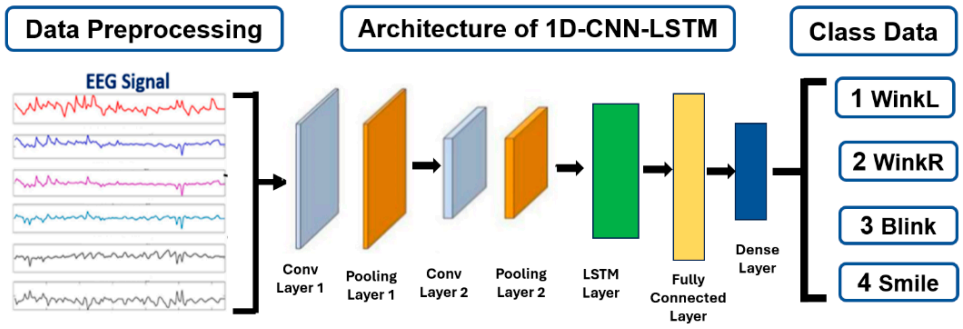


Figure 7. The architecture of algorithm 1D-CNN-LSTM [34].

Table 3. Characteristics of the 1D-CNN + LSTM structure [30].

Layer Type	Filter Shape	Input Size
Convolution (ReLu)	3 × 3 × 64	14 × 14 × 64
Max-Pooling	2 × 2	12 × 12 × 64
Convolution (ReLu)	3 × 3 × 32	6 × 6 × 64
Max-Pooling	2 × 2	4 × 4 × 32
LSTM	100	2 × 2 × 32
Fully Connected	32 × 4	2 × 2 × 100
Dense (Sigmoid)	4	4 × 1

The combination of Convolutional Neural Networks (CNNs) and Long Short-Term Memory networks (LSTMs) offers several advantages in the processing of signals. CNNs excel at extracting local features from inputs by utilizing convolution operations and pooling layers. They are particularly adept at capturing patterns in spatial dimensions. On the other hand, LSTMs possess exceptional capabilities when it comes to processing sequential data, such as time series or EEG signals, due to their ability to handle long dependencies through gated recurrent units.

4.3. Fusion of EEG Signals Decision and Image Processing Decision

Implementing layer fusion for decision-making involves combining the outputs of multiple layers or models to arrive at a final decision. This fusion process can be achieved using various techniques, such as voting, averaging, or weighted averaging. Here is an example of implementing layer fusion for decision-making.

In this work, we first obtained the predictions from each model for the given input data. Next, we performed fusion to combine these predictions and made a final decision based on the fusion result.

Our goal was to increase the recognition accuracy of these emotions (Smile, WinkL, WinkR, and Blink) by merging webcam images with EEG data. Webcams function as a visual supplement to the EEG data, capturing the dynamic interaction of facial emotions and eyebrow movements. The integration of the image processing classifier and the EEG classifier is depicted in Figure 8. Before performing fusion between two of these classifications, we worked on two classification algorithms: 1D-CNN-LSTM for processing EEG signals from the EMOTIV Insight and 3D-CNN for analyzing pictures.

We use voting-based fusion to count the occurrences of each prediction and select the one with the highest count as the fused prediction. Based on the fused prediction, we made a decision by mapping it to the corresponding class label.

In a deep learning algorithm, the concatenation layer is used to combine the outputs of multiple layers or branches of the network. This can improve the network's ability to learn complex relationships by allowing it to access features learned at different levels of abstraction. Regularization techniques, such as dropout and Webcam/EEG signal regularization, are used to prevent overfitting of the network to the training data. Exclusion randomly removes some units from the network during training, forcing the network to

learn redundant representations, which can improve its ability to generalize to the new data. Overall, concatenation and regularization techniques are important tools in designing deep learning algorithms that can learn complex relationships and generalize well to new data.

