Multi-Task Heterogeneous Ensemble Learning-Based Cross-Subject EEG Classification Under Stroke Patients

Minji Lee, *Member, IEEE*, Hyeong-Yeong Park, *Student Member, IEEE*, Wanjoo Park, *Member, IEEE*, Keun-Tae Kim, Yun-Hee Kim, and Ji-Hoon Jeong, *Associate Member, IEEE*

***Abstract*— Robot-assisted motor training is applied for neurorehabilitation in stroke patients, using motor imagery (MI) as a representative paradigm of brain-computer interfaces to offer real-life assistance to individuals facing movement challenges. However, the effectiveness of training with MI may vary depending on the location of the stroke lesion, which should be considered. This paper introduces a multi-task electroencephalogram-based heterogeneous ensemble learning (MEEG-HEL) specifically designed for cross-subject training. In the proposed framework, common spatial patterns were used for feature extraction, and the features according to stroke lesions are shared and selected through sequential forward floating selection. The heterogeneous ensembles were used as classifiers. Nine patients with chronic ischemic stroke participated, engaging in MI and motor execution (ME) paradigms involving finger tapping. The classification criteria for the multi-task were established in two ways, taking into account the characteristics of stroke patients. In the cross-subject session, the first involved a direction recognition task for two-handed classification, achieving a performance of 0.7419 (±0.0811) in MI and 0.7061 (±0.1270) in ME. The second task focused on motor assessment for lesion location, resulting in a performance of 0.7457 (±0.1317) in MI and 0.6791 (±0.1253) in ME. Comparing the specific-subject session, except for ME on the motor assessment task, performance on both tasks was significantly higher than the cross-subject session. Furthermore, classification performance was similar to or statistically higher in cross-subject sessions compared to baseline models. The proposed MEEG-HEL holds promise in improving the practicality of neurorehabilitation in clinical settings and facilitating the detection of lesions.**

***Index Terms*— Stroke, electroencephalography, motor imagery, cross-subject training, multi-task heterogeneous ensemble learning.**

I. Introduction

MOTOR imagery (MI) refers to a paradigm of imagining movement without actual muscle movement [1]. This is widely used in the brain-computer interface (BCI), which is used to decode brain activity as a command to control external devices [2], [3]. In particular, electroencephalography (EEG) is widely used for BCI because it is relatively inexpensive, easy to move, and has a high time resolution compared to other neuroimaging tools [4]. Furthermore, this paradigm can aid in the neurorehabilitation of stroke patients. In fact, robot-assisted BCI training improves motor rehabilitation in stroke patients [5], [6]. This is because the neural pathway during MI is similar to motor execution (ME) [7]; therefore, it seems to help motor rehabilitation after stroke as the neural pathway related to the sensorimotor region is activated.

MI has been applied to BCI training to improve motor function [8]. In stroke patients with motor dysfunction, direct movement is inconvenient, so they imagine moving a specific body part. In addition, in robot-assisted MI training, it was most common to classify between the imagination of the right and left hands [9]. Various methods have been devised for effective discrimination of bilateral MI, including extracting feature vectors through a common spatial pattern (CSP) or distinguishing brain waves by frequency band for feature extraction [10], [11]. However, for effective motor rehabilitation, it is very important to consider the location of the lesion in the imagination of both hands because it is necessary to correct the intensity of the robot-assisted BCI training according to the location of the lesion. On the other

Manuscript received 20 November 2023; revised 9 April 2024; accepted 19 April 2024. Date of publication 29 April 2024; date of current version 6 May 2024. This work was supported in part by Chungbuk National University BrainKorea21 Chungbuk Information Technology Education and Research Center (BK21) Program (2023) and in part by the National Research Foundation of Korea (NRF) Grant funded by the Korea Government [Ministry of Science and ICT (MSIT)] under Grant RS-2023-00252624. *(Corresponding authors: Yun-Hee Kim; Ji-Hoon Jeong.)*

This work involved human subjects or animals in its research. Approval of all ethical and experimental procedures and protocols was granted by the Institutional Review Board (IRB) of Samsung Medical Center under Approval No. SMC 2013-02-091.

Minji Lee is with the Department of Biomedical Software Engineering, The Catholic University of Korea, Bucheon, Gyeonggi 14662, South Korea (e-mail: minjilee@catholic.ac.kr).

Hyeong-Yeong Park is with the Department of Computer Science, Chungbuk National University, Cheongju, Chungbuk 28644, South Korea (e-mail: hyeong.y.park@chungbuk.ac.kr).

Wanjoo Park is with the Engineering Division, New York University Abu Dhabi, Abu Dhabi, United Arab Emirates (e-mail: wanjoo@nyu.edu).

Keun-Tae Kim is with the College of Information Science, Hallym University, Chuncheon, Gangwon 24252, South Korea (e-mail: ktkim@hallym.ac.kr).

Yun-Hee Kim is with the Department of Physical and Rehabilitation Medicine, Sungkyunkwan University School of Medicine, Suwon, Gyeonggi 16419, South Korea, and also with the Myongji Choonhey Rehabilitation Hospital, Seoul 07378, South Korea (e-mail: yunkim@skku.edu).

Ji-Hoon Jeong is with the Department of Computer Science, Chungbuk National University, Cheongju, Chungcheongbuk 28644, South Korea, and also with the Research Institute for Computer and Information Communication, Cheongju, Chungbuk 28644, South Korea (e-mail: jh.jeong@chungbuk.ac.kr).

Digital Object Identifier 10.1109/TNSRE.2024.3395133

hand, cross-modal brain activation was mainly neurological evidence that was widely used in two-hand classification [12], [13]. However, it is difficult to use these neurophysiological features in stroke patients. This is because when the stroke patient moves or imagines the paretic fingers, the sensorimotor rhythm is not active due to lesions in the sensorimotor region; instead, activity occurs in the surrounding area due to brain reorganization [14]. Therefore, for effective motor rehabilitation in stroke patients, the location of the lesion should be considered in both-hand classification.

With the advent of multi-task learning, it has become possible to learn multiple tasks together. This aims to produce better generalization performance of all tasks by utilizing useful information contained in multiple related tasks as a method of machine learning [15]. This is different from multi-class classification, in which multiple classes are trained within one task. Multi-tasks are clearly different tasks, but this method can help improve the performance of each task based on task relatedness. Multi-task learning is also used in EEG signals, [16], [17]. It was used primarily in two ways: (i) to explore the important shared characteristics of related tasks and (ii) to overcome the lack of labeled EEG data [18]. In this respect, classifying both hands and the location of the lesion in stroke patients during MI are completely single tasks, but two tasks can be considered together.

For EEG-based BCIs, many studies have focused on specific-subject learning, which trains and tests classifiers within one subject [19]. This approach mainly considers the individual brain activity patterns of each subject and creates different classifiers for each subject. Therefore, the classification performance in a specific-subject approach is high [20]. On the other hand, EEG signals have the characteristic of inter-subject and inter-session variabilities. Brain dynamics that change over the short and long term generate variability between sessions, which eventually affects an individual's brain response, resulting in variability between subjects [21]. These are also related to different sources of variability [22], which eventually leads to inconsistent BCI performance [23].

In this respect, cross-subject approaches continue to be attempted due to their lack of practicality. This aims to develop classification models that can be generalized across different individuals, regardless of their unique brain activity patterns. This approach is often used in terms of 'inter-subject learning' [24] or 'subject-independent learning' [25]. Note that we described it as cross-subject learning.

From the perspective of cross-subject EEG classification, several models have been proposed to classify MI in healthy individuals. Autthasan et al. [25] proposed MIN2Net, which uses deep metric learning in a multi-task autoencoder to learn feature representation and classify patterns simultaneously. Consequently, it achieved an F1-score of 72.6% in two-handed MI. In addition, Liu et al. [26] designed a compact multi-branch one-dimensional convolutional neural network (CNN). The accuracy of 63.3% in the cross-subject approach was reported in healthy controls for both handed classifications. In this respect, deep learning (DL) has been used to increase the generalization performance of the model and convenience with a cross-subject approach.

As the model is specific to each subject, it may not be readily applicable to new stroke patients compared to healthy controls. To achieve a better generalization of MI classification in stroke patients, cross-subject classification methods have also been explored for stroke patients. Raza et al. [27] obtained classification accuracy while performing left and right MI in stroke patients. They used EEGNet, which is a typical methodology for classifying EEG signals, with a CNN [28]. The performance was 67.0% which is close to the specific-subject performance of 70.3% [27]. However, stroke patients may have different MI patterns depending on the location of stroke lesions [6], and it is difficult to obtain a large number of trials [1], [29]; therefore, this DL approach could be unreasonable, as in healthy individuals. Therefore, a more careful approach to cross-subject classification is required when considering the characteristics of stroke patients.

