An Upper-Limb Rehabilitation Exoskeleton System Controlled by MI Recognition Model With Deep Emphasized Informative Features in a VR Scene

Zhichuan Tang, Hang Wang, Zhixuan Cui, Xiaoneng Jin, Lekai Zhang, Yuxin Peng, *Member, IEEE*, and Baixi Xing

Abstract— The prevalence of stroke continues to increase with the global aging. Based on the motor imagery (MI) brain–computer interface (BCI) paradigm and virtual reality (VR) technology, we designed and developed an upper-limb rehabilitation exoskeleton system (VR-ULE) in the VR scenes for stroke patients. The VR-ULE system makes use of the MI electroencephalogram (EEG) recognition model with a convolutional neural network and squeeze-and-excitation (SE) blocks to obtain the patient's motion intentions and control the exoskeleton to move during rehabilitation training movement. Due to the individual differences in EEG, the frequency bands with optimal MI EEG features for each patient are different. Therefore, the weight of different feature channels is learned by combining SE blocks to emphasize the useful information frequency band features. The MI cues in the VR-based virtual scenes can improve the interhemispheric balance and the neuroplasticity of patients. It also makes up for the disadvantages of the current MI-BCIs, such as single usage scenarios, poor individual adaptability, and many interfering factors. We designed the offline training experiment to evaluate the feasibility of the EEG recognition strategy, and designed the online control experiment to verify the effectiveness of the VR-ULE system. The results showed that the MI classification method with MI cues in the VR scenes improved the accuracy of MI classification (86.49% ± 3.02%); all subjects performed two types of rehabilitation training tasks under their own models trained in the offline training experiment, with the highest average completion rates of 86.82% ± 4.66% and 88.48% ± 5.84%. The VR-ULE system can efficiently help stroke patients with hemiplegia complete upper-limb rehabilitation training tasks, and provide the new methods and strategies for BCI-based rehabilitation devices.

Index Terms— Rehabilitation exoskeleton, brain–comput-er interface, virtual reality, convolutional neural networks, squeeze-and-excitation block, motor imagery.

I. INTRODUCTION

THE prevalence of stroke continues to rise with the global aging. Stroke patients with hemiplegia have neurological damage caused by the massive death of brain cells, resulting in varying degrees of upper-limb motion disorders [1]. Rehabilitation exoskeletons based on brain–computer interface (BCI) technology have become a more common rehabilitation treatment plan for stroke patients in different rehabilitation periods [2].

BCI technology realizes communication between the human brain and external electronic devices by decoding the features of the electroencephalogram (EEG) in the cerebral cortex. As a new means of expression and interaction for motor intention, BCI has been widely used in the rehabilitation training of stroke patients at different stages [3]. Barsotti et al. [4] designed a set of upper-limb exoskeletons based on MI-BCI to rehabilitate the grasping ability of poststroke patients. Soekadar et al. [5] designed a noninvasive brain/neural hand exoskeleton to assist stroke patients in daily motions such as eating and drinking. As a bridge for direct communication between the human brain and external devices, BCI has been widely used in stroke rehabilitation treatment [6].

As one of the main paradigms of BCI technology, MI has been widely used in rehabilitation therapy of cerebral motor function in stroke patients [7]. Through MI training, the motor nerve conduction pathways of stroke patients can be repaired

or reconstructed. The MI of different motions is mapped to the EEG changes in the corresponding regions of the cerebral cortex, and decoding different EEG features can distinguish different motions [8]. For example, in unilateral hand MI, mu rhythms (8-13 Hz) and beta rhythms (13-30 Hz) of the motor sensory area on the opposite side of the brain will decrease in power, while mu rhythms and beta rhythms in the ipsilateral motor sensory area will increase in power. This phenomenon is called event-related desynchronization (ERD) and event-related synchronization (ERS) [7]. BCI technology uses various computer algorithms to classify these different ERD/ERS patterns and convert them into control signals for external devices. Tang et al. [8] proposed a BMI system based on ERD/ERS and used for upper-limb exoskeleton control, achieving high classification accuracy. Liu et al. [9] proposed an ERD/ERS-based BCI control system and verified its effectiveness by operating a two-arm multi-finger robotic to complete tasks. Based on ERD/ERS, Li et al. [10] proposed a BCI hybrid control strategy that combines EEG and EMG signals to achieve flexible and stable control of the lower limb dynamic exoskeleton.

Convolutional Neural Networks (CNN) as a representative algorithm of deep learning have been widely used in computer vision, natural language processing, and other fields [11]. Conventional EEG data processing relies on the experience of researchers for complex data preprocessing and feature extraction. However, human-operated preprocessing and feature extraction will reduce the accuracy and reliability of classification results [11]. And the correlations between EEGs of different channels are easily ignored during the feature extraction process [12]. The CNN model can automatically extract features from the original input signal and obtain deeper and more distinguishable feature information through local receptive fields, weight sharing, and down-sampling, which reducing the subjectivity and incompleteness of feature selection caused by human factors [13], [14], [15]. Amin et al. [16] proposed an attention-based CNN model to learn the importance of different features of MI data and obtained good results when they applied it to the BCI IV 2a dataset. Roy [17] proposed a Multi-Scale (MS) CNN which can extract the distinguishable features of several non-overlapping canonical frequency bands of EEG signals from multiple scales for MI-BCI classification. Zhao et al. [18] proposed a multi-branch 3D-CNN classification strategy and the 3D representation is generated by transforming EEG signals into a sequence of 2D array which preserves spatial distribution of sampling electrodes. Li et al. [19] proposed an end-to-end EEG decoding framework, which employs raw multi-channel EEG as inputs, to boost decoding accuracy by the channel-projection mixed-scale convolutional neural network aided by amplitude-perturbation data augmentation. However, due to significant individual differences between subjects, such as the optimal time period and frequency band of ERD/ERS changes [20], [21], [22]. It is not good enough to use conventional recognition methods to perform shallow temporal or spectral feature learning on MI features. Therefore, due to the influence of individual differences among stroke patients, the refinement and weight assignment of deep

features is another research interest that could improve the accuracy of MI-EEG decoding deep learning models.

Squeeze-and-Excitation Networks (SENet) as a channel-based attention mechanism, treats each feature channel as a whole and use global information to automatically "learn" the importance of different feature channels in the training process, thereby suppressing the relatively unimportant features in the training classification process and boosting the most discriminative and information-rich features to improve the accuracy of the model [23]. Sun et al. [24] proposed a CNN with sparse spectrotemporal decomposition (SSD) for MI-EEG classification, which adopted SE to adaptively recalibrate the channel direction. Zhang et al. [25] proposed an orthogonal CNN fused with SE blocks to perform feature recalibration across different EEG channels. Inspired by SE, we merged SE blocks into the CNN model, enabling the model to automatically obtain the weights of each feature channel (EEG features of different time and frequency bands), adaptively weighted the feature maps generated by the original feature fusion layer, and improve the proportion of useful features in the current task. This approach can solve the problem of optimal features of EEG signals from different subjects located in different frequency bands, and train a MI recognition and classification model with high recognition accuracy suitable for specific users.

The current rehabilitation strategies based on MI-BCI mainly improve the MI recognition accuracy of subjects by improving the feature extraction algorithms and neglecting the impact of MI signal strength on recognition accuracy [26]. Therefore, in order to maximize the activation of the subjects' motor nerves and improve their signal strength, virtual rehabilitation technology combining MI-BCI technology and virtual reality (VR) technology is applied in the field of stroke rehabilitation [27]. VR technology has solved the problem of poor immersion and multiple external environmental interference factors (sound, light) in conventional rehabilitation training strategies (by observing cues on computer screens) [28]. VR technology can provide an immersive training environment, improve the interhemispheric activation balance (IHAB) and enhance the cortical connectivity between the primary sensorimotor cortex (SM1), the primary motor cortex (M1), and the supplementary motor area (SMA) on both sides of the subject during motion induction. The VR scene can provide real-time feedback in each training task, achieve more comprehensive MI training, improve rehabilitation efficiency, shorten the rehabilitation period, and enhance the patient's initiative and adaptability in rehabilitation [28]. Jang et al. [29] demonstrated a shift in cortical organization of the affected limb from the ipsilateral hemisphere to the contralateral hemisphere after the VR intervention. Mekbib et al. [30] revealed that unilateral and bilateral limb mirroring exercises in an immersive virtual environment may stimulate MNs in the damaged brain areas and may facilitate functional recovery of the affected upper extremities post-stroke. However, the current VR approaches use single-scenario rehabilitation, the inter-individual adaptability is poor [27]. At the same time, the conventional rehabilitation training strategies lack visual feedback based on motor intention, but at the neural
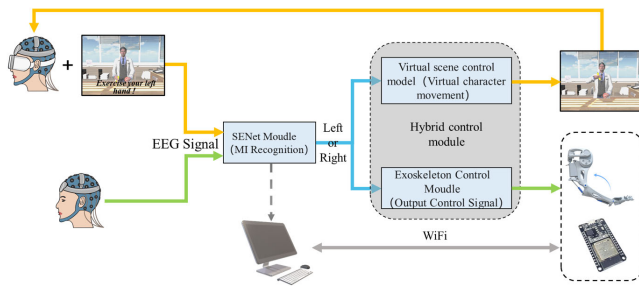
Fig. 1. The architecture diagram of the upper-limb rehabilitation exoskeleton system (VR-ULE).

