Transferring a deep learning model from healthy subjects to stroke patients in a motor imagery brain-computer interface

To cite this article before publication: Aarthy Nagarajan *et al* 2023 *J. Neural Eng.* in press https://doi.org/10.1088/1741-2552/ad152f

Transferring a Deep Learning Model from Healthy Subjects to Stroke Patients in a Motor Imagery Brain-Computer Interface

Aarthy Nagarajan[1], Neethu Robinson[1], Kai Keng Ang[2], Karen Sui Geok Chua[3], Effie Chew[4], Cuntai Guan[1]

[1] School of Computer Science and Engineering, Nanyang Technological University, 50 Nanyang Ave, Singapore 639798, Singapore

[2] Institute for Infocomm and Research, Agency of Science, Technology and Research (A*STAR), 1 Fusionopolis Way, Singapore 138632, Singapore

[3] Department of Rehabilitation Medicine, Tan Tock Seng Hospital, 11 Jln Tan Tock Seng, Singapore 308433, Singapore

[4] National University Health System, 1E Kent Ridge Road, Singapore 119228, Singapore

E-mail: aarthy001@e.ntu.edu.sg, nrobinson@ntu.edu.sg, kkang@i2r.a-star.edu.sg, Karen_Chua@ttsh.com.sg, Effie_CHEW@nuhs.edu.sg, ctguan@ntu.edu.sg

November 2023

Abstract.

Objective: Motor imagery (MI) brain-computer interfaces (BCI) based on electroencephalogram (EEG) have been developed primarily for stroke rehabilitation, however, due to limited stroke data, current deep learning methods for cross-subject classification rely on healthy data. This study aims to assess the feasibility of applying MI-BCI models pre-trained using data from healthy individuals to detect MI in stroke patients. **Approach:** We introduce a new transfer learning approach where features from two-class MI data of healthy individuals are used to detect MI in stroke patients. We compare the results of the proposed method with those obtained from analyses within stroke data. Experiments were conducted using Deep ConvNet and state-of-the-art subject-specific machine learning MI classifiers, evaluated on OpenBMI two-class MI-EEG data from healthy subjects and two-class MI versus rest data from stroke patients. **Main Results:** Results of our study indicate that through domain adaptation of a model pre-trained using healthy subjects' data, an average MI detection accuracy of 71.15% ($\pm$12.46%) can be achieved across 71 stroke patients. We demonstrate that the accuracy of the pre-trained model increased by 18.15% after transfer learning ($p < 0.001$). Additionally, the proposed transfer learning method outperforms the subject-specific results achieved by Deep ConvNet and FBCSP, with significant enhancements of 7.64% ($p < 0.001$) and 5.55% ($p < 0.001$) in performance, respectively. Notably, the healthy-to-stroke transfer learning approach achieved similar performance to stroke-to-stroke transfer learning, with no significant difference ($p > 0.05$). Explainable AI analyses using transfer models determined channel relevance patterns that indicate contributions from the bilateral motor, frontal, and parietal regions of the cortex towards MI detection in stroke patients. **Significance:** Transfer

learning from healthy to stroke can enhance the clinical use of BCI algorithms by overcoming the challenge of insufficient clinical data for optimal training.

Keywords: MI-BCI for stroke rehabilitation, transfer learning, healthy-to-stroke transfer, explainable AI.

1. Introduction

The diverse technological landscape of motor imagery (MI) based electroencephalogram (EEG) brain-computer interfaces (BCI) has garnered significant attention from both academic [1] and industry stakeholders [2]. The development and progress of EEG-BCI systems necessitate a multidisciplinary approach, drawing on expertise from fields such as neuroscience [3], psychology [4], computer science [5], biomedical [6] and rehabilitative engineering [7], among others. The MI-BCI also holds the potential for versatile applications across various user groups, including both healthy individuals [8] and patient populations [9]. However, the predominant focus of MI-BCI has been in the medical domain, notably stroke rehabilitation [10]. Though several clinical trials have been performed using MI-BCI for upper limb stroke rehabilitation [11] [12], the computational studies involving stroke data are limited.

MI-BCIs, which entail the mental imagination of hand movements, may involve real-time feedback and can operate in either online or offline models. A traditional MI-BCI protocol typically incorporates closed-loop training to enhance its effectiveness. In this protocol, the user is instructed to mentally imagine specific limb movements, such as moving their impaired or target hand or foot, based on auditory or visual cues presented to them. EEG or other neural signals are recorded during these mental imagery tasks. These neural signals are then decoded by BCI algorithms to identify patterns associated with the imagined movements. Importantly, the system provides real-time feedback to the user based on the decoded neural activity. This feedback allows the user to adjust their mental imagery strategy. As the user refines their mental imagery and the BCI system adapts to their unique neural patterns, a closed-loop training loop is established [13]. Several research studies [14] [15] [16] [17] have demonstrated the effectiveness of a comprehensive mutual learning approach in MI-BCI training, incorporating machine, subject, and application levels. These insights were derived from participation in the Cybathlon BCI race [18].

Conventional MI-BCI algorithms were developed using machine learning (ML) techniques such as common spatial patterns (CSP) [19] and filter-bank CSP (FBCSP) [20] for feature extraction. These features were then combined with classifiers like linear discriminant analysis (LDA) [21] or random forest (RF) [22] to perform MI classification. However, recent MI classifiers have shifted towards deep learning (DL),

which leverages a multi-layered neural network structure to enable nonlinear end-to-end learning. In [23], Sakhavi et al. proposed one of the first DL methods for MI classification using a convolutional neural network (CNN). Nonetheless, they first create an envelope representation of the input EEG based on the FBCSP method, which is then used as an input to the CNN. In contrast, other state-of-the-art (SOA) models, such as Deep ConvNet [24] and EEGNet [25], utilize raw EEG to classify MI, thus removing the need for feature engineering. The advent of deep neural networks (DNNs) in MI-BCI has piqued the interest of BCI researchers, leading to several investigations into their functioning [26][27]. These DL algorithms have demonstrated promising results in classifying MI using EEG data, more so with subject-independent modeling [28][27], where networks pre-trained using numerous multi-subject EEG trials are evaluated on a new target subject.

MI-BCIs can be conducted using different methods of brain data collection such as invasive methods involving direct placement of electrodes within the brain, partially invasive methods such as electrocorticography (ECoG) [29], and non-invasive methods such as EEG [7], magnetoencephalography (MEG) [30], functional magnetic resonance imaging (fMRI) [31], and functional near-infrared spectroscopy (fNIRS) [32][33][34]. Among the non-invasive methods, EEG is the most commonly utilized for acquiring brain signals due to its high temporal resolution, user-friendliness, and cost-effectiveness. Despite the convenience and cost-effectiveness of using non-invasive scalp sensors to record MI-EEG data, prolonged experimentation can cause fatigue and attention deficits in users [35], particularly in stroke patients. Consequently, most EEG-BCI computational research involving subject-independent modeling has been carried out on data from healthy participants [36][37][28]. The few studies involving stroke data have focused primarily on subject-specific analyses using a small training set [38][39][40]. Furthermore, these studies have predominantly utilized ML-based classification methods.

Chowdhury et al. [39] developed an adaptive classifier that employs online covariate shift detection to address subject-specific EEG nonstationarity. The authors utilized CSP for spatial filtering and support vector machines (SVM) for classification. They demonstrated performance improvement with adaptation by retraining the classifier. The method was evaluated separately for a healthy group and a stroke group, each consisting of only ten participants. Similarly, Irimia et al. [40] compared MI-BCI control performance between healthy subjects and stroke patients using CSP for feature extraction and LDA for classification. Their results were based on data collected from experimental sessions ranging from 10 to 24 across five participants. Their research findings indicated that individuals who had experienced a stroke demonstrated the ability to operate an MI-BCI with an accuracy level comparable to that of healthy participants. Furthermore, it was observed that stroke patients exhibited improvements in their motor function through BCI control, irrespective of their performance level.

Some recent studies have used DL to explore stroke MI-EEG data. For example, Cheng et al. [38] utilized a DNN to extract spatial and spectral patterns to uncover

plasticity mechanisms in the impaired cortices of stroke patients. Their results are based on data collected from five stroke patients over two months, and are subject-specific. Previous research has examined SOA MI classifiers based on DL, such as the Deep ConvNet, for subject-independent classification using data from healthy individuals, and has demonstrated promising outcomes [28][27]. However, Raza et al. recently evaluated the efficacy of EEGNet for cross-subject classification using data from 10 hemiparetic stroke patients and found no significant difference between cross-subject and within-subject classification [41]. Their results emphasize the need for adaptation when working with stroke data, due to the high variability in this population caused by the damage sustained in the motor areas. While there have been promising results for MI-BCI control among stroke patients in existing literature [40][42], there is a dearth of studies exploring generic models for the stroke group, except for a few that are conducted using small databases. This is likely due to the scarcity of an adequate MI database that contains a sufficient number of stroke patient trials to develop robust subject-independent models.

Research on BCI for upper limb stroke rehabilitation has revealed that there is motor activity in the contralesional hemisphere during stroke MI, similar to what is seen in healthy individuals. Ang et al. [42] performed a clinical study to investigate and compare the MI-related spatial patterns of hemiparetic stroke patients with that of healthy subjects, using CSP. Their findings illustrated comparable patterns between the two groups, which focused on bilateral hemispheres. Shu et al. conducted a study involving 24 stroke patients and 10 healthy individuals, in which they proposed two physiological variables - laterality index (LI) and cortical activation strength (CAS) - based on event-related spectral perturbation (ERSP) to predict the performance of MI-BCI in stroke patients [43]. The analysis of spectral power changes using ERSP revealed cortical activations in the contralesional hemisphere during paretic hand MI. In a recent publication [44], Mansour et al. proposed the use of contralesional BCI as an effective approach for patients with high motor impairment. The similarity observed in motor activity between stroke patients and healthy individuals, particularly in the contralesional hemisphere of stroke patients, provides a basis for the potential transfer of knowledge acquired from healthy individuals to improve the effectiveness of BCI for upper limb rehabilitation purposes.

We attempt to address this through the following main contributions.

(i) Our novel proposal suggests utilizing transfer learning from healthy individuals to stroke patients as a means to address the issue of limited availability of stroke data.

(ii) We propose an experimental approach to transfer a generic model, developed using left-hand (LH)/right-hand (RH) MI data obtained from healthy subjects, for detecting MI in stroke patients relative to their affected limb.

(iii) A novel channel swapping method is introduced to pre-train effective cross-subject models tailored to the affected limb of the stroke patient.

(iv) For the same set of target stroke patients, we compare the transfer learning performance of healthy-to-stroke (H-to-S) models with that of stroke-to-stroke (S-to-S) models and with the subject-specific performance.

(v) We assess the efficacy of our proposed approach using Deep ConvNet [24], which is a SOA MI-BCI model. In addition, we conduct subject-specific classification using established ML models, specifically CSP+LDA, CSP+RF, FBCSP+LDA, and FBCPS+RF. To obtain our experimental results, we used the two-class MI data collected from 54 healthy subjects in the OpenBMI dataset [45], as well as the MI versus rest data collected from 71 chronic stroke patients by Ang et al. in [11] and [12].

(vi) We compare the spatial relevance patterns in stroke patient groups using explainable artificial intelligence (XAI) analyses, performed with S-to-S and H-to-S models. To conduct the relevance analyses, we used DeepLift, an XAI framework presented by Shrikumar et al. in [46].

(vii) We compare the S-to-S and H-to-S models by analyzing their adaptation speed and data requirements. Specifically, we measure the number of epochs and the amount of adaptation data necessary to achieve optimal accuracy in each model.

The paper is organized as follows. Section 2 presents related work. In Section 3, we introduce our proposed method for transfer learning from healthy to stroke, provide architectural details of the Deep ConvNet and SOA ML models, and details of model training, optimization, and adaptation settings. Section 4 describes the two datasets used, the conventional subject-specific and subject-independent stroke (S-to-S) classification, the flow of our transfer learning experiments, and additional analyses. We report and discuss the results in Sections 5 and 6, before concluding in Section 7.

2. Related Work

In recent years, there has been a growing number of studies that examine cross-subject models in the field of MI-BCI. This is attributed to the superior predictive power of cross-subject models based on DL, compared to subject-specific models. The former have been trained on a large number of trials from multiple subjects, leading to their ability to perform well on new subjects without the need for extensive calibration time. Despite these benefits, most of the research in this area has been limited to healthy subjects, with only a few studies using data from a limited number of stroke patients.

2.1. Transfer learning in healthy subjects

Several studies have explored the use of transfer learning in MI-BCI, with a focus on healthy subjects. One approach involves a combination of ML models and data or classifier alignment methods. For instance, Azab et al. [36] introduced a weighted transfer learning method that utilizes previously recorded EEG data from multiple subjects and a few trials from the target subject to train a classifier. They incorporated

regularization into the classifier objective function to ensure that the parameters learned for a new subject closely resemble those of similar subjects. The effectiveness of their approach was demonstrated by testing it on BCI competition datasets IV 2a and III IVa, as well as a small dataset consisting of six stroke patients and sixteen healthy individuals. The results showed that the approach was particularly effective for low-performing subjects.

Zheng et al. investigated the transferability of CSP features learned from data collected using traditional MI commands to new commands [47]. Although their approach showed promising results, it was derived from data collected from only five healthy subjects. Xu et al. [37] proposed a transfer learning approach based on the Riemannian tangent space and the Riemannian alignment, which utilizes labelled samples from both the source and the target subjects. Their approach was tested on a dataset of 52 healthy subjects and a BCI competition dataset, achieving high accuracy, particularly for targets with fewer or no labelled samples. Finally, another transfer learning method for MI-BCI based on the Riemannian framework was introduced by He et al. [48]. They proposed label alignment, which aligns the EEG covariance matrices of the source and target subjects by aligning their label sets. By evaluating their method using the BCI competition datasets IV 1 and 2a, they showed a significant improvement compared to the Euclidean alignment approach.

Recently, researchers have investigated transfer learning approaches for MI-BCI using DNNs and domain adaptation techniques. For instance, Wang et al. proposed an unsupervised domain adaptation method that aligns Euclidean space data to improve MI classification performance using CNNs [49]. However, they used CSP features instead of end-to-end learning, thereby defeating the purpose of using DNNs.

In [50] Begiełło et al. used CNNs to investigate the performance of cross-subject classification with and without transfer learning. They fine-tuned the weights of a model that was pre-trained on data from 105 subjects in the Physionet MI dataset, by utilizing target data gathered from novice BCI users. Their results showed a significant improvement after transfer learning.

Ju et al. [51] introduced a privacy-preserving DNN architecture for MI-EEG classification based on the federated learning framework. Their algorithm extracts common discriminative features from multi-subject EEG using a single-trial covariance matrix, and applies domain adaptation techniques to optimize both classification and domain loss. Their network consists of a manifold reduction layer, common embedded space, tangent projection layer, and a federated layer. Their method outperformed other SOA subspace, covariance, and DL methods when evaluated using the Physionet two-class MI dataset [52].

In [28], Zhang et al. explored five different subject adaptive transfer learning schemes using the Deep ConvNet MI-BCI model [24] and OpenBMI MI data [45], and the results were compared with subject-independent and subject-specific baselines. In each subject adaptive scheme, the authors performed the adaptation of a subset of model parameters using different amounts of adaptation data of the target subject at different

learning rates. The performance of the subject-independent framework trained using large amounts of multi-subject data was around 32% higher than that of the subject-specific framework. Nevertheless, the authors performed further fine-tuning using target subject's data to address the problem of inter-subject EEG variability.

2.2. Transfer learning in stroke patients

The number of studies exploring transfer learning using stroke data is limited. Cao et al. proposed a scheme to perform transfer calibration based on the performance of existing methods [53]. They evaluated intra- and inter-subject transfer learning calibrations using data from seven stroke patients and found that their scheme benefitted low-precision sessions the most. In [54], Xu et al. studied transfer learning with a combination of healthy and stroke data, collected from eleven healthy and five stroke patients. They used EEGNet for transfer learning based on fine-tuning and achieved an average accuracy of 66.36%. However, they did not investigate transfer learning between the healthy and stroke groups.

This study undertakes a distinct initiative to explore transfer learning from healthy subjects to stroke patients for the purpose of MI detection using DL models. Additionally, we compare the performances of S-to-S transfer and H-to-S transfer using two large-scale databases of MI trials from healthy subjects and stroke patients. Furthermore, we introduce XAI analyses to identify spatial relevance patterns in stroke patient groups using transfer-learned models.

3. Methodology

In this section, we outline our proposed strategy for transfer learning from healthy to stroke. In addition, we present the Deep ConvNet MI classifier used for evaluating our proposed method and the corresponding training parameters. We also introduce the ML-based classifiers for MI, used for performing subject-specific MI detection using stroke data.

3.1. Proposed Transfer Learning from Healthy to Stroke

Our approach for transfer learning from healthy to stroke involves pre-training models with MI data from healthy individuals, which can subsequently be utilized to identify MI versus rest state in stroke patients.

3.1.1. Model Pre-training using Data from Healthy Subjects: In order to pre-train the "healthy" model, the data from healthy subjects must align with the corresponding task from the target stroke dataset. The data is epoched and processed to achieve this, after which, all the trials are aggregated to create the training set for implementing the model. Additional information regarding this process is available in Section 4.

Channel Swapping. The dataset from healthy subjects contains MI trials performed with both the left and right hands. To create effective pre-trained models for stroke patients, it is crucial to leverage all available MI trials in the dataset. Since the brain signal is nonstationary and is subject to variability [55], a large amount of data is required to effectively train a robust pre-trained model for stroke patients. However, it is important to ensure that the pre-trained models for stroke patients are tailored to specific affected limbs. Prior research in the BCI literature has demonstrated that MI tasks corresponding to a specific limb are typically associated with contralateral event-related desynchronization (ERD) in the sensorimotor cortex [56] [57]. Therefore, before incorporating trials from a different limb into the training set, spatial patterns must be aligned with respect to the corresponding affected limb. This alignment is achieved through channel swapping, a process in which data from channels in the left and right hemispheres of all mismatched MI trials, along with their corresponding rest trials, are interchanged. Subsequently, these swapped trials are combined with the remaining matching trials to create the training set for developing the pre-trained model for stroke patients. Figures 1(a) and 1(b) illustrate the specifics of the channel swapping procedure schematically with respect to EEG data. By applying this channel swapping method, two distinct pre-trained models can be created using all the trials from healthy subjects: one model for patients with a left affected limb and another for patients with a right affected limb.