In this paper, we propose a cross-subject MI classification framework for stroke patients using multi-task ensemble learning with two tasks: (i) direction recognition (DR) for both hands and (ii) motor assessment (MA) for lesion sides in stroke patients. Stroke patients used both hands to imagine or perform tapping with the four fingers, excluding the thumb. For stroke patients, ensemble learning is more effective than DL because the number of trials is small [1]. Consequently, we designed a heterogeneous ensemble learning using seven baseline models. The proposed framework can ultimately help establish more effective rehabilitation strategies by providing a comprehensive understanding of MI capabilities, which, in turn, can lead to better rehabilitation of stroke patients.

The brief contributions of this study are as follows: (i) We proposed a multi-task heterogeneous ensemble learning framework using two tasks: DR and MA for stroke patients. The proposed heterogeneous ensemble component utilizes a variety of classifier types within a single framework. In particular, this approach allows for covering non-linear relationships within the EEG data of the feature space where a small number of samples. (ii) To improve generalizability to new individuals or larger patient populations, we perform cross-subject EEG classification. This approach is particularly beneficial in the context of EEG data, where there is considerable variability between subjects; thus, the proposed framework effectively learns to recognize patterns that are consistent across individuals, improving its ability to accurately classify EEG signals for new subjects. (iii) To the best of our knowledge, we first attempted to investigate whether the distinguished patterns were horizontal for both hand classification and vertically distinct for lesion locations in stroke patients. Due to its attempt, the proposed framework incorporates a shared layer with a feature weighting mechanism that selectively emphasizes features based on their relevance to both tasks. Furthermore, the targeted feature selection and weighting strategy improves the model's ability to focus on the most informative aspects of the EEG signals for both tasks, enhancing classification accuracy.

The structure of the paper is as follows: Section I offers an introduction that highlights the importance of the study and related works. Section II details the datasets used, the proposed methodology, and the statistical analysis. Section III

examines the experimental results and addresses the complexities involved in the discussion. Finally, Section IV provides a conclusion.

II. Materials and Methods

A. Participants

Nine chronic stroke patients (55.00 ± 5.36 years; F = 3) participated in this study. Data were previously published by Lee et al. [30]. Table I shows the clinical information for stroke patients. Mini-Mental State Examination (MMSE) is an index that can indicate cognitive function [31] and the range of MMSE of subjects who participated in the experiment is 28 to 30. The cognitive functions of the subjects who participated in this experiment are normal in that MMSE is considered normal if the maximum value is 30 and 25 or higher [31]. We measured an MMSE to explore whether stroke patients who participated in this experiment had no cognitive problems enough to participate in the finger tapping experiment. The Fugl-Meyer Assessment of Upper Limb (FMA_UL) is an indicator of motor function [7], [32], and the FMA_UL range of subjects included in this paper is from 17 to 56. Since our application field is focused on helping motor rehabilitation through BCI training, subjects with varying degrees of motor impairment were included to enable scenarios for BCI training. In addition, the inclusion criteria for stroke patients were (i) unilateral lesions after the first ischemic stroke and (ii) chronic stroke stage more than three months after the onset of the stroke. The exclusion criteria for stroke patients were (i) intracranial mental insertion, (ii) claustrophobia, and (iii) use of pacemakers. The study was approved by the Institutional Review Board (IRB) of Samsung Medical Center (SMC 2013-02-091) and was carried out with the written consent of all patients prior to participating in the study.

B. Experimental Protocols

The subjects performed a sequential finger-tapping task with both hands. The task was divided into ME and MI paradigms. The MI paradigm was performed after correctly understanding the task while performing the ME paradigm first. This task followed a block design with 20 blocks per task. In each block, when a dot (cue) appeared on the screen, the patients tapped their fingers sequentially from their ring to their little finger. The cue was presented on the screen for 0.8 s, followed by a break time of 0.5 s. At the end of each block, an evaluation step indicated whether the subject understood and performed the task correctly. For example, in a particular block, if the subject started with the ring finger and the dot appeared on the screen three times, the little finger was tapped during the evaluation step to obtain the correct answer. In other words, it was difficult to confirm whether the movement was imagined correctly. Therefore, we focused on what the subject could imagine as accurately as possible. Furthermore, it is important to ensure that the subject imagined it accurately because the MI classification performance is higher with only properly imagined trials [1]. This task was performed with the unaffected hand, which had better motor function, followed by the affected hand. A more detailed description of this task is provided in [30].

C. Data Acquisition & Preprocessing

EEG data was measured using the Neuro-Prax®EEG system (NeuroConn GmbH, Germany) with a sampling rate of 4,000 Hz. The signals were collected using surface Ag/AgCl electrodes (Easy Cap, Woerthsee-Etterschla, Germany) according to the international 10-20 system and referenced to an electrode of the right earlobe.

The EEG signals were preprocessed using the EEGLAB toolbox [33] in MATLAB R2022a (MathWorks, USA). Continuous data were down-sampled at 1,000 Hz and filtered between 1 and 60 Hz using a basic finite impulse response filter. In addition, the EEG data were re-referenced using the common average reference and interpolated to remove noise. Finally, we selected the 27 channels (Fp1, Fp2, F7, F3, Fz, F4, F8, Fc5, Fc1, Fc2, Fc6, T3, C3, Cz, C4, T4, Cp5, Cp1, Cp2, Cp6, T5, P3, Pz, P4, T6, O1, O2), and segmented the EEG data into 0∼0.8 s (800 samples). Therefore, the pre-processed data are configured as times (800) × channels (27) × trials.

There were 80 trials per paradigm, consisting of 20 trials per finger, excluding the thumb. However, due to the performance of this paradigm, some data in the block that did not receive the correct answer were excluded. In particular, in stroke patients, whether movement was imagined properly has a significant influence on the performance of movement imagination [1]. In this sense, depending on the performance results, each subject had a different number of trials. As a result, 58.86 ± 22.76 tasks were used in this experiment. Furthermore, for Sub01 and Sub03, there was no tapping in the ME paradigm due to the difficulty in moving the little finger. Therefore, nine patients in the MI paradigm and seven patients in the ME paradigm were analyzed, the latter excluding two patients.

D. Multi-Task EEG-Based Heterogeneous Ensemble Learning (MEEG-HEL)

We proposed a multi-task ensemble learning framework for cross-subject MI classification using DR and MA tasks in stroke patients. Fig. 1 describes the overview of the proposed framework for the cross-subject approach, which trains a model using existing patient data and only tests on a new stroke patient. For classification, a CSP was used for feature extraction. The features using shared layers were selected by feature ranking. Finally, heterogeneous ensemble learning was performed by giving different weights for each feature.

1) Multiple Task: We divided the multiple tasks according to two criteria applicable to stroke patients. The first task was the DR to classify left-handed (Lt) and right-handed (Rt) neural commands in BCI. As mentioned earlier, BCI can be used directly for rehabilitation training and applied to various devices that help stroke patients in real life. The second task was MA to classify the affected hand (AH) and the unaffected hand (UAH) according to the direction of the stroke lesion. This task is a useful criterion to help stroke patients easily and quickly distinguish the direction of the lesions before using expensive and limited neuroimaging devices. Hence, there are two classes for each task according to each criterion.

2) Feature Extraction: CSP is a spatial filtering technique specifically designed for BCI applications [34]. It aims to

TABLE I

CLINICAL CHARACTERISTICS OF STROKE PATIENTS [30]

No	Sex	Age (years)	Lesion	Type of stroke	Periods of duration (months)	FMA_UL	MMSE
Sub01	M	53	Rt. BG	infarction	62	32	28
Sub02	M	49	Rt. Medial medulla	infarction	49	44	30
Sub03	M	58	Lt. Medial medulla	infarction	63	31	30
Sub04	F	56	Rt. CR	infarction	51	56	30
Sub05	F	46	Lt. BG	infarction	17	54	30
Sub06	M	59	Rt. BG	infarction	56	48	29
Sub07	M	52	Rt. Pontine	infarction	37	45	30
Sub08	M	60	Lt. BG	infarction	49	54	30
Sub09	F	62	Rt. Stratoscapsular	infarction	53	17	28
Mean ± SD	F = 3	55.00 ± 5.36		infarction = 9	49.14 ± 15.32	42.33 ± 13.12	29.44 ± 0.88

FMA-UL = Fugl-Mayer assessment of upper limb; MMSE = mini-mental state exam; F = female; M = male; Rt = right; Lt = left; BG = basal ganglia; CR = corona radiata.