Fig. 2. The lightweight wearable exoskeleton hardware subsystem.

mechanism level, visual feedback based on motor intention can activate the mirror neuron system, promote brain plasticity changes and functional reorganization, and contribute to the recovery of motor function [31]. Therefore, we used virtual character arm motions to cue patients for simulated movements in VR, and the patient's mirror neurons were activated. Then, the patient performed MI, and the computer decoded the EEG and converted it into control commands for the virtual character to achieve visual feedback of the motion intention. Patients continuously adjusted the MI process based on feedback results.

In this paper, we developed a virtual reality upper-limb rehabilitation exoskeleton system (VR-ULE). VR-ULE used an SE block-based CNN model and a VR scene to improve the MI recognition accuracy of stroke patients. VR-ULE includes a wearable exoskeleton hardware subsystem and an MI recognition software subsystem. The wearable exoskeleton subsystem is used to assist the patient's limb movement. The software subsystem for MI recognition is composed of a VR scenes cues control module, CNN+SE module, and an online hybrid control module. The VR scenes cues module is used to provide patients with visual cues and feedback with mirror operation intention. CNN+SE module is used to automatically analyze the importance of EEG features in different time periods and frequency bands. The SE blocks is used to emphasize important features and suppress nonimportant features through adaptive weighting. The online hybrid control module is used to receive the patient's MI signals and provide virtual feedback signals in the early and middle stages of the patient's motor neuron rehabilitation. In the later stages of rehabilitation, the online hybrid control module is used to and control the upper-limb exoskeleton robotic arm to assist patients in muscle group strength rehabilitation training. We designed an offline training experiment to obtain the MI EEG data of different subjects and trained the CNN+SE classification model. We also designed an online control experiment to evaluate and validate our proposed rehabilitation strategy and training system. The contributions of this study include:

(1) Based on the theory of neural plasticity, we independently designed an upper-limb rehabilitation exoskeleton system for stroke patients at different rehabilitation training stages to perform active movement and passive movement;

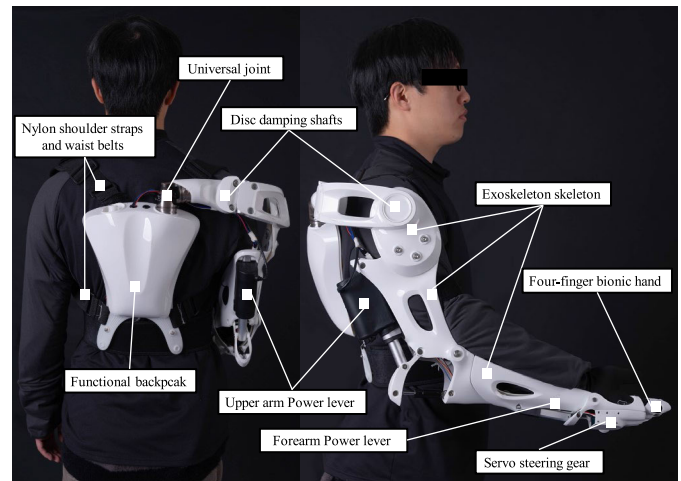(2) The SE module based on channel attention mechanism in the CNN model were used to obtain the frequency band differences of EEG signals between individuals and the CNN+SE model effectively improved the accuracy of MI recognition and classification;

(3) Based on the mirror neuron theory, we built three VR rehabilitation training scenes (lifting dumbbells, tasting fruits, and feeding pets) with virtual motion guidance to improve the immersion experience and avoid some environmental interference factors in conventional screen cues.

II. SE-BASED CLASSIFICATION STRATEGY OF MI

A. System Introduction

The VR-ULE consists of a wearable exoskeleton hardware subsystem and a MI recognition software subsystem. The system framework is shown in Fig 1.

The wearable exoskeleton subsystem consists of a self-designed and developed functional backpack and an upper-limb exoskeleton robotic arm. The functional backpack is used to store various hardware control modules, and the upper-limb exoskeleton robotic arm is used to assist the patient's limb movement. The MI recognition software subsystem consists of a VR scene cue module, a CNN+SE module, and an online hybrid control module. The VR scene cue module is used to provide patients with a strongly immersive virtual MI cue during the CNN+SE model training stage. The CNN+SE module it is used to amplify the strong-response frequency band of EEG during the patient's MI, accurately identify and classify the patient's motor intentions. The online hybrid control module is used to receive the MI signal from patients during rehabilitation, provide feedback signals in different periods of the patient's rehabilitation, and control the motion of virtual character or upper-limb exoskeleton robotic arm motion.

1) Wearable Exoskeleton Subsystem: We independently designed and produced a lightweight wearable exoskeleton hardware subsystem, as shown in Fig 2. The wearable exoskeleton hardware subsystem consists of a functional backpack, an upper-limb exoskeleton robotic arm skeleton, two power levers, a push lever drive board, a servo motor gear, a single-chip microcomputer, a power module pack, a lithium battery, a four-finger bionic hand, a disc damping shaft, and a
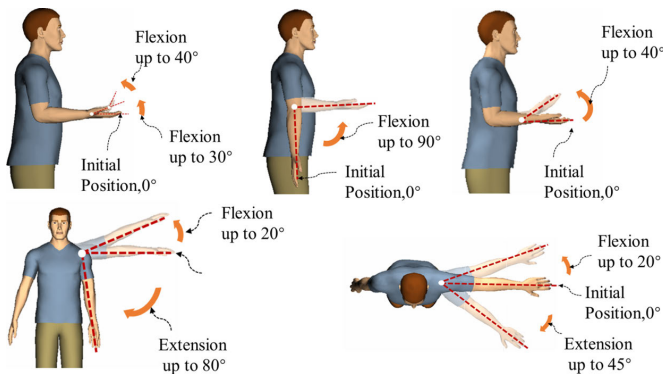
Fig. 3. Exoskeleton degrees of freedom display.

Fig. 4. VR training scene diagram (cue).

universal joint. Among them, the functional backpack made of 3D printing materials which is light, small, and comfortable. The inside of the backpack is used to store the push lever drive board, servo motor gear, single-chip microcomputer, power module pack, and lithium battery. Nylon shoulder straps and waist belts can be adjusted according to the patient's body shape. The upper-limb exoskeleton arm can simulate the motion of a healthy arm. Bionic arm skeleton can fit the patient's arm, and the power lever can provide motion thrust to assist the patient's limb movement, which realized a series of rehabilitation training motions with five degrees of freedom, including grasping, wrist flexion, elbow flexion, and shoulder abduction (Fig 3). The four joints of the exoskeleton arm are connected by detachable binding screws, which can be smoothly rotated by simply inserting them into the holes. The four-finger bionic finger module is located at the end of the upper-extremity exoskeleton, and the knuckles use the short-range rope-driven motion mode of the servo motor to achieve a more natural finger grasping effect. The upper-limb exoskeleton arm is mounted on the backpack through the universal joint behind the right shoulder. The backpack is equipped with a lithium battery (12 V, 2400 mA), a lever driver board (L298N), and a microcontroller (ESP-WROOM-32, Shenzhen Yusong Chuangda Electronics Co. Ltd, China). The entire wearable exoskeleton hardware subsystem is sewn onto the inner nylon fabric, it is fixed to the patient's chest and waist by multiple elastic nylon straps, and all its drive levers are driven by lever driver board control. The single-chip microcomputer has a 4-MB storage space, which can communicate with the computer through WiFi to receive and convert the MI signal identified and classified by the CNN model with SE into feedback control signals to control the motion of the exoskeleton.

2) MI Recognition Software Subsystem: The MI recognition software subsystem consists of a VR scene cue module, a CNN module with SE, and an exoskeleton control module. It is used for the identification, classification, and control signal output of the patient's MI signal.

a) VR scene cue module: We designed and built three types of VR training scenes for MI: lifting dumbbells, tasting fruits, and feeding pets. These are used to provide patients with virtual MI cues (Fig 4). While ensuring the sense of immersion, it increases the interest in rehabilitation
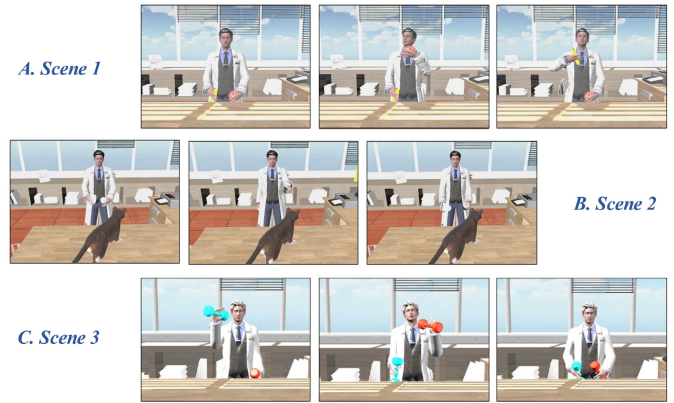
training, better stimulates the enthusiasm of patients for training, improves rehabilitation efficiency, and shortens the rehabilitation cycle. When a patient wears a VR head-mounted display (VIVE-P130, HTC, Inc.) for offline MI training, the VR scene only provides virtual motion cues, but during online MI training, the patient sees the virtual motion cues and then imagines the movement. After the MI signal is recognized and classified by the CNN model, the online hybrid control module outputs it back to the virtual scene to control the motion of the virtual characters. If the MI is wrong or fails, the virtual scene will display the recognition results to provide feedback to the patient to prompt the patient continuously correct or strengthen the MI.