3.1.2. Domain Adaptation based Transfer Learning: In this section, the concept of transfer learning is briefly introduced through notations and definitions. Transfer learning is an idea that was inspired by the natural ability of humans to transfer knowledge to each other [58]. Transfer learning allows researchers to reuse pre-trained models constructed with existing data to tackle prediction tasks with new data, resulting in improved efficiency, generalizability and performance [59].

A MI-BCI domain, from a transfer learning perspective, is characterized using Equation 1. A domain in MI-BCI is defined as a feature space that comprises attributes extracted from MI-EEG signals of multiple subjects, denoted by $\mathbf{X}$, along with its corresponding marginal probability distribution represented by $\mathrm{P}(X)$, where $X \in \mathbf{X}$. Two MI-BCI domains belonging to different subject groups are deemed distinct if they have either dissimilar MI-EEG feature spaces ($\mathbf{X_t} \neq \mathbf{X_s}$) or divergent marginal probability distributions ($P(X_t) \neq P(X_s)$).

$$\mathcal{D} = \{\mathbf{X}, P(X)\} \tag{1}$$

An MI-BCI task, such as MI classification or MI detection, is defined by Equation 2. A task $\mathcal{T}$, which belongs to an MI-BCI domain $\mathcal{D}$, includes a label space $\mathbf{Y}$ that represents the various classes of the MI-BCI task and a predictive function $\mathbf{f}(.)$ that can identify the class of an input MI-EEG signal, based on the knowledge obtained from the training data.

$$\mathcal{T} = \{\mathbf{Y}, \mathbf{f}(.)\} \tag{2}$$

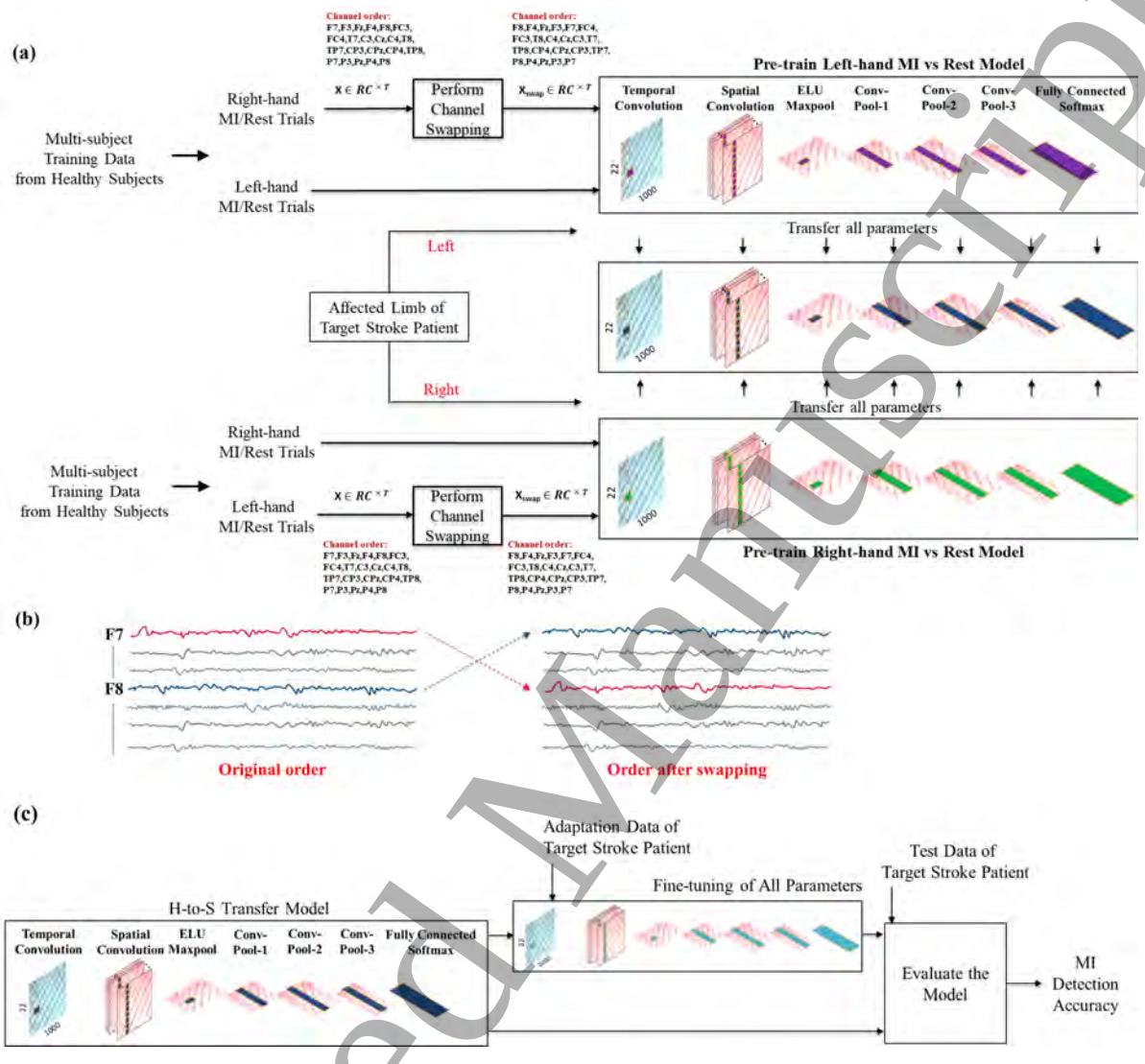

Figure 1: Pictorial representation of the proposed healthy-to-stroke (H-to-S) transfer learning method. Subfigure (a) demonstrates model pre-training using data from healthy subjects, based on the affected limb of the stroke patient. Subfigure (b) is an example of channel swapping between F7 and F8. Subfigure (c) demonstrates model adaptation and evaluation using data from the target patient.

Looking at it from a probabilistic standpoint, the predictive function $\mathbf{f}(.)$, also referred to as a "Model", can be defined as the probability of a specific MI class occurring given an MI-EEG feature space, represented as $P(Y \mid X)$. Therefore, the MI-BCI task $\mathcal{T}$ can be restated as Equation 3.

$$\mathcal{T} = \{\mathbf{Y}, P(Y \mid X)\} \tag{3}$$

The fundamental goal of transfer learning in MI-BCI is to enhance the learning of the conditional probability distribution $P(Y_t \mid X_t)$ in the target MI-BCI domain $\mathcal{D}_t$, by using the knowledge obtained from the corresponding MI-BCI task $\mathcal{T}_s$ in the source domain $\mathcal{D}_s$.

Inter-subject variabilities in BCI can negatively impact model performance due to disparities in feature spaces between the source and target domains. To mitigate this issue, domain adaptation based transfer learning methods are commonly utilized [51][28]. In this study, the differences in MI-related neurophysiological patterns between the healthy and patient groups of subjects exacerbate the inter-subject disparities. To address this, we applied domain adaptation, where the model pre-trained using healthy subjects' data was adapted using the target stroke patient's training set to enhance model performance. This involves knowledge transfer by sharing parameters from the source domain model to learn the target domain model. The entire procedure for the proposed H-to-S transfer learning is pictorially described in Figures 1(a), 1(b), and 1(c).

3.1.3. Model Evaluation using Data from Stroke Patients: During evaluation, one of the two models, pre-trained using data from healthy subjects, is chosen depending on the target patient's affected limb. We then evaluate the performance of the selected pre-trained model using the test set of the target patient. We further performed domain adaptation based transfer learning of the pre-trained model using the training data of the target stroke patient, and evaluated the adapted model. The resulting subject-wise and mean accuracies were recorded for further analysis.

3.2. Model Architecture

In this section, the architectures of the different SOA classifiers used in the experiments are discussed.

3.2.1. Architecture of State-of-the-art MI Classifiers: All experiments were conducted employing the following SOA ML and DL-based MI classifiers. The CSP technique [19], a widely adopted data-driven feature extraction approach, was utilized. CSP aims to spatially enhance EEG data by extracting the most discriminative features for classification. By calculating a spatial filtering matrix W using two-class data, CSP projects the EEG data into a discriminative subspace where one class exhibits maximum variance while minimizing variance from the other class. Four such spatial filters are derived for filtering. Subsequently, log-variance based features are extracted from the filtered data.

The FBCSP technique [20], a benchmark algorithm for EEG-BCI in MI classification and based on CSP [19], operates within the realm of ML. FBCSP effectively extracts features that are spectro-spatially relevant, and its efficacy has been demonstrated across multiple MI-BCI studies [60][11][12]. This methodology entails segmenting the initial EEG data into nine distinct narrow frequency bands, followed by subjecting these bands to spatial filtration through the CSP algorithm. Consequently, four notably distinctive CSP filters are derived from each band. The log-variance of the CSP-filtered EEG data serves as the feature. From a pool of 36 features, derived from the combination of 9 frequency bands with 4 features per band, the Mutual Information

Table 1: Deep ConvNet [24] Architecture and Training Parameters

Input:	(22,1000), where 22 = No. of channels, 1000 = No. of time samples			
Training parameters:	Learning rate - 0.001, No. of epochs - 200			
Block	**Layer**	**No. of Filters**	**Size**	**Output Shape**
1	Conv2D	25	(1,10)	(25,22,991)
	Conv2D	25	(22,1)	(25,1,991)
	BatchNorm2D ELU MaxPooling, Stride		 (1,3), (1,3)	 (25,1,330)
2	Dropout (p = 0.5) Conv2D BatchNorm2D ELU MaxPooling, Stride	50	 (1,10) (1,3), (1,3)	 (50,1,107)
3	Dropout (p = 0.5) Conv2D BatchNorm2D ELU MaxPooling, Stride	100	 (1,10) (1,3), (1,3)	 (100,1,32)
4	Dropout (p = 0.5) Conv2D BatchNorm2D ELU MaxPooling, Stride	200	 (1,10) (1,3), (1,3)	 (200,1,7)
5	Conv2D LogSoftmax	2	(1,7)	 (2,1)

based Best Individual Features with Parzen Window (MIBIFPW) algorithm [61] is employed to select the most informative four features. Subsequently, a support vector machine (SVM) classifier is trained utilizing the selected features to effectively categorize individual trials into one of two classes. For this classification task, an epsilon-support vector regression using a radial basis function kernel variant of SVM is adopted [62].

Due to the recent progress of deep learning [63], several neural network based EEG-BCI classification models have been introduced and are reported to be performing better than the machine learning counterparts [5]. In our study on MI-BCI, we showcase our innovative transfer learning approach from healthy to stroke, utilizing the Deep ConvNet - a top-performing CNN classifier for MI detection - developed by Schirrmeister et al. in [24]. The study conducted by Kaishuo et al. in [28] using the OpenBMI dataset indicated that the Deep ConvNet performed well in transfer learning among healthy subjects. The architecture of the Deep ConvNet comprises a temporal convolution and a spatial convolution, followed by max-pooling, three convolution-max-pooling blocks, and a fully-connected softmax classification layer. The Deep ConvNet incorporates the exponential linear unit (ELU) [64] as the activation function in every layer. We used

the Pytorch framework [65] to implement the Deep ConvNet model. Table 1 outlines the architectural specifics of the Deep ConvNet tailored to the datasets employed within this study. For subject-independent classification, only Deep ConvNet was employed, as DL models have demonstrated higher predictive power when trained with large amounts of data. This enables them to perform well on new subjects without requiring extensive calibration time. Subject-specific analyses were carried out using ML models as well as Deep ConvNet.

3.3. Training and Adaptation Parameters

The Deep ConvNet was trained using the Adam optimizer [66] and the negative log-likelihood loss function to update the model weights. In addition, Batch Normalization [67] and Dropout [68] were performed for each convolution-max-pooling block. The training was conducted for 200 epochs, and the final model was selected for evaluation and further analysis. The learning rate used during training was fixed at 0.001. These parameters were kept consistent throughout all training and adaptation experiments.

4. Experiments

This section presents an overview of the OpenBMI two-class MI-EEG dataset and the MI versus rest stroke dataset. We elaborate on the techniques employed for data preparation and division in our experiments. In addition, we provide the specifics of the conventional subject-specific and subject-independent classification within the stroke group. Furthermore, we delve into the supplementary experiments and analyses carried out to enhance our understanding of transfer learning from healthy to stroke. The computations were carried out using multiple GPU and CPU resource clusters provided by the School of Computer Science and Engineering at NTU and the National Supercomputing Centre in Singapore.

4.1. Datasets and Data Preparation

In this section, we provide a detailed description of the datasets used in our transfer learning experiments to obtain meaningful insights.

4.1.1. OpenBMI MI Dataset: The OpenBMI MI-EEG dataset, which is an open-source dataset published by Lee et al. in [45], was chosen as our dataset of healthy subjects. The OpenBMI dataset contains two-class MI data involving the left-hand and right-hand grasping tasks. The MI-EEG dataset from OpenBMI, which is approximately 60 GB in size, contains recordings from 54 healthy individuals, aged between 24 and 35 years, collected using 62 Ag/AgCl electrodes at a sampling frequency of 1000 Hz. Each subject provided EEG data from two sessions, with 200 MI trials per session, of which 100 trials belonged to each class. The sessions were divided into an offline phase (phase 1), for constructing the classifier, and an online test phase (phase 2) with visual

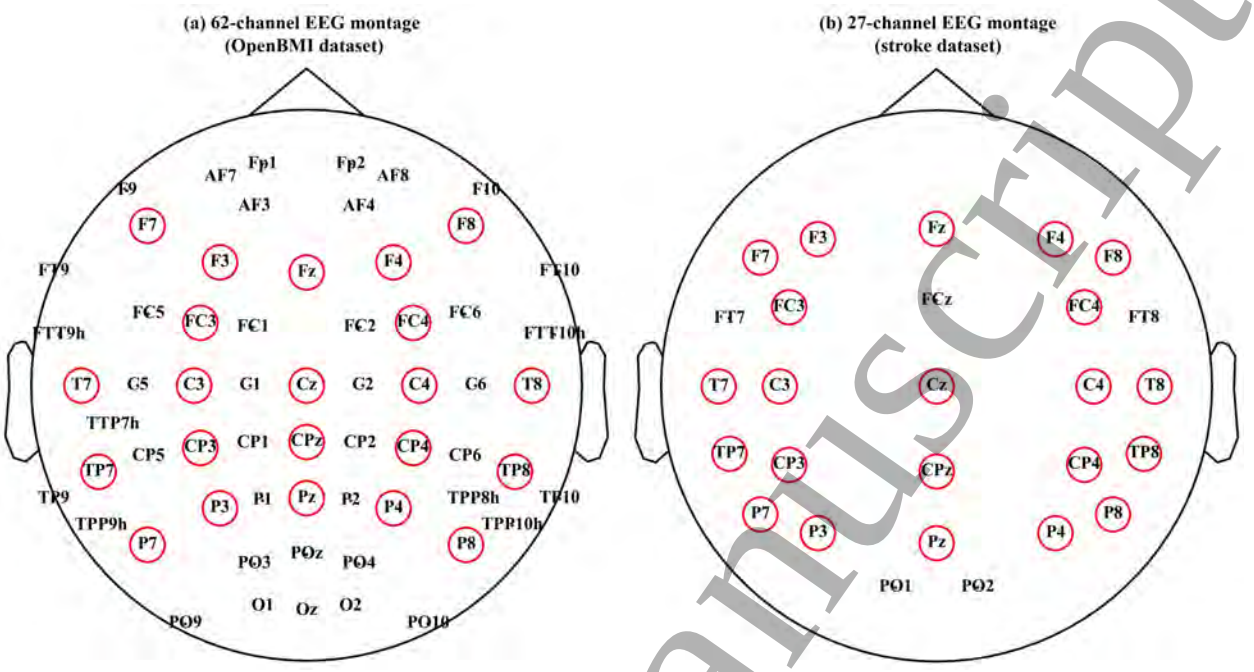

Figure 2: (a) The 62-channel EEG montage of the OpenBMI dataset, and (b) the 27-channel EEG montage of the stroke datasets. The 22 common channels between the two datasets are highlighted using red circles.

feedback. At the beginning of each trial, a fixation mark was displayed to prepare the subject, followed by a visual cue (a left or right arrow) for 4 seconds, during which the corresponding MI task was performed. The screen remained blank for approximately 6 seconds after each trial, during which the subject was allowed to rest. We used post-cue MI data from 0 to 4 seconds and downsampled it to 250 Hz for our experiments.

The rest-state trials are typically not included while implementing MI-EEG classifiers for healthy subjects. In the context of MI-BCI for healthy individuals, the classification of two distinct MI tasks is more prevalent than MI versus rest classification. Therefore, for the purpose of this study, we specially epoched rest-state segments from every MI trial in the dataset. More details regarding this procedure are provided in Section 4.1.4.