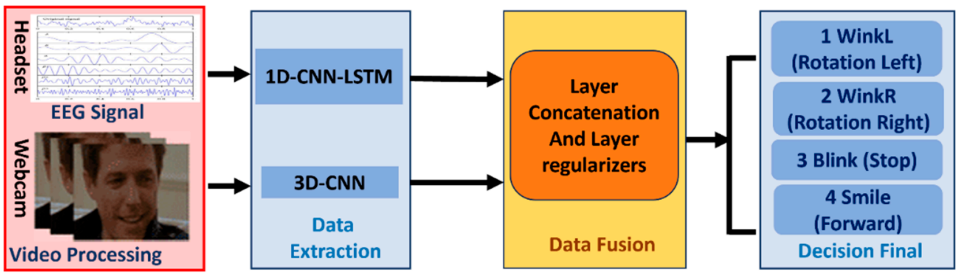


Figure 8. The structure of our model for the fusion of EEG signals and images.

We employed two layers; the concatenation layer typically refers to a neural network layer that combines or fuses information from multiple sources or modalities. This is common in multimodal deep learning when working with data from different sources such as text, images, and audio and wanting to combine them for a specific task.

We first defined two sets of features as input layers (input_features1 for data from the Webcam and input_features2 for data from EEG signals). We then used the regularizer layer to combine the outputs of multiple layers, effectively fusing the information from the two sets.

Additionally, it was important to adjust the number of features, activation functions, and other hyper-parameters based on our use case (4 features in the first set (Webcam), 4 features in the second set (EEG Signals)). Synchronization was achieved with both the headset and webcam operating at the same frequency; the webcam frequency as well as the headset frequency remained the same.

5. Experimental Results and Discussion

5.1. Evaluation Metrics

At the experimental level, we carried out the data preprocessing and signal visualization with the Python software V 3.8 (Jupyter Notebook, Anaconda) and used a PC with 16 GB RAM and an Intel CPU (GeForce GTX 1080) by the company Intel in Santa Clara, California, United States.

Preparing the data for classification requires the following first step. Two portions of the EEG and image dataset were isolated: the first was designated as the training data, or "80%", and the second as the test data, or "20%", which was further subdivided into inputs and outputs. This is consistent with the videos' emotional moods and the EEG readings. These are able to accept a value of either 0 or 1.

We distinguished four validation measures that can be used to assess a classification algorithm.

A model's precision is its capacity to recognize only pertinent items. Equation (3) provides the expression of the percentage of accurate positive predictions:

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}} \tag{3}$$

The capacity of a model to locate all pertinent cases is correlated with recall. Equation (4) provides the expression of the percentage of true positives found among all pertinent field truths:

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}} \tag{4}$$

The accuracy is calculated according to the following Equation (5):

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}} \tag{5}$$

F1-score is calculated by Equation (6):

$$\text{F1\_score} = \frac{2 * \text{Precision} * \text{Recall}}{\text{Precision} * \text{Recal}} \quad (6)$$

where TP = True Positives, TN = True Negatives, FP = False Positives, and FN = False Negatives.

5.2. Evaluation Results

To evaluate the performance of our proposal we proceeded in three steps: First, we performed an evaluation for the EEG. Next, we evaluated image processing performance and finally, we carried out assessments for the fusion of these two sources. Next, we will present the fusion results and compare the results of our technique with the existing techniques. To prove the efficiency of our proposal/technique we will conduct a comparative study with existing techniques.

5.2.1. Webcam Only Command Results

The first step consists of identifying the thresholds to be used. Indeed, we repeated the same tests for a single user and calculated these values using Equations (1) and (2). For eye state detection, the threshold was (0.15) whereas it was (0.35) for mouth state detection.

As shown in Figure 9a, if the EAR of the mouth is equal to (0.37) which exceeds the threshold (0.35), the system detects the emotional expression of a smile. Conversely, in the scenario depicted in Figure 9b, the ratio of the eye is (0.19) which falls over the threshold (0.15), leading to the identification of the blinking emotion.