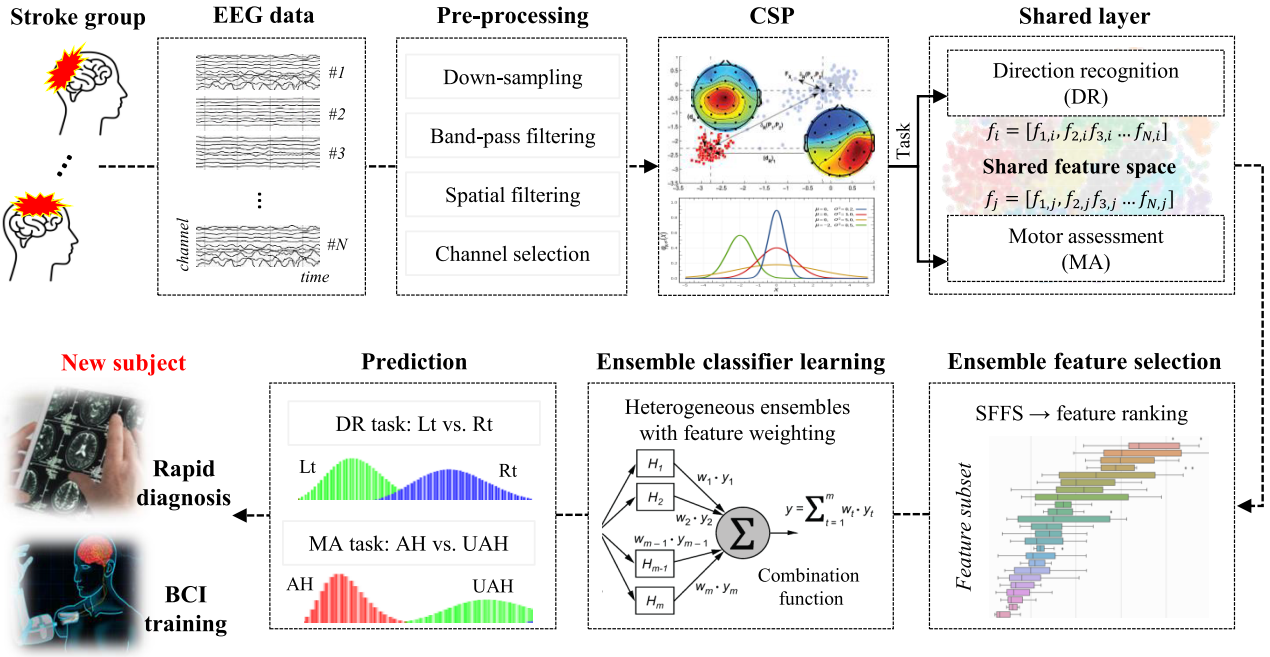

Fig. 1. Overview of the EEG-based MI system using the cross-subject approach. In the proposed framework, spatial features are extracted through CSP, and features according to stroke lesions are considered to share features for DR and MA. These features are selected through SFFS and predicted through heterogeneous ensembles. As a result, new subjects can efficiently provide BCI training for rehabilitation and lesion location for rapid diagnosis. EEG = electroencephalography; CSP = common spatial pattern; SFFS = sequential forward floating selection; Lt = left hand; Rt = right hand; AH = affected hand; UAH = unaffected hand; BCI = brain-computer interface.

identify spatial filters that improve discriminative information between different classes (e.g., Lt vs. Rt movement) in EEG signals. EEG signals are formalized as

$$\{E_n\}_{n=1}^{N} \in \mathbb{R}^{ch \times time} \tag{1}$$

where N is th number of trials, ch is the number of channels, and $time$ is the range of time domain. For each class, the covariance matrix is computed by averaging the covariance matrices of all epochs belonging to that class. The average covariance matrices transform into a whitening transformation matrix P using the eigenvector matrix U.

$$E_n \in \mathbb{R}^{ch \times time} \mapsto x_n \in \mathbb{R}^{d}, \tag{2}$$

$$X \in \mathbb{R}^{d \times N} \tag{3}$$

The projection matrix is represented as W, and the input EEG can be transformed into uncorrelated components Z with the projection matrix, the input EEG data X are then reconstructed using the inverse matrix W^{-1}.

$$W = U^T P, \tag{4}$$

$$Z = WX, \tag{5}$$

$$X = W^{-1} Z \tag{6}$$

The generalized eigenvalue problem is solved using covariance matrices to determine spatial filters that maximize the variance between the two classes while minimizing the variance within each class. Spatial filters are applied to the raw EEG data to obtain CSP features (f). We calculated a transformation matrix using spatial patterns consisting of the logarithmic variances of the first and last three columns of the inverse weight matrix [35], [36].

$$f_{DR} = \{f_{1,DR}, f_{2,DR}, f_{3,DR}, \ldots, f_{N,DR}\} \tag{7}$$

$$f_{MA} = \{f_{1,MA}, f_{2,MA}, f_{3,MA}, \ldots, f_{N,MA}\} \tag{8}$$

3) Shared Layer With Feature Weighting: In the context of multi-task learning that incorporates CSP features in the shared layer, we construct a learning architecture that unifies the spatial patterns of distinction derived from CSP according to each task (i.e., DR and MA tasks). Furthermore, the combined features of the shared layer operate by enhancing the connectivity between the CSP features relevant to both DR and MA tasks, thus facilitating a cohesive learning environment that significantly improves the overall performance of the model. This is achieved by employing a sophisticated weighting mechanism for each CSP feature, ensuring that the most relevant features are prioritized during the learning process. Therefore, the CSP features identified for each task are spread through a technique that uses variables to multiply weights, focusing on mutual features. These shared features were formulated as depicted in the subsequent equation.

$$f_{Multi} = \left\{ f_{1,Multi}, f_{2,Multi}, f_{3,Multi}, \ldots, f_{N,Multi} \right\} \quad (9)$$

For example, the value of $f_{1,Multi}$ was calculated for $f_{1,DR} \times f_{1,MA}$ to determine the combined feature weight, and $f_{1,Multi}$ performed several iterations to configure to $f_{N,Multi}$. By this shared layer, the learning framework can train on combined features, ultimately leading to improved overall performance through the effective utilization of shared information relevant to all tasks.

4) Ensemble Feature Selection: Sequential feature selection algorithms are a set of greedy search techniques utilized to reduce an initial feature space of dimensionality d to a subspace of dimensions k, where k is less than d. These algorithms automatically identify a subset of features that are most relevant to the given problem, employing a sophisticated strategy that integrates various feature selection techniques [37], [38]. The primary objectives of this approach are to improve the reliability and adaptability of the selected features. One of the methods within this category is sequential forward floating selection (SFFS), in systematically constructing and refining a specific feature subset [39], [40]. Compared to simpler sequential forward selection (SFS) algorithms, floating variants like SFFS include additional steps to exclude or include features after their initial selection, allowing for the exploration of a larger number of feature subset combinations. It is essential to highlight that these exclusion or inclusion steps are conditional and only executed if the criterion function deems the resulting feature subset as 'improved' after the removal or addition of a specific feature. Therefore, the SFFS algorithm takes the entire set of features (with a feature space of N dimensions) as input, commencing with an empty set referred to as the 'null set,' indicating an initial k value of 0, where k denotes the subset's size. The representation of a feature subset is as follows:

$$X_k = \left\{ x_j | j = 1, 2, \ldots, k; x_j \in f_{Multi} \right\}, \quad (10)$$

$$where\ k = (0, 1, 2, \ldots, N) \quad (11)$$

The process of inclusion comprises adding the feature from the feature space that leads to the most significant enhancement in performance for the feature subset, as evaluated by the criterion function. The criterion function $J(x)$ serves as a quantitative measure to evaluate the significance of each feature (or set of features) to enhance the model performance. $J(x)$ is essential to identify the most informative and relevant CSP features, facilitating the construction of an optimal feature subset for the model. In the inclusion phase, a feature x is added to the current feature subset X_k if it maximizes the function $J(X_k + x)$, indicating a substantial improvement in the model performance with the inclusion of x. Y represents the entire feature space, and X_k denotes the current set of selected features.

$$x^+ = argmax J(X_k + x), where\ x \in Y - X_k, \quad (12)$$

$$X_{k+1} = X_k + x^+, \quad (13)$$

$$k = k + 1 \quad (14)$$

On the contrary, the conditional exclusion phase involves removing a feature x from X_k if its exclusion improves the value of the function of the criterion $J(X_k - x)$, suggesting that the model's performance is better without x. This iterative process continues until a stopping criterion is met, such as reaching a predetermined number of features in the subset. Therefore, if no performance enhancement is discernible, the process returns to the inclusion step.

$$x^- = argmax J(X_k - x), where\ x \in X_k, \quad (15)$$

$$if\ J(X_k - x) > J(X_k), \quad (16)$$

$$X_{k-1} = X_k - x^-, \quad (17)$$

$$k = k - 1 \quad (18)$$

These processes of inclusion and conditional exclusion persist until the termination criterion is satisfied, ceasing when the subset size k equals the desired number of features.

The approach involves various iterations to adjust and test the CSP feature [34], with each iteration adapting different configurations and settings for the parameters. Following each iteration of CSP feature extraction, the SFFS method is employed to carefully select the CSP features that are most informative, adjusting for the nuances of each specific configuration. The selected features from each CSP configuration are then thoughtfully combined into what is described as an ensemble feature set. This set incorporates the most relevant spatial patterns from the EEG data, adept at handling the variations caused by different CSP parameters. The utility of this set of ensemble features is particularly notable in tasks that involve classification, where it markedly improves the performance of the model.