4.1.2. Stroke Dataset 1: The stroke dataset 1 was originally collected and documented by specific authors of this study and was featured in the publication by Ang et al. [12]. This dataset includes MI versus rest trials acquired from 37 stroke patients, aged between 21 and 70 years, during their MI-BCI screening session for post-stroke motor rehabilitation. The patients who were recruited had experienced their first subcortical stroke at least nine months prior to recruitment, with a moderate to severe impairment of upper extremity function as represented by a subscore of 11-45 on the Fugl-Meyer Motor Assessment (FMMA) [69]. The data consists of 160 4s trials for each patient, with 80 trials belonging to each class. These trials were collected across 4 runs, with each

Table 2: Data Preparation Details

Dataset	No. of subjects	MI Trials per subject	Rest trials per subject	Trials removed per subject	Channels selected	Total no. of trials
OpenBMI [45]	54	200 Left-hand 200 Right-hand	400	None	22 channels in common with stroke dataset	= 54*800 = 43,200
Stroke Dataset 1 [12]	37	80	80	4*	22 channels in common with OpenBMI dataset	= 37*156 = 5,772
Stroke Dataset 2 [11]	34	80	80	2*	22 channels in common with OpenBMI dataset	= 34*158 = 5,372

*Trials were removed due to data quality issues.

run comprising 40 trials. The EEG data was collected using 27 channels at a sampling frequency of 250 Hz and bandpass filtered by the acquisition hardware between 0.05-40 Hz. Out of the 37 stroke patients in this dataset, 20 of them experienced impairment in their left hand and the remaining 17 experienced impairment of their right hand as a result of the stroke.

4.1.3. Stroke Dataset 2: The stroke dataset 2, previously recorded by specific authors of this research in a study conducted by Ang et al. [11], is similar to stroke dataset 1, and is also a part of the post-stroke BCI motor rehabilitation screening session. The participants enrolled in this study, ranging in age from 21 to 80 years, had experienced their initial clinical stroke, which was verified through neuroimaging, at least four months prior to recruitment. These individuals had sustained moderate to severe impairment of upper extremity function, as evaluated by an FMMA score ranging from 10 to 50. The dataset consists of 160 MI of hand open-and-close versus rest trials collected from each of the 34 stroke patients, over 2 runs of 80 trials each. The data collection protocol is the same as that of stroke dataset 1. Among the 34 stroke patients, 20 of them experienced impairment in their left hand, while the remaining 14 experienced impairment in their right hand as a result of the stroke.

4.1.4. Data Preparation and Division: Since the purpose of this study is transfer learning from healthy to stroke, we used all the MI trials present in the OpenBMI dataset to pre-train the models. In addition, we required resting-state trials. Therefore, for each LH or RH MI trial in the dataset, we extracted the resting-state segment that occurs at the end of the trial. We ensured that the epoched rest segment had the same number of samples as the MI segment. Hence, for each subject in the dataset, after epoching the rest trials, we obtained a total of 800 trials, comprising LH MI and the corresponding rest trials and RH MI and the corresponding rest trials. The model was pre-trained using 400 MI trials and 400 rest trials each from the 54 healthy subjects. No validation data was used during training.

The combined stroke dataset, obtained by merging stroke datasets 1 and 2 from [11] and [12], was approximately 60 GB in size and included MI versus rest trials from 71 stroke patients. We noticed issues with data quality in the last trial of every run, during our preliminary analysis of both datasets. As a result, we excluded these trials, leaving each subject in stroke dataset 1 with 156 trials and each subject in stroke dataset 2 with 158 trials.

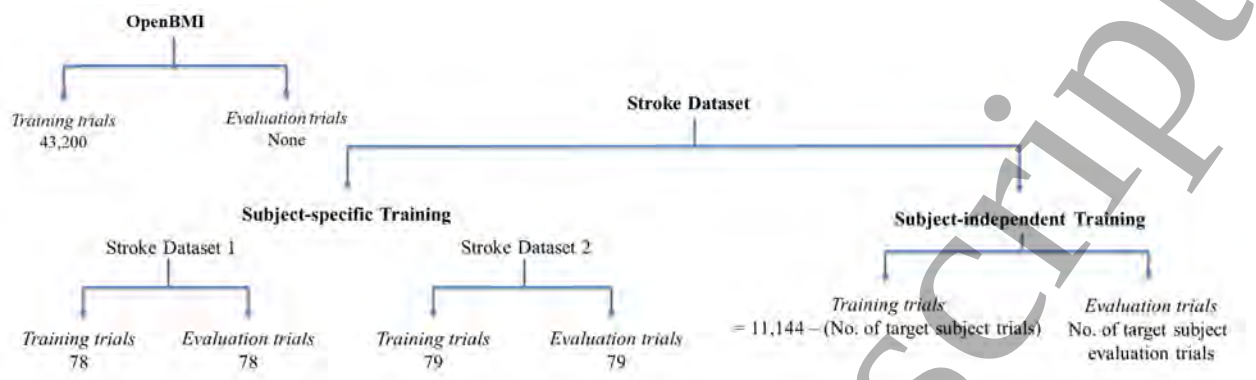

Figure 3: Schematic Illustration of Data Division

For each target patient, we set aside 50% of the data, i.e. 78 trials for patients from stroke dataset 1 and 79 for patients from stroke dataset 2, for training subject-specific models and for adaptation purposes. The remaining trials were kept exclusively for model evaluation. For training the subject-independent MI model for each target patient in the combined stroke dataset, we used all the trials from the remaining 70 stroke patients. No validation data was used for training.

The Deep ConvNet performs classification using raw EEG data. However, in the case of FBCSP, a prerequisite was the segmentation of EEG data into nine distinct frequency bands. Previous studies have established that the relevant MI-associated data within EEG primarily resides in spectrally confined ranges, specifically the mu (8-12 Hz), low beta (12-20 Hz), and high beta (22-32 Hz) bands [70]. Consider an individual raw EEG data trial, denoted as $x \in \mathbb{R}^{C\times T}$, where C denotes the count of EEG channels, T represents the time instances, and N_c represents the total number of unique classes. Each view in an FBCSP corresponds to a focused narrow-band EEG representation, achieved by employing spectral filtration on the initial EEG data x using a filter bank $F = f_{i\,i=1}^{N_b}$, comprising N_b temporal filters. This filtration procedure results in the localization of time series data along the third dimension of x into the spectral domain.

Therefore, we have the equation:

$$x_{FB} = F \otimes x \quad \in \mathbb{R}^{N_b\times C\times T} \tag{4}$$

Here, the symbol $\otimes$ denotes the operation of bandpass filtering.

The filter bank F can comprise an arbitrary number of filters with different cutoff frequencies. However, in line with [20], we applied a filter bank with $N_b = 9$ filters. Each filter had a bandwidth of 4Hz and non-overlapping frequency bands ranging from 4Hz to 40Hz (4-8, 8-12, ..., 36-40 Hz). The Chebyshev Type II filter was employed for filtering, with a transition bandwidth of 2Hz and a stopband ripple of 30dB. This selection of the filter bank was based on the conventional division of EEG into neurologically significant spectral bands proposed in the FBCSP algorithm. It has been demonstrated to achieve high classification accuracies across multiple studies [20][61].

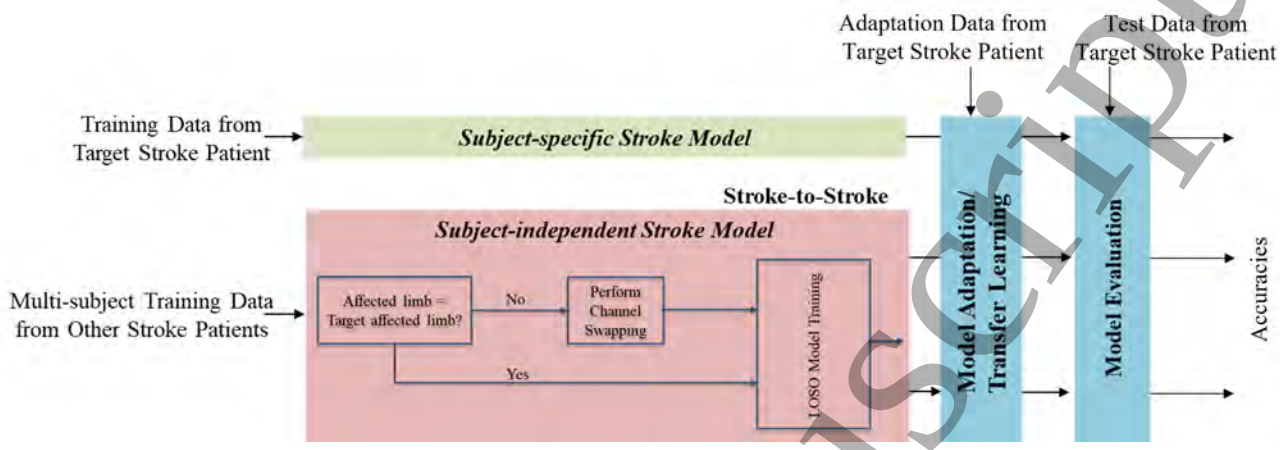

Figure 4: Illustration of our experimental procedure involving the stroke-to-stroke transfer learning and subject-specific classification. Two types of models were obtained from our experiments: (i) subject-specific model using the training data from target stroke patient, and, (ii) subject-independent stroke model using multi-subject training data from remaining stroke patients. The subject-independent models were evaluated with and without adaptation using data from the target patient.

Finally, to account for the differences in EEG channel layout between the OpenBMI dataset and the stroke dataset, we identified the common channels between the two datasets. Figure 2(a) illustrates the 62-channel EEG montage of the OpenBMI dataset, and Figure 2(b) illustrates the 27-channel montage of the stroke dataset. We used data from the 22 common channels between the two datasets, highlighted using red circles, to conduct all our experiments and analyses.

As both OpenBMI and the stroke datasets were referenced to the nasion, no additional preprocessing of the data was needed. The raw EEG data from both datasets were used in all our experiments. It is important to note that the test set of the target stroke patient remains consistent across all analyses to ensure a fair comparison. Table 2 and Figure 3 provide the specifics of the data preparation and division steps.

4.2. Classification within the Stroke Dataset

Here, we present two types of analyses conducted on the stroke dataset, which will be used as benchmarks for comparison with our proposed H-to-S transfer learning method. Figure 4 depicts the framework of these experiments, which comprises the training and evaluation of conventional subject-specific and subject-independent stroke models.

4.2.1. Transfer Learning from Stroke to Stroke: The first baseline for comparison with our proposed method is the S-to-S transfer performance. In this baseline approach, we train subject-independent stroke models for MI versus rest classification using the leave-one-subject-out cross-validation (LOSO-CV) method. We record subject-wise accuracies and the average accuracy across all patients for comparison purposes. We further adapt

each subject-independent stroke model using the training data of the target stroke patient, following a procedure similar to that used in the proposed H-to-S transfer learning. The results were compiled for further analysis.

To train the "S-to-S" model for each test/target patient, the training data includes all trials from the remaining patients in the stroke dataset. However, since stroke patients in our dataset have unilateral affected limbs, care should be taken while using trials of a training subject whose affected limb is not the same as the target affected limb. Thus, we interchange data from channels in the left and right hemispheres of all MI and the corresponding rest trials of patients with a different affected limb compared to the target. This is similar to the channel swapping procedure outlined in Section 3.1.1 for the proposed H-to-S transfer. Subsequently, we aggregated the trials of all subjects to develop the MI versus rest S-to-S model that is most suitable for evaluation using the test data of the target stroke subject.

4.2.2. Subject-specific Classification using Stroke Data: For comparison with our proposed transfer learning from healthy to stroke, an important baseline is the subject-specific model performance using stroke data. This also enables us to evaluate the subject-specific performance of DL compared to that of SOA ML methods. To create subject-specific models, we adopt the hold-out method, where a portion of the subject-specific data is reserved for testing and the remaining trials are used to train the model. We record subject-wise accuracies, as well as the average accuracy across all stroke patients, to compare with our proposed method.

4.3. Investigation on the Impact of Channel Swapping

Several experiments were conducted to examine the effects of channel swapping during S-to-S and H-to-S transfer. For this purpose, we created S-to-S and H-to-S transfer models without channel swapping and evaluated their performances for comparison with the proposed approach with channel swapping. Moreover, two different scenarios of using training data from healthy subjects, without channel swapping, to develop pre-trained models for H-to-S classification were also explored. Please note that we did not perform adaptation in this analysis, as the primary aim was to assess the impact of channel swapping on performance. Figure 5 is a visual depiction of H-to-S and S-to-S transfer without channel swapping.

4.3.1. Healthy-to-Stroke Transfer Without Channel Swapping We evaluated H-to-S transfer without channel swapping, under two different scenarios of usage of training data from healthy subjects.

Using Two-class MI Data from Healthy Subjects to Train One Model: During model pre-training using healthy subjects' data for the classification using stroke data, we proposed a technique in which we exchanged data from channels in the left and right

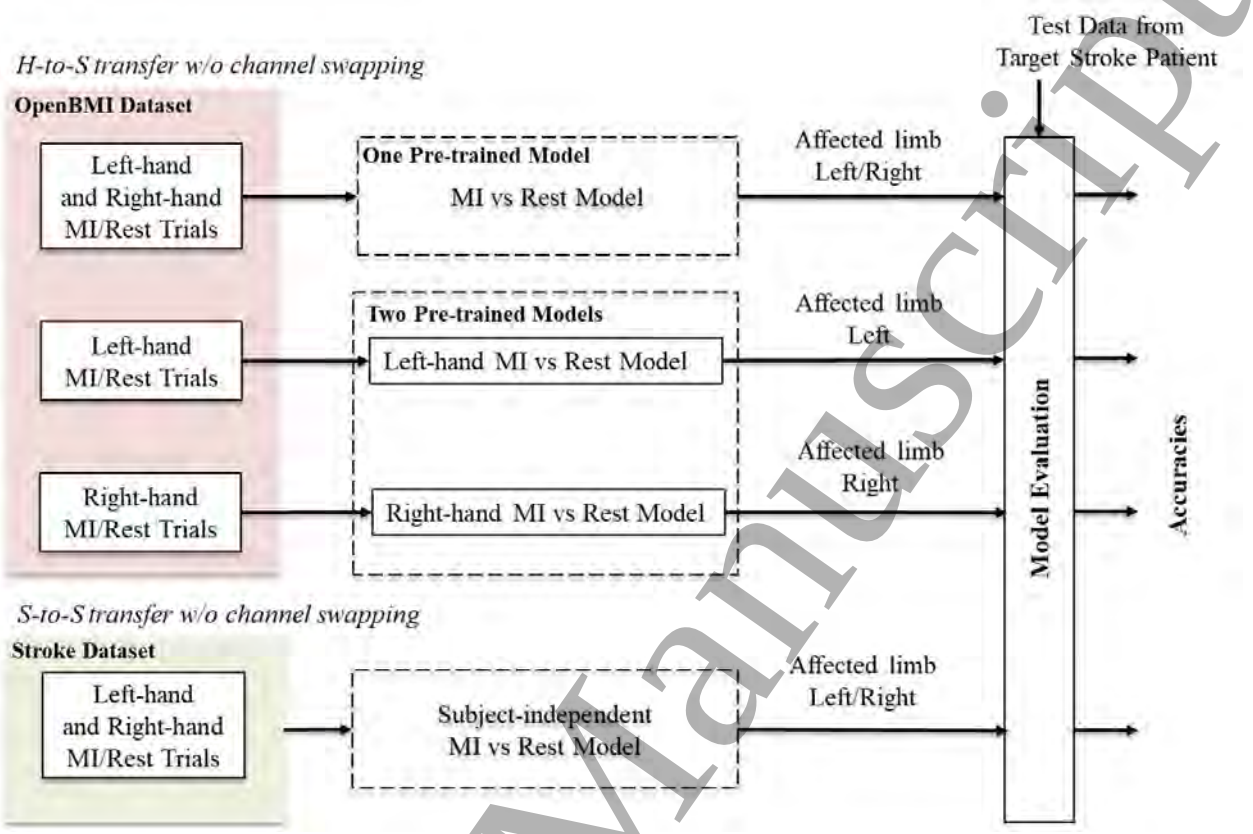

Figure 5: Illustration of Healthy-to-Stroke (H-to-S) and Stroke-to-Stroke (S-to-S) Transfer Without Channel Swapping.

hemisphere in MI trials that did not involve the target affected limb. This technique was motivated by the fact that two-class MI tasks in individuals often exhibit contrasting neural oscillatory patterns [56]. To demonstrate the necessity and significance of channel swapping in our pre-training strategy, we conducted a second analysis where we repeated the H-to-S classification experiments by using all trials without any channel swapping during the pre-training phase. That is, 800 trials of each of the 54 subjects are combined without any channel swapping to train a single MI versus rest classification model that is evaluated on stroke data. We compared the results of this analysis with those obtained from the proposed technique with channel swapping.

Using Unilateral MI Data from Healthy Subjects to Train Two Distinct Models: As part of our second investigation on healthy to stroke transfer, we trained two distinct models using data from healthy subjects. One model was trained on LH MI and the corresponding rest trials, while the other model was trained on RH MI and associated rest state trials. No channel swapping was required in this approach. During the evaluation step, one of the two models, pre-trained using healthy subjects' data, is selected based on the target patient's affected limb, similar to the proposed method. The MI versus rest classification results obtained from this analysis are compared with

those of our proposed method.

4.3.2. Stroke-to-Stroke Transfer Without Channel Swapping In our transfer learning baseline study, using subject-independent stroke model, we swapped the data from channels in the left and right hemispheres for those patients in the training set for whom the affected limb did not correspond to the target affected limb. We also performed a second analysis, where we repeated the LOSO-CV experiments using all trials from the stroke training set without any channel swapping to train a MI versus rest classification model. We compared the results of this analysis with those obtained from the proposed baseline for S-to-S transfer with channel swapping.

We refrained from conducting an S-to-S transfer analysis that uses unilateral MI data, due to an unequal number of patients with left versus right affected limbs. Selecting trials from the stroke training set based on the affected limb would have been impractical in our study and would have introduced bias in the subject-independent models due to different sizes of training data.

