A machine learning approach to identify the engagement of a brain network targeted by non-invasive brain-stimulation

Authors: Shinde A. B.[1, 2, 3 *], Mohapatra S.[2,4*], Schlaug G.[1, 2, 3]

*Contributed Equally

Institutions:
[1]Department of Biomedical Engineering, University of Massachusetts Amherst, Amherst, MA -01003
[2]Institute of Applied Life Sciences, University of Massachusetts Amherst, Amherst, MA - 01003
[3]Department of Neurology, Baystate Medical Center and UMass Chan Medical School - Baystate Campus, Springfield, MA - 01107
[4]Manning College of Information & Computer Sciences University of Massachusetts Amherst, Amherst, MA - 01003

Keywords
concurrent tDCS-fMRI, Brain stimulation, Dynamic functional connectivity, Machine learning

Corresponding author
Gottfried Schlaug
Email: gschlaug@gmail.com

Acknowledgement
This research was supported by an NIH BrainInitiative grant (RO1MH111874). Dr. Schlaug also acknowledges support from U01NS102353. SM acknowledges the support from IALS summer internship program. We are thankful for the generous support that Klaus Schellhorn from neuroConn has provided to us with making a state-of-the-art multi-channel MR DC stimulator available to us and for being available for any trouble shooting over the years. We would also like to thank Elena Bliss (MR technologists UMass Amherst) for her support with the concurrent tDCS-MR acquisitions.

Abstract (Limit To 175 words)

Transcranial direct current stimulation (tDCS) can noninvasively modulate behavior, cognition, and brain functions depending on polarity and dose of the stimulation as well as electrode montage. Concurrent tDCS-fMRI presents a novel way to explore the parameter space of non-invasive brain stimulation and to inform if a targeted brain region or a network of spatially distributed brain regions are engaged. In this study, a multi-electrode montage targeting nodal regions of the Arcuate Fasciculus Network (AFN) was compared to a single electrode montage and a no-stimulation control condition. Machine learning methods were applied to classify rs-fMRI data with the help of dynamic functional connectivity (DFC) values into three different categories achieving an overall accuracy of 92% with the highest obtained for multi-electrode condition. The top 5 ROI pairs contributing towards the classification were identified using model performance parameters and were mainly within the right AFN. A proof-of-concept real-time feedback system to identify AFN engagement using DFC values to predict accuracy of stimulation electrode placement is proposed.

Introduction

Experimental studies revealed evidence that transcranial direct current stimulation (tDCS) modulates brain activity and affects behavior, mood, and cognition that draws on targeted brain regions[1–4]. tDCS applies a constant electric current to targeted brain regions through their scalp access points typically using one or more anodal electrodes and one or more cathodal electrodes[5,6]. The current induces subthreshold changes across the neuronal membrane potentially modulating the neuronal firing depending on the polarity and dose of the current as well as the montage of electrodes. In general, brain regions targeted with anodal tDCS shows increased excitability while cathodal tDCS decreases it [3,7–9].

Concurrent tDCS-MRI has been used to reveal neural correlates of stimulation varying dose and montage to test whether or not targeted brain regions can be engaged, and their connections be modulated[5,10]. Several MRI methods such as regional blood flow and metabolic measures as well as resting state functional magnetic resonance imaging (rs-fMRI) and their cross-correlations across multiple brain regions have been used[11].

Here we are using machine learning approaches to (1) determine if a brain has been engaged by targeted non-invasive brain-stimulation, and (2) assess differential effects depending on electrode montage using concurrent tDCS-fMRI techniques. We chose a challenging but accessible network of brain regions that spans the frontal, parietal, and temporal lobes and is connected by a structural backbone in the underlying white matter, the Arcuate Fasciculus (Figure 2b). Two different electrode montages targeting this Arcuate Fasciculus Network, one with multiple electrodes and one with a single electrode (SE) are compared to a No-Stimulation (NS) control condition. Artificial intelligence (AI) methods are used to: 1) classify subjects 4D rs-fMRI brain images into three different categories corresponding to the 3 conditions mentioned above, 2) determine the electrode montage and ROI pairs that lead to the most accurate classification, and 3) develop a real-time feedback system that would inform the experimenter and potentially the subject whether a particular montage did engage a targeted brain network and modulated the activity between regions.