b) CNN+SE module: The EEG is a signal with spatiotemporal characteristics, and its feature extraction process needs to consider temporal and spatial features [32]. Because there are significant individual differences in the frequency characteristics and spatial characteristics of EEG signals among different subjects during MI., by incorporating the improved SE blocks into the CNN model, it is beneficial to train a CNN model that meets the specified user, thereby improving the recognition accuracy of MI. To this end, we independently designed and built a CNN model with an embedded SE block for the correct identification and classification of the patient's MI signals, as shown in Fig 5. The entire CNN+SE model consisted of 10 layers: the first layer was the input layer; the second and third layers were convolutional layers, which constituted the feature extraction part; the fourth layer was the feature fusion layer; the fifth layer was the SE blocks layer, which constituted the frequency-band channel-weight learning part; the sixth layer was the feature weighting layer, which was used to weight the output feature map of the fourth layer; the seventh layer was the average pooling layer, which was used for down sampling; the eighth and ninth layers were fully connected layers; the 10th output layer constituted the classification part; and the 11th layer was the output layer, which we used to output the classification result.

c) Description of each network layer of CNN+SE: Input layer (L1) - 200 22×750 matrices of input samples per channel $L_{N,T}$ where N is 22, representing 22 EEG channels, and T is 750, representing the sampling time point in each channel.
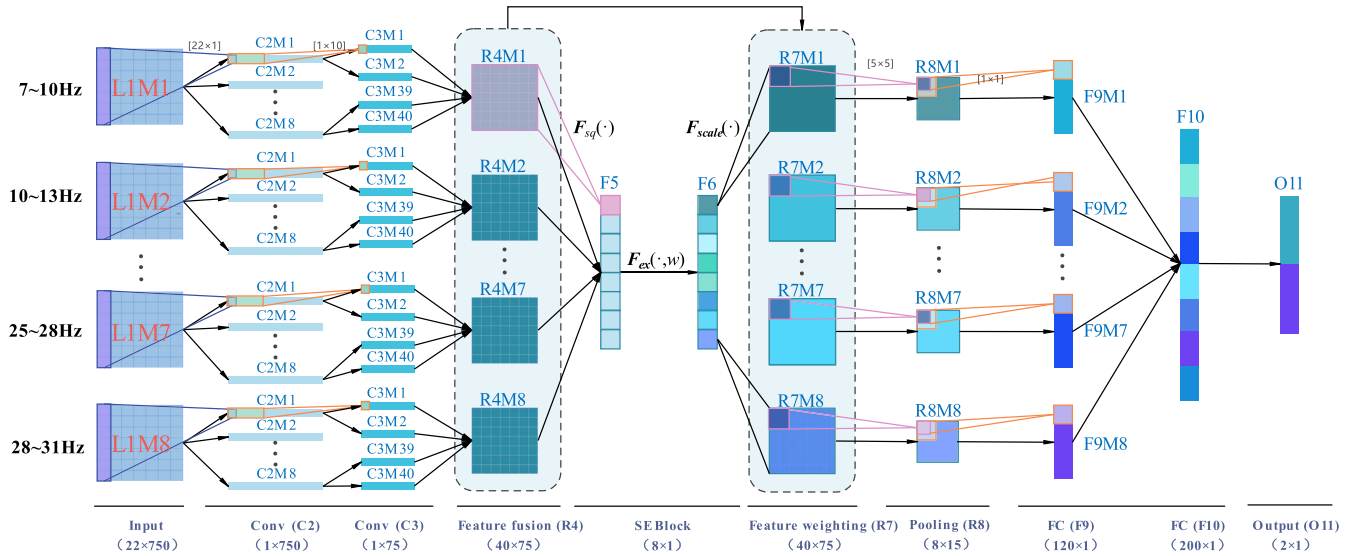
Fig. 5. Framework diagram of MI recognition model. The input EEG signal samples are divided into multiple different frequency bands in multiple channels, the weight of different feature channels is learned by combining SE blocks to emphasize the useful information frequency band features.
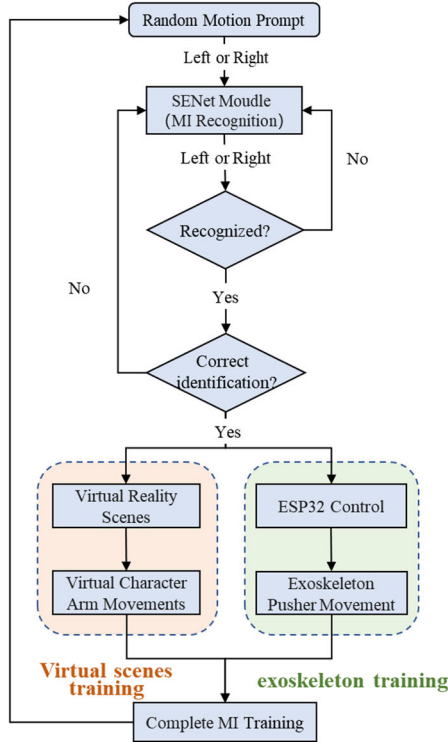

Fig. 6. Overall Control-flow diagram of the VR-ULE system.

Convolutional layer (C2) - This convolutional layer verifies the input EEG signals through eight convolution kernels of size $[22 \times 1]$ $L_{N,T}$ $L_{N,T}$ and performs spatial convolution. Eight frequency band channels are extracted to output eight feature maps of size $[1 \times 750]$. This process can be expressed as

$$C_m^2(j) = f(\sum_{i=1}^{i \leq 22} L_{i,j} \times k_m^2 + b_m^2(j)), \qquad (1)$$

where $C_m^2(j)$ is the output feature map of the C2 layer, the superscript 2 represents the number of layers, the subscript m represents the mth feature map, j represents the jth neuron

in the feature map, k_m^2 is the convolution kernel of $[22 \times 1]$, and $b_m^2(j)$ is the bias.

Convolutional layer (C3) - This convolutional layer performs temporal convolution on the input EEG through five convolution kernels of size $[1 \times 10]$. Eight frequency band channels are extracted to output 40 $[1 \times 75]$ feature maps. The process can be expressed as

$$C_m^3(j) = f\left(\sum_{i=1}^{i \leq 10} C_m^2((j-1) \times 10 + i) \times k_m^2 + b_m^2(j)\right), \qquad (2)$$

where k_m^2 is the convolution kernel of $[1 \times 10]$ and $b_m^2(j)$ is the bias.

Feature fusion layer (R4) - This layer splices 40 feature maps of size $[1 \times 75]$ output by the C3 layer of each frequency band channel to form a $[40 \times 75]$ feature map, with a total of eight feature maps of size $[40 \times 75]R4M_c$ output from eight frequency bands.

SE Blocks layer (SE) - this layer inputs feature maps $R4M_c$ with the size of $[40 \times 75 \times 8]$. In the SE layer, the squeeze operation is first performed to compress the input feature map tensor in space, that is, global average pooling is performed on the input feature maps in turn, and the results are fully connected. The output is a $[1 \times 1 \times 8]$ feature map F_5 to achieve the purpose of compressing and integrating the original 8 feature maps and shielding spatial distribution information. At the same time, by extracting the overall information of the eight feature channels, the underlying network can also obtain the global receptive field. This process can be expressed as

$$F_5 = F_{sq}(R4M_c) = \frac{1}{H \times W} \sum_{i=1}^{H} \sum_{j=1}^{W} R4M_c(i, j). \qquad (3)$$

Then, the excitation operation is performed to learn the nonlinear interactions between the eight feature channels and limit the complexity of the model by using two fully connected

layers with activation functions and no bias. These two fully-connected layers are dimensioned down and up, respectively, to form the structure of a bottleneck. The process can be expressed as

$$(F_6, R4M_c) = \sigma\left(g\left(F_6, R4M_c\right)\right) = \sigma\left(R4M_2\delta\left(F_6 \cdot R4M_1\right)\right).$$
(4)

i. e, first F_5 is multiplied by $R4M_1$ in a fully connected layer operation, then multiplied by a rectified linear unit (ReLU) layer to keep the output dimension unchanged. Then by multiplying the result by $R4M_2$ in a fully connected layer operation, and then through a ReLU, and so on to $R4M_8$, a $[1 \times 1 \times 8]$ feature map is output through the sigmoid function F_6.

Feature weighting layer (R7) - This layer performs channel-wise multiplication using the weights obtained by the excitation operation to perform channel-by-channel adaptive weighting on the original eight feature maps. That is, it multiplies the eight feature maps in the initial input SE by the eight weights in F_6. Finally, eight feature maps of size $[40 \times 75]$ are obtained to achieve feature weighting. The process can be expressed as

$$\widetilde{R7M}_c = F_{scale}\left(R4M_c, F_6\right) = R4M_c \cdot F_6.$$
(5)

Pooling layer (R8) - This layer performs average pooling of the output of the R7 layer in 5×5 regions with a stride of 5, and the output is eight feature maps of size $[8 \times 15]$. Fully connected layer (F9) - This layer is used as a fully connected layer. The eight feature maps output by the R8 layer are fully connected to obtain eight feature maps with a size of $[120 \times 1]$. This process can be expressed as

$$F_m^9(j) = f\left(\sum \widetilde{R7M}_c((j-1) \times 10 + 1) \times k_m^9 + b_m^9(j)\right),$$
(6)

where k_m^9 is the convolution kernel of $[1 \times 1]$; $b_m^9(j)$ is the bias;

Fully connected layer (F10) - This layer fully connects the eight feature maps output from the F9 layer to form a classification part of size $[960 \times 1]$, containing 200 neurons:

$$F_m^{10}(j) = f\left(\sum_{i=1}^{i \leq 8} \sum_{p=1}^{p \leq 120} F_i^9(p)\,\omega_i^{10}(p) + b^{10}(j)\right),$$
(7)

where $\omega_i^{10}(p)$ is the connection weight from the neurons in the F9 layer to the neurons in the F10 layer, and $b^{10}(j)$ is the bias.