4.4. XAI based Interpretation of Transfer Learning from Stroke to Stroke and Healthy to Stroke

We conducted an XAI analysis using the DeepLift framework [46] to better understand the neurophysiological basis of decision-making in the transfer learned models. This analysis involved backpropagating the network's output to identify the contributions of EEG input from specific channels. The DeepLift method calculates relevance scores by comparing the activation of each neuron to its reference activation and assigning scores based on the resulting difference [46].

$$\sum_{i=1}^{n} C_{\Delta x_i \Delta t} = \Delta t \tag{5}$$

This approach is mathematically expressed as equation 5, which also satisfies the summation-to-delta property. Specifically, given a target output neuron t, a set of intermediate neurons required to compute t denoted as x_i, and the difference-from-reference denoted as Δt, DeepLift assigns relevance scores $C_{\Delta x_i \Delta t}$ to Δx_i, representing the amount of difference-from-reference in t that can be attributed to the difference-from-reference of x_i.

We used the test set of target stroke patients for the DeepLift analysis and computed the DeepLift relevance scores using the Captum XAI library [71].

4.4.1. Procedure: To calculate the channel-wise relevance scores, the first step involves creating a DeepLift criterion instance for the model. Then, individual trials are passed through the "attribute" function of the criterion to obtain class-based attributions. The attribute function carries out backward propagation of the output score through all layers of the model sequentially, and produces relevance scores for each neuron in the

underlying layers. The output attributions have the same dimensions as the EEG input trial (N_c, N_t), and are averaged across time samples to obtain a single score for each channel in the input. These relevance scores are averaged class-wise across all trials and further normalized between 0 and 1, for each subject.

The analysis was performed using subject-specific stroke models, as well as the S-to-S and H-to-S models, with and without adaptation. Please note that the S-to-S and H-to-S models utilized in this analysis are the actual baseline and proposed methods, described in Sections 4.2.1 and 3.1, respectively. The analysis was performed at three levels: the first level demonstrates average relevance patterns across all stroke patients, regardless of their affected limb, while the second level illustrates differences between the two groups of stroke patients associated with left and right affected limb, and the third one highlights individual differences in relevance patterns for some representative patients. To conduct the first two levels of analysis, the attribution scores were averaged across all patients or respective patient groups.

4.5. Comparing S-to-S versus H-to-S Transfer Learning in terms of Adaptation Speed and Adaptation Efficiency

To further compare the S-to-S and H-to-S transfer models, we conducted experiments to evaluate their transfer learning performance under varying amounts of adaptation data and training epochs. The model's adaptation efficiency was determined by measuring its performance with different amounts of adaptation data, while the speed of adaptation was determined by evaluating the model's performance for varying numbers of training epochs. We conducted the adaptation experiments for both models, by incrementally increasing the amount of adaptation data by 10%, while the number of epochs was fixed at 200. Furthermore, during the initial adaptation of the model using 100% adaptation data, detailed in Sections 3.1.3 and 4.2.1, we saved the model parameters at 20-epoch intervals. This allowed us to evaluate the performance of models trained for epochs ranging from 20-200, with an interval of 20.

5. Results

Our study utilized both Deep ConvNet and SOA ML classifiers for MI to obtain a range of results. Specifically, we present the average accuracies of subject-specific models, and S-to-S and H-to-S models, with and without adaptation. We also analyzed the subject-wise classification accuracies for all three types of models. In order to evaluate the statistical significance of improvements across all comparisons, a two-sided Wilcoxon signed-rank test was employed with a confidence level of 0.05. We would like to reiterate that in each analysis performed, every target stroke patient was evaluated using the same test set. Furthermore, we report the results from additional analyses using S-to-S and H-to-S models pre-trained without channel swapping. We provide visualizations of channel-wise relevances across all stroke patients and patient groups. Finally, we report

on the differences between the S-to-S and H-to-S models in terms of speed and efficiency of adaptation.

Table 3: Accuracies (in %) of Subject-Specific Machine Learning and Deep Learning Stroke Models.

Subject ID	**FBCSP + LDA**	**FBCSP + RF**	**CSP + LDA**	**CSP + RF**	**Deep ConvNet**
1	56.41	58.97	56.41	52.56	50.00
2	42.31	44.87	51.28	53.85	43.59
3	93.59	92.31	93.59	93.59	87.18
4	50.00	44.87	46.15	51.28	52.56
5	67.95	71.79	64.10	56.41	62.82
6	82.05	73.08	47.44	47.44	79.49
7	67.95	66.67	65.38	65.38	65.38
8	78.21	73.08	66.67	61.54	93.59
9	64.10	69.23	71.79	64.10	60.26
10	60.26	70.51	57.69	58.97	74.36
11	91.03	85.90	62.82	66.67	58.97
12	66.67	50.00	43.59	47.44	48.72
13	55.13	50.00	53.85	51.28	48.72
14	57.69	60.26	46.15	50.00	50.00
15	93.59	91.03	75.64	67.95	83.33
16	61.54	58.97	52.56	48.72	61.54
17	50.00	42.31	43.59	46.15	55.13
18	57.69	71.79	53.85	55.13	58.97
19	58.97	69.23	57.69	43.59	70.51
20	57.69	58.97	50.00	56.41	75.64
21	61.54	65.38	57.69	51.28	50.00
22	79.49	46.15	52.56	65.38	60.26
23	58.97	53.85	55.13	48.72	69.23
24	50.00	70.51	56.41	51.28	75.64
25	78.21	79.49	60.26	60.26	70.51
26	64.10	64.10	60.26	56.41	61.54
27	89.74	93.59	61.54	62.82	85.90
28	50.00	60.26	46.15	44.87	55.13
29	57.69	65.38	41.03	35.90	62.82
30	92.31	96.15	70.51	67.95	78.21
31	71.79	66.67	79.49	75.64	75.64
32	64.10	66.67	52.56	53.85	66.67
33	55.13	55.13	42.31	44.87	56.41
34	69.23	55.13	47.44	47.44	43.59
35	64.10	65.38	47.44	55.13	55.13
36	42.31	38.46	48.72	52.56	52.56
37	58.97	73.08	61.54	58.97	71.79
38	67.95	62.82	44.87	38.46	50.00
39	69.62	72.15	48.10	46.84	67.09
40	55.13	55.13	58.97	61.54	57.69
41	50.63	43.04	51.90	46.84	50.63
42	67.09	69.62	59.49	54.43	58.23
43	56.41	65.38	57.69	64.10	71.79
44	69.62	74.68	55.70	55.70	72.15
45	55.70	58.23	59.49	59.49	50.63
46	69.23	82.05	73.08	65.38	66.67
47	40.51	51.90	58.23	48.10	63.29
48	64.10	57.69	57.69	57.69	67.95
49	73.42	78.48	70.89	63.29	74.68
50	82.05	82.05	56.41	56.41	61.54
51	55.70	55.70	50.63	54.43	64.56
52	71.79	69.23	50.00	47.44	62.82
53	59.49	58.23	50.63	49.37	62.03
54	58.23	53.16	49.37	51.90	60.76
55	65.38	65.38	51.28	58.97	67.95
56	74.36	71.79	50.00	60.26	67.95
57	60.26	58.97	56.41	51.28	62.82
58	48.72	46.15	55.13	51.28	50.00

59	65.82	65.82	45.57	53.16	53.16
60	77.22	74.68	45.57	51.90	63.29
61	73.08	73.08	65.38	73.08	75.64
62	63.29	54.43	41.77	41.77	49.37
63	97.44	87.18	48.72	52.56	82.05
64	56.41	50.00	51.28	51.28	56.41
65	76.92	65.38	62.82	64.10	70.51
66	67.09	63.29	45.57	43.04	65.82
67	62.82	78.21	47.44	48.72	61.54
68	94.94	89.87	82.28	81.01	55.70
69	48.10	45.57	48.10	41.77	48.10
70	50.00	51.28	48.72	50.00	61.54
71	88.46	87.18	75.64	75.64	83.33
Mean	**65.60±13.46**	**65.31±13.58**	**56.00±10.41****	**55.54±10.01****	**63.51±11.14**

CSP = common spatial patterns, FBCSP = filter-bank CSP, LDA = linear discriminant analysis, and RF = random forest. The subject-specific accuracies were obtained using the hold-out analysis. The ** ($p<0.001$) indicate that the average accuracies are significantly lower than that of FBCSP+LDA, which is the best-performing model with an average accuracy of 65.60%.

5.1. Average Accuracy of Conventional Subject-Specific Classification using Stroke Data

The subject-specific classification accuracies of both the ML models and the Deep ConvNet establish a baseline for our study. The evaluation of subject-specific classification was executed through a hold-out analysis. Among the different SOA models, FBCSP demonstrated the highest average subject-specific performance. Specifically, it achieved an accuracy of 65.60% utilizing the LDA classifier and 65.31% with the RF classifier, across 71 stroke patients.

Subsequently, the Deep ConvNet achieved an average subject-specific accuracy of 63.51%, a value that did not significantly deviate from the accuracy achieved by FBCSP+LDA ($p = 0.18$). In contrast, the CSP model exhibited a substantial performance deficit compared to the other two models. The disparity was statistically significant ($p < 0.001$), with the CSP model attaining accuracies of 56.00% and 55.54% when employing the LDA and RF classifiers, respectively. A comprehensive breakdown of the subject-specific accuracies for the ML models and the Deep ConvNet can be found in Table 3.

5.2. Average Accuracy of the Proposed Healthy-to-Stroke Transfer Learning and the Stroke-to-Stroke Transfer Learning

Table 4 shows that our proposed method achieved an average accuracy of 53.00% before adaptation and 71.15% after adaptation, using Deep ConvNet. The results illustrate that the pre-adaptation performance of our proposed method is considerably lower than the subject-specific performances of FBCSP and Deep ConvNet, with a p-value less than 0.001, and the corresponding average pre-adaptation accuracy of subject-independent stroke models with $p < 0.05$. Although there is a significant difference ($p < 0.05$) between the precision of the H-to-S and the S-to-S models before transfer learning, the adaptation performances of the two models are similar, with no significant difference. Furthermore, our proposed H-to-S transfer learning approach exhibited significantly higher performance than the corresponding H-to-S classification without adaptation

($p < 0.001$), and the subject-specific classification accuracies of FBCSP and Deep ConvNet ($p < 0.001$).

Table 4: Accuracies (in %) of Deep ConvNet Stroke-to-Stroke (S-to-S) and Healthy-to-Stroke (H-to-S) Models, With (w) and Without (w/o) Adaptation.

Subject ID	**Stroke-to-Stroke (S-to-S)**		**Healthy-to-Stroke (H-to-S)**	
	w/o adaptation	**w adaptation**	**w/o adaptation**	**w adaptation**
1	51.28205128	73.07692308	47.43589744	65.38461538
2	44.87179487	47.43589744	50	52.56410256
3	61.53846154	92.30769231	55.12820513	93.58974359
4	51.28205128	53.84615385	51.28205128	48.71794872
5	57.69230769	73.07692308	47.43589744	76.92307692
6	50	83.33333333	57.69230769	75.64102564
7	56.41025641	71.79487179	46.15384615	60.25641026
8	53.84615385	94.87179487	24.35897436	91.02564103
9	46.15384615	73.07692308	38.46153846	69.23076923
10	52.56410256	74.35897436	58.97435897	80.76923077
11	73.07692308	98.71794872	42.30769231	88.46153846
12	48.71794872	50	57.69230769	61.53846154
13	56.41025641	55.12820513	42.30769231	52.56410256
14	50	67.94871795	56.41025641	57.69230769
15	55.12820513	100	57.69230769	96.15384615
16	51.28205128	65.38461538	51.28205128	58.97435897
17	50	52.56410256	53.84615385	56.41025641
18	48.71794872	61.53846154	38.46153846	60.25641026
19	48.71794872	79.48717949	56.41025641	80.76923077
20	51.28205128	71.79487179	43.58974359	65.38461538
21	57.69230769	66.66666667	48.71794872	60.25641026
22	50	64.1025641	41.02564103	61.53846154
23	57.69230769	79.48717949	46.15384615	71.79487179
24	50	76.92307692	50	79.48717949
25	51.28205128	76.92307692	37.17948718	73.07692308
26	48.71794872	71.79487179	37.17948718	66.66666667
27	61.53846154	93.58974359	57.69230769	92.30769231
28	51.28205128	55.12820513	52.56410256	60.25641026
29	70.51282051	88.46153846	42.30769231	88.46153846
30	50	83.33333333	39.74358974	83.33333333
31	57.69230769	84.61538462	42.30769231	76.92307692
32	53.84615385	62.82051282	52.56410256	61.53846154
33	55.12820513	56.41025641	42.30769231	57.69230769
34	61.53846154	71.79487179	47.43589744	51.28205128
35	64.1025641	87.17948718	42.30769231	83.33333333
36	56.41025641	58.97435897	51.28205128	55.12820513
37	51.28205128	75.64102564	48.71794872	71.79487179
38	60.25641026	56.41025641	60.25641026	61.53846154
39	67.08860759	83.5443038	64.55696203	82.27848101
40	53.84615385	61.53846154	61.53846154	64.1025641
41	50.63291139	56.96202532	70.88607595	60.75949367
42	49.36708861	77.21518987	75.94936709	83.5443038
43	50	75.64102564	50	65.38461538
44	67.08860759	77.21518987	63.29113924	78.48101266
45	59.49367089	60.75949367	49.36708861	62.02531646
46	64.1025641	74.35897436	51.28205128	67.94871795
47	53.16455696	45.56962025	67.08860759	65.82278481
48	50	64.1025641	44.87179487	70.51282051
49	56.96202532	78.48101266	58.2278481	77.21518987
50	53.84615385	89.74358974	60.25641026	80.76923077
51	48.10126582	67.08860759	63.29113924	67.08860759
52	58.97435897	84.61538462	56.41025641	82.05128205
53	60.75949367	51.89873418	59.49367089	70.88607595
54	54.43037975	65.82278481	46.83544304	68.35443038
55	60.25641026	74.35897436	55.12820513	58.97435897
56	78.20512821	89.74358974	74.35897436	89.74358974
57	62.82051282	64.1025641	48.71794872	56.41025641

58	52.56410256	47.43589744	53.84615385	58.97435897
59	68.35443038	68.35443038	48.10126582	69.62025316
60	51.89873418	77.21518987	50.63291139	78.48101266
61	47.43589744	76.92307692	67.94871795	91.02564103
62	51.89873418	64.55696203	64.55696203	74.6835443
63	65.38461538	93.58974359	66.66666667	89.74358974
64	57.69230769	52.56410256	60.25641026	61.53846154
65	76.92307692	80.76923077	66.66666667	88.46153846
66	48.10126582	81.01265823	73.41772152	81.01265823
67	56.41025641	64.1025641	50	62.82051282
68	56.96202532	89.87341772	46.83544304	88.60759494
69	56.96202532	59.49367089	59.49367089	59.49367089
70	50	52.56410256	58.97435897	56.41025641
71	57.69230769	93.58974359	57.69230769	89.74358974
Mean	**55.85±7.13**$^{\vee}$	**71.76±13.77**$^{\wedge\dagger}$	**53.00±9.91**$^{\vee *}$	**71.15±12.46**$^{\wedge\dagger}$

The S-to-S accuracies were obtained using the LOSO-CV method. The $^{\vee}$ $^{\wedge}$ ($p<0.001$) indicate that the accuracies are significantly lower/higher than the subject-specific accuracy of FBCSP and Deep ConvNet (Table 3). The † ($p<0.001$) indicates that the accuracies with (w) adaptation are significantly higher than the corresponding without (w/o) adaptation accuracy. The * ($p<0.05$) indicates significantly lower performance compared to that of the corresponding S-to-S model.

The subject-independent accuracy of Deep ConvNet using stroke data is a second baseline for comparison. The subject-independent classification was performed using LOSO-CV analysis. The subject-independent models, also known as S-to-S models, exhibited an average performance of 55.85%, which was found to be significantly lower than the subject-specific performances of FBCSP and Deep ConvNet with a p-value less than 0.001, however, not significantly different from the subject-specific accuracy of CSP. To address disparities between the source and target domains, the study utilized adaptation of S-to-S models, as outlined in Section 4.2.1. After adaptation, the average subject-independent classification accuracy of Deep ConvNet using stroke data increased by 15.91% to reach 71.76% ($p < 0.001$). Moreover, the S-to-S transfer learned models demonstrated significantly better average accuracy compared to the average subject-specific classification accuracies of both Deep ConvNet and FBCSP ($p < 0.001$). The results of H-to-S and S-to-S transfer learning are presented in Table 4.

5.3. Subject-wise Classification Accuracies of Deep ConvNet using the Proposed Healthy-to-Stroke Transfer Learning and the Baseline Methods

Figure 6 illustrates the accuracies of Deep ConvNet at the subject level, from hold-out, S-to-S, and H-to-S analysis. The accuracies of S-to-S and H-to-S models prior to adaptation are denoted by dotted lines, while the transfer learning performances of the two models are depicted with solid lines. The x-axis denotes the subject ID, while the y-axis shows the corresponding accuracies. The blue line represents the subject-specific accuracies, and the subject IDs on the x-axis are arranged based on their corresponding subject-specific performance. Tables 3 and 4 present a comprehensive list of subject-wise classification accuracies obtained using the three methods.