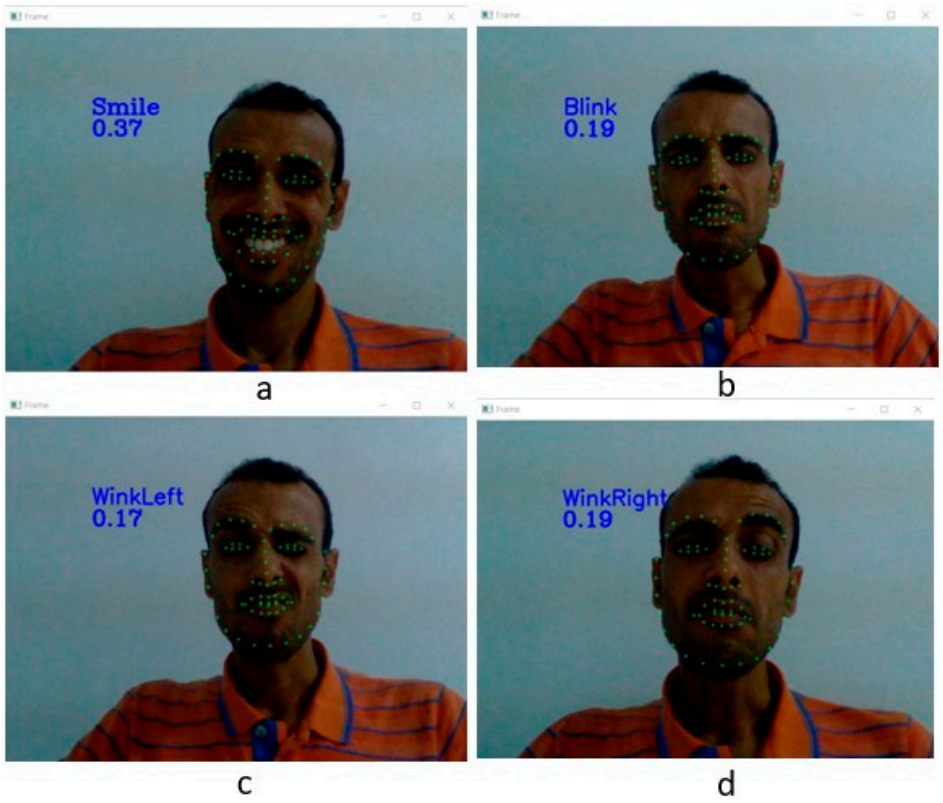


Figure 9. Experimental result of facial expressions with a webcam: (**a**) emotion Smile, (**b**) emotion Blink, (**c**) emotion wink Left, (**d**) emotion wink Right.

5.2.2. EEG Signals Only Command Results

The results of facial expressions inferred from the EEG signals are presented in Figure 10. They correspond to four distinct expressions, labeled from one to four, namely "Wink Left" (1), "Wink Right" (2), "Blink" (3), and "Smile" (4). Each expression is associated

with values ranging from 0 to 1, which can be interpreted as percentages, representing the intensity or level of the respective expression.

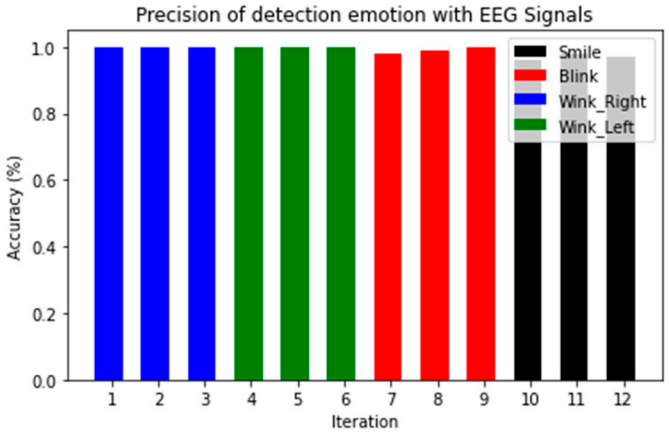


Figure 10. Experimental result of facial expressions with EEG.