Output layer (O11) - This layer is the output layer, containing two neurons, representing a binary classification problem. The process can be expressed as

$$O_m^{11}(j) = f\left(\sum_{i=1}^{i \leq 200} F_m^{10}(i)\,\omega^5(i) + b^5(j)\right),$$
(8)

where $\omega^5(i)$ is the connection weight of the neurons in the F10 layer to the neurons in the O11 layer, and $b^5(j)$ is the bias.

Online hybrid control model: The online hybrid control module converts the classification signal identified by the MI recognition software subsystem into a VR-scene character motion control signal or an exoskeleton motion control signal. First, the VR scene control module in the MI recognition software subsystem will randomly generate left- and right-hand MI motion cues, and the stroke patients will then try MI within a certain time. The trained CNN+SE model will acquire the subject's MI EEG data and perform identification and classification. The classification results are identified by the online hybrid control module according to the training task and converted into a continuous control signal output. The Control-flow diagram of the VR-ULE system is shown in Fig 6.

III. EXPERIMENT

For the MI classification strategy based on the combination of VR and SE blocks in our proposed VR-ULE, we designed offline training experiments and online control experiments to test the effectiveness of the strategy. In the offline training experiment, we trained two types of CNN+SE models that were cued by the conventional experimental scene and the VR scene for each subject in order to do comparative verification. In the online control experiment, we first selected the highest-accuracy classification model trained in the offline experiment for different subjects. Then, analogous to the rehabilitation stage of brain motor neurons in the pre-rehabilitation stage of stroke patients, the subjects will perform MI based on the VR scene to control the virtual characters to complete the corresponding virtual tasks. At the same time, analogous to the upper-limb muscle group strength training stage of stroke patients in the later stage of rehabilitation, the subjects independently perform MI according to the task requirements to achieve different control of the exoskeleton system and complete corresponding tasks. Finally, the completion results of the two types of analogy experiments are evaluated. We also chose three methods, conventional CNN [11], MRA+LDA [33], and CSP+SVM [34], to train the MI recognition model on the same training set, and then these models were tested using the same test set.

A. Subjects and Dataset Preparation

For the experiment, we recruited 20 healthy subjects (age: 22 ± 1.21 years), all right-handed (as assessed by the Edinburgh Handedness Questionnaire) [35]. At the same time, we also recruited one mild stroke patient and one moderate stroke patient to participate in the experiment. All the subjects participated in the EEG experiment for the first time and were not told any experimental hypotheses. Each subject signed an informed consent form before the experiment. The experimental procedure was reviewed and approved by the Human Ethics Review Committee of Zhejiang University of Technology.

EEG signals were acquired with the ActiveTwo64 channel EEG signal acquisition system (BioSemi, Netherlands). Twenty-three channels of EEG data (Fz, FC3, FC1, FCz, FC2, FC4, FC6, C5, C3, C1, Cz, C2, C4, C6, CP3, CP1, CPz, CP2,
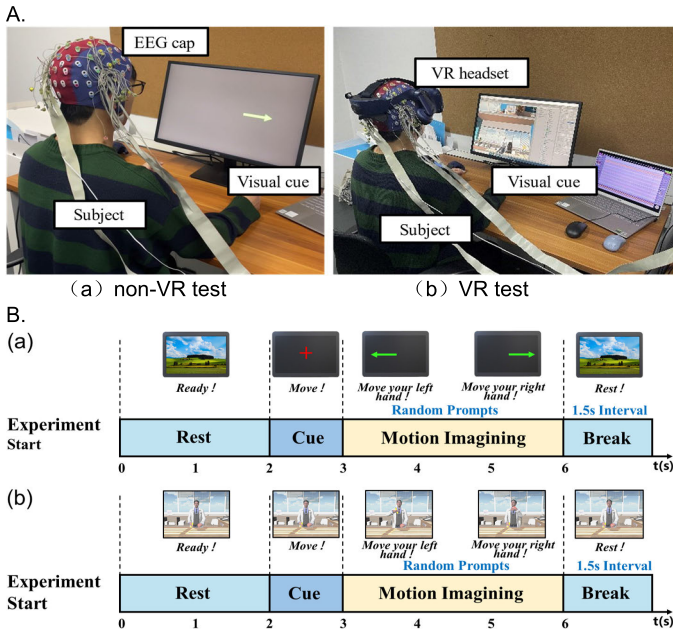
Fig. 7. The offline training experimental scenario is shown in A, where (a) is training without VR and (b) is training with VR; the timing diagram of one trial for two types of experiments are shown in B, where (a) is timing diagram of one trial without VR and (b) is timing diagram of one trial with VR.

CP4, P1, Pz, P2, POZ) were collected. The reference electrode was placed at the mastoid of the left ear; the ground electrode was replaced by two independent electrodes, common-mode sense (CMS) and driven right leg (DRL). Before placing the electrodes, a conductive gel was used to reduce the impedance between the electrodes and the scalp. The sampling frequency was set to 250 Hz, the cutoff frequency of the high-pass filter was 0.05 Hz, the cutoff frequency of the low-pass filter was 200 Hz, and the power frequency notch was 50 Hz. After placing all the electrodes, the subjects sat in front of the computer screen and put their hands on the table naturally. The subjects were asked to avoid blinking and unnecessary head or body motions as much as possible. The collected data were divided into a training set, validation set, and test set in a 3:1:1 ratio.

B. Offline Training Experiment

Each subject needed to complete 200 trials in each of the non-VR tests and VR tests throughout the offline training experiment (Fig 7), including 100 imaginary left-hand movements and 100 right-hand movements. The sequence diagram of the non-VR test is shown in Fig 7B-a. Each trial lasted for 8 s. The screen was blank for the first 2 s, and then a "+" character appeared in the center of the screen to remind the subjects that the trial was about to start. From 3 s to 6 s, the "+" character on the screen changed to a randomly generated leftward or rightward arrow and the subjects imagined their left-hand movement or the right-hand movement according to the arrow point. There was a random interval of 2 s to 5 s between each trial. There was a 3-min rest time between every 20-trial set to avoid fatigue.

The sequence diagram of the VR test is shown in Fig 7B-b. Each trial lasted for 8 s. The screen was blank for the first 2 s, and then the word "ready!" appeared in the center of the VR scene to remind the subjects that the trial was about to start. From 3 s to 6 s, the left- or right-hand movement randomly appeared in the VR scene. The subjects imagined their left hand or right-hand movement according to the body movements in the VR scene. There was a random interval of 2 s to 5 s between each trial, and there was a 3-min rest time between every 20-trial set to prevent fatigue. Each subject was tested with one set of data, including 200 non-VR trials and 200 VR trials, with seven subjects totaling seven sets of data and 14 parts.

The data of each subject's non-VR test and VR test were cropped, and after removing the data corresponding to the subjects' rest period, the EEG data in the frequency band of 7 Hz to 31 Hz were obtained by filtering, and the extracted EEG data were further processed for frequency band separation of 7–10, 10–13, 13–16, 16–19, 19–22, 22–25, 25–28, and 28–31 Hz. At the same time, the window length of the data segment was defined as 7.5 s. The CNN+SE preprocessing code intercepted the 3 s after the MI cue as the model input, so each input sample was composed of a 22-channel $\times$ 750-sampling-time-point (3 s time period $\times$ 250 Hz sampling rate) matrix, and each subject finally had 200 matrices of size $[22 \times 750]$ corresponding to the non-VR test and VR test. Finally, 200 matrix data of size $[22 \times 750]$ for each subject's non-VR test and VR test were randomly divided into five copies, three of which were used to train each subject's MI model, one copy is used as a validation set for training the model and one of which was used as a test set for evaluating the trained model. The EEG of the C3 and C4 channels in all trials of each subject was subjected to the superposition average calculation (ERD/ERS), which can be expressed as

$$\frac{ERD}{ERS}\% = \frac{EEG_A - EEG_R}{EEG_R} \times 100\%. \tag{9}$$

The sequence diagram and brain topography of ERD/ERS were observed, and the time period when the ERD/ERS pattern appeared and ended in each trial was recorded.

We compared this model with three other types models, CNN, MRA+LDA, and CSP+SVM, which we trained on the same training set and then used the same test set to test these models.

C. Online Control Experiment

The online control experiment had two parts: the VR scene online cue control test and the exoskeleton online control test. The MI model with high accuracy trained by each subject in the offline training experiment was used for the control signal input of the VR scene and the exoskeleton in the online test. The experimental scenario is shown in Fig 8.

In the online control test of the VR scene, the subjects did not wear the exoskeleton equipment, only the VR head-mounted display and the EEG cap, and sat in front of the screen. The subjects rested their hands on the table. The screen was used to display the head-mounted display in real-time.

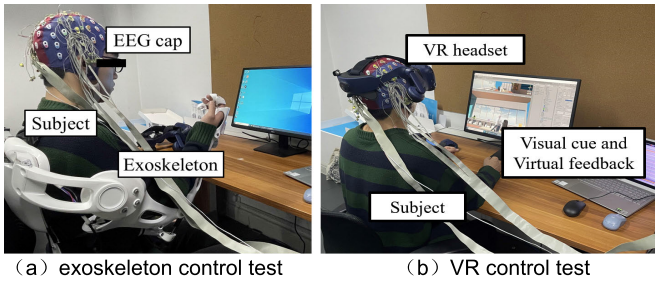（a）exoskeleton control test   （b）VR control test

Fig. 8. The Online control experimental scenario, where (a) is exoskeleton control test and (b) is VR control test.