The line plot in Figure 6 shows that the pre-adaptation performances of both S-to-S and H-to-S models fall short of the subject-specific model performance of Deep ConvNet for most subjects. Interestingly, for a few subjects, the H-to-S and S-to-S models perform similarly or even better than the subject-specific performance without the need for adaptation based transfer learning. After transfer learning, the H-to-S

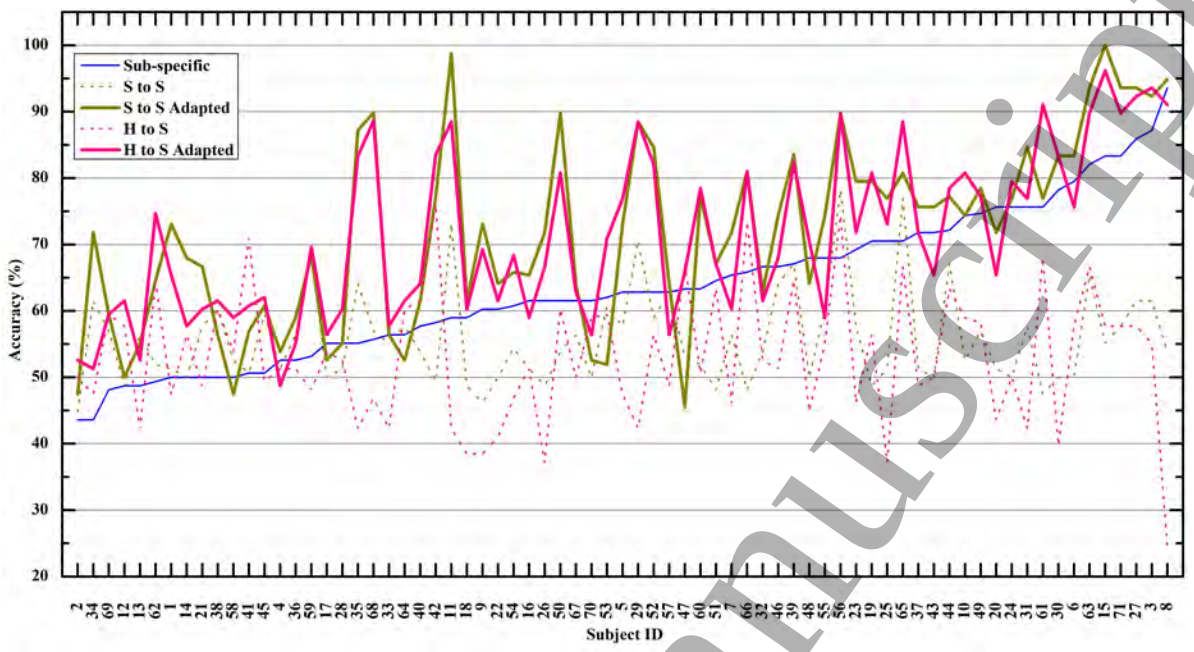

Figure 6: Subject-wise classification accuracies of Deep ConvNet using subject-specific, stroke-to-stroke (S to S), and healthy-to-stroke models (H to S). The subjects are sorted based on their subject-specific model performance.

and S-to-S models perform comparably to each other ($p > 0.05$), however, significantly better than the subject-specific models ($p < 0.001$). These results indicate that models pre-trained on data from healthy subjects may be a viable option for detecting MI in stroke patients, obviating the need for subject-independent stroke models. Collecting sufficient stroke data for training subject-independent stroke models is often challenging, and using limited amounts of data can lead to suboptimal performance. Therefore, the proposed H-to-S transfer, which utilizes abundantly available data from healthy subjects, presents a practical alternative.

5.4. Results from the Investigation on the Impact of Channel Swapping

As detailed in Section 4.3, we conducted additional experiments to examine the importance of channel swapping in our proposed approach.

To explore this, two variations of the approach were developed that did not involve channel swapping. The first variation employed all trials of healthy subjects to pre-train a single model, while the second variation used unilateral MI data to pre-train two separate models. The results of these experiments are presented in Table 5.

The results reveal that pre-training a single model on all trials without channel swapping leads to a significant 2% reduction ($p < 0.05$) in H-to-S classification performance, highlighting the importance of implementing channel swapping when combining two-class MI data from healthy subjects to pre-train models. Utilizing two distinct models trained on unilateral MI data without channel swapping did not

Table 5: Average Accuracy (in %) of Healthy-to-Stroke and Stroke-to-Stroke Models Without Channel Swapping.

Healthy-to-Stroke (H-to-S)		Stroke-to-Stroke (S-to-S)
One Pre-trained Model	Two Distinct Pre-trained Models	
51.00±6.54*	52.63±9.25	55.40±8.44

The * ($p<0.05$) indicates significantly lower performance compared to the corresponding performance using the proposed method.

differ significantly from our proposed approach, but exhibited a slight 0.37% decrease in performance. Despite using only half the number of trials as our proposed approach, this variant produced similar results, indicating the potential advantages of building models specific to each limb using data from healthy individuals to enhance classification accuracy in stroke patients.

Table 5 also includes the results of the baseline S-to-S classification approach without channel swapping. In this analysis, all trials in the training set were used to train the subject-independent stroke model, without implementing channel swapping based on the target affected limb. The reported findings reveal that there were no significant differences between the performance of this approach and that of the actual baseline. This variant achieved an average accuracy of 55.40%, which is marginally lower than the actual baseline S-to-S accuracy of 55.85%, but the difference is not significant. These findings differ from those observed for H-to-S classification in a similar scenario, and suggest that it may not be essential to implement channel swapping when combining MI trials from stroke patients with different affected limbs to train subject-independent stroke models.

5.5. Average Relevance Patterns in Stroke Patients Across Subject-Specific Stroke, Stroke-to-Stroke and Healthy-to-Stroke Models

To elucidate the neurophysiological underpinnings of our transfer learning findings, we employed the DeepLift XAI framework [46] to conduct an XAI analysis on the subject-specific models as well as the H-to-S and S-to-S models using Deep ConvNet.

The average channel-wise relevances from various models were visualized as topoplots in Figure 7 (a)-(e). In line with XAI conventions, we denote low relevance as negative relevance and consider channels with positive relevance as having a positive contribution to the decision for a particular class, while channels with negative relevance are deemed to support the decision against a certain class.

In Figure 7(a), the relevance patterns identified via subject-specific classification exhibit considerable diffusion, lacking localization around the motor area for the MI class [70]. From a neurophysiological perspective, these patterns are not convincing.

In Figure 7(b), the S-to-S models reveal notable localization of high positive relevance in the left and right motor areas, encompassing central and centro-parietal channels C3, C4, CP3, and CP4, for the MI class. Additionally, frontal channel Fz appears to positively impact MI detection. Conversely, the parietal area channels

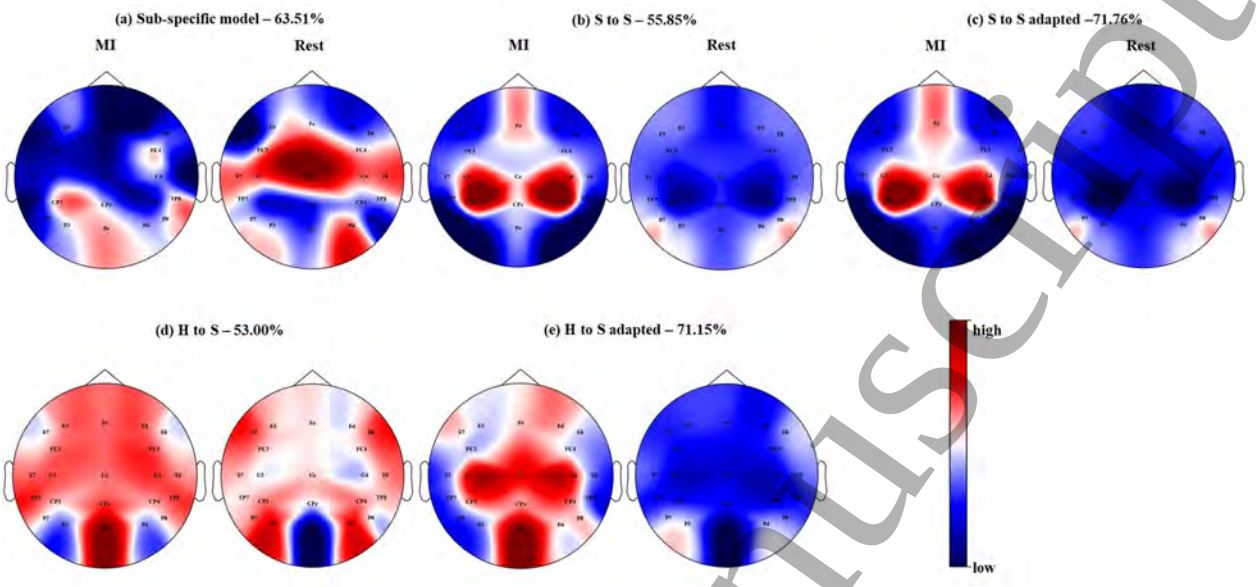

Figure 7: Average relevances of channels for MI vs Rest classes, using (a) subject-specific, (b) stroke-to-stroke (S to S), (c) S to S adapted, (d) healthy-to-stroke (H to S) and (e) H to S adapted models.

The channel relevance scores were obtained using DeepLift[46].

exhibit negative relevance for the MI class. For the rest-state, the relevance patterns complement those for the MI class, meaning that the channels with positive relevance for the MI class display negative relevance for the rest class and vice versa.

Following transfer learning, the S-to-S relevance patterns display similarities to those present prior to transfer learning, albeit with some noteworthy differences, as can be observed in Figure 7(c). Notably, the relevance localization remains in the left and right central areas, although it is now more focused towards centro-parietal channels than central channels. Furthermore, the positive relevance originating from the frontal channel is more pronounced after adaptation. As for the rest-state, the relevance patterns are complementary, with greater negative relevance observed in the left and right centro-parietal channels, along with the frontal and parietal channels, Fz and Pz, respectively.

In Figure 7(d), the pre-adaptation H-to-S classification models exhibit diffuse relevance patterns for both the MI and rest classes. However, the middle parietal electrode Pz displays strong positive relevance for the MI class, while exhibiting the opposite effect for the rest class. Furthermore, the left and right parietal electrodes, P3 and P4, show negative relevance for the MI class and positive relevance for the rest state.

After adaptation-based transfer learning, the H-to-S models display average relevance patterns that are more prominently localized around the left and right central areas, akin to the S-to-S models (Figure 7(e)). Furthermore, the parietal channel Pz exhibits a strong contribution, which was evident even before adaptation. Additionally,

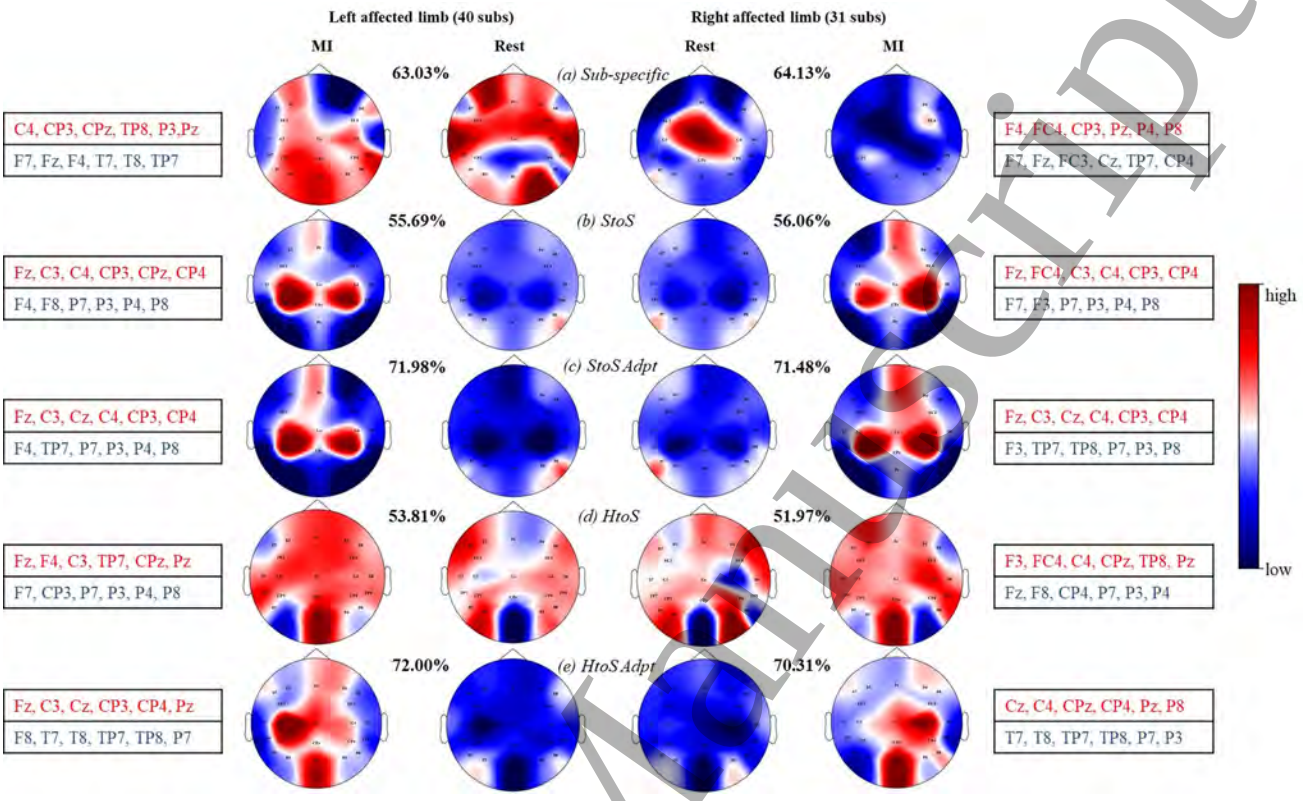

Figure 8: Group-wise average relevance topoplots, comparing the relevance patterns between patient groups with left affected limb versus right affected limb, of (a) subject-specific, (b) stroke-to-stroke (StoS), (c) stroke-to-stroke adapted (StoS Adpt), (d) healthy-to-stroke (HtoS), and (e) healthy-to-stroke adapted (HtoS Adpt) models. The top 30% positively (red color font) and negatively (blue color font) relevant channels for MI prediction are listed beside the corresponding models.

the areas that positively contribute to the MI class exhibit negative contribution to the rest class, as anticipated.

5.6. Group-wise Average Relevance Patterns in Stroke Patients Across Subject-Specific Stroke, Stroke-to-Stroke and Healthy-to-Stroke Models

To better understand the stroke relevance patterns specific to the affected limb, we visualize the channel-wise average relevances by patient group, with left and right affected limbs separately, in Figure 8. The figure illustrates the two-class relevance patterns and the average accuracy for stroke patients with left and right affected limbs, across subject-specific models, S-to-S, and H-to-S models of Deep ConvNet. In addition, we investigated the top 30% of channels that had a positive or negative impact on the MI class within each of the two groups, across various models. The combined stroke dataset consists of 40 patients with an affected left limb and 31 with an affected right limb. To facilitate the discussion of results, we have designated the group with the left affected limb as LAL and the group with the right affected limb as RAL. Additionally,

Table 6: Group-wise Average Classification Accuracies of Patients with Left Affected Limb (LAL) and Right Affected Limb (RAL) Across Subject-specific, Stroke-to-Stroke (S-to-S) and Healthy-to-Stroke (H-to-S) Models.

	Subject-specific	S-to-S w/o adaptation	S-to-S w adaptation	H-to-S w/o adaptation	H-to-S w adaptation
LAL	63.03	55.69	71.98	53.81	72.00
RAL	64.13	56.06	71.48	51.97	70.31

we have employed the shortened forms LAL-MI/RAL-MI to denote patterns associated with the MI class in each group.

The group-level relevance patterns of subject-specific models are visualized in Figure 8(a). These patterns are distributed across various regions of the cortex, irrespective of the group. We also note that there are non-complementary relevance patterns between the two classes in both groups. The relevance patterns observed in RAL are more distinct and also closely resemble the full average plots presented in Figure 7(a).

The group-wise channel relevances in the S-to-S and H-to-S transfer models reveal certain distinct patterns. Specifically, the S-to-S classification model demonstrates very similar relevance patterns between LAL-MI and RAL-MI, indicating that specific cortical areas are more active during stroke MI, regardless of the affected limb [72][44]. This can be observed in Figure 8(b). This differs from the observation in healthy individuals, where brain activation areas are usually contralateral to the MI limb due to ERD [56]. To clarify, for both LAL-MI and RAL-MI, the S-to-S models receive predominant positive contributions from both hemispheres, indicating bilateral impact during stroke MI. However, the ipsilateral (contralesional) contribution is more pronounced than the contralateral influence in both groups. Furthermore, we observe some influence from the frontal channel (Fz) on the classification of MI for both groups. The negatively relevant channels in both groups are similar, although there are some notable differences. The left and right parietal channels negatively contribute to MI detection in both groups. However, the frontal channels contralateral to LAL (F4, F8) negatively impact LAL-MI, whereas those contralateral to RAL (F7, F3) negatively contribute to RAL-MI. The patterns for the rest class are complementary to those of the MI class in both groups.

After applying transfer learning, there are notable updates in the relevance patterns of S-to-S models, which are visualized in Figure 8(c). One main difference is the inclusion of the central channel Cz as a top positively contributing channel for both LAL-MI and RAL-MI. Additionally, after adaptation, the similarity in positively contributing areas of the cortex for both groups becomes more evident. Despite the adaptation, stronger positive contributions from the contralesional hemisphere are still present. Previously, only the frontal and parietal channels were part of top negatively contributing channels to MI in both groups. However, after adaptation, we also observed the influence of

temporal channels on the classification against MI for both LAL and RAL. These updated relevance patterns of S-to-S models, after transfer learning, form the basis for significant improvement in classification performance.

In the H-to-S models, there are noticeable differences in the relevance patterns between LAL and RAL, compared to the S-to-S models. Prior to transfer learning (Figure 8(d)), Pz is the channel with the highest positive relevance in both LAL-MI and RAL-MI, followed by several other areas predominantly on the respective ipsilateral hemispheres. Negative contributions are also observed, primarily from the parietal area, with P3 for LAL-MI and P4 for RAL-MI being the most pronounced. Furthermore, some channels from the left hemisphere's frontal and parietal areas exhibit negative relevance towards LAL-MI, while the corresponding channels on the right hemisphere have negative relevance towards RAL-MI. However, despite these contributions, the H-to-S MI detection models did not achieve high performance, without transfer learning.