To illustrate the effectiveness of our model in discerning and categorizing facial expressions based on EEG signals, we can examine emotion number three, which corresponds to a smile. To ensure the accuracy and consistency of our findings, we conducted a comprehensive set of 400 measurements for each emotion category (1, 2, 3, and 4). In our experiments, we obtained a precision value of 0.96 on the EEG for this particular emotion. The results of these measurements were consistently and accurately classified, with minimal discrepancies. This outcome provides strong evidence for the reliability of our model, confirming its ability to accurately identify and categorize facial expressions using EEG signals.

5.2.3. Fusion Command Results

The study findings, depicted in Figure 11, demonstrated a consistent correlation among the results obtained from the webcam, EEG signals, and their integrated fusion in terms of accurately detecting and recognizing emotions. Additionally, the ear (action unit for the recognition of expression) of emotion was also found to align consistently with the outcomes obtained from the EEG signals and the webcam modalities. This convergence of results across multiple modalities lends further support to the robustness and reliability of our approach in capturing and interpreting emotional states.

```
wink left webcam 0.17701287476043434
wink left EEG 0.15701287476043435
wink left fusion
wink right webcam 0.18224610976434305
wink right EEG 0.16224610976434306
wink right fusion
Blink webcam 0.1782813276648998
Blink EEG 0.2082813276648998
Blink fusion
Smile webcam 0.3877551020408163
Smile EEG 0.4577551020408163
Smile fusion
```

Figure 11. Example of experimental result of fusion.

The alignment observed between facial expressions captured by the webcam and neural activity monitored through EEG signals highlights the effectiveness of combining these two modalities in the context of controlling a wheelchair.

Confusion matrices are a valuable tool for evaluating classification models by summarizing the relationship between predicted and actual labels. They are structured as N × N

tables, with one axis representing predicted labels and the other axis representing true labels. In a multiple classification problem with N classes, confusion matrices provide insights into the model's classification accuracy and revealing its strengths and weaknesses. In our multiple classification problem, N = 4.

Figure 12 shows the confusion matrix of the fusion algorithm for our classification. Where the green color represents correct matrices and the green color represents incorrectly classified matrices. It shows that out of the 400 samples, the actual states of the emotion Wink Left are number 1 in this figure, and the model (fusion) was correctly classified (393) and misclassified (7). Similarly, for these emotions (Wink Right "2", Blink "3", and Smile "4"), 400 samples each (395, 396, 400) were correctly classified, however 5, 4, and 0 were incorrectly classified.

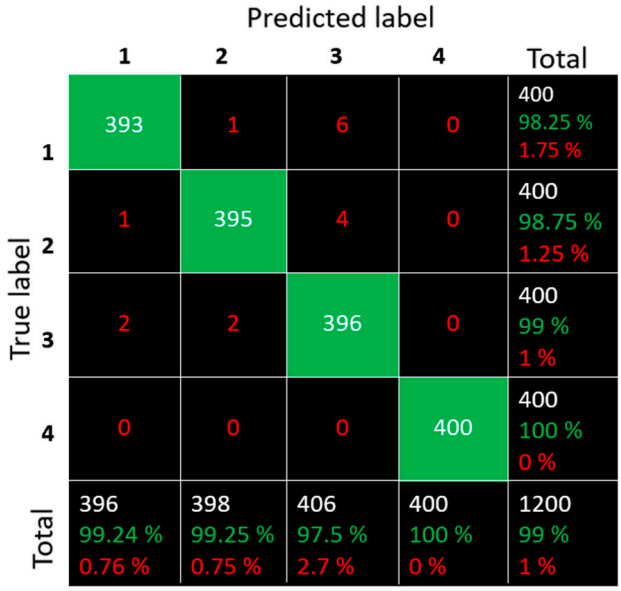


Figure 12. Confusion matric of fusion.