In the VR scene, the MI instruction "Please move your left hand" or "Please move your right hand" appeared once every 10 s. At the same time, the virtual characters in the VR scene made the same actions to guide the subjects, and the subjects performed the corresponding MI according to the instructions. The PC saved the EEG data collected 2 s before and 5.5 s after the MI instructions. The data processing code preprocessed the saved data and input it into the MI model trained by the offline training experiment for classification. The classification result was converted into a control signal and input to the VR scene c control module to control the virtual character arm in the VR scene to move and complete the corresponding task. Each subject performed 20 MI tasks with the left hand and 20 with the right hand for each type of scene, so a total of 120 MI tasks were performed.

In the exoskeleton control test, the subjects wore the exoskeleton on the right hand and sat in front of the table with the EEG cap on. The MI EEG signals in the left and right hands were mapped as control signals for the input exoskeleton to perform the task action or not, respectively. The subjects made an MI trial every 8 s according to the task requirements, and there was a 8-s rest between two trials. All EEG data were saved in the PC during the test. Similarly, the data processing code preprocessed the saved data and input them into the MI model trained by the offline training experiment. The classification results were converted into control signals and input to the exoskeleton control module to control the motion of the exoskeleton. A total of 30 MI trials were designed as a full test.

IV. RESULTS

A. Public Dataset Results

To verify the effectiveness of the SE blocks in the SE-VR-based MI classification strategy, we applied it to the public dataset 1 in BCI Competition IV for model training and verification, and compared it with CNN, MRA+LDA, and CSP+SVM. For this dataset, seven subjects selected two types of movements from the left hand, right hand and foot, and to perform 100 MI trials. Record the EEG data of 64 EEG channels for each subject at a sampling rate of 1000 Hz. More details on the experimental paradigm corresponding to this dataset can be obtained from the following website: https://www.bbci.de/competition/iv/desc_1.html.

The classification results of the four MI classification models for different subjects in the public dataset are shown

TABLE I
CLASSIFICATION ACCURACY OF FOUR CLASSIFICATION MODELS FOR EACH SUBJECT IN THE PUBLIC DATASET

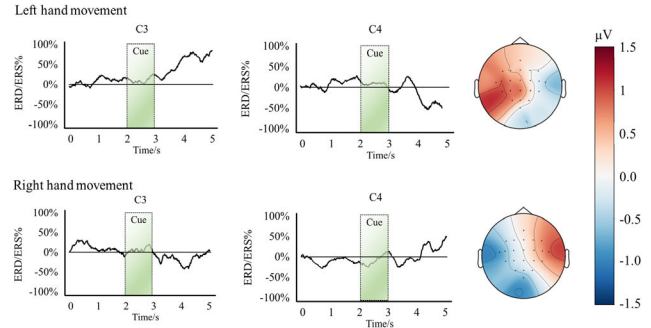| Subject	CNN+SE	CNN	MRA+LDA	CSP+SVM
1	87.69%	82.27%	85.17%	84.96%
2	85.47%	81.86%	84.29%	81.51%
3	88.16%	84.48%	85.55%	83.72%
4	88.07%	83.67%	84.81%	84.49%
5	86.89%	83.06%	80.93%	79.96%
6	87.31%	82.93%	84.60%	80.83%
7	89.15%	84.96%	86.48%	85.67%
Average±std	87.53%±1.07%	83.32%±1.04%	84.55%±1.62%	83.02%±2.07%

Fig. 9. ERD/ERS time course from 0 s to 5 s and EEG topography from 0 s to 6 s for left- and right-hand MI of all subjects in all trials.

in Table I. The average classification accuracy of the four MI classification models was CNN+SE 87.53% ± 1.07%; CNN 83.32% ± 1.04%; MRA+LDA 84.55% ± 1.62%; and CSP+SVM 83.02% ± 2.07%.

B. Results of Offline Training Experiment

ERD/ERS Analysis: After the EEG data of the C3 and C4 channels of each subject in all tests were separated, the superimposed average calculation of the ERD/ERS phenomenon was performed [36]. The ERD/ERS sequence diagram from 0 to 5 s and brain topography from 0 to 6 s for a single trial is shown in Fig 9. As seen from the Fig, the ERD/ERS pattern appeared in each trial for a time period ranging from 3.5 s to 4.5 s.

Analysis of the training results of the non-VR tests and VR tests: The two types EEG data of each subject were trained in four classification models, and models with good convergence were obtained, and the training loss curve of each subject was recorded. The best-converging model came from subject 06. The training loss curve of top three models is shown in Fig 10.

Analysis of the classification results of the non-VR tests and VR tests: After using the four classification models for training of each subject's non-VR tests and VR tests EEG data, the model with the best convergence was selected, and the classification test was performed. The resulting classification accuracy is shown in Table II. From the data in the table, it can be seen that compared with a single screen cue, the VR scene to provide MI cues obtained higher classification accuracy with the same classification model, and CNN+SE achieved higher classification accuracy than CNN, MRA+LDA, and CSP+SVM (non-VR:78.13% ± 2.60 and

TABLE II
THE TEST CLASSIFICATION ACCURACY OF EACH SUBJECT'S NON-VR TEST AND VR TEST DATA IN THE FOUR CLASSIFICATION MODELS

| Subject	CNN+SE		CNN		MRA+LDA		CSP+SVM	
	non-VR test	VR test	non-VR test	VR test	non-VR test	VR test	non-VR test	VR test
1	80.64%	87.23%	77.17%	80.04%	73.27%	79.64%	74.36%	72.11%
2	74.56%	78.62%	69.78%	73.55%	70.03%	78.57%	69.62%	73.71%
3	75.13%	84.29%	71.93%	77.27%	69.97%	80.36%	70.94%	80.33%
4	80.90%	91.64%	75.96%	77.61%	71.95%	82.50%	75.72%	88.70%
5	77.35%	88.56%	68.33%	75.46%	74.61%	81.07%	67.21%	80.03%
6	80.11%	87.13%	74.46%	79.36%	73.50%	82.14%	69.92%	75.35%
7	71.69%	84.92%	65.68%	80.11%	66.82%	76.43%	66.80%	68.92%
8	77.42%	87.62%	67.26%	75.81%	72.10%	80.93%	71.03%	73.90%
9	79.63%	90.05%	81.57%	87.63%	69.04%	79.02%	74.44%	78.81%
10	80.22%	89.86%	76.90%	80.27%	70.31%	78.96%	69.76%	77.91%
11	77.12%	85.76%	70.28%	73.39%	73.32%	84.92%	64.70%	72.05%
12	73.90%	91.32%	67.70%	72.79%	78.61%	92.76%	70.74%	73.66%
13	78.51%	84.67%	71.11%	79.63%	64.85%	70.75%	64.39%	70.90%
14	78.42%	84.31%	72.82%	74.60%	70.04%	79.63%	69.08%	80.03%
15	80.36%	89.37%	67.85%	79.63%	69.38%	78.40%	69.20%	76.27%
16	81.37%	87.43%	75.73%	74.39%	71.27%	80.37%	66.80%	75.49%
17	80.49%	84.29%	72.37%	75.58%	69.42%	77.07%	64.31%	74.48%
18	79.81%	84.55%	68.06%	72.26%	70.03%	78.98%	64.04%	74.98%
19	78.52%	87.74%	69.33%	77.84%	67.63%	72.06%	71.63%	70.09%
20	76.46%	85.06%	65.17%	71.94%	65.39%	72.77%	69.04%	71.30%
Patient 1	78.34%	86.44%	65.41%	72.67%	66.27%	74.14%	65.03%	72.91%
Patient 2	73.66%	81.94%	60.05%	73.33%	60.74%	70.03%	69.31%	74.85%
Average±std	77.94%±2.65%	86.49%±3.02%	71.80%±4.11%	76.60%±3.70%	69.93%±3.72%	78.70%±4.89%	69.00%±3.29%	75.31%±4.30%
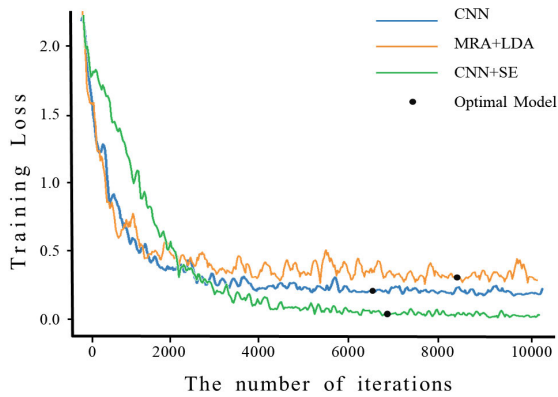
Fig. 10.   The model training loss curve obtained by the training of top three classification models on the VR tests data of Subject 06.

VR:86.72% ± 2.99%). In addition, for the CNN+SE model with VR, the model classification accuracy of two patients were lower than the average model classification accuracy of 20 healthy subjects (86.72% ± 2.99%), and the model recognition accuracy of Patient 1 (86.44%) was higher than that of Patient 2 (81.94%).

To better evaluate the classification accuracy of the four classification models, the three average indicators, precision, recall, and F score, were introduced in Table III. Based on the data in Table II and Table III, we conducted an ANOVA on the two types of MI cues (VR or non-VR), four types of classification models (CNN+SE, CNN, MRA+LDA and CSP+SVM), and two types of MI classes (Left or Right hand) to evaluate their interactions and their impact on classification accuracy. The results indicate that when the confidence level is set to 95%, there is no interaction between the MI class and the classification model, and there is no interaction between the MI cue and the classification model (all $p > 0.01$). The classification model has a significant effect on classification accuracy ($F = 63.984$, $p < 0.01$), and the MI cue has a significant effect on classification accuracy ($F = 152.328$, $p < 0.01$), but the MI class has no significant effect on the classification accuracy ($F = 0.02$, $p > 0.01$).