Following transfer learning, the positive relevance patterns of H-to-S models are more localized on the motor region, with an ipsilateral (contralesional) predominance corresponding to the affected limb, and the middle parietal region. Figure 8(e) illustrates the aforementioned observation. The positive relevance patterns for LAL-MI in H-to-S models are enhanced after transfer learning, with additional positive relevances observed in the frontal channel Fz and the contralateral central channels, similar to those observed in S-to-S models. The patterns of the rest class display negative relevances in most areas of the cortex, which are more pronounced in the ipsilateral and middle parietal channels, complementing the patterns observed in the respective MI classes of the two groups.

The average results obtained by LAL and RAL groups across subject-specific, S-to-S and H-to-S models of Deep ConvNet are shown in Table 6. There are no notable discrepancies in the average accuracies between the two groups for any of the models examined in this study. However, it should be noted that we did not conduct a statistical significance test for these values, owing to the dissimilarity in the sample size between the two groups.

Finally, we also explored the subject-level patterns of a selected number of stroke patients from each group, using only the adapted transfer models. The subject-level patterns emphasize the commonalities as well as notable differences in the relevances between the S-to-S and H-to-S transfer models. These representative patterns also depict the impact of bilateral motor areas, frontal and parietal areas of the cortex for MI detection. Visualization of feature distributions using t-SNE, shows discriminative and well-separable features between the MI versus rest classes, more so in the transfer models. More details on the relevance analysis of selected patients are provided in Appendix A, Figures A1 and A2.

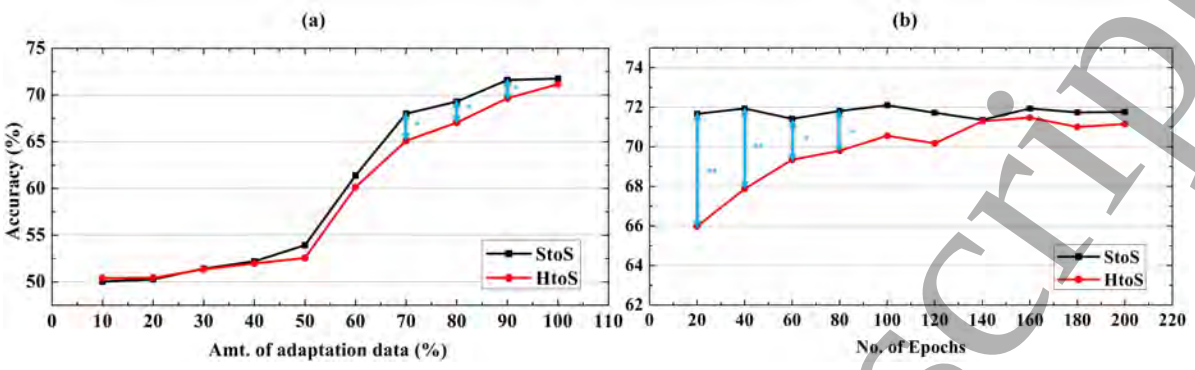

Figure 9: Comparison of (a) adaptation efficiency and (b) adaptation speed between stroke-to-stroke (StoS) and healthy-to-stroke (HtoS) models. The * and ** indicate significant differences between the two types of transfer learning, where * is for $p < 0.05$ and ** is for $p < 0.001$

5.7. Comparison of Adaptation Efficiency and Adaptation Speed in Stroke-to-Stroke versus Healthy-to-Stroke Models

In order to compare the S-to-S and H-to-S transfer models, we performed additional experiments to gather information about their speed and efficiency of transfer learning. The results of these adaptation experiments, which included varying amounts of adaptation data (adaptation efficiency) and different number of training epochs (adaptation speed), are presented in Figure 9, with Figure 9(a) showing the former and Figure 9(b) showing the latter.

The efficiency plot in Figure 9(a) indicates that the S-to-S and H-to-S transfer learning exhibit comparable levels of performance when the adaptation data is varied from 10% to 60% of the total data. After this point, the S-to-S model outperforms the H-to-S model significantly ($p < 0.05$) when the adaptation data size is 70%, 80%, and 90% of the total amount. However, there is no significant difference between the two models when the entire adaptation set is used for transfer learning. For the S-to-S transfer model, a minimum of 90% of the total data is required to achieve an accuracy comparable to that obtained using the entire adaptation data. In contrast, the H-to-S model requires the full adaptation set to achieve optimal accuracy.

Figure 9(b) shows the results of the adaptation speed experiment, and indicates that the S-to-S models outperformed the H-to-S models ($p < 0.05$) when using fewer epochs for transfer learning, ranging from 20 to 80 epochs. However, for 100 or more epochs, there are no significant differences in accuracies between the two models. These findings suggest that S-to-S models can adapt faster than H-to-S models. Specifically, S-to-S transfer learning can achieve similar accuracy in just 20 epochs compared to H-to-S transfer learning, which requires at least 140 epochs to achieve an accuracy that is not significantly different from that obtained using 200 epochs.

6. Discussion

In this section, we summarize the noteworthy findings stemming from our research and emphasize the innovative components and comparative benefits of the concept of transfer learning from the healthy to the stroke. Our study aimed to assess the feasibility of using an MI-BCI model pre-trained with data from healthy individuals to detect MI in stroke patients. This approach was explored as a potential remedy for the inadequate stroke data availability that hinders the implementation of effective cross-subject models for stroke patients [41] [54].

Our initial experiments using Deep ConvNet involving the baseline subject-specific and S-to-S classification approaches, as well as the proposed H-to-S classification method, yielded modest outcomes of 63.51%, 55.85%, and 53.00%, respectively. It is to be noted here that the ML-based subject-specific model using FBCSP+LDA achieved the highest average subject-specific performance of 65.60%. Nevertheless, the performance of subject-specific Deep ConvNet was not significantly different from that of FBCSP. The S-to-S transfer demonstrated some superiority over H-to-S transfer, prior to adaptation, by achieving a 2.85% higher accuracy ($p < 0.05$). Nevertheless, both S-to-S and H-to-S transfer techniques exhibited significantly lower accuracy than the corresponding subject-specific average of Deep ConvNet, with a p-value of less than 0.001. These findings underscored the necessity of domain adaptation-based transfer learning for the S-to-S and H-to-S models, to tackle the problem of feature distributional differences between the source and target domains. The suboptimal results of these transfer models prior to domain adaptation are in line with the findings from BCI literature pertaining to cross-subject classification using stroke data [41] [54].

After applying domain adaptation-based transfer learning, we noted that the performance of our H-to-S transfer approach was 71.15%, which is at par with the result of S-to-S transfer, which is 71.76%, with no significant difference. Both S-to-S and H-to-S transfer models demonstrated average accuracies that were significantly superior to the average subject-specific performance, and their respective accuracies before adaptation, as reported in Tables 3 and 4. In both instances, the p-value was less than 0.001. Our S-to-S and H-to-S transfer learning, applied across 71 stroke patients, exhibit superior performance compared to the transfer learning outcomes reported in [53] and [54]. As discussed in Section 2, Cao et al. [53] utilized ML techniques for transfer calibration with data from seven stroke patients and achieved an average accuracy of 59.16% in inter-subject transfer, while Xu et al. [54] employed EEGNet for transfer learning with a combination of data from eleven healthy individuals and five stroke patients to obtain an average accuracy of 66.36%. Moreover, the comparable results of H-to-S and S-to-S transfer learning are instrumental in emphasizing the feasibility and effectiveness of transferring a motor imagery detection model developed with data from healthy individuals to stroke patients.

We also analyzed the individual accuracies of stroke patients utilizing the three types of motor imagery versus rest classification models of Deep ConvNet. Our objective

was to highlight the fact that even at the subject level, the performances of the S-to-S and H-to-S transfer learned models were similar and significantly better than the corresponding subject-specific accuracies. Figure 6 depicts these findings in a line graph. Although healthy individuals and stroke patients may exhibit variations in neural oscillations during MI [73][74][75], we demonstrated that through transfer learning, one can obtain a comparable level of motor imagery classification accuracy using the H-to-S transfer model as with the S-to-S transfer model for most patients. We observed that in few patients, the accuracy obtained using the adapted H-to-S model was better compared to that of the adapted S-to-S model, and vice versa.

We conducted additional experiments to validate the necessity of channel swapping in H-to-S and S-to-S classification. The results of these experiments are presented in Table 5. We observed that, in the case of S-to-S transfer, the subject-independent stroke models showed little change in average performance when no channel swapping was performed. In contrast, for our proposed method, we observed a significant 2% reduction in average accuracy ($p < 0.05$) when bilateral MI trials were combined and used for training without any channel swapping. The results obtained using two distinct models pre-trained with unilateral MI data were not significantly different, but marginally lower than that of our proposed method.

From the additional analyses, we can infer several important insights. Firstly, the lack of significant differences in the results of S-to-S analysis without channel swapping of mismatched trials may be attributed to the fact that stroke patients exhibit bilateral brain activations during MI regardless of the affected limb, as identified in our relevance analysis. These observations are consistent with prior reports in the BCI literature [72][44]. Consequently, there may be low variation in features with respect to the MI limb. Nevertheless, the H-to-S analyses results emphasize the importance of producing a model pre-trained using healthy subject trials specific to the affected limb of the target stroke patient, with or without channel swapping. Contrasting neural oscillatory patterns observed during two-class MI tasks in healthy individuals may underlie the observed effects [56]. However, performing channel swapping allows us to utilize all the trials available in the dataset for training, which can result in better accuracies. Nevertheless, for a comprehensive understanding, it is imperative to delve into the effectiveness of channel swapping in further studies. This is particularly important considering the observed comparable, and in some cases superior, performance of the H-to-S models trained with unilateral trials and the S-to-S models without the incorporation of channel swapping, in comparison to their counterparts that do employ channel swapping.

We further aimed to deduce the neurophysiological interpretations of our transfer learning results using stroke data. To achieve this, we performed an XAI analysis to examine the subject-specific and transfer models, the results of which are presented in Figures 7 and 8. Our XAI analysis, performed using DeepLift, revealed interesting insights into the similarities and differences between the relevance patterns derived from S-to-S and H-to-S transfer models. Through our average relevance plots across

all patients, we observed the diffusive nature of relevance patterns in subject-specific models. This observation may be indicative of their suboptimal performance, likely due to the limited amount of training data used in their construction. Moreover, we found that S-to-S models showed bilateral MI activations before and after adaptation [44][72][42]. We also observed more pronounced relevance in the centro-parietal area of the contralateral (ipsilesional) hemisphere, post transfer learning in S-to-S models. Furthermore, the S-to-S models showed a strong influence of the middle frontal and a mild one from the middle parietal area of the cortex towards stroke MI. A previous study by Lee et al. [76] reported a similar finding of MI-related neural activity located primarily in the frontal and sensorimotor cortices among stroke patients. Notably, they detected this activity in the beta frequency band and predominantly in the ipsilesional hemisphere. The H-to-S transfer models, on the other hand, showed distributed channel contributions prior to transfer learning, except for the high positive relevance seen in the middle parietal electrode Pz, and negative relevance in bilateral parietal electrodes, for MI detection. Post transfer learning, H-to-S models exhibited clear bilateral localization around the central and centro-parietal areas of the cortex, similar to the S-to-S models, with continued emphasis in the middle parietal electrode and some positive relevance also found in the frontal region. These relevance patterns are aligned with findings reported in BCI literature between the healthy and the stroke groups. For example, in [42], Ang et al. identified similar spatial patterns in healthy subjects and stroke patients, with focused activity appearing on both ipsilateral and contralateral hemispheres with respect to the limb involved in MI. In the study by Lee et al. [76], the authors observed a pattern of neural activity over the frontal, motor, and parietal areas of the cortex in healthy controls during MI performance. More recently, in a cross-subject channel selection study using OpenBMI dataset, Nagarajan et al. [27] identified the influence of motor, parietal, and occipital channels in MI-EEG classification among healthy subjects. However, the bilateral nature of MI relevance found in both S-to-S and H-to-S transfer models raised questions regarding how these average patterns are influenced by the respective relevances pertaining to stroke MI of the LAL versus the RAL.

We analyzed relevance patterns in stroke patients with LAL versus RAL to further strengthen the insights derived from average relevance patterns. We did not observe any notable differences in average accuracy between the LAL and the RAL groups, consisting of 40 patients and 31 patients respectively, for all models considered. The findings discussed for the group-wise relevance analysis are visualized in Figure 8. In both groups, the subject-specific models continued to show distributed relevance patterns across different cortical regions. Both S-to-S and H-to-S models demonstrate bilateral influence from the central motor region, impact of the frontal channel Fz, and the middle parietal channel Pz for MI detection in stroke patients. These are in line with the findings from [44], [72], and [76]. In both LAL and RAL groups, S-to-S transfer models showed ipsilateral localization around both central and centro-parietal areas and contralateral localization in the centro-parietal area of the cortex (Figure 8(b)), which was consistently observed even after adaptation (Figure 8(c)). However, we observed that adaptation-

based transfer learning further strengthened the emphasis on the frontal channel Fz and middle central channel Cz. It is notable that the bilateral influence is seen mainly on the centro-parietal area of the cortex. This is similar to the observation from the corresponding average relevance plot in Figure 7(c). Before transfer learning, H-to-S models showed positive localization in the middle parietal electrode Pz and negative relevances in the bilateral parietal electrodes, along with some positive relevance seen in the ipsilateral central and frontal regions, for both LAL-MI and RAL-MI (Figure 8(d)). These patterns were further strengthened with adaptation-based transfer learning, as can be noticed in Figure 8(e). Additionally, we observed bilateral patterns in the central region, especially in LAL-MI, similar to those observed in S-to-S models. The influence of the middle parietal region observed in H-to-S transfer contrasts with the middle frontal influence seen during S-to-S transfer. The parietal influence specific to transfer from models pre-trained using data from healthy subjects is in line with findings from the literature reported for healthy subjects during MI performance [27][76].

The aforementioned points suggest that the relevance patterns observed in H-to-S transfer models show similarity to those observed in S-to-S transfer models, in terms of contribution from bilateral motor region, thus justifying the comparable performance of the two types of transfer for stroke patients. However, we also found that for certain stroke patients, H-to-S transfer models provide better accuracy, possibly due to additional neurophysiological attributions provided by the model that are not present in S-to-S models. Our proposed H-to-S transfer learning reveals interesting neurophysiological interactions and associations, derived using knowledge from both healthy and stroke MI-EEG, that can otherwise remain unknown. The identified associative patterns also underlie the good MI detection performance in stroke patients achieved through our proposed method. Our analysis of both types of transfer suggests that the bilateral motor, middle frontal, and middle parietal regions are highly relevant for MI detection in stroke patients [76][72][44][27].

Finally, our analyses comparing the efficiency and speed of adaptation between the two transfer models indicate similarities between the H-to-S and S-to-S transfer learning with respect to adaptation data requirements. However, S-to-S transfer learning outperformed H-to-S transfer learning in terms of adaptation speed. The S-to-S transfer learning can achieve similar adaptation performance as using the entire adaptation set with 90% of the data, which is not the case for H-to-S transfer. However, the difference in data size is only 10% and further improvements to the adaptation efficiency of the proposed method will be explored in future investigations. The S-to-S transfer learning requires only 20 epochs to achieve a similar performance as that obtained using 200 epochs. On the other hand, H-to-S transfer models need at least 140 epochs to achieve a similar accuracy. The slower adaptation speed of H-to-S transfer models can be addressed by enhancing the availability of computing resources.

Limitations and Future Work: Our study on transfer learning from healthy to stroke patients has demonstrated promising results, with comparable performance to subject-independent stroke models and high interpretability. However, there are

also some limitations that need to be addressed. Our method utilizes the model adaptation approach to address feature distribution differences between the source and target domains. However, it relies on labelled data from the target domain, making it a supervised domain adaptation technique. In our future work, we will explore unsupervised domain adaptation methods to improve the robustness and efficiency of H-to-S transfer learning. Additionally, online domain adaptation is a critical concept to consider for clinical applications of BCI for stroke rehabilitation.

In this transfer learning study, we utilized the OpenBMI dataset [45] which includes two-class MI-EEG data from 54 healthy subjects, and stroke datasets [11] and [12] which together consist of MI versus rest data from 71 stroke patients. However, to gather more generalizable insights on transfer learning from healthy to stroke, it is essential to conduct evaluations of the proposed method using multiple datasets in future work. Additionally, it is crucial to validate our transfer learning method using other DL-based BCI models. Furthermore, real-time evaluations of our proposed method will be conducted in the future. Exploring the viability of transferring models from healthy to stroke populations in the context of MI-BCI using alternative neural data modalities like ECoG, MEG, or fNIRS presents an interesting avenue for future investigation.

7. Conclusion

The objective of this research was to suggest and examine the practicality of transfer learning from healthy individuals to stroke patients, which is crucial for tackling the obstacle of scarce stroke data for building EEG-BCI for stroke rehabilitation. Deep learning techniques were employed to assess and explain our proposed transfer learning approach from healthy subjects to stroke patients. To pre-train models for a stroke patient with a specific affected limb, we introduced a new channel swapping method. Domain adaptation-based transfer learning was executed to enhance the MI detection performance in stroke patients. Furthermore, the results are compared with subject-specific classification accuracies obtained using Deep ConvNet and SOA ML classifiers.

The key findings obtained from our transfer learning experiments that employed Deep ConvNet, evaluated on OpenBMI and stroke datasets, are as follows: (i) the average accuracies of MI detection of both healthy-to-stroke and stroke-to-stroke models, without transfer learning, were found to be close to the chance level in our study. The healthy-to-stroke model exhibited an accuracy that was 2.85% lower ($p < 0.05$) than that of the stroke-to-stroke model (ii) the healthy-to-stroke model achieved an average accuracy of 71.15% after transfer learning, which was significantly higher than its corresponding pre-adaptation accuracy ($p < 0.001$) as well as the subject-specific classification accuracy ($p < 0.001$) (iii) the performance of the proposed healthy-to-stroke transfer learning approach was comparable to that of the stroke-to-stroke transfer learning method, with no significant difference observed ($p > 0.05$) (iv) according to the XAI-based channel relevance analysis, the detection of MI in stroke patients is influenced by the bilateral motor, frontal, and parietal regions of the cortex (v) finally,

the adaptation experiments demonstrated comparable adaptation efficiency of both types of transfer. However, the proposed healthy-to-stroke transfer learning exhibited a slower adaptation rate when compared to stroke-to-stroke transfer. The outcomes underscore the importance of our investigation in evaluating the viability of utilizing MI-BCI models pre-trained with data from healthy individuals for stroke patients. This approach holds the potential for addressing the scarcity of stroke data in the development of BCIs for upper limb stroke rehabilitation, as well as in achieving subject-independent deep learning models for stroke patients that are robust, interpretable, and efficient.