5.3. Comparison with Previous Works

The fusion method provides higher accuracy compared to EEG signals or a webcam. The accuracy of the fusion method is always higher than that of the webcam and EEG signals, as shown in Figure 13.

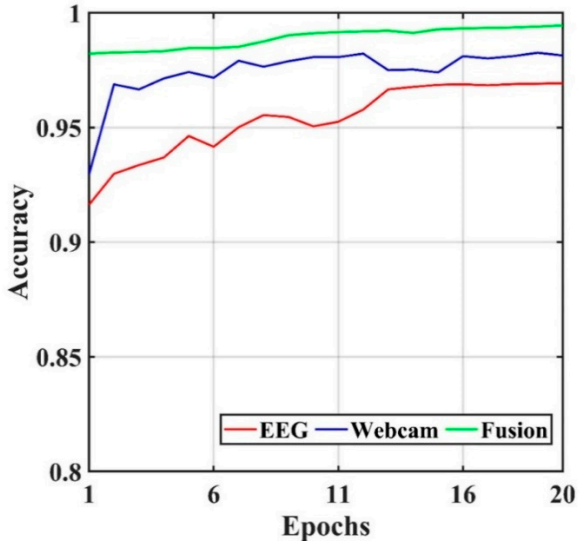


Figure 13. Validation accuracy.

For example, in epoch 11, the accuracy of fusion is up to 0.99 whereas the accuracy of EEG signals and webcam are limited to 0.95; 0.98, respectively.

The combination of facial expressions and neural signals has the potential to improve wheelchair control systems, offering intuitive interaction and precise control. This fusion of parameters enhances responsiveness and personalization, empowering individuals to navigate their surroundings with increased ease and independence.

A comparison of the proposed methodology with deep learning and other classification algorithms from the literature [35–39] is presented in Table 4. The comparison focuses on studies that were also evaluated in terms of precision, recall, accuracy, and F-score.

Table 4. Performance comparison of our proposed approach with state-of-the-art approaches.

Metric	Precision	Recall	Accuracy	F-Score
Roots, K. et al. [35]	74.45	74.47	74.5	74.46
Amin, S. U et al. [36]	84.1	83.8	84.1	84
Shankar, A. et al. [37]	93	-	93.05	-
Lin, L. C. et al. [38]	74.23	-	74.66	-
Goel, S. et al. [39]	98.15	-	98.21	-
Proposed	**99.05**	**99.03**	**99**	**98.97**

We have found that our proposed algorithm based on a fusion of CNN-LSTM outperforms the other works in both accuracy and precision. These results show the advantage of our networks in predicting the four emotions based on EEG signals and video recognition.

Our network architecture essentially represents a comprehensive method of emotion identification, seamlessly integrating the perceptive capabilities of image analysis with the nuanced insights derived from EEG signal processing. Late fusion creates a cooperative synergy that enables our model to overcome single-modality methods' constraints.

6. Conclusions

This work lays the groundwork for the creation of cutting-edge assistive technology and opens up new directions for investigating the potential of human–computer interaction. We can gain a better knowledge of human cognition and behavior by combining the power of facial expressions with EEG signals, which will ultimately improve the quality of life for people who have mobility disabilities.

In this paper, we proposed a smart wheelchair control system designed to aid individuals with physical impairments in their mobility. Our approach integrates fusion between decision modes from an EEG signals sensor and a webcam images sensor, achieving an outstanding accuracy level of up to 99% in emotion recognition. The incorporation of CNN and LSTM architectures through a fusion algorithm exhibits superior performance, surpassing single-modality methodologies. This comprehensive approach not only advances the field of human–computer interaction but also contributes to assistive technologies, thereby improving the quality of life of individuals with mobility impairments.

In future endeavors, we aim to validate our study's findings by implementing our application on an embedded system and deploying it in real-world scenarios, particularly in the control of wheelchairs. Furthermore, we plan to introduce an additional control modality, such as voice control, expanding the versatility and practicality of our proposed system.