Methodology

Participants

A total of thirty-three (33) subjects were recruited in the greater Boston area as well as in the greater Amherst area. Subjects (n = 12) in the greater Boston area were scanned on a 3T GE scanner at Beth Israel Deaconess Medical Center (BIDMC) while subjects recruited in the greater Amherst/Pioneer Valley region (n = 21) were scanned on a 3T Siemens wide bore scanner. All subjects were right-handed according to the Edinburgh Handedness Inventory (Oldfield, 1971) and had no history of neurologic or psychiatric conditions and had no contraindications to undergo MRI or tDCS as verified by a safety checklist. The Institutional Review Boards of Beth Israel Deaconess Medical Center and University of Massachusetts Amherst approved this study, and all subjects gave written informed consent. All subjects underwent one mock session prior to the concurrent tDCS-MRI sessions in which they were familiarized with the setup and were given a test stimulation of the 4mA stimulation to make sure that they were able to tolerate the intensity of stimulation and were comfortable in the MR scanner environment. Subjects were then assigned to one of three concurrent tDCS-MRI sessions as their first real tDCS-MRI session, which was followed by either one of the remaining two sessions. Not all subjects were able to participate in all sessions. To ensure that there was not a systematic difference between the BIDMC and UMass MRI sessions due to factors not controlled for, an initial analysis compared the BIDMC no-stimulation (NS) sessions (n = 5) with the first 5 UMass NS sessions. A broad resting-state interhemispheric functional connectivity measure was calculated as the simple average of all possible interhemispheric ROI connections for every session, and a t-test was conducted. This test did not show a significant difference between sites, with $p>0.05$, and this result did not change as more UMass NS sessions were added to the now unequal-n t-test. Therefore, the BIDMC and UMass data were used as a single dataset for all further analyses.

Electrode placement

Prior to a subject entering the MRI room, we placed MR compatible rubber electrodes on each subject's scalp using the 10-20 EEG system as a guide for the initial scalp targets. In the multielectrode (ME) condition, we placed three

round anodal electrodes (diameter of 3cm each) over the scalp position of the supramarginal gyrus (SMG), the posterior inferior frontal gyrus (IFG) and the posterior STG/MTG junctional region. The IFG electrode location was placed halfway between C6 and F8, the SMG electrode location was placed one third of the distance between C6 and CP4, and the STG/MTG electrode location was placed halfway between CP6 and TP8. A round cathodal electrode (diameter of 5cm) was placed over the contralateral supra-orbital region (corresponding to FP1). In the single electrode (SE) condition, we placed a round electrode (diameter of 4cm) over either the SMG, the IFG, or the posterior STG/MTG junction and the cathodal electrode again over the left fronto-orbital region (FP1). After cleaning the targeted scalp locations with alcohol, rubber electrodes lathered with approximately 2mm thick layer of Ten20 neurodiagnostic electrode paste (Weaver and company, Aurora, CO, USA) were placed. Electrodes were held in place using a self-adhesive bandage and hypoallergenic medical tape. Electrode connectors were adjusted to avoid any crossover between wires. The no-stimulation (NS) data was acquired while participants had either single or multi electrode stimulation montage placed over their head and were informed that stimulation might be applied but was not actually applied in that sessions. Seven participants out of 33 participants from the no-stimulation (NS) group, participated in an MRI session where no electrodes were placed over the participant's scalp; these participants were naive to the stimulation at the time.