C. Results of Online Control Experiment

In the online experiment, the task completion rate of each subject is shown in Table IV. The task success rate was defined as the percentage of times a task was completed correctly. It can be seen from the online experimental results that the success rates of virtual scene tasks 1 (lifting dumbbells) was higher than virtual scene task 2 (tasting fruits) and 3 (feeding pets), and the success rate of the exoskeleton control task was higher than the virtual scene tasks (88.48% ± 5.84%). And there is no significant difference in the success rate between the four types of tasks of Patient 1 (87.50%, 85.00%, 80.00%, 86.67%) and the average success rate of the four types of tasks of 20 healthy subjects (87.00% ± 4.78%, 85.25% ± 4.74%, 80.13% ± 4.71%, 88.83% ± 5. 99%). The success rate of the four types of tasks of Patient 2 (87.00% ± 4.78%, 85.25% ± 4.74%, 80.13% ± 4.71%, 88.83% ± 5.99%) is lower than the average success rate of the four types of tasks of 20 healthy subjects.

V. DISCUSSION

Our designed offline training experiments (MI recognition and classification model training) and online control experiments (VR-scene character and exoskeleton arm movement control tasks) fully verify the effectiveness of our proposed VR-ULE system.

The offline training experiment results show that adding SE blocks into the CNN can promote the accuracy of MI EEG classification in public datasets. And we got a classification accuracy of 77.94% ± 2.65% in our own

TABLE III
THE TEST CLASSIFICATION PRECISION, RECALL, AND F SCORE OF THE VR TESTS DATA
FOR EACH SUBJECT IN THE FOUR CLASSIFICATION MODELS

Subject		CNN+SE		CNN		MRA+LDA		CSP+SVM	
		Left hand	Right hand	Left hand	Right hand	Left hand	Right hand	Left hand	Right hand
1	Precision	0.8740	0.8706	0.8006	0.8002	0.7943	0.7985	0.7216	0.7206
	Recall	0.8700	0.8746	0.8000	0.8008	0.8000	0.7928	0.7200	0.7222
	F-Score	0.8720	0.8726	0.8003	0.8005	0.7971	0.7957	0.7208	0.7214
2	Precision	0.7840	0.7884	0.7334	0.7376	0.7833	0.7882	0.7357	0.7385
	Recall	0.7900	0.7824	0.7400	0.7310	0.7900	0.7814	0.7400	0.7342
	F-Score	0.7870	0.7854	0.7367	0.7343	0.7866	0.7848	0.7379	0.7363
3	Precision	0.8449	0.8409	0.7742	0.7712	0.8058	0.8014	0.8053	0.8013
	Recall	0.8400	0.8458	0.7700	0.7754	0.8000	0.8072	0.8000	0.8066
	F-Score	0.8424	0.8434	0.7721	0.7733	0.8029	0.8043	0.8026	0.8039
4	Precision	0.9134	0.9194	0.7740	0.7783	0.8300	0.8200	0.8847	0.8893
	Recall	0.9200	0.9128	0.7800	0.7722	0.8218	0.8283	0.8900	0.8840
	F-Score	0.9167	0.9161	0.7770	0.7752	0.8259	0.8241	0.8873	0.8867
5	Precision	0.8822	0.8890	0.7570	0.7523	0.8111	0.8103	0.8005	0.8001
	Recall	0.8900	0.8812	0.7500	0.7592	0.8100	0.8114	0.8000	0.8006
	F-Score	0.8861	0.8851	0.7535	0.7557	0.8106	0.8108	0.8002	0.8004
6	Precision	0.8723	0.8703	0.7957	0.7915	0.8223	0.8205	0.7503	0.7566
	Recall	0.8700	0.8726	0.7900	0.7972	0.8200	0.8228	0.7449	0.7618
	F-Score	0.8711	0.8715	0.7929	0.7943	0.8211	0.8216	0.7476	0.7592
7	Precision	0.8486	0.8498	0.8018	0.8004	0.7666	0.7620	0.6889	0.6895
	Recall	0.8500	0.8484	0.8000	0.8022	0.7600	0.7686	0.6900	0.6884
	F-Score	0.8493	0.8491	0.8009	0.8013	0.7633	0.7653	0.6894	0.6890
8	Precision	0.8785	0.8741	0.7571	0.7591	0.8170	0.8017	0.7385	0.7395
	Recall	0.8673	0.8847	0.7600	0.7562	0.8020	0.8168	0.7400	0.7380
	F-Score	0.8729	0.8794	0.7586	0.7576	0.8094	0.8092	0.7393	0.7387
9	Precision	0.9100	0.8912	0.8735	0.8791	0.7903	0.7901	0.7870	0.7892
	Recall	0.8911	0.9101	0.8800	0.8726	0.7900	0.7904	0.7900	0.7862
	F-Score	0.9005	0.9005	0.8768	0.8758	0.7902	0.7902	0.7885	0.7877
10	Precision	0.8559	0.8593	0.8043	0.8011	0.7894	0.7898	0.7786	0.7796
	Recall	0.8600	0.8552	0.8000	0.8054	0.7900	0.7892	0.7800	0.7782
	F-Score	0.8579	0.8572	0.8022	0.8032	0.7897	0.7895	0.7793	0.7789
11	Precision	0.8559	0.8593	0.7357	0.7321	0.8486	0.8498	0.7207	0.7203
	Recall	0.8600	0.8552	0.7300	0.7378	0.8500	0.8484	0.7200	0.7210
	F-Score	0.8579	0.8573	0.7329	0.7349	0.8493	0.8491	0.7204	0.7206
12	Precision	0.9159	0.9106	0.7269	0.7289	0.9256	0.9297	0.7350	0.7382
	Recall	0.9100	0.9164	0.7300	0.7258	0.9300	0.9252	0.7400	0.7332
	F-Score	0.9129	0.9135	0.7285	0.7273	0.9278	0.9274	0.7375	0.7357
13	Precision	0.8380	0.8548	0.7941	0.7985	0.7065	0.7085	0.7086	0.7094
	Recall	0.8438	0.8494	0.8000	0.7926	0.7100	0.7050	0.7100	0.7080
	F-Score	0.8409	0.8521	0.7971	0.7955	0.7082	0.7068	0.7093	0.7087
14	Precision	0.8404	0.8456	0.7440	0.7480	0.7941	0.7985	0.8005	0.8001
	Recall	0.8351	0.8507	0.7500	0.7420	0.8000	0.7926	0.8000	0.8006
	F-Score	0.8377	0.8481	0.7470	0.7450	0.7971	0.7955	0.8002	0.8004
15	Precision	0.8977	0.8897	0.7941	0.7985	0.7863	0.7817	0.7641	0.7613
	Recall	0.8911	0.8964	0.8000	0.7926	0.7800	0.7880	0.7600	0.7654
	F-Score	0.8944	0.8930	0.7971	0.7955	0.7831	0.7849	0.7621	0.7633
16	Precision	0.8865	0.8626	0.7458	0.7420	0.7999	0.8076	0.7574	0.7524
	Recall	0.8614	0.8875	0.7400	0.7478	0.8100	0.7974	0.7500	0.7598
	F-Score	0.8738	0.8748	0.7429	0.7449	0.8049	0.8025	0.7537	0.7561
17	Precision	0.8535	0.8326	0.7612	0.7504	0.7611	0.7803	0.7397	0.7499
	Recall	0.8317	0.8543	0.7525	0.7592	0.7755	0.7661	0.7475	0.7422
	F-Score	0.8424	0.8434	0.7568	0.7548	0.7682	0.7731	0.7436	0.7460
18	Precision	0.8342	0.8573	0.7284	0.7171	0.7956	0.7842	0.7497	0.7499
	Recall	0.8586	0.8327	0.7100	0.7352	0.7800	0.7996	0.7500	0.7496
	F-Score	0.8462	0.8448	0.7191	0.7261	0.7877	0.7918	0.7499	0.7497
19	Precision	0.8754	0.8794	0.7832	0.7738	0.7282	0.7132	0.7013	0.7005
	Recall	0.8800	0.8748	0.7700	0.7868	0.7129	0.7285	0.7000	0.7018
	F-Score	0.8777	0.8771	0.7765	0.7802	0.7204	0.7208	0.7006	0.7012
20	Precision	0.8596	0.8418	0.7120	0.7269	0.7267	0.7288	0.7143	0.7117
	Recall	0.8416	0.8598	0.7273	0.7117	0.7300	0.7254	0.7100	0.7160
	F-Score	0.8505	0.8507	0.7196	0.7192	0.7283	0.7271	0.7121	0.7139
Patient 1	Precision	0.8676	0.8612	0.7298	0.7237	0.7496	0.7334	0.7215	0.7369
	Recall	0.8600	0.8688	0.7200	0.7334	0.7327	0.7503	0.7374	0.7210
	F-Score	0.8638	0.8650	0.7249	0.7285	0.7410	0.7418	0.7293	0.7289
Patient 2	Precision	0.8190	0.8198	0.7375	0.7291	0.7004	0.7002	0.7553	0.7418
	Recall	0.8200	0.8188	0.7327	0.7339	0.7000	0.7006	0.7426	0.7545
	F-Score	0.8195	0.8193	0.7351	0.7315	0.7002	0.7004	0.7489	0.7481
Average± std	Precision	0.8658±0.0316	0.8641±0.0298	0.7666±0.0366	0.7654±0.0376	0.7878±0.0486	0.7863±0.0497	0.7527±0.0431	0.7535±0.0432
	Recall	0.8663±0.0379	0.8671±0.0315	0.7651±0.0382	0.7669±0.0364	0.7861±0.0498	0.7880±0.0484	0.7528±0.0433	0.7533±0.0431
	F-Score	0.8659±0.0331	0.8656±0.0301	0.7658±0.0372	0.7661±0.0368	0.7870±0.0490	0.7871±0.0489	0.7527±0.0431	0.7534±0.0430

measured experimental dataset. The reason is that the CNN will learn the temporal and spatial features of the subject's MI, and the SE blocks performs feature weighting operations on the EEG data of the subjects in different frequency bands to learn strong features of the MI EEG frequency band of each subject. The advantage of this method is that while avoiding individual differences, the final classification result of the model will also be output based on the weights of all frequency bands, avoiding transiently missing a signal in a band of the subject during the online test affecting the