Acknowledgment

This work was partially supported by the RIE2020 AME Programmatic Fund, Singapore (No. A20G8b0102). The computational work for this article was partially performed on the resources provided by the National Supercomputing Centre, Singapore (https://www.nscc.sg).

References

[1] Mak J N and Wolpaw J R 2009 Clinical Applications of Brain—Computer Interfaces: Current State and Future Prospects *IEEE Reviews in Biomedical Engineering* **2** 187–199

[2] Douibi K, Bars S L, Lemontey A, Nag L, Balp R and Breda G 2021 Toward eeg-based bci applications for industry 4.0: Challenges and possible applications *Frontiers in Human Neuroscience* **15**

[3] Mudgal S K, Sharma S K, Chaturvedi J and Sharma A 2020 Brain computer interface advancement in neurosciences: Applications and issues *Interdisciplinary Neurosurgery* **20** 100694 URL `https://www.sciencedirect.com/science/article/pii/S2214751920300098`

[4] Abiri R, Borhani S, Sellers E W, Jiang Y and Zhao X 2019 A comprehensive review of eeg-based brain-computer interface paradigms *Journal of Neural Engineering* **16**

[5] Craik A, He Y and Contreras-Vidal J L 2019 Deep learning for electroencephalogram (eeg) classification tasks: a review *Journal of Neural Engineering* **16**

[6] Chai R, Naik G R, Ling S H and Nguyen H T 2017 Hybrid brain-computer interface for biomedical cyber-physical system application using wireless embedded eeg systems *BioMedical Engineering Online* **16**

[7] Lazarou I, Nikolopoulos S, Petrantonakis P C, Kompatsiaris I and Tsolaki M 2018 Eeg-based brain–computer interfaces for communication and rehabilitation of people with motor impairment: A novel approach of the 21st century *Frontiers in Human Neuroscience* **12** 14

[8] Lécuyer A, Lotte F, Reilly R B, Leeb R, Hirose M and Slater M 2008 Brain-computer interfaces, virtual reality, and videogames *Computer* **41** 66 – 72

[9] Mulder T 2007 Motor imagery and action observation: Cognitive tools for rehabilitation *Journal of Neural Transmission* vol 114 p 1265–1278

[10] Ang K K, Guan C, Chua K S G, Ang B T, Kuah C, Wang C, Phua K S, Chin Y Z and Zhang H 2010 Clinical study of neurorehabilitation in stroke using eeg-based motor imagery brain-computer interface with robotic feedback *2010 Annual International Conference of the IEEE Engineering in Medicine and Biology Society, EMBC'10* pp 5549–52

[11] Ang K K, Guan C, Phua K S, Wang C, Zhou L, Tang K Y, Joseph G J E, Kuah C W K and Chua K S G 2014 Brain-computer interface-based robotic end effector system for wrist and hand

rehabilitation: Results of a three-armed randomized controlled trial for chronic stroke *Frontiers in Neuroengineering* **7** 30

[12] Ang K K, Guan C, Phua K S, Wang C, Zhao L, Teo W P, Chen C, Ng Y S and Chew E 2015 Facilitating effects of transcranial direct current stimulation on motor imagery brain-computer interface with robotic feedback for stroke rehabilitation *Archives of Physical Medicine and Rehabilitation* **96** S79–87

[13] Pfurtscheller G and Neuper C 2001 Motor imagery and direct brain-computer communication *Proceedings of the IEEE* **89**(7)

[14] Perdikis S, Tonin L, Saeedi S, Schneider C and del R Millán J 2018 The cybathlon bci race: Successful longitudinal mutual learning with two tetraplegic users *PLoS Biology* **16**(5)

[15] Tortora S, Beraldo G, Bettella F, Formaggio E, Rubega M, Felice A D, Masiero S, Carli R, Petrone N, Menegatti E and Tonin L 2022 Neural correlates of user learning during long-term bci training for the cybathlon competition *Journal of NeuroEngineering and Rehabilitation* **19**(1) 69

[16] Hehenberger L, Kobler R J, Lopes-Dias C, Srisrisawang N, Tumfart P, Uroko J B, Torke P R and Müller-Putz G R 2021 Long-term mutual training for the cybathlon bci race with a tetraplegic pilot: A case study on inter-session transfer and intra-session adaptation *Frontiers in Human Neuroscience* **15** 635777

[17] Robinson N, Chouhan T, Mihelj E, Kratka P, Debraine F, Wenderoth N, Guan C and Lehner R 2021 Design considerations for long term non-invasive brain computer interface training with tetraplegic cybathlon pilot *Frontiers in Human Neuroscience* **15** 648275

[18] Novak D, Sigrist R, Gerig N J, Wyss D, Bauer R, Götz U and Riener R 2018 Benchmarking brain-computer interfaces outside the laboratory: The cybathlon 2016 *Frontiers in Neuroscience* **11**(JAN) 756

[19] Blankertz B, Tomioka R, Lemm S, Kawanabe M and Müller K R 2008 Optimizing spatial filters for robust EEG single-trial analysis *IEEE Signal Processing Magazine* **25** 41 – 56

[20] Ang K K, Chin Z Y, Zhang H and Guan C 2008 Filter Bank Common Spatial Pattern (FBCSP) in brain-computer interface *Proceedings of the International Joint Conference on Neural Networks*

[21] FISHER R A 1936 The use of multiple measurements in taxonomic problems *Annals of Eugenics* **7**(2) 179–188

[22] Breiman L 2001 Random forests *Machine Learning* **45**(1) 5–32

[23] Sakhavi S, Guan C and Yan S 2018 Learning temporal information for brain-computer interface using convolutional neural networks *IEEE Transactions on Neural Networks and Learning Systems* **29** 5619–5629

[24] Schirrmeister R T, Springenberg J T, Fiederer L D J, Glasstetter M, Eggensperger K, Tangermann M, Hutter F, Burgard W and Ball T 2017 Deep learning with convolutional neural networks for EEG decoding and visualization *Human Brain Mapping* **38** 5391–5420 (*Preprint* arXiv: 1703.05051)

[25] Lawhern V J, Solon A J, Waytowich N R, Gordon S M, Hung C P and Lance B J 2018 EEGNet: A compact convolutional neural network for EEG-based brain-computer interfaces *Journal of Neural Engineering* **15** (*Preprint* arXiv:1611.08024)

[26] Nagarajan A, Robinson N and Guan C 2021 Investigation on robustness of eeg-based brain-computer interfaces *43rd Annual International Conference of the IEEE Engineering in Medicine and Biology Society. IEEE Engineering in Medicine and Biology Society. Annual International Conference* **2021** 6334–6340

[27] Nagarajan A, Robinson N and Guan C 2023 Relevance-based channel selection in motor imagery brain-computer interface *Journal of Neural Engineering* **20**

[28] Zhang K, Robinson N, Lee S W and Guan C 2021 Adaptive transfer learning for eeg motor imagery classification with deep convolutional neural network *Neural Networks* **136** 1–10

[29] Xu F, Rong F, Miao Y, Sun Y, Dong G, Li H, Li J, Wang Y and Leng J 2021 Representation learning for motor imagery recognition with deep neural network *Electronics (Switzerland)* **10**(2) 112

[30] Reichert C, Dürschmid S, Kruse R and Hinrichs H 2016 An efficient decoder for the recognition of event-related potentials in high-density meg recordings *Computers* **5**(2)

[31] Sitaram R, Caria A, Veit R, Gaber T, Rota G, Kuebler A and Birbaumer N 2007 Fmri brain-computer interface: A tool for neuroscientific research and treatment *Computational Intelligence and Neuroscience* **2007**

[32] Hong K S, Khan M J and Hong M J 2018 Feature extraction and classification methods for hybrid fnirs-eeg brain-computer interfaces *Frontiers in Human Neuroscience* **12** 246

[33] Buccino A P, Keles H O and Omurtag A 2016 Hybrid eeg-fnirs asynchronous brain-computer interface for multiple motor tasks *PLoS ONE* **11**(1)

[34] Chiarelli A M, Croce P, Merla A and Zappasodi F 2018 Deep learning for hybrid eeg-fnirs brain-computer interface: Application to motor imagery classification *Journal of Neural Engineering* **15**(3)

[35] Talukdar U, Hazarika S M and Gan J Q 2019 Motor imagery and mental fatigue: inter-relationship and eeg based estimation *Journal of Computational Neuroscience* **46** 55–76

[36] Azab A M, Mihaylova L, Ang K K and Arvaneh M 2019 Weighted transfer learning for improving motor imagery-based brain–computer interface *IEEE Transactions on Neural Systems and Rehabilitation Engineering* **27** 1352–1359

[37] Xu Y, Huang X and Lan Q 2021 Selective cross-subject transfer learning based on riemannian tangent space for motor imagery brain-computer interface *Frontiers in Neuroscience* **15**

[38] Cheng D, Liu Y and Zhang L 2018 Exploring motor imagery eeg patterns for stroke patients with deep neural networks *ICASSP, IEEE International Conference on Acoustics, Speech and Signal Processing - Proceedings* vol 2018-April pp 2561–2565

[39] Chowdhury A, Raza H, Meena Y K, Dutta A and Prasad G 2018 Online covariate shift detection-based adaptive brain-computer interface to trigger hand exoskeleton feedback for neuro-rehabilitation *IEEE Transactions on Cognitive and Developmental Systems* **10** 1070–1080

[40] Irimia D C, Ortner R, Poboroniuc M S, Ignat B E and Guger C 2018 High classification accuracy of a motor imagery based brain-computer interface for stroke rehabilitation training *Frontiers Robotics AI* **5** 130

[41] Raza H, Chowdhury A and Bhattacharyya S 2020 Deep learning based prediction of eeg motor imagery of stroke patients' for neuro-rehabilitation application *Proceedings of the International Joint Conference on Neural Networks* pp 1–8

[42] Ang K K, Guan C, Chua K S G, Ang B T, Kuah C, Wang C, Phua K S, Chin Z Y and Zhang H 2008 A clinical evaluation on the spatial patterns of non-invasive motor imagery-based brain-computer interface in stroke *Proceedings of the 30th Annual International Conference of the IEEE Engineering in Medicine and Biology Society, EMBS'08 - "Personalized Healthcare through Technology"* pp 4174–7

[43] Shu X, Chen S, Yao L, Sheng X, Zhang D, Jiang N, Jia J and Zhu X 2018 Fast recognition of bci-inefficient users using physiological features from eeg signals: A screening study of stroke patients *Frontiers in Neuroscience* **12** 93

[44] Mansour S, Giles J, Ang K K, Nair K P, Phua K S and Arvaneh M 2022 Exploring the ability of stroke survivors in using the contralesional hemisphere to control a brain–computer interface *Scientific Reports* **12**

[45] Lee M H, Kwon O Y, Kim Y J, Kim H K, Lee Y E, Williamson J, Fazli S and Lee S W 2019 EEG dataset and OpenBMI toolbox for three BCI paradigms: An investigation into BCI illiteracy *GigaScience* **8**

[46] Shrikumar A, Greenside P and Kundaje A 2017 Learning important features through propagating activation differences *34th International Conference on Machine Learning, ICML 2017* vol 70 p 3145–3153

[47] Zheng X, Li J, Ji H, Duan L, Li M, Pang Z, Zhuang J, Rongrong L and Tianhao G 2020 Task transfer learning for eeg classification in motor imagery-based bci system *Computational and Mathematical Methods in Medicine* **2020** 1–11

[48] He H and Wu D 2020 Transfer learning for brain-computer interfaces: A euclidean space data alignment approach *IEEE Transactions on Biomedical Engineering* **67** 399–410

[49] Wang X, Yang R and Huang M 2022 An unsupervised deep-transfer-learning-based motor imagery eeg classification scheme for brain–computer interface *Sensors* **22** 2241

[50] Begiełło F, Tokovarov M and Plechawska-Wójcik M 2020 Transfer learning approach in classification of bci motor imagery signal *International Conference on Computer Information Systems and Industrial Management)* vol 12133 LNCS p 3–14

[51] Ju C, Gao D, Mane R, Tan B, Liu Y and Guan C 2020 Federated transfer learning for eeg signal classification *Proceedings of the 42nd Annual International Conference of the IEEE Engineering in Medicine and Biology Society, EMBS* vol 2020-July pp 3040–3045

[52] Moody G B 2000 Physionet: Research resource for complex physiologic signals *Circulation* **101** E215–20

[53] Cao L, Chen S, Jia J, Fan C, Wang H and Xu Z 2021 An inter- and intra-subject transfer calibration scheme for improving feedback performance of sensorimotor rhythm-based bci rehabilitation *Frontiers in Neuroscience* **14**

[54] Xu F, Miao Y, Sun Y, Guo D, Xu J, Wang Y, Li J, Li H, Dong G, Rong F, Leng J and Zhang Y 2021 A transfer learning framework based on motor imagery rehabilitation for stroke *Scientific Reports* **11**

[55] Krumpe T, Baumgärtner K, Rosenstiel W and Spüler M 2017 Non-stationarity and inter-subject variability of eeg characteristics in the context of bci development *Graz Brain-Computer Interface Conference*

[56] Pfurtscheller G and da Silva F L 1999 Event-related eeg/meg synchronization and desynchronization: basic principles *Clinical Neurophysiology* **110** 1842–1857

[57] Kilavik B E, Zaepffel M, Brovelli A, MacKay W A and Riehle A 2013 The ups and downs of beta oscillations in sensorimotor cortex *Experimental Neurology* **245** 15–26 ISSN 0014-4886 special Issue: Neuronal oscillations in movement disorders URL `https://www.sciencedirect.com/science/article/pii/S0014488612003767`

[58] Pan S J and Yang Q 2010 A survey on transfer learning *IEEE Transactions on Knowledge and Data Engineering* **22** 1345–1359

[59] Weiss K, Khoshgoftaar T M and Wang D D 2016 A survey of transfer learning *Journal of Big Data* **3**

[60] Ang K K, Guan C, Chua K S G, Ang B T, Kuah C W K, Wang C, Phua K S, Chin Z Y and Zhang H 2011 A large clinical study on the ability of stroke patients to use an eeg-based motor imagery brain-computer interface *Clinical EEG and Neuroscience* **42** 253–8

[61] Ang K K, Chin Z Y, Zhang H and Guan C 2012 Mutual information-based selection of optimal spatial-temporal patterns for single-trial EEG-based BCIs *Pattern Recognition* **45** 2137–2144

[62] Chang C C and Lin C J 2011 Libsvm: A library for support vector machines *ACM Transactions on Intelligent Systems and Technology* **2** 1–27

[63] Krizhevsky A, Sutskever I and Hinton G E 2017 ImageNet classification with deep convolutional neural networks vol 60 p 84–90

[64] Clevert D A, Unterthiner T and Hochreiter S 2016 Fast and accurate deep network learning by exponential linear units (ELUs) *4th International Conference on Learning Representations, ICLR 2016 - Conference Track Proceedings* (*Preprint* arXiv:1511.07289)

[65] Paszke A, Gross S, Massa F, Lerer A, Bradbury J, Chanan G, Killeen T, Lin Z, Gimelshein N, Antiga L, Desmaison A, Kopf A, Yang E, DeVito Z, Raison M, Tejani A, Chilamkurthy S, Steiner B, Fang L, Bai J and Chintala S 2019 Pytorch: An imperative style, high-performance deep learning library *Proceedings of the 33rd International Conference on Neural Information Processing Systems* p 8026–8037

[66] Kingma D P and Ba J L 2015 Adam: A method for stochastic optimization **abs/1412.6980** 1–15

[67] Ioffe S and Szegedy C 2015 Batch normalization: Accelerating deep network training by reducing internal covariate shift *32nd International Conference on Machine Learning, ICML 2015* vol 1

p 448–456

[68] Srivastava N, Hinton G, Krizhevsky A, Sutskever I and Salakhutdinov R 2014 Dropout: A simple way to prevent neural networks from overfitting *Journal of Machine Learning Research* **15** 19291958

[69] Fugl-Meyer A R, Jääskö L, Leyman I A, Olsson S and Steglind S 1975 The post-stroke hemiplegic patient. 1. a method for evaluation of physical performance. *Scandinavian journal of rehabilitation medicine* **7** 13–31

[70] Pfurtscheller G, Brunner C, Schlögl A and da Silva F H L 2006 Mu rhythm (de)synchronization and eeg single-trial classification of different motor imagery tasks *NeuroImage* **31** 153–9

[71] Kokhlikyan N, Miglani V, Martin M, Wang E, Alsallakh B, Reynolds J, Melnikov A, Kliushkina N, Araya C, Yan S and Reblitz-Richardson O 2020 Captum: A unified and generic model interpretability library for pytorch (*Preprint* arXiv:2009.07896)

[72] Li M, Liu Y, Wu Y, Liu S, Jia J and Zhang L 2014 Neurophysiological substrates of stroke patients with motor imagery-based brain-computer interface training *International Journal of Neuroscience* **124** 403–15

[73] Rossiter H E, Boudrias M H and Ward N S 2014 Do movement-related beta oscillations change after stroke? *Journal of Neurophysiology* **112** 2053–8

[74] Li H, Huang G, Lin Q, Zhao J, Fu Q, Li L, Mao Y, Wei X, Yang W, Wang B, Zhang Z and Huang D 2020 Eeg changes in time and time-frequency domain during movement preparation and execution in stroke patients *Frontiers in Neuroscience* **14** 827

[75] Dionísio A, Gouveia R, Castelhano J, Duarte I C, Santo G C, Sargento-Freitas J and Castelo-Branco M 2022 The neurophysiological impact of subacute stroke: Changes in cortical oscillations evoked by bimanual finger movement *Stroke Research and Treatment* **2022**

[76] Lee M, Park C H, Im C H, Kim J H, Kwon G H, Kim L, Chang W H and Kim Y H 2016 Motor imagery learning across a sequence of trials in stroke patients *Restorative Neurology and Neuroscience* **34** 635–45

[77] Maaten L V D and Hinton G 2008 Visualizing data using t-sne *Journal of Machine Learning Research* **9** 2579–2605

Appendix A. Analysis of Relevance Patterns and Feature Distributions for Selected Stroke Patients Across Subject-Specific Stroke, Stroke-to-Stroke, and Healthy-to-Stroke Models

In addition to conducting group-wise relevance pattern analyses, we also explored the subject-level patterns of a selected number of stroke patients who demonstrated good performance using our proposed H-to-S transfer learning. We also used t-SNE [77], an unsupervised visualization algorithm that reduces dimensionality to visualize the feature distribution of the respective model. Well-separated groups of data points, associated with different classes, indicate that the model has good classification efficacy. Figure A1 displays the relevance topoplots of these selected stroke patients, which are labelled with their corresponding subject ID and affected limb. Furthermore, the accuracies for these individual subjects using the three models are also provided. Figure A2 contains t-SNE feature distribution plots for the same set of subjects, across all three models. We would like to note that only the adapted or transfer learned models were used to visualize these selected patient patterns, as they outperformed their corresponding non-adapted counterparts in stroke MI detection.