Concurrent tDCS-fMRI

Subjects were placed inside the MRI bore in supine position and electrodes were connected to the connector box. The stimulation signal was transferred from the MR control room to the MRI suite through the RF filter panel as shown in Figure 1a. A T1-weighted 3D MPRAGE sequence with sagittal acquisition and 2mm slice thickness (resolution= $1.0 \times 1.0 \times 2.0\ mm^3$; TR/TE 1490/3.36 ms; flip angle = 9 degrees; matrix = 256×256; field of view = $256 \times 256\ mm^2$) was obtained in a short scanning session to confirm the anatomical accuracy of the scalp-placed electrodes. If the electrodes were placed inaccurately (meaning that they did not overlay the target region), then the subject was moved out of the MRI bore and the electrode position was adjusted, and the corrected placement was again confirmed with a fast T1-weighted acquisition. When electrodes were accurately placed, a 24-min rs-fMRI scan using a gradient-echo echo planar imaging sequence (TR of 3s, TE of 31.0 ms, flip angle of 90 degrees, and voxel size of $2.5 \times 2.5 \times 2.5\ mm^3$) was obtained. The stimulator was turned ON and OFF during this 24-minute scan with ON periods being 4-minutes long (with a 30 second ramp-up and ramp-down) and OFF periods 3-minutes long, flanking the ON-stimulation at the beginning and the end (see Figure 1). The total applied tDCS current was 4mA in the SE and ME montages; in the SE montage, all 4mA was applied through a single electrode while in the ME montage, 2mA was applied to the posterior STG/MTG junction and 1mA each was applied to the SMG and posterior IFG electrodes.

Applying tDCS

The stimulation was applied using the multichannel MR compatible Neuroconn DC-MC device (Neurocare, Germany). The Neuroconn stimulator device was placed in the MR control room where the stimulation signals were generated. Three channels were programmed to concurrently generate 1mA and 2mA currents for the multi-electrode condition or 4mA in a single channel output for the single electrode condition (Figure 1a). In each of the concurrent tDCS-MRI sessions, we alternated between the stimulation OFF and ON conditions, starting with the OFF condition. Three 4-minute epochs of stimulation were applied during the 24-minute resting state fMRI. Figure 1b shows the timing diagram of the stimulation and no-stimulation epochs during one rs-fMRI acquisition of 24 minutes.

Dynamic functional connectivity matrices

Although we recorded rs-fMRI with three stimulation epochs interspersed by non-stimulation epochs, here we focus on the first stimulation epoch, especially the correlation matrices generated from 4D rs-fMRI image sets acquired after the first stimulation onset. ML models designed in this case rely on feature vectors to perform classification. Hence, we used a vector representation of a dynamic functional connectivity (DFC) matrix as input feature vector. DFC is a method of calculating variations in functional connectivity between iterative samples of two regions generating a DFC matrix for a selected window length, then moving the time window by a step size to calculate the next DFC matrix. A DFC matrix for a selected time window is calculated in two steps- 1) extract (or derive) the raw fMRI signal from the