TABLE IV
THE SUCCESS RATE OF EACH SUBJECT'S FOUR
TYPES OF ONLINE EXPERIMENTAL TASKS

Subject	Virtual task 1	Virtual task 2	Virtual task 3	Exoskeleton control task
1	85.00%	87.50%	77.50%	90.00%
2	82.50%	77.50%	70.00%	76.67%
3	80.00%	82.50%	75.00%	80.00%
4	95.00%	92.50%	87.50%	96.67%
5	92.50%	85.00%	80.00%	93.33%
6	85.00%	92.50%	75.00%	86.67%
7	82.50%	82.50%	75.00%	83.33%
8	87.50%	87.50%	82.50%	86.67%
9	95.00%	90.00%	85.00%	93.33%
10	95.00%	90.00%	82.50%	96.67%
11	87.50%	87.50%	72.50%	86.67%
12	92.50%	92.50%	77.50%	96.67%
13	85.00%	85.00%	75.00%	83.33%
14	82.50%	77.50%	77.50%	80.00%
15	90.00%	85.00%	85.00%	93.33%
16	85.00%	85.00%	82.50%	93.33%
17	80.00%	77.50%	77.50%	90.00%
18	82.50%	80.00%	77.50%	93.33%
19	87.50%	85.00%	82.50%	93.33%
20	87.50%	82.50%	77.50%	83.33%
Patient 1	87.50%	85.00%	80.00%	86.67%
Patient 2	82.50%	85.00%	77.50%	83.33%
Average±std	86.82%±4.66%	85.23%±4.52%	80.00%±4.52%	88.48%±5.84%

classification results. Relevant studies have obtained results consistent with ours. For example, Sun et al. [24] used a deep learning framework called SSDSE-CNN integrating the SE blocks for MI-EEG classification, and the highest classification accuracy obtained was 79.3% ± 6.9%. Li et al. [37] proposed a novel temporal-spectral-based SE feature fusion network for MI-EEG decoding, and the highest classification accuracy was 84.49% on the public dataset.

Our offline experimental results also show that when the subjects tried MI, compared with the boring screen cues, the use of VR cues was more helpful for training a network model, with higher classification accuracy. All four classification models were verified, in which CNN+SE obtained a classification accuracy of 86.49% ± 3.02%. The reason is that the VR scene brings the subjects a more immersive experience, avoids the interference of many external factors, and makes it easier for the subjects to concentrate, and the arm movements of the characters in the virtual scene will guide the subjects to quickly generate corresponding responses, improving their IHAB while activating connections between more areas of the cerebral cortex, cueing patients to produce more pronounced MI EEG features. In related research on stroke rehabilitation, Sip et al. [38] applied the Virtual Mirror Hand 1.0 procedure to the treatment of hand functional recovery after stroke and compared it with the classic mirror therapy, finding that applying VR to the rehabilitation of stroke patients was feasible. Nath et al. [39] developed a VR task library for upper-limb rehabilitation of poststroke patients and concluded that VR therapy can improve the clinical symptoms of chronic stroke patients.

Our online control experiments showed that the average success rate of exoskeleton control task was 88.48% ± 5.84%, which was higher than that of virtual character arm movement tasks in VR scenes. The reason is that the MI command in the exoskeleton control task uses real task actions to improve patients' perception and motion mechanisms [40]. Patients can perform more concrete MI based on the

obtained perception experience, which can improve the model recognition accuracy.

In the offline training experiment, the model classification accuracy of two patients were lower than the average model classification accuracy of 20 healthy subjects (86.72% ± 2.99%), and the model recognition accuracy of Patient 1 (86.44%) was higher than that of Patient 2 (81.94%). The reason is that different degrees of stroke can cause varying degrees of damage to the patient's motor neurons, thereby affecting the patient's ERD/ERS patterns during MI and reduce the classification performance [41], [42].

One limitation of this study is that the subjects in our experiment lack diversity. In future studies, we will conduct more experiments on stroke patients of different ages, genders, and rehabilitation stages to verify the effectiveness of our proposed rehabilitation strategy.

VI. CONCLUSION

In this paper, based on the MI-BCI paradigm and VR technology, we designed and developed a VR-ULE that can be used for the rehabilitation of stroke patients with hemiplegia. The system obtains the patient's motion intention through the MI EEG identification strategy based on a CCN and SE blocks, and it controls the execution of VR-ULE rehabilitation training actions. The SE module makes up for the shortcoming that different subjects have differences in MI frequency domain characteristics. The MI indication based on the VR scene strengthens the MI EEG of the subjects and makes up for the shortcomings of the current MI-BCI rehabilitation strategies, such as a single rehabilitation scene, poor individual adaptability, and many external environmental interfering factors. Our results show that compared with the conventional classification strategy, the proposed MI EEG recognition method (CNN+SE) can improve the MI classification accuracy. The VR-ULE system can more efficiently help stroke patients complete upper-limb rehabilitation training tasks through a more reasonable MI identification strategy and an immersive experience of VR scenes, all of which improve the patients' autonomous rehabilitation.

REFERENCES

[1] F. Y. Wu and H. H. Asada, "Implicit and intuitive grasp posture control for wearable robotic fingers: A data-driven method using partial least squares," *IEEE Trans. Robot.*, vol. 32, no. 1, pp. 176–186, Feb. 2016, doi: 10.1109/TRO.2015.2506731.

[2] F. Grimm, A. Walter, M. Spüler, G. Naros, W. Rosenstiel, and A. Gharabaghi, "Hybrid neuroprosthesis for the upper limb: Combining brain-controlled neuromuscular stimulation with a multi-joint arm exoskeleton," *Frontiers Neurosci.*, vol. 10, p. 367, Aug. 2016, doi: 10.3389/fnins.2016.00367.

[3] A. Remsik et al., "A review of the progression and future implications of brain-computer interface therapies for restoration of distal upper extremity motor function after stroke," *Exp. Rev. Med. Devices*, vol. 13, no. 5, pp. 445–454, May 2016, doi: 10.1080/17434440.2016.1174572.

[4] M. Barsotti et al., "A full upper limb robotic exoskeleton for reaching and grasping rehabilitation triggered by MI-BCI," in *Proc. IEEE Int. Conf. Rehabil. Robot. (ICORR)*, Singapore, Aug. 2015, pp. 49–54, doi: 10.1109/ICORR.2015.7281174.

[5] S. R. Soekadar et al., "Hybrid EEG/EOG-based brain/neural hand exoskeleton restores fully independent daily living activities after quadriplegia," *Sci. Robot.*, vol. 1, Dec. 2016, Art. no. eaag3296, doi: 10.1126/scirobotics.aag3296.

[6] P. D. E. Baniqued et al., "Brain-computer interface robotics for hand rehabilitation after stroke: A systematic review," *J. NeuroEng. Rehabil.*, vol. 18, no. 1, p. 15, Jan. 2021, doi: 10.1186/s12984-021-00820-8.

[7] G. Pfurtscheller and C. Neuper, "Motor imagery activates primary sensorimotor area in humans," *Neurosci. Lett.*, vol. 239, pp. 65–68, Dec. 1997, doi: 10.1016/S0304-3940(97)00889-6.

[8] Z. Tang, S. Sun, S. Zhang, Y. Chen, C. Li, and S. Chen, "A brain-machine interface based on ERD/ERS for an upper-limb exoskeleton control," *Sensors*, vol. 16, no. 12, p. 2050, Dec. 2016, doi: 10.3390/s16122050.

[9] Y. Liu et al., "Motor-imagery-based teleoperation of a dual-arm robot performing manipulation tasks," *IEEE Trans. Cognit. Develop. Syst.*, vol. 11, no. 3, pp. 414–424, Sep. 2019, doi: 10.1109/TCDS.2018.2875052.

[10] Z. Li et al., "Hybrid brain/muscle signals powered wearable walking exoskeleton enhancing motor ability in climbing stairs activity," *IEEE Trans. Med. Robot. Bionics*, vol. 1, no. 4, pp. 218–227, Nov. 2019, doi: 10.1109/TMRB.2019.2949865.

[11] Z. Tang, C. Li, and S. Sun, "Single-trial EEG classification of motor imagery using deep convolutional neural networks," *Optik*, vol. 130, pp. 11–18, Feb. 2017, doi: 10.1016/j.ijleo.2016.10.117.

[12] X. Xiao and Y. Fang, "Motor imagery EEG signal recognition using deep convolution neural network," *Frontiers Neurosci.*, vol. 15, Mar. 2021, Art. no. 655599, doi: 10.3389/fnins.2021.655599.