Author Contributions: Conceptualization, L.Z., N.B.A. and M.J.; methodology, N.B.A.; software, L.Z.; validation, L.Z., J.K. and M.J.; formal analysis, L.Z.; resources, L.C.A.; data curation, L.Z.; writing—original draft preparation, L.Z., N.B.A. and J.K.; writing—review and editing, M.J. and L.D.; visualization, L.Z.; supervision, N.B.A.; funding acquisition, L.D. and L.C.A. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by Computer Embedded System Laboratory and Laboratory Innovative Technologies Laboratory.

Institutional Review Board Statement: Not applicable.

Data Availability Statement: Data are contained within the article.

Acknowledgments: During the preparation of this work, the authors used the tool Grammarly for grammar checking and English language enhancement.

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

Abbreviations

EPW	Electric Powered Wheelchairs
BCI	Brain–Computer Interface
EEG	Electroencephalogram
LSTM	Long Short-Term Memory
CNN	Convolutional Neural Network
WHO	World Health Organization
SSVEP	Steady-State Visually Evoked Potentials
LORETA	Low-Resolution Electromagnetic Tomography of the Brain
BCI-OSC	Brain–Computer Interface Open Sound Control
EAR	Eye Aspect Ratio
HOG	Histogram of Oriented Gradients
TP	True Positives
TN	True Negatives
FP	False Positives
FN	False Negatives

References

1. Global Report on Health Equity for Persons with Disabilities. 2022. Available online: https://www.who.int/teams/health-product-policy-and-standards/assistive-and-medical-technology/assistive-technology/wheelchair-services (accessed on 10 January 2024).
2. Bayer, E.M.; Smith, R.S.; Mandel, T.; Nakayama, N.; Sauer, M.; Prusinkiewicz, P.; Cris, K. Integration of transport-based models for phyllotaxis and midvein formation. *Genes Dev.* **2009**, *23*, 373–384. [CrossRef] [PubMed]
3. Rebsamen, B.; Guan, C.; Zhang, H.; Wang, C.; Teo, C.; Ang, M.H.; Burdet, E. A brain controlled wheelchair to navigate in familiar environments. *IEEE Trans. Neural Syst. Rehabil. Eng.* **2010**, *18*, 590–598. [CrossRef] [PubMed]
4. Puanhvuan, D.; Yodchanan, W. Semi-automatic P300-based brain-controlled wheelchair. In Proceedings of the ICME International Conference on Complex Medical Engineering IEEE, Kobe, Japan, 1–4 July 2012.
5. Al-Qaysi, Z.T.; Zaidan, B.B.; Zaidan, A.A.; Suzani, M.S. A review of disability EEG based wheelchair control system: Coherent taxonomy, open challenges and recommendations. *Comput. Methods Programs Biomed.* **2018**, *164*, 221–237. [CrossRef]
6. Swee, S.K.; Kiang, K.D.T.; You, L.Z. EEG controlled wheelchair. In Proceedings of the International Conference on Mechanical, Manufacturing, Modeling and Mechatronics IC4M, Kuala Lumpur, Malaysia, 27–29 February 2016.
7. Pires, C.P.; Sarkar, S.; Carvalho, L. Innovation in services–how different from manufacturing. *Serv. Ind. J.* **2008**, *28*, 1339–1356. [CrossRef]
8. Ortner, R.; Guger, C.; Prueckl, R.; Grünbacher, E.; Edlinger, G. SSVEP based brain-computer interface for robot control. In Proceedings of the International Con-ference Computers Helping People with Special Needs ICCHP, Vienna, Austria, 14–16 July 2010.
9. Dar, M.N.; Akram, M.U.; Khawaja, S.G.; Pujari, A.N. CNN and LSTM-based emotion charting using physiological signals. *Sensors* **2020**, *20*, 4551. [CrossRef] [PubMed]
10. Yan, Z.; Hu, L.; Chen, H.; Lu, F. Computer Vision Syndrome: A widely spreading but largely unknown epidemic among computer users. *Comput. Hum. Behav.* **2008**, *24*, 2026–2042. [CrossRef]
11. Paszkiel, S.; Paszkiel, S. Using the LORETA Method for Localization of the EEG Signal Sources in BCI Technology. In *Analysis and Classification of EEG Signals for Brain–Computer Interfaces*; Springer: Berlin/Heidelberg, Germany, 2020; Volume 852, pp. 27–32.
12. Santos, E.M.; San-Martin, R.; Fraga, F.J. Comparison of LORETA and CSP for Brain-Computer Interface Applications. In Proceedings of the International Multi-Conference on Systems, Signals & Devices (SSD), Sousse, Tunisia, 22–25 March 2021.
13. Manoilov, P. Eye-blinking artefacts analysis. In Proceedings of the International Conference on Computer Systems and Technologies, University of Ruse, Ruse, Bulgaria, 14–15 June 2007.
14. Zhao, W.; Li, C.; Chen, X.; Gui, W.; Tian, Y.; Lei, X. EEG spectral analysis in insomnia disorder: A systematic review and meta-analysis. *Sleep Med. Rev.* **2021**, *59*, 101457. [CrossRef] [PubMed]
15. Yong, X.; Fatourechi, M.; Ward, R.K.; Birch, G.E. Automatic artefact removal in a self-paced hybrid brain-computer interface system. *J. Neuroeng. Rehabil.* **2012**, *9*, 50. [CrossRef] [PubMed]