5) Ensemble Classifier Learning: Heterogeneous ensemble methods utilize a variety of base learning algorithms to intentionally introduce diversity within the ensemble [41]. In this study, the heterogeneous ensemble is composed of seven base estimators: a shallow neural network (SNN), kernel support vector machine (KSVM), subspace discriminant classifier (SDC), logistic kernel regression (LKR), bagging decision tree (DT), k-nearest neighbors (k-NN), and kernel naïve Bayes (KNB). It is crucial to emphasize that these base estimators undergo independent training without interdependence.

- Shallow neural network: The architecture of the SNN typically consists of an input layer, two hidden layers, and an

output layer [42]. In this study, SNN has one hidden layer with five units and uses a rectified linear unit activation function with an output layer as follows.

$$f(x) = max(0, x) \tag{19}$$

The network takes 10 input features and outputs the probability of the input belonging to multiple classes in the classification task.

• Kernel support vector machine: It is an extension of the standard linear SVM [43] that allows for non-linear decision boundaries between classes. The Gaussian kernel is given by the following formula.

$$K(x_i, x_j) = exp(-\frac{\left\|x_i, x_j\right\|^2}{2\sigma^2}) \tag{20}$$

where x_i and x_j are two data points in the original feature space. $\left\|x_i, x_j\right\|$ represents the Euclidean distance between x_i and x_j in the original feature space. σ is a hyperparameter known as kernel width.

• Subspace discriminant classifier: SDC operates in a subspace, making it particularly useful for dealing with high-dimensional data from linear discriminant analysis. The SDC aims to find a low-dimensional subspace in which the classes are well separated, leading to improved performance in classification tasks for dealing with high-dimensional datasets [44].

• Logistic kernel regression: It builds on the principles of both logistic regression and kernel methods to handle non-linear relationships in data through kernel transformations [45]. The feature vector x is replaced by the transformed feature vector $\Phi(x)$ using the kernel function:

$$P(y = 1|x) = \frac{1}{1 + exp(-w^T \Phi(x))} \tag{21}$$

where Φ(x) denotes the feature vector transformed using the kernel function.

• Bagging decision tree: It is an ensemble learning technique that combines the power of decision trees with the principles of bagging to improve predictive performance and reduce overfitting [46].

• k-nearest neighbors: Given a new data point x, the k-NN algorithm finds the k closest data points in the training set (neighbors) based on a chosen distance metric and assigns the majority class to predict the new data point $\bar{x}$. The prediction of $\bar{x}$ is as follows.

$$argmax_{c_i} \sum_{u \in N_k(x)} \delta(y_i, c_i) \tag{22}$$

where $N_k(x)$ is the set of k nearest neighbors of x from the training dataset. c_i represents the ith class label in the dataset. $\delta(y_i, c_i)$ is the Kronecker delta function, which is equal to 1 if $y_i = c_i$ and 0 otherwise [47].

• Kernel naïve Bayes: It is an extension of the naïve Bayes algorithm to make the assumption of feature independence given the class label [43]. C is the class label, and vector of n features x is defined as x = $(x_1, x_2, \ldots, x_n)$ as follows.

$$P(C|x) = \frac{P(C) \times P(x|C)}{P(x)} \tag{23}$$

where $P(C|x)$ is the posterior probability of class label C given feature vector x. $P(C)$ is the prior probability of the class C. $P(x|C)$ is the likelihood of the feature vector x given class C. $P(x)$ is the probability of observing the feature vector x throughout the data set.

In addition, the heterogeneous ensemble strategy was modified to integrate an entropy weighting approach for each estimator [48]. This involves utilizing entropy as the evaluation metric to gauge the value of each base estimator. Entropy, which serves as a measure of uncertainty or impurity in a set, reflects higher disorder in a more chaotic set. Information entropy is influenced by two factors: (i) the number of distinct values that the variable can take and (ii) the uncertainty associated with each value [40]. Consequently, lower entropies are favored and to achieve this, it is crucial to ensure that the weights assigned to base classifiers are inversely proportional to their respective entropies. To compute the entropy over the validation set, the base estimators contribute, and an ensemble of seven base estimators generates a vector of predicted labels.

$$y_{pred} = \{y_1, y_2, y_3, y_4, y_5, y_6, y_7\} \tag{24}$$

This set has m predictions of y = $Class_1$ and n predictions of $Class_2$. These label counts can be equivalently expressed as label probabilities: the probability of predicting y = $Class_1$ is $P(y = Class_1)$, and the probability of predicting y = $Class_2$ is $P(y = Class_2)$. With these label probabilities, we can compute the entropy over this set of base estimator predictions as follows.

$$\begin{aligned} Entropy_f = &-P(y = Class_1) log_2 P(y = Class_1) \\ &-P(y = Class_2) log_2 P(y = Class_2) \end{aligned} \tag{25}$$

$Entropy_{f,k}$ is the validation entropy of the kth estimator. The weight of each base estimator is set to term as follows.

$$W_{Enpk} = \frac{\frac{1}{Entropy_{f,k}}}{\sum_{k=1}^{7} \left(\frac{1}{Entropy_{f,k}}\right)} \tag{26}$$

The entropy of a base estimator is determined exclusively by its predicted labels. This entropy metric gauges the uncertainty linked to an estimator's predictions, where a lower entropy signifies greater confidence. As a result, the weights allocated to individual base estimators are reciprocally related to their respective entropies. This adjustment ensures that the accuracy-based weight term ($a_{f,k}$) is adapted for the final weight calculation.

$$W_{Acck} = \frac{a_{f,k}}{\sum_{k=1}^{7} a_{f,k}} \tag{27}$$

This computation ensures that a classifier's weight is proportionate to its accuracy, and the weights collectively sum to 1. The final prediction is then calculated as a weighted sum of the individual predictions:

$$\begin{aligned} y_{final} = &w_{Enpk,1} \cdot w_{Acck,1} \cdot y_1 + w_{Enpk,2} \cdot w_{Acck,2} \cdot y_2 + \\ &\cdots w_{Enpk,6} \cdot w_{Acck,6} \cdot y_6 + w_{Enpk,7} \cdot w_{Acck,7} \cdot y_7 \end{aligned} \tag{28}$$

Indeed, higher weights are allocated to models demonstrating superior accuracy or specialized knowledge, underscoring their increased impact on the final decision of the ensemble. The procedure encompasses multiplying the model predictions by their corresponding weights and summing them, resulting in a consolidated decision.

Hence, each classifier within the heterogeneous ensemble is indeed trained independently using the same inputs, which consist of features extracted and selected through the CSP analysis and the SFFS algorithm. Following independent training of each classifier, weight calculations for these classifiers are performed on the same validation set. Therefore, MEEG-HEL was adapted with an entropy weighting approach combined with accuracy-based weighting to integrate the predictions of the base classifiers into the final ensemble decision. In particular, the entropy of each classifier's predictions on the validation set is used to assess the uncertainty associated with each classifier's predictions. Lower entropy values indicate more confident and consistent predictions by a classifier, leading to higher weights being assigned to such classifiers. In addition, the accuracy of each classifier on the validation set further influences the final weighting, with higher-accuracy classifiers receiving more weight in the ensemble's final decision. This weighting strategy ensures that both the confidence (as measured by entropy) and the accuracy of each classifier's predictions are taken into account when aggregating the ensemble's final output.

E. Performance Metrices

Accuracy serves as a fundamental performance metric to assess the effectiveness of a machine learning model in classification tasks. Classification accuracy represents the ratio of the number of accurate predictions to the total number of input samples [49]. We measured the accuracy of 2-class classification at each task.

F. Model Training

1) Cross-Subject Session: Training models across cross-subjects could consider individual variations and improve generalization across diverse datasets. This method entails consolidating data from multiple subjects to create a comprehensive dataset, allowing the model to use a more diverse and representative sample of the study population. During the training phase, the model learns from the combined dataset by incorporating data points from various subjects.

In the specific context of leave-one-subject-out cross-validation (LOSO-CV), accuracy gauges how well the model generalizes to unseen subjects. LOSO-CV systematically excludes one subject from the training set at a time, evaluating the model predictions on that subject's data [50]. This cross-validation variant is tailored for situations where data exhibit specific-subject characteristics or dependencies. In LOSO-CV, the dataset is partitioned into subsets based on subjects, and each fold involves using one subject's data as the validation set while the remaining data serves as the training set. This process is iterated for each subject, generating as many folds as there are subjects in the dataset. To calculate the accuracy of a specific fold in LOSO-CV, model predictions are compared with the ground truth labels, known for the left-out subject. Therefore, in our experiment, we trained with 8 subjects' data except for one subject and tested with the other subject. In addition, since there are a total of 9 subjects, this process was repeated 9 times.

2) Specific-Subject Session: In the specific-subject training scenario, the data for each subject is partitioned into distinct training and test sets. For example, within the dataset of a single subject, 80% of the data is designated to train the proposed model, leaving the remaining 20% to evaluate the performance of the model. Furthermore, we employ a 5-fold cross-validation strategy for a fair performance evaluation. The training process adheres to conventional machine learning procedures, involving tasks such as data pre-processing, feature extraction, and the application of classification algorithms.