[13] V. J. Lawhern, A. J. Solon, N. R. Waytowich, S. M. Gordon, C. P. Hung, and B. J. Lance, "EEGNet: A compact convolutional neural network for EEG-based brain-computer interfaces," *J. Neural Eng.*, vol. 15, no. 5, Oct. 2018, Art. no. 056013, doi: 10.1088/1741-2552/aace8c.

[14] G. A. Altuwaijri and G. Muhammad, "A multibranch of convolutional neural network models for electroencephalogram-based motor imagery classification," *Biosensors*, vol. 12, no. 1, p. 22, Jan. 2022, doi: 10.3390/bios12010022.

[15] S. U. Amin, H. Altaheri, G. Muhammad, W. Abdul, and M. Alsulaiman, "Attention-inception and long- short-term memory-based electroencephalography classification for motor imagery tasks in rehabilitation," *IEEE Trans. Ind. Informat.*, vol. 18, no. 8, pp. 5412–5421, Aug. 2022, doi: 10.1109/TII.2021.3132340.

[16] S. U. Amin, H. Altaheri, G. Muhammad, M. Alsulaiman, and W. Abdul, "Attention based inception model for robust EEG motor imagery classification," in *Proc. IEEE Int. Instrum. Meas. Technol. Conf. (IMTC)*, Glasgow, U.K., May 2021, pp. 1–6, doi: 10.1109/I2MTC50364.2021.9460090.

[17] A. M. Roy, "An efficient multi-scale CNN model with intrinsic feature integration for motor imagery EEG subject classification in brain-machine interfaces," *Biomed. Signal Process. Control*, vol. 74, Apr. 2022, Art. no. 103496, doi: 10.1016/j.bspc.2022.103496.

[18] X. Zhao, H. Zhang, G. Zhu, F. You, S. Kuang, and L. Sun, "A multi-branch 3D convolutional neural network for EEG-based motor imagery classification," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 27, no. 10, pp. 2164–2177, Oct. 2019, doi: 10.1109/TNSRE.2019.2938295.

[19] Y. Li, X.-R. Zhang, B. Zhang, M.-Y. Lei, W.-G. Cui, and Y.-Z. Guo, "A channel-projection mixed-scale convolutional neural network for motor imagery EEG decoding," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 27, no. 6, pp. 1170–1180, Jun. 2019, doi: 10.1109/TNSRE.2019.2915621.

[20] G. Dai, J. Zhou, J. Huang, and N. Wang, "HS-CNN: A CNN with hybrid convolution scale for EEG motor imagery classification," *J. Neural Eng.*, vol. 17, no. 1, Jan. 2020, Art. no. 016025, doi: 10.1088/1741-2552/ab405f.

[21] O. P. Idowu, A. E. Ilesanmi, X. Li, O. W. Samuel, P. Fang, and G. Li, "An integrated deep learning model for motor intention recognition of multi-class EEG signals in upper limb amputees," *Comput. Methods Programs Biomed.*, vol. 206, Jul. 2021, Art. no. 106121, doi: 10.1016/j.cmpb.2021.106121.

[22] Y. Zhang, C. S. Nam, G. Zhou, J. Jin, X. Wang, and A. Cichocki, "Temporally constrained sparse group spatial patterns for motor imagery BCI," *IEEE Trans. Cybern.*, vol. 49, no. 9, pp. 3322–3332, Sep. 2019, doi: 10.1109/TCYB.2018.2841847.

[23] J. Hu, L. Shen, S. Albanie, G. Sun, and E. Wu, "Squeeze-and-excitation networks," 2019, *arXiv:1709.01507*.

[24] B. Sun, X. Zhao, H. Zhang, R. Bai, and T. Li, "EEG motor imagery classification with sparse spectrotemporal decomposition and deep learning," *IEEE Trans. Autom. Sci. Eng.*, vol. 18, no. 2, pp. 541–551, Apr. 2021, doi: 10.1109/TASE.2020.3021456.

[25] J. Zhang, R. Yao, W. Ge, and J. Gao, "Orthogonal convolutional neural networks for automatic sleep stage classification based on single-channel EEG," *Comput. Methods Programs Biomed.*, vol. 183, Jan. 2020, Art. no. 105089, doi: 10.1016/j.cmpb.2019.105089.

[26] T. P. Luu, Y. He, S. Brown, S. Nakagome, and J. L. Contreras-Vidal, "Gait adaptation to visual kinematic perturbations using a real-time closed-loop brain-computer interface to a virtual reality avatar," *J. Neural Eng.*, vol. 13, no. 3, Jun. 2016, Art. no. 036006, doi: 10.1088/1741-2560/13/3/036006.

[27] L. Ferrero, M. Ortiz, V. Quiles, E. Iáñez, and J. M. Azorín, "Improving motor imagery of gait on a brain-computer interface by means of virtual reality: A case of study," *IEEE Access*, vol. 9, pp. 49121–49130, 2021, doi: 10.1109/ACCESS.2021.3068929.

[28] P. Xie et al., "Research on rehabilitation training strategies using multimodal virtual scene stimulation," *Frontiers Aging Neurosci.*, vol. 14, Jun. 2022, Art. no. 892178, doi: 10.3389/fnagi.2022.892178.

[29] S. H. Jang et al., "Cortical reorganization and associated functional motor recovery after virtual reality in patients with chronic stroke: An experimenter-blind preliminary study," *Arch. Phys. Med. Rehabil.*, vol. 86, no. 11, pp. 2218–2223, Nov. 2005, doi: 10.1016/j.apmr.2005.04.015.

[30] D. B. Mekbib et al., "Proactive motor functional recovery following immersive virtual reality-based limb mirroring therapy in patients with subacute stroke," *Neurotherapeutics*, vol. 17, no. 4, pp. 1919–1930, Oct. 2020, doi: 10.1007/s13311-020-00882-x.

[31] J. Huang, M. Lin, J. Fu, Y. Sun, and Q. Fang, "An immersive motor imagery training system for post-stroke rehabilitation combining VR and EMG-based real-time feedback," in *Proc. 43rd Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. (EMBC)*, Nov. 2021, pp. 7590–7593, doi: 10.1109/EMBC46164.2021.9629767.

[32] F. Lotte, M. Congedo, A. Lécuyer, F. Lamarche, and B. Arnaldi, "A review of classification algorithms for EEG-based brain–computer interfaces," *J. Neural Eng.*, vol. 4, no. 2, pp. 1–13, Jun. 2007, doi: 10.1088/1741-2560/4/2/R01.

[33] P. Martin-Smith, J. Ortega, J. Asensio-Cubero, J. Q. Gan, and A. Ortiz, "A label-aided filter method for multi-objective feature selection in EEG classification for BCI," in *Advances in Computational Intelligence*, vol. 9094, I. Rojas, G. Joya, and A. Catala, Eds. Cham, Switzerland: Springer, 2015, pp. 133–144, doi: 10.1007/978-3-319-19258-1_12.

[34] H. Sun, Y. Xiang, Y. Sun, H. Zhu, and J. Zeng, "On-line EEG classification for brain-computer interface based on CSP and SVM," in *Proc. 3rd Int. Congr. Image Signal Process.*, Yantai, China, Oct. 2010, pp. 4105–4108, doi: 10.1109/CISP.2010.5648081.

[35] R. C. Oldfield, "The assessment and analysis of handedness: The Edinburgh inventory," *Neuropsychologia*, vol. 9, no. 1, pp. 97–113, 1971, doi: 10.1016/0028-3932(71)90067-4.

[36] G. Pfurtscheller and F. H. L. da Silva, "Event-related EEG/MEG synchronization and desynchronization: Basic principles," *Clin. Neurophysiol.*, vol. 110, no. 11, pp. 1842–1857, 1999, doi: 10.1016/S1388-2457(99)00141-8.

[37] Y. Li, L. Guo, Y. Liu, J. Liu, and F. Meng, "A temporal-spectral-based squeeze-and-excitation feature fusion network for motor imagery EEG decoding," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 29, pp. 1534–1545, 2021, doi: 10.1109/TNSRE.2021.3099908.

[38] P. Sip, M. Kozlowska, D. Czysz, P. Daroszewski, and P. Lisinski, "Perspectives of motor functional upper extremity recovery with the use of immersive virtual reality in stroke patients," *Sensors*, vol. 23, no. 2, p. 712, Jan. 2023, doi: 10.3390/s23020712.

[39] D. Nath et al., "Clinical effectiveness of non-immersive virtual reality tasks for post-stroke neuro-rehabilitation of distal upper-extremities: A case report," *J. Clin. Med.*, vol. 12, no. 1, p. 92, Dec. 2022, doi: 10.3390/jcm12010092.

[40] J. S. Tutak, "Virtual reality and exercises for paretic upper limb of stroke survivors," *Tehnicki Vjesnik*, vol. 2, vol. 24, pp. 451–458, Sep. 2017, doi: 10.17559/TV-20161011143721.

[41] K. M. Oostra, A. Van Bladel, A. C. L. Vanhoonacker, and G. Vingerhoets, "Damage to fronto-parietal networks impairs motor imagery ability after stroke: A voxel-based lesion symptom mapping study," *Frontiers Behav. Neurosci.*, vol. 10, p. 5, Feb. 2016, doi: 10.3389/fnbeh.2016.00005.

[42] F. Pichiorri et al., "Brain-computer interface boosts motor imagery practice during stroke recovery," *Ann. Neurol.*, vol. 77, no. 5, pp. 851–865, May 2015, doi: 10.1002/ana.24390.