16. Divjak, M.; Bischof, H. *Eye Blink Based Fatigue Detection for Prevention of Computer Vision Syndrome*; Machine Vision Applications MVA: Yokohama, Japan, 2009.
17. Frølich, L.; Winkler, I.; Müller, K.R.; Samek, W. Investigating effects of different artefact types on motor imagery BCI. In Proceedings of the Inter-national Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), Milan, Italy, 25–29 August 2015.
18. Çınar, S.; Acır, N. A novel system for automatic removal of ocular artefacts in EEG by using outlier detection methods and independent component analysis. *Expert Syst. Appl.* **2017**, *68*, 36–44. [CrossRef]
19. Benda, M.; Volosyak, I. Peak detection with online electroencephalography (EEG) artifact removal for brain–computer interface (BCI) purposes. *Brain Sci.* **2019**, *9*, 347. [CrossRef] [PubMed]
20. Li, J.; Struzik, Z.; Zhang, L.; Cichocki, A. Feature learning from incomplete EEG with denoising autoencoder. *Neurocomputing* **2015**, *165*, 23–31. [CrossRef]
21. Brinker, T.J.; Hekler, A.; Utikal, J.S.; Grabe, N.; Schadendorf, D.; Klode, J.; Von Kalle, C. Skin cancer classification using convolutional neural networks: Systematic review. *J. Med. Internet Res.* **2018**, *20*, e11936. [CrossRef] [PubMed]
22. Yadav, S.S.; Jadhav, S.M. Deep convolutional neural network based medical image classification for disease diagnosis. *J. Big Data* **2019**, *6*, 113. [CrossRef]
23. Yan, R.; Ren, F.; Wang, Z.; Wang, L.; Zhang, T.; Liu, Y.; Zhang, F. Breast cancer histopathological image classification using a hybrid deep neural network. *Methods* **2020**, *173*, 52–60. [CrossRef] [PubMed]
24. Ngo, B.-V.; Nguyen, T.-H.; Ngo, V.-T.; Tran, D.-K.; Nguyen, T.-D. Wheelchair navigation system using EEG signal and 2D map for disabled and elderly people. In Proceedings of the International Conference on Green Technology and Sustainable Development (GTSD), Ho Chi Min, Vietnam, 27–28 November 2020.
25. Majewski, P.; Pawuś, D.; Szurpicki, K.; Hunek, W.P. Toward Optimal Control of a Multivariable Magnetic Levitation System. *Appl. Sci.* **2022**, *12*, 674. [CrossRef]
26. Ghorbel, A.; Ben Amor, N.; Abid, M. GPGPU-based parallel computing of viola and jones eyes detection algorithm to drive an intelligent wheelchair. *J. Signal Process. Syst.* **2022**, *94*, 1365–1379. [CrossRef]
27. Sokół, S.; Pawuś, D.; Majewski, P.; Krok, M. The Study of the Effectiveness of Advanced Algorithms for Learning Neural Networks Based on FPGA in the Musical Notation Classification Task. *Appl. Sci.* **2022**, *12*, 9829. [CrossRef]