G. Statistical Analysis

Prior to commencing the statistical analysis, we performed normal and homoskedastic tests, considering the small sample size. The Shapiro-Wilk test, a common tool for normality validation, confirmed the satisfaction of the null hypothesis (H0), indicating normality. Levene's test validated that the assumption of homoskedasticity was met. A two-way analysis of variance (ANOVA) was executed to compare classification performances, using the paradigm (ME vs. MI) and the session (cross-subject vs. specific-subject) as factors. Post hoc analysis consisted of paired or two-sample t-tests for binary comparisons. Subsequently, to assess classification performance between sessions and tasks (DR vs. MA), a two-way ANOVA was conducted, supplemented by paired t-tests. Furthermore, one-way ANOVA was employed to compare the classification performance between the proposed method and other models. Furthermore, we performed Pearson's correlations to investigate the relationship between EEG decoding performance and FMA_UL across subjects. Significance levels were set at $p = 0.05$, with the Bonferroni correction applied.

III. Experimental Results

A. Classification Performance in Cross-Subject Session

1) DR Task: Table II presents the classification performance for the DR task in both the cross-subject and specific-subject sessions. It is difficult to distinguish between both hands in situations where the location of each lesion is different for each stroke patient without a flip-flop for the data. However, it is very important to recognize both movements without flip-flops of EEG data in order to utilize BCI without the training stage. The cross-subject session and the grand average classification performance for the classification of Lt vs. Rt in the ME and MI paradigms were 0.7061 and 0.7419, respectively. Sub08 demonstrated the highest performance with an accuracy of 0.8670 in the MI paradigm. In addition, the MI paradigm could not observe statistically significant differences in performance compared to the ME paradigm.

We evaluated the performance of the decoding model using a confusion matrix, which provided valuable information on the reliability of the results through true positive (TP), true negative (TN), false positive (FP), and false negative (FN)

TABLE II
CLASSIFICATION PERFORMANCES OF DR TASK USING PROPOSED MEEG-HEL

Lt vs. Rt					
Cross-subject session			Specific-subject session		
Test subject	ME	MI	Subject	ME	MI
Sub01	-	0.7690	Sub01	-	0.9090
Sub02	0.5570	0.8330	Sub02	0.8970	0.8930
Sub03	-	0.6670	Sub03	-	0.7620
Sub04	0.6270	0.8070	Sub04	0.7670	0.9090
Sub05	0.6370	0.7610	Sub05	0.9680	0.7310
Sub06	0.9590	0.6150	Sub06	0.9030	0.9800
Sub07	0.6510	0.6910	Sub07	0.7100	0.7100
Sub08	0.6940	0.8670	Sub08	0.8330	0.9800
Sub09	0.8180	0.6670	Sub09	0.9570	0.8890
Average	0.7061	0.7419	Average	0.8621	0.8626
± SD	± 0.1270	± 0.0811	± SD	± 0.0896	± 0.1026

TABLE III
CLASSIFICATION PERFORMANCES OF MA TASK USING PROPOSED MEEG-HEL

AH vs. UAH					
Cross-subject session			Specific-subject session		
Test subject	ME	MI	Subject	ME	MI
Sub01	-	0.7690	Sub01	-	0.8180
Sub02	0.5950	0.6390	Sub02	0.7240	0.7500
Sub03	-	0.6570	Sub03	-	0.8100
Sub04	0.6400	0.7370	Sub04	0.9670	0.9550
Sub05	0.6000	0.5670	Sub05	0.9680	0.7690
Sub06	0.5410	0.9620	Sub06	0.9680	0.9800
Sub07	0.8550	0.6670	Sub07	0.7100	0.7420
Sub08	0.8870	0.9720	Sub08	0.8750	0.9800
Sub09	0.6360	0.7410	Sub09	0.9570	0.9800
Average	0.6791	0.7457	Average	0.8813	0.8649
± SD	± 0.1253	± 0.1317	± SD	± 0.1084	± 0.1064

values. Fig. 2(a) illustrates the results for both Lt and Rt classes. In the ME paradigm, the average TP and FN values were 0.8109 and 0.8682, respectively. These values indicate the accuracy of correctly predicting the Lt class and the rate of correctly predicting the Rt class as Rt, respectively. Surprisingly, the TN value, which represents the rate of correctly predicting the Rt class, was 0.1891, indicating poor performance in distinguishing Rt from Lt.

In contrast, in the MI paradigm, the TP and FN values were 0.8730 and 0.9532, respectively. These numbers reveal the effectiveness of correctly identifying the Rt class and the rate of misclassifying the Rt class as Lt. Interestingly, the average probability of predicting the Rt class was significantly high, suggesting that the MI paradigm showed promising results in distinguishing Rt from Lt. The contrasting performances between the ME and MI paradigms highlight the importance of employing the appropriate approach in this classification task. Further investigation and optimization techniques are required to improve the accuracy and robustness of the classification model, particularly in the context of the prediction of Rt and Lt. Furthermore, exploring the factors that contribute to the notable difference in predicting Rt between the two paradigms could offer valuable information on brain signals and improve our overall understanding of the underlying neural processes.

2) MA Task: We successfully demonstrated the classification performance of the cross-subject session for AH vs. UAH classification. As shown in Table III, the grand average decoding performances in the cross-subject session were 0.6791 and 0.7457 for the ME and MI paradigms, respectively. In particular, the MI paradigm showed a significant improvement of approximately 7% over the ME paradigm. Sub08 was the best performer in this experiment, achieving a high decoding performance of 0.9720 in the MI paradigm.

As shown in Fig. 2(b), the TP and FN values for the UAH and AH classes in the ME paradigms were 0.7730 and 0.8012, respectively. Furthermore, the TN value, representing the probability of correctly predicting UAH when the actual class was AH, was 0.2270, which is significantly higher than that of the other misclassifications. This suggests that the model performed relatively better in distinguishing UAH instances in the ME paradigm. In contrast, in the MI paradigm, the TP and FN values were even higher at 0.8420 and 0.9212, respectively. The significantly high probability of predicting UAH indicates the promising capability of the model to accurately identify UAH instances in this paradigm.

B. Classification Performance in Specific-Subject Session

1) DA Task: Both the ME and MI paradigms achieved high classification performances, with accuracies of 0.8621 and 0.8626, respectively. However, there was no significant differences were observed between the accuracies of the two paradigms. Furthermore, the subjects who showed the best performance in the two paradigms differed in specific-subject sessions.

2) MA Task: The grand average decoding performances for AH vs. UAH were 0.8813 and 0.8649 for the ME and MI paradigms, respectively. Interestingly, there was no statistically significant difference between the ME and MI paradigms, indicating only a marginal distinction. These findings highlight the potential of using the MI paradigm for AH vs. UAH classification, as they demonstrated superior performance compared with the ME paradigm in the cross-subject session. However, in the specific-subject session, both the ME and MI paradigms showed comparable performances, suggesting that individual differences might be involved in classification performance. Additionally, the remarkable performance of Sub08 warrants further investigation to explore the unique neural patterns that contributed to these exceptional results.

C. Comparison of Two Multi-Tasks

We classified the same data using two tasks suitable for stroke patients. As a result, the proposed framework learned well according to each criterion and, in this respect, we compared the differences according to the criteria.

1) Performance: We compared the classification performance in each task throughout the session (cross-subject vs. specific-subject) and task (DR vs. MA). There was a performance difference when considering the session and the task for both ME ($Chi - square = 10.59$, $p = 0.001$) and MI ($Chi - square = 10.79$, $p = 0.001$). In ME, the specific-subject session performance was significantly higher than the cross-subject session performance according to the

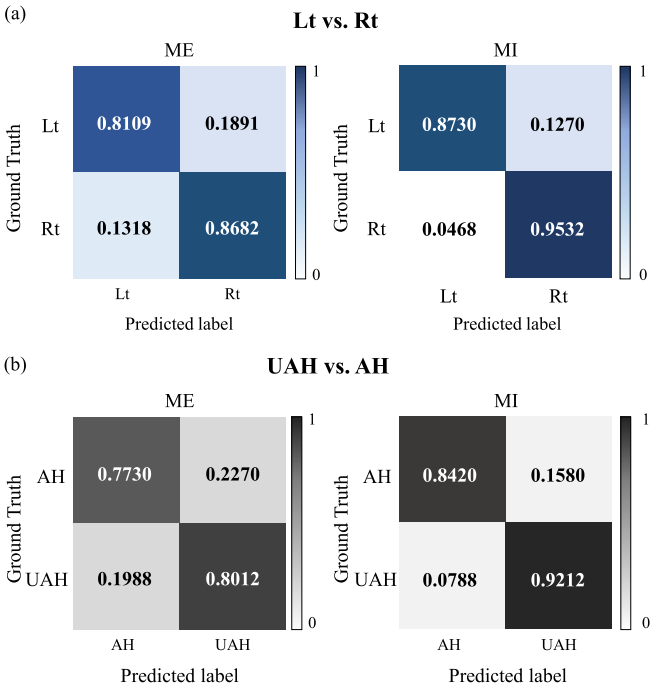

Fig. 2. Confusion matrix according to each ME and MI paradigm in cross-subject session. (a) DR task (Lt vs. Rt classification). (b) MA task (AH vs. UAH classification).