Appendix A.0.1. Subject 3: Subject 3, a stroke patient with left affected limb, obtained high accuracies of 87.18%, 92.31%, and 93.59%, respectively, in the subject-specific, S-to-S, and H-to-S classification. Nonetheless, the relevance patterns observed from the subject-specific model were found to lack neurophysiological plausibility compared to the transfer models.

The S-to-S transfer learned model shows higher MI contributions in motor electrodes in the ipsilateral hemisphere (C3 and CP3) and parietal electrode in the contralateral hemisphere (P8), along with mild positive relevances seen in the frontal channel Fz and contralateral motor channels. On the other hand, the H-to-S adapted model achieved a similar, in fact slightly higher accuracy compared to the S-to-S model, but indicates high positive relevance for MI mainly on parietal electrodes, P7, P8, and Pz, with a mild positive relevance observed on the ipsilateral central channel C3.

These differences in channel relevances between the two transfer models, along with their high performances, suggest that the H-to-S models can consider cortical areas different from those of S-to-S transfer models as relevant for MI detection, yet achieve a similarly high performance. However, it is worth noting that P8 was identified as one of the strongest positively contributing electrodes for MI detection by both S-to-S and H-to-S models, highlighting the influence of the parietal region of the brain during stroke MI. Furthermore, we cannot deny the impact of the frontal area of the cortex, as is evident from the relevance and the accuracy observed in the subject-specific model.

Furthermore, the corresponding t-SNE feature distributions for subject 3 in all three models show highly discriminative and well-separable features between the MI versus rest classes. The separability is especially more pronounced in the transfer models.

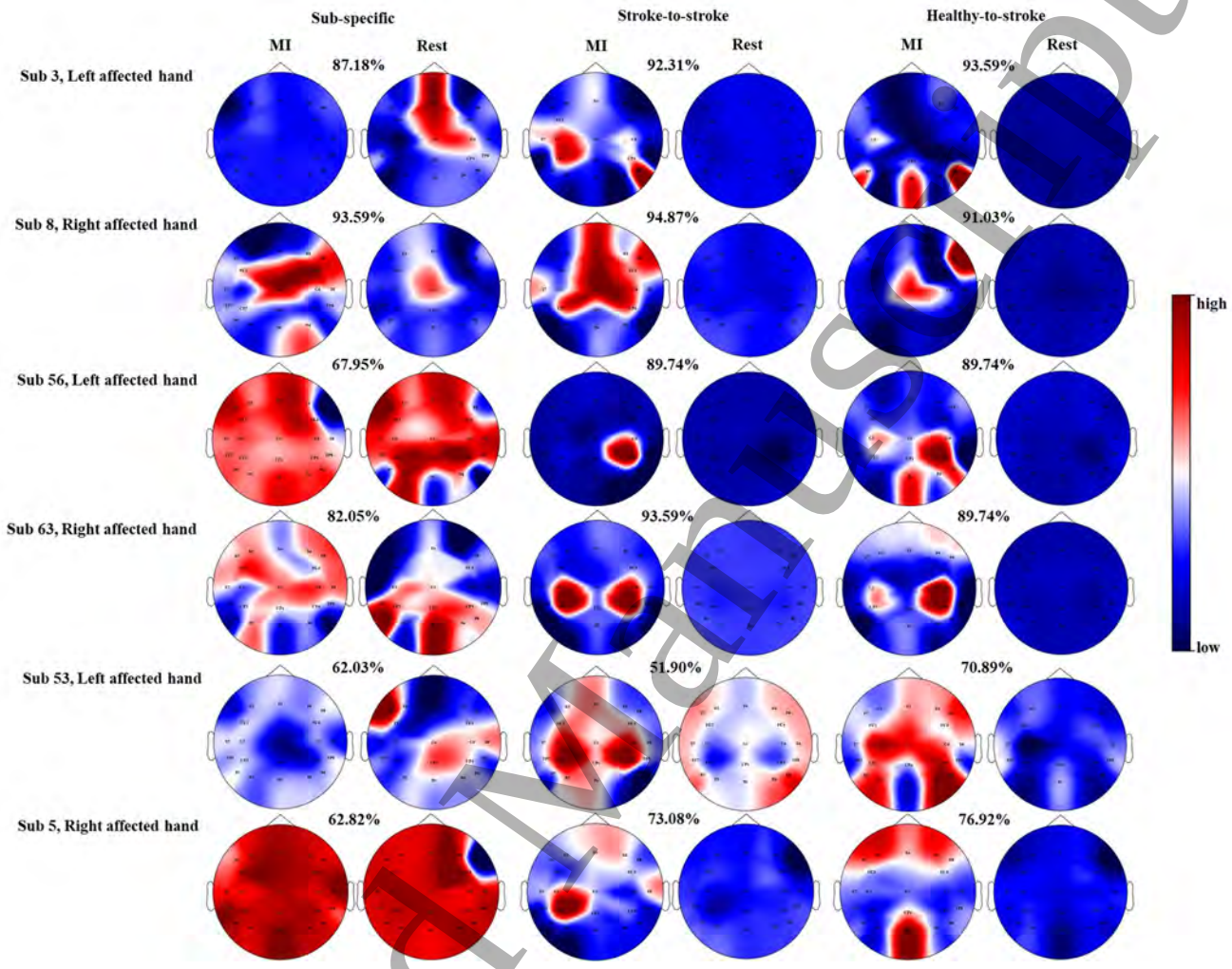

Figure A1: Relevance patterns in selected representative stroke patients, across subject-specific, stroke-to-stroke, and healthy-to-stroke models. Only the transfer learned stroke-to-stroke and healthy-to-stroke models were used in this analysis.

Appendix A.0.2. Subject 8: Subject 8, who had a stroke affecting the right limb, achieved impressive accuracies of 93.59%, 94.87%, and 91.03% using the subject-specific, S-to-S, and H-to-S classification models. The subject-specific model reveals positive relevance for MI in the central motor channel Cz, frontal-central channels FC4 and FC3, frontal channel F8, and the parietal channel P4.

The S-to-S transfer learned model exhibits greater MI contributions in the motor electrodes of both hemispheres (CP3, CP4, C4, Cz) and frontal electrodes Fz, F8, and FC4. The H-to-S model attained an accuracy that was 3.84% lower than that of the S-to-S model, and demonstrates high positive relevance for MI on the central electrode Cz and the ipsilateral frontal electrode F8.

The subject-specific and S-to-S models achieved higher accuracies compared to H-to-S model, likely due to their attribution of positive relevance to channels from bilateral hemispheres, albeit from different regions. It is noteworthy that all three

models identified F8 and Cz as positively contributing electrodes for MI detection in subject 8.

Additionally, the t-SNE feature distributions for subject 8 across all three models display a clear separation of features between the MI and rest classes.

Appendix A.0.3. Subject 56: Subject 56, a stroke patient with left affected limb, achieved accuracies of 67.95%, 89.74%, and 89.74% in subject-specific, S-to-S, and H-to-S classification, respectively. The relevance patterns observed from the subject-specific model are diffuse, with positive relevance for MI being more pronounced in the parietal and frontal areas. However, the S-to-S model exhibits clearly localized positive relevance for MI in the contralateral centro-parietal electrode CP4. On the other hand, the H-to-S model achieved the same accuracy as the S-to-S model but displays high positive relevance not only in CP4 but also in the middle and contralateral parietal electrodes, Pz and P8, with mild positive relevance also observed in the ipsilateral central channel C3. These findings restate the impact of bilateral motor channels and the middle parietal electrode, Pz, on MI detection in stroke patients.

The observed differences in channel relevances between the S-to-S and H-to-S transfer models, despite their equal accuracies, reiterate the fact that the H-to-S models may identify cortical regions not considered relevant by the S-to-S models for MI detection. Nonetheless, both transfer models identified the contralateral central channel C4 as the most relevant for MI detection, consistent with typical observations during MI in healthy individuals. Additionally, the importance of the parietal region of the cortex cannot be disregarded, as evidenced by the relevance patterns in the subject-specific and H-to-S models.

Moreover, the t-SNE feature distribution for subject 56 is less distinguishable in the subject-specific model compared to the two transfer models, reflecting the performance disparity in MI detection among them.

Appendix A.0.4. Subject 63: Subject 63, who had a stroke affecting the right limb, achieved high accuracies of 82.05%, 93.59%, and 89.74% in subject-specific, S-to-S, and H-to-S classifications, respectively. However, the relevance patterns observed in the subject-specific model are distributed across various cortical areas and lack neurophysiological plausibility compared to the transfer models.

The S-to-S model exhibits greater MI contributions in the motor electrodes of the bilateral hemispheres, C3, C4, CP3, and CP4. However, the H-to-S model obtained an accuracy 3.85% lower than the S-to-S model and identifies positive relevance for MI primarily on the ipsilateral central electrodes C4 and CP4, with some mild relevance from the contralateral central electrodes C3 and CP3.

The differences in channel relevances observed between the two transfer models, along with their respective performances, once again emphasize the importance of motor areas in both hemispheres for MI detection in stroke patients. It is noteworthy that the

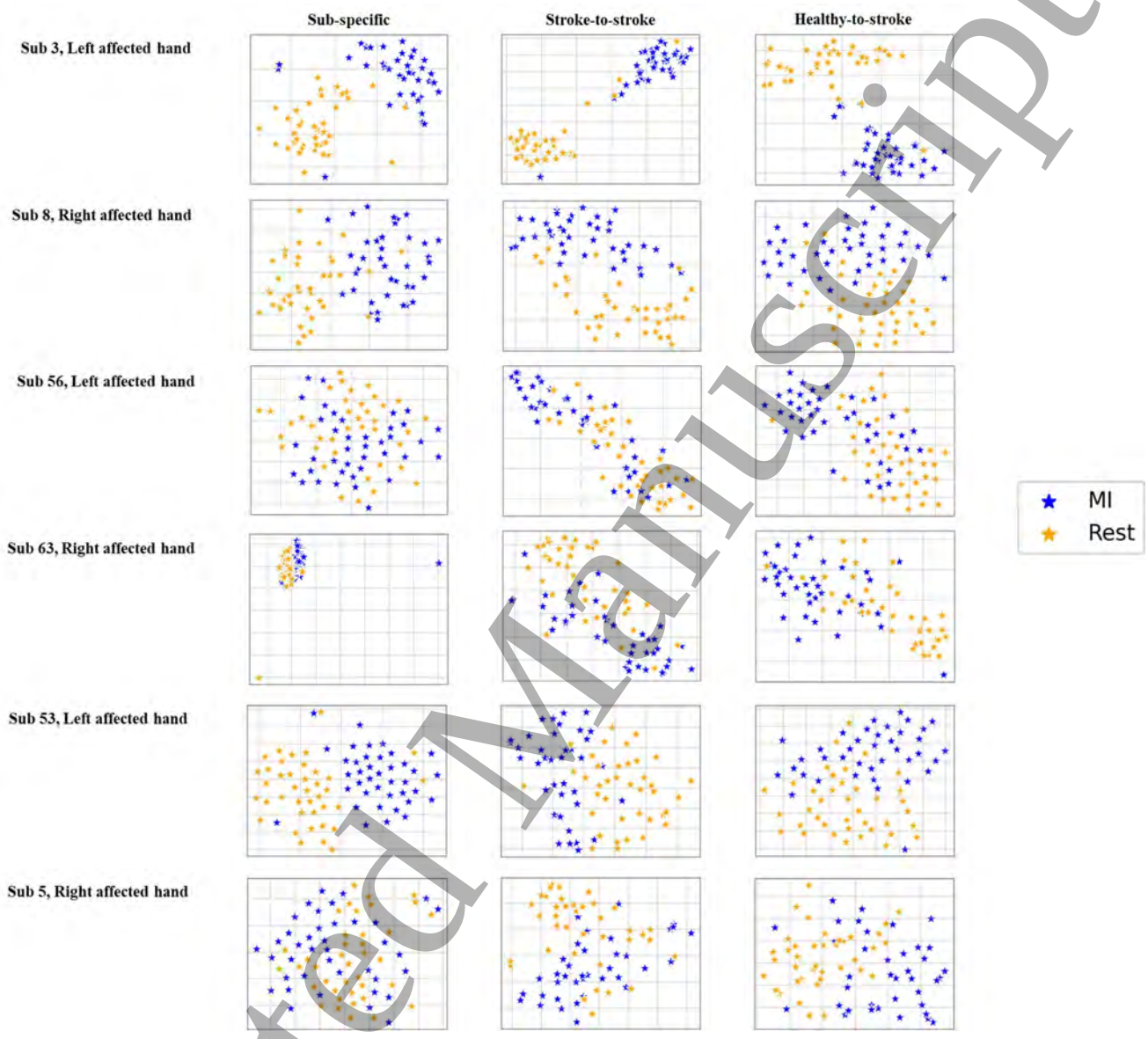

Figure A2: Feature distribution plots for selected representative stroke patients using t-SNE [77]. Only the adapted stroke-to-stroke and healthy-to-stroke models were used in this analysis.

relevance patterns of the subject-specific model highlight contributions from the parietal areas, as observed typically in H-to-S models.

Subject 63's t-SNE feature distribution in the subject-specific model reveals a fair separation of features between MI and rest classes, although the transfer models exhibit greater separability.

Appendix A.0.5. Subject 53: Subject 53, who had a left affected limb, achieved accuracies of 62.03%, 51.90%, and 70.89% using subject-specific, S-to-S, and H-to-S models, respectively. The relevance patterns observed from the subject-specific model

did not exhibit any substantial positive contribution for MI detection from any cortical region.

In contrast, the S-to-S adapted model shows strong MI contributions in bilateral motor electrodes, C3, CP3, C4, and CP4, particularly in the centro-parietal electrodes CP3 and CP4, with some positive relevance observed in the middle frontal and the middle parietal areas of the cortex (Fz and Pz). Additionally, the rest class patterns complement the MI class patterns. Despite this, the S-to-S model was unsuccessful in detecting MI for subject 53 and only achieved chance-level accuracy.

On the other hand, the H-to-S transfer learned model achieved good performance compared to the S-to-S model, with high positive relevance for MI coming from bilateral central channels, particularly the ipsilateral central channel C3 and contralateral centro-parietal channel CP4, as well as central channel Cz, bilateral parietal channels P7, P3, P4, and P8, and frontal channels Fz, F4, and F8. Thus, H-to-S classification once again highlights the bilateral influence during MI performance in stroke patients, in both motor and parietal areas of the cortex.

Interestingly, unlike subject 56, who also had a left affected limb, subject 53 shows a pronounced negative relevance in the parietal channel Pz. Furthermore, the localization pattern shows that while the ipsilateral relevance is from both central and centro-parietal channels, in the contralateral hemisphere, the relevance is mostly from the centro-parietal electrode CP4 than the central electrode C4. This observation is similar to what was noticed in the S-to-S group-wise average relevance patterns shown in Figure 8.

The performance differences and channel relevances observed in the two transfer models suggest that the parietal influence identified by the H-to-S model could be a significant factor contributing to the superior MI detection performance of our proposed method in subject 53. Furthermore, the t-SNE feature distributions for subject 53 reveal good separability of MI versus rest class features in both the subject-specific and H-to-S adapted models, when compared to the S-to-S adapted model.

Appendix A.0.6. Subject 5: Subject 5, who had a right affected limb due to stroke, achieved accuracies of 62.82%, 73.08%, and 76.92% in the subject-specific, S-to-S, and H-to-S classification scenarios, respectively. The relevance patterns observed in the subject-specific model are not well-defined.

In contrast, the S-to-S model shows high MI relevances in the contralateral centro-parietal electrode CP3, along with mild positive relevances in frontal channels Fz and F4, and temporal channel T8. Meanwhile, the H-to-S model demonstrated a 3.84% higher accuracy than the S-to-S model, with negative relevance for MI in bilateral central channels and positive relevances from bilateral frontal channels. The model also shows a strong positive contribution from the parietal electrode Pz, which was also observed for other stroke patients evaluated using the H-to-S model.

The varied perspectives of the two transfer models in determining the neurophysiological basis for MI detection in stroke patients are highlighted by the

differences in channel relevances. The t-SNE feature distribution plots, in Figure A2, illustrate comparatively better class separability using the H-to-S transfer model for subject 5.