28. Fogelton, A.; Benesova, W. Eye blink detection based on motion vectors analysis. *Comput. Vis. Image Underst.* **2016**, *148*, 23–33. [CrossRef]
29. Cojocaru, D.; Manta, L.F.; Pană, C.F.; Dragomir, A.; Mariniuc, A.M.; Vladu, I.C. The design of an intelligent robotic wheelchair supporting people with special needs, includ-ing for their visual system. *Healthcare* **2021**, *10*, 13. [CrossRef] [PubMed]
30. Zaway, L.; Ben Amor, N.; Ktari, J.; Jallouli, M.; Chrifi-Alaoui, L.; Delahoche, L. Fusion with EEG signals and Images for closed or open eyes detection using deep learning. In Proceedings of the International Conference on Design, Test and Technology of Integrated Systems (DTTIS), Tunis, Tunisia, 1–4 November 2023.
31. Song, F.; Tan, X.; Liu, X.; Chen, S. Eyes closeness detection from still images with multi-scale histograms of principal oriented gradients. *Pattern Recognit.* **2014**, *47*, 2825–2838. [CrossRef]
32. Maharana, K.; Mondal, S.; Nemade, B. A review: Data pre-processing and data augmentation techniques. *Glob. Transit. Proc.* **2022**, *3*, 91–99. [CrossRef]
33. Kim, K.W.; Hong, H.G.; Nam, G.P.; Park, K.R. A study of deep CNN-based classification of open and closed eyes using a visible light camera sensor. *Sensors* **2017**, *17*, 1534. [CrossRef] [PubMed]
34. Zaway, L.; Chrifi-Alaoui, L.; Amor, N.B.; Jallouli, M.; Delahoche, L. Classification of EEG Signals using Deep Learning. In Proceedings of the International Multi-Conference on Systems, Signals & Devices (SSD), Setif, Algeria, 6–10 May 2022.
35. Roots, K.; Muhammad, Y.; Muhammad, N. Fusion convolutional neural network for cross-subject EEG motor imagery classification. *Computers* **2020**, *9*, 72. [CrossRef]
36. Amin, S.U.; Alsulaiman, M.; Muhammad, G.; Bencherif, M.A.; Hossain, M.S. Multilevel weighted feature fusion using convolutional neural networks for EEG motor im-agery classification. *IEEE Access* **2019**, *7*, 18940–18950. [CrossRef]
37. Shankar, A.; Khaing, H.K.; Dandapat, S.; Barma, S. Analysis of epileptic seizures based on EEG using recurrence plot images and deep learning. *Biomed. Signal Process. Control* **2021**, *69*, 102854. [CrossRef]
38. Lin, L.C.; Chang, M.Y.; Chiu, Y.H.; Chiang, C.T.; Wu, R.C.; Yang, R.C.; Ouyang, C.S. Prediction of seizure recurrence using electroencephalogram analysis with multiscale deep neural networks before withdrawal of antiepileptic drugs. *Pediatr. Neonatol.* **2022**, *63*, 283–290. [CrossRef] [PubMed]
39. Goel, S.; Agrawal, R.; Bharti, R.K. Automated detection of epileptic EEG signals using recurrence plots-based feature extraction with transfer learning. *Soft Comput.* **2023**, *28*, 2367–2383. [CrossRef]