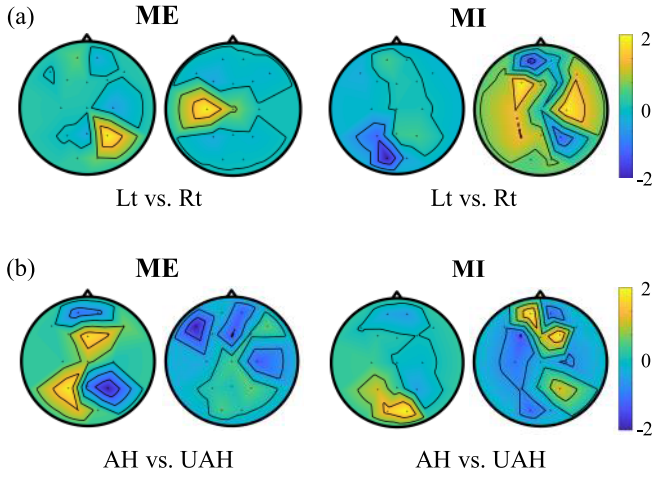

Fig. 3. Representation of CSP pattern according to each ME and MI paradigm in cross-subject session. (a) DR task (Lt vs. Rt classification). (b) MA task (AH vs. UAH classification).

DR ($t = -2.92$, $p = 0.035$). However, there were no differences in performance were observed between the two sessions in the MA. In MI, the performance of the specific-subject session was higher than that of the cross-subject session for both DR ($t = -3.11$, $p = 0.015$) and MA ($t = -4.07$, $p = 0.003$). In summary, in ME, there was no difference in performance between the specific-subject and cross-subject sessions when comparing performance from the MA perspective. This is believed to be due to differences in the task for stroke patients. If the DR is concentrated only in the left or right directions where it actually moves, the lesion side is divided into ipsilesional and contralesional sides, where it is not in the left or right directions. In other words, even with the same data, the classification characteristics vary according to the task. This difference was more pronounced in ME, where the muscles actually move.

We additionally investigated the relationship between DR and MA and their relevance to MI and ME classifications. In each MI and ME, there was no significant difference in classification performance between DR and MA in the cross-subject approach. This is probably because the criteria are clear in DR and MA, and the model itself has learned according to each criterion, so there seems to be no relationship between performance in DR and MA.

2) Spatial Feature: We compared the CSP for the classification of DR and lesion location as shown in Fig. 3. In the classification of DR, the Lt class has a deactivated pattern in the left parietal region, while in the Rt class, an activation pattern appears in both sensorimotor cortexes. However, it can be seen that the active patterns are clearly classified into Lt and Rt classes. In ME, the left sensorimotor cortex is activated in the Rt class and the right parietal region is activated in the Lt class. Similarly, the two classes are clearly distinguished. Because stroke patients have motor function deficits, motor-related neural pathways may be damaged, resulting in a somewhat different pattern from CSP in both hands of healthy controls [1]. However, it is important to have a contrasting CSP in both hemispheres for DR. In other words, the pattern is divided horizontally for DR in stroke patients.

On the contrary, the classification of lesion sides did not show CSP distinguished by both hemispheres. In the AH class, both MI and ME showed activation in the parietal region. However, in the UAH class, MI showed activity and ME showed inactivity in both frontal regions. In other words, unlike DR, CSP is divided into front and rear axes. In other words, no cross-modal activation pattern was found, which seems plausible for MA classification because some subjects moved their right hand and some moved their left hand in the AH class. In fact, this phenomenon is believed to occur in the frontal and parietal regions during motor preparation or planning following a stroke [52], [53]. Thus, the pattern is vertically divided for the MA.

3) Relationship With FMA_UL: We analyzed the correlation between EEG decoding performance and FMA_UL across subjects. Classification performance included all the accuracy from two sessions and two tasks. The FMA_UL had no significant relationship with any decoding performance in both tasks. This measure is a representative indicator of motor function and is used as a representative factor to predict motor impairment. However, the ME and MI paradigms used in this experiment are used for BCI training, which is possible even if the motor disability is severe, so there is no difference in performance according to motor function. In the end, this aspect may be the reason why this paradigm is widely employed in motor rehabilitation.

D. Comparison of Classification Performances

We compared the classification performance of the proposed MEEG-HEL with that of existing comparative models, as shown in Table IV and Table V. The comparative models were selected as widely used approaches for analyzing EEG in

TABLE IV
COMPARISON OF CLASSIFICATION PERFORMANCE USING THE BASELINE MODELS AND PROPOSED MEEG-HEL IN DA TASK (LT VS. RT)

Lt vs. Rt								
Test subject	ME				MI			
	SVM [44]	SNN [52]	Ensemble [1]	MEEG-HEL	SVM [44]	SNN [52]	Ensemble [1]	MEEG-HEL
Sub01	-	-	-	-	0.5540	0.6241	0.6540	0.7690
Sub02	0.5190	0.5190	0.5700	0.5570	0.4310	0.5420	0.4720	0.8330
Sub03	-	-	-	-	0.6470	0.7060	0.6670	0.6670
Sub04	0.4570	0.5200	0.5470	0.6270	0.3310	0.7560	0.6432	0.8070
Sub05	0.5230	0.5250	0.5750	0.6370	0.4780	0.4788	0.5670	0.7610
Sub06	0.8590	0.5680	0.8540	0.9590	0.5380	0.3850	0.4230	0.6150
Sub07	0.5660	0.6580	0.5180	0.6510	0.5680	0.6300	0.6050	0.6910
Sub08	0.6120	0.6610	0.6450	0.6940	0.6000	0.5670	0.4670	0.8670
Sub09	0.8180	0.5090	0.8640	0.8180	0.3330	0.6670	0.3310	0.6670
Average	0.6220	0.5657	0.6533	**0.7061**	0.4978	0.5951	0.5366	**0.7419**
± SD	± 0.1442	± 0.0618	± 0.1349	± 0.1270	± 0.1066	± 0.1089	± 0.1115	± 0.0811

TABLE V
COMPARISON OF CLASSIFICATION PERFORMANCE USING THE BASELINE MODELS AND PROPOSED MEEG-HEL IN MA TASK (AH VS. UAH)

AH vs. UAH								
Test subject	ME				MI			
	SVM [44]	SNN [52]	Ensemble [1]	MEEG-HEL	SVM [44]	SNN [52]	Ensemble [1]	MEEG-HEL
Sub01	-	-	-	-	0.6540	0.6540	0.5380	0.7690
Sub02	0.5190	0.5700	0.5300	0.5950	0.4860	0.4720	0.4720	0.6390
Sub03	-	-	-	-	0.3530	0.4960	0.3530	0.6570
Sub04	0.5730	0.5470	0.5470	0.6400	0.2980	0.6250	0.3140	0.7370
Sub05	0.6000	0.5000	0.5500	0.6000	0.3420	0.5250	0.5220	0.5670
Sub06	0.5410	0.6760	0.5410	0.5410	0.7690	0.4620	0.9430	0.9620
Sub07	0.8250	0.5410	0.5300	0.8550	0.3700	0.5310	0.5310	0.6670
Sub08	0.3870	0.3550	0.6940	0.8870	0.9340	0.5330	0.5330	0.9720
Sub09	0.5640	0.7450	0.6360	0.6360	0.2960	0.3330	0.4770	0.7410
Average	0.5727	0.5620	0.5754	**0.6791**	0.5002	0.5146	0.5203	**0.7457**
± SD	± 0.1212	± 0.1158	± 0.0592	± 0.1253	± 0.2187	± 0.0883	± 0.1680	± 0.1317

stroke patients. SVM [43] and SNN [51], the models with top accuracy in the estimators, were retrained, and the ensemble voting model [1] was added for comparison.

In the comparison between Lt and Rt classification, the cross-subject session demonstrated statistically significant differences and outperformed the other models in a statistically significant manner (Table IV). The classification performance was exhibited in the proposed model with accuracy rates of 0.7061 in the ME paradigm. The Ensemble, SVM, and SNN exhibited significant disparities in classification performance compared to the proposed model (Ensemble: $t = -2.54$, $p = 0.038$; SVM: $t = -4.72$, $p = 0.002$; SNN: $t = -2.84$, $p = 0.025$). Similarly, in the MI paradigm, MEEG-HEL achieved an accuracy of 0.7419. On the contrary, the classification performances showed 0.4978, 0.5951, and 0.5366, respectively. In addition, Ensemble ($t = -5.10$, $p < 0.001$), SVM ($t = -5.42$, $p < 0.001$), and SNN ($t = -3.69$, $p = 0.005$) showed statistically significant differences.

Regarding MA classification, similar but slightly different results were observed (Table V). In the ME paradigm, the classification performances were 0.5727, 0.5620, and 0.5754 for Ensemble, SVM, and SNN, respectively. Furthermore, the performance of the comparative models was significantly lower than that of the proposed model on average ($t = -2.69$, $p = 0.031$). During the cross-subject session in the MI paradigm, the MEEG-HEL model achieved an accuracy of 0.7457, while Ensemble, SVM, and SNN achieved accuracies of 0.5203, 0.5002, and 0.5146, respectively. This significant difference in performance demonstrates that the proposed model outperforms the existing models in this particular scenario (Ensemble: $t = -5.05$, $p < 0.001$; SVM: $t = -5.91$, $p < 0.001$; SNN: $t = -4.58$, $p = 0.001$).

The comparable or superior performance of the proposed MEEG-HEL can be attributed to the unique characteristics of stroke patients. Acquiring a substantial number of trials from stroke patients presents challenges due to mobility problems and shorter attention spans, as emphasized in [1].

E. Discussion & Future Work

This study had several limitations. First, the number of data from stroke patients was few samples. Given the limited data available on stroke patients, dividing the dataset to allocate a portion for validation results in a further reduction in the training set samples for each classifier within the multi-task EEG-based heterogeneous ensemble learning (MEEG-HEL) framework. This reduction of training samples can pose a risk of underfitting, where the models may not learn the underlying patterns in the data sufficiently to generalize well to unseen data. Underfitting occurs when a model is too simple to capture the complexity of the data set, often because it has too few samples to learn from, leading to poor performance in both the training and validation sets. However, the MEEG-HEL framework employs several strategies to mitigate the risk of underfitting and enhance the model's ability to learn effectively from a small dataset.

We adopted various strategies to avoid this problem; heterogeneous ensemble learning, feature selection, entropy-based weighting, and cross-subject training approach. Although these strategies do not eliminate the risk of underfitting completely, the framework is designed to mitigate its impact and enhance the model's ability to learn from a limited amount of data. However, it remains important to improve model performance and consider additional strategies, such as data augmentation or transfer learning, to further address challenges associated with small datasets in stroke patient studies. Eventually, even within the task, the number of trials per class is not large; therefore, it may not be suitable for deep learning, which requires large amounts of data. Despite the difficulty in obtaining long-term data for stroke patients, it is worth trying to collect a large amount of data in the future.

Another approach to solving the problem of data scarcity is data augmentation, which is suitable for the MI dataset [54]. This method effectively augments hard-to-obtain datasets, enabling deep learning methods to achieve better performance. In addition, the nine stroke patients who participated in this experiment can be divided into 6 supratentorial and 3 infratentorial depending on the location of the lesion. The number of subjects itself is not large, so it is difficult to directly conclude remarks on the diversity of stroke lesions and BCI performance. Nevertheless, the lesion location as one of the lesion-related information was included in our study. Depending on the lesion location, the classification of both hands shows a vertically divided pattern and appears differently. Therefore, this information could be an important factor in promoting generalizability by complementing cross-subject BCI.

Second, our study was limited to binary classification, which necessitates extension into multi-class classifications. In terms of application to BCI, classification of both hands is also less practical, and in terms of the patient's lesion direction, it will be more important to go one step further and determine not only the direction but also the lesion location. Consequently, future research should focus on developing a multi-class cross-subject framework for stroke patients, enhancing the practicality and specificity of BCI applications in clinical settings. In addition, for specific-subject sessions, we plan to conduct additional experiments to explore the impact of varying the number of EEG channels and the length of EEG signals on classification performance in future work. This will involve adjusting these parameters within our dataset and evaluating the resultant changes in the model's ability to accurately classify motor imagery tasks. Hence, we aim to identify optimal settings that balance the trade-offs between computational efficiency, model complexity, and performance metrics.

Finally, the process of verifying generalization using publicly available datasets was not thoroughly explored. This was due in part to the absence of publicly accessible datasets specific to stroke patients, which highlights the innovative aspect of our study, namely, the implementation of cross-subject learning using the LOSO-CV. Additional validation efforts are required, utilizing public datasets from other stroke patient studies to further substantiate our findings.

IV. Conclusion

In this paper, we proposed the MEEG-HEL framework during MI in stroke patients using a cross-subject approach. The results showed that there is no statistically significant difference in classification performance between ME and MI. However, it exhibited a much higher performance than the other estimated models. The classification performance was divided into two tasks: (i) DR and (ii) MA, and it was found that both-handed classification showed a horizontally divided pattern, while the classification of the injury side showed a vertically divided pattern. In this respect, it is important to classify the lesion location by changing the perspective of stroke patients. In addition, since patients have different levels of motor rehabilitation in AH and UAH depending on the degree of motor damage, it is significant to classify them by stroke lesion directions in the current state of BCI. This makes it possible to provide effective rehabilitation training depending on the lesion location. These frameworks could help with neurorehabilitation through BCI in stroke patients and the detection of the lesion location using EEG signals. Furthermore, these findings could provide insights into the mechanisms underlying neural plasticity and neurorehabilitation after a stroke.

References

[1] M. Lee, J.-H. Jeong, Y.-H. Kim, and S.-W. Lee, "Decoding finger tapping with the affected hand in chronic stroke patients during motor imagery and execution," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 29, pp. 1099–1109, 2021.

[2] J.-H. Jeong, J.-H. Cho, B.-H. Lee, and S.-W. Lee, "Real-time deep neurolinguistic learning enhances noninvasive neural language decoding for brain–machine interaction," *IEEE Trans. Cybern.*, vol. 53, no. 12, pp. 7469–7482, Dec. 2023.

[3] L. Chen et al., "Adaptive asynchronous control system of robotic arm based on augmented reality-assisted brain–computer interface," *J. Neural Eng.*, vol. 18, no. 6, Dec. 2021, Art. no. 066005.

[4] A. M. Ladda, F. Lebon, and M. Lotze, "Using motor imagery practice for improving motor performance—A review," *Brain Cognition*, vol. 150, Jun. 2021, Art. no. 105705.

[5] R. Mane, T. Chouhan, and C. Guan, "BCI for stroke rehabilitation: Motor and beyond," *J. Neural Eng.*, vol. 17, no. 4, Aug. 2020, Art. no. 041001.

[6] W. Park, G. H. Kwon, Y.-H. Kim, J.-H. Lee, and L. Kim, "EEG response varies with lesion location in patients with chronic stroke," *J. NeuroEng. Rehabil.*, vol. 13, no. 1, pp. 1–10, Dec. 2016.

[7] M. Lee, Y.-H. Kim, and S.-W. Lee, "Motor impairment in stroke patients is associated with network properties during consecutive motor imagery," *IEEE Trans. Biomed. Eng.*, vol. 69, no. 8, pp. 2604–2615, Aug. 2022.

[8] R. Mane et al., "Prognostic and monitory EEG-biomarkers for BCI upper-limb stroke rehabilitation," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 27, no. 8, pp. 1654–1664, Aug. 2019.

[9] V. K. Benzy, A. P. Vinod, R. Subasree, S. Alladi, and K. Raghavendra, "Motor imagery hand movement direction decoding using brain computer interface to aid stroke recovery and rehabilitation," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 28, no. 12, pp. 3051–3062, Dec. 2020.

[10] J. Jiang, C. Wang, J. Wu, W. Qin, M. Xu, and E. Yin, "Temporal combination pattern optimization based on feature selection method for motor imagery BCIs," *Frontiers Hum. Neurosci.*, vol. 14, p. 231, Jun. 2020.

[11] Y. Pei et al., "A tensor-based frequency features combination method for brain–computer interfaces," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 30, pp. 465–475, 2022.

[12] Y. Yang, C. Ye, X. Guo, T. Wu, Y. Xiang, and T. Ma, "Mapping multimodal brain connectome for brain disorder diagnosis via cross-modal mutual learning," *IEEE Trans. Med. Imag.*, vol. 43, no. 1, pp. 108–121, Jan. 2024.

[13] K. X. Khor et al., "Portable and reconfigurable wrist robot improves hand function for post-stroke subjects," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 25, no. 10, pp. 1864–1873, Oct. 2017.

[14] M. A. Dimyan and L. G. Cohen, "Neuroplasticity in the context of motor rehabilitation after stroke," *Nat. Rev. Neurol.*, vol. 7, no. 2, pp. 76–85, 2011.

[15] Y. Zhang and Q. Yang, "A survey on multi-task learning," *IEEE Trans. Knowl. Data Eng.*, vol. 34, no. 12, pp. 5586–5609, Mar. 2021.

[16] Y. Zhang et al., "Improving EEG decoding via clustering-based multitask feature learning," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 33, no. 8, pp. 3587–3597, Aug. 2022.

[17] X. Liu, L. Lv, Y. Shen, P. Xiong, J. Yang, and J. Liu, "Multiscale space-time-frequency feature-guided multitask learning CNN for motor imagery EEG classification," *J. Neural Eng.*, vol. 18, no. 2, Apr. 2021, Art. no. 026003.

[18] Q. Zheng, Y. Wang, and P. A. Heng, "Multitask feature learning meets robust tensor decomposition for EEG classification," *IEEE Trans. Cybern.*, vol. 51, no. 4, pp. 2242–2252, Apr. 2021.

[19] J.-H. Jeong, N.-S. Kwak, C. Guan, and S.-W. Lee, "Decoding movement-related cortical potentials based on subject-dependent and section-wise spectral filtering," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 28, no. 3, pp. 687–698, Mar. 2020.

[20] S.-B. Lee, H.-J. Kim, H. Kim, J.-H. Jeong, S.-W. Lee, and D.-J. Kim, "Comparative analysis of features extracted from EEG spatial, spectral and temporal domains for binary and multiclass motor imagery classification," *Inf. Sci.*, vol. 502, pp. 190–200, Oct. 2019.

[21] S. Saha, K. I. U. Ahmed, R. Mostafa, L. Hadjileontiadis, and A. Khandoker, "Evidence of variabilities in EEG dynamics during motor imagery-based multiclass brain–computer interface," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 26, no. 2, pp. 371–382, Feb. 2018.

[22] K. Mahjoory, V. V. Nikulin, L. Botrel, K. Linkenkaer-Hansen, M. M. Fato, and S. Haufe, "Consistency of EEG source localization and connectivity estimates," *NeuroImage*, vol. 152, pp. 590–601, May 2017.

[23] M. Lee, J.-G. Yoon, and S.-W. Lee, "Predicting motor imagery performance from resting-state EEG using dynamic causal modeling," *Frontiers Hum. Neurosci.*, vol. 14, p. 321, Aug. 2020.

[24] S. Pérez-Velasco, E. Santamaría-Vázquez, V. Martínez-Cagigal, D. Marcos-Martínez, and R. Hornero, "EEGSym: Overcoming inter-subject variability in motor imagery based BCIs with deep learning," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 30, pp. 1766–1775, 2022.

[25] P. Autthasan et al., "MIN2Net: End-to-end multi-task learning for subject-independent motor imagery EEG classification," *IEEE Trans. Biomed. Eng.*, vol. 69, no. 6, pp. 2105–2118, Jun. 2022.

[26] X. Liu, S. Xiong, X. Wang, T. Liang, H. Wang, and X. Liu, "A compact multi-branch 1D convolutional neural network for EEG-based motor imagery classification," *Biomed. Signal Process. Control*, vol. 81, Mar. 2023, Art. no. 104456.

[27] H. Raza, A. Chowdhury, and S. Bhattacharyya, "Deep learning based prediction of EEG motor imagery of stroke patients' for neuro-rehabilitation application," in *Proc. Int. Joint Conf. Neural Netw. (IJCNN)*, Jul. 2020, pp. 1–8.

[28] V. Lawhern, A. Solon, N. Waytowich, S. M. Gordon, C. Hung, and B. J. Lance, "EEGNet: A compact convolutional neural network for EEG-based brain–computer interfaces," *J. Neural Eng.*, vol. 15, no. 5, 2018, Art. no. 056013.

[29] J.-H. Jeong et al., "Multimodal signal dataset for 11 intuitive movement tasks from single upper extremity during multiple recording sessions," *GigaScience*, vol. 9, no. 10, Oct. 2020, Art. no. giaa098.

[30] M. Lee et al., "Motor imagery learning across a sequence of trials in stroke patients," *Restorative Neurol. Neurosci.*, vol. 34, no. 4, pp. 635–645, Aug. 2016.

[31] D. W. Molloy and T. I. M. Standish, "A guide to the standardized mini-mental state examination," *Int. Psychogeriatrics*, vol. 9, no. S1, pp. 87–94, Dec. 1997.

[32] L. Formstone, W. Huo, S. Wilson, A. McGregor, P. Bentley, and R. Vaidyanathan, "Quantification of motor function post-stroke using novel combination of wearable inertial and mechanomyographic sensors," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 29, pp. 1158–1167, 2021.

[33] A. Delorme and S. Makeig, "EEGLAB: An open source toolbox for analysis of single-trial EEG dynamics including independent component analysis," *J. Neurosci. Methods*, vol. 134, no. 1, pp. 9–21, Mar. 2004.

[34] K. Keng Ang, Z. Yang Chin, H. Zhang, and C. Guan, "Filter bank common spatial pattern (FBCSP) in brain-computer interface," in *Proc. IEEE Int. Joint Conf. Neural Netw.*, Jun. 2008, pp. 2390–2397.

[35] K. K. Ang, Z. Y. Chin, C. Wang, C. Guan, and H. Zhang, "Filter bank common spatial pattern algorithm on BCI competition IV datasets 2a and 2b," *Frontiers Neurosci.*, vol. 6, p. 39, Mar. 2012.

[36] B. Blankertz, R. Tomioka, S. Lemm, M. Kawanabe, and K.-R. Müller, "Optimizing spatial filters for robust EEG single-trial analysis," *IEEE Signal Process. Mag.*, vol. 25, no. 1, pp. 41–56, Dec. 2007.

[37] F. J. Ferri, P. Pudil, M. Hatef, and J. Kittler, "Comparative study of techniques for large-scale feature selection," in *Machine Intelligence and Pattern Recognition*, vol. 16. Amsterdam, The Netherlands: Elsevier, 1994, pp. 403–413.

[38] P. Pudil, J. Novovičová, and J. Kittler, "Floating search methods in feature selection," *Pattern Recognit. Lett.*, vol. 15, no. 11, pp. 1119–1125, Nov. 1994.

[39] Z. Qiu, J. Jin, H.-K. Lam, Y. Zhang, X. Wang, and A. Cichocki, "Improved SFFS method for channel selection in motor imagery based BCI," *Neurocomputing*, vol. 207, pp. 519–527, Sep. 2016.

[40] H. Yang, S. Huang, S. Guo, and G. Sun, "Multi-classifier fusion based on MI–SFFS for cross-subject emotion recognition," *Entropy*, vol. 24, no. 5, p. 705, May 2022.

[41] J. Zhuang, K. Geng, and G. Yin, "Ensemble learning based brain–computer interface system for ground vehicle control," *IEEE Trans. Syst., Man, Cybern., Syst.*, vol. 51, no. 9, pp. 5392–5404, Sep. 2021.

[42] A. Craik, Y. He, and J. L. Contreras-Vidal, "Deep learning for electroencephalogram (EEG) classification tasks: A review," *J. Neural Eng.*, vol. 16, no. 3, Jun. 2019, Art. no. 031001.

[43] H. U. Amin, W. Mumtaz, A. R. Subhani, M. N. M. Saad, and A. S. Malik, "Classification of EEG signals based on pattern recognition approach," *Frontiers Comput. Neurosci.*, vol. 11, p. 103, Nov. 2017.

[44] F. Lotte et al., "A review of classification algorithms for EEG-based brain–computer interfaces: A 10 year update," *J. Neural Eng.*, vol. 15, no. 3, Jun. 2018, Art. no. 031005.

[45] M. Mousavi, L. R. Krol, and V. R. de Sa, "Hybrid brain-computer interface with motor imagery and error-related brain activity," *J. Neural Eng.*, vol. 17, no. 5, Oct. 2020, Art. no. 056041.

[46] O. Aydemir and T. Kayikcioglu, "Decision tree structure based classification of EEG signals recorded during two dimensional cursor movement imagery," *J. Neurosci. Methods*, vol. 229, pp. 68–75, May 2014.

[47] J. Shin, "Random subspace ensemble learning for functional near-infrared spectroscopy brain-computer interfaces," *Frontiers Hum. Neurosci.*, vol. 14, p. 236, Jul. 2020.

[48] J. N. van Rijn, G. Holmes, B. Pfahringer, and J. Vanschoren, "The online performance estimation framework: Heterogeneous ensemble learning for data streams," *Mach. Learn.*, vol. 107, no. 1, pp. 149–176, 2018.

[49] S. Siuly, Y. Li, and Y. Zhang, *EEG Signal Analysis and Classification: Techniques and Applications*. Berlin, Germany: Springer, 2017.

[50] F. Fahimi, Z. Zhang, W. B. Goh, T.-S. Lee, K. K. Ang, and C. Guan, "Inter-subject transfer learning with an end-to-end deep convolutional neural network for EEG-based BCI," *J. Neural Eng.*, vol. 16, no. 2, Apr. 2019, Art. no. 026007.

[51] R. T. Schirrmeister et al., "Deep learning with convolutional neural networks for EEG decoding and visualization," *Hum. Brain Mapp.*, vol. 38, no. 11, pp. 5391–5420, 2017.

[52] S. B. Lim, S. Peters, C.-L. Yang, L. A. Boyd, T. Liu-Ambrose, and J. J. Eng, "Frontal, sensorimotor, and posterior parietal regions are involved in dual-task walking after stroke," *Frontiers Neurol.*, vol. 13, Jun. 2022, Art. no. 904145.

[53] J. S. Siegel et al., "Re-emergence of modular brain networks in stroke recovery," *Cortex*, vol. 101, pp. 44–59, Apr. 2018.

[54] Y. Pei et al., "Data augmentation: Using channel-level recombination to improve classification performance for motor imagery EEG," *Frontiers Hum. Neurosci.*, vol. 15, Mar. 2021, Art. no. 645952.