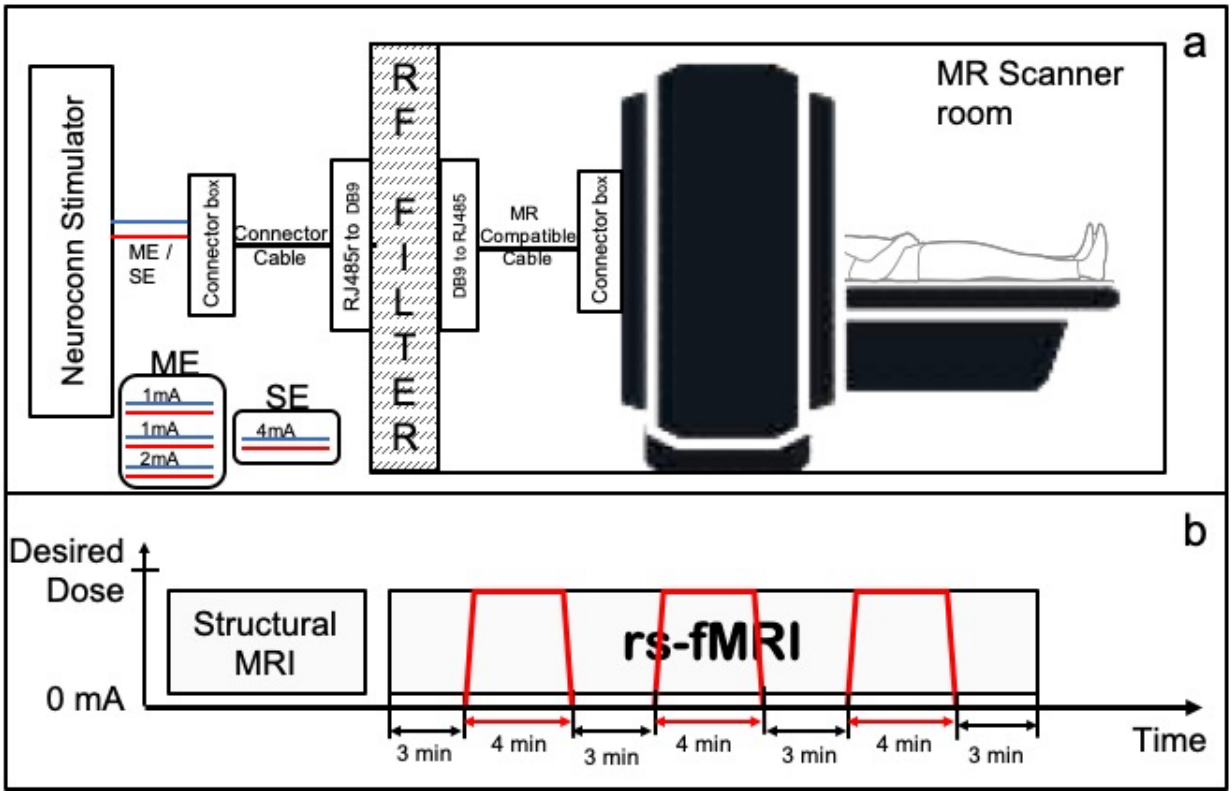

Figure 1: a) Concurrent tDCS-fMRI setup. b) Timing diagram for MRI experiment. Red colored lines show the application of tDCS during the 24-minute long rs-fMRI scan.

time-course for a region/s of interest (ROI) and 2) correlate that time-course signal from each ROI with all of the other ROIs of interest in a correlation matrix. We selected 1 minute as a window period and the step size of one acquisition volume (approximately 3 seconds). Figure 3 shows the DFC calculation, window size and step size.

A DFC for the no-stimulation conditions was calculated using the onset time for the pretend-stimulation as 3 minutes after the start of the acquisition to mimic the onset times of the SE and ME conditions. Data from the first 2 minutes of acquisitions after this real or pretend onset time was used to generate the DFC matrices and consequently the NS featured class vectors. To identify the best stimulation onsets from the rs-fMRI data, we explored three different groups of data: first 5, first 10, and first 20 DFC matrices after the stimulation onset. These DFC matrices were used to generate the featured class vectors which were then grouped into two categories: single electrode stimulation onset, and multi-electrode stimulation onset. The groups of data from the stimulation sessions were then individually grouped with the NS featured class vectors resulting in a cumulative dataset of three classes.

Selection of ROI pairs

The standard Harvard-Oxford atlas[11–14] with 112 regions of interest was used to calculate a 112×112 DFC correlation matrix. Because of the structure of the correlation matrix, each value is replicated and the values on the diagonal axis are all '1.' We converted each DFC matrix into a vector of 1×6216 unique ROI pairs.

The large size of the feature vector can increase time requirements for preprocessing and classification. To develop a real time feedback system and reduce the data processing time, we devised a strategy to reduce the number of anatomical ROIs. Fifteen (15) spherical ROIs sampling brain regions of the perisylvian AF-network and 4 spherical ROIs corresponding to the ILF-network were selected on the right as well as the left hemisphere. This resulted in a 38 x 38 DFC correlation matrix with a feature vector size of 1×703. To reduce the size of the feature vector even further, we also tested a total of 26, 22, 16 ROIs corresponding to the vector sizes 1×325, 1×231, and 1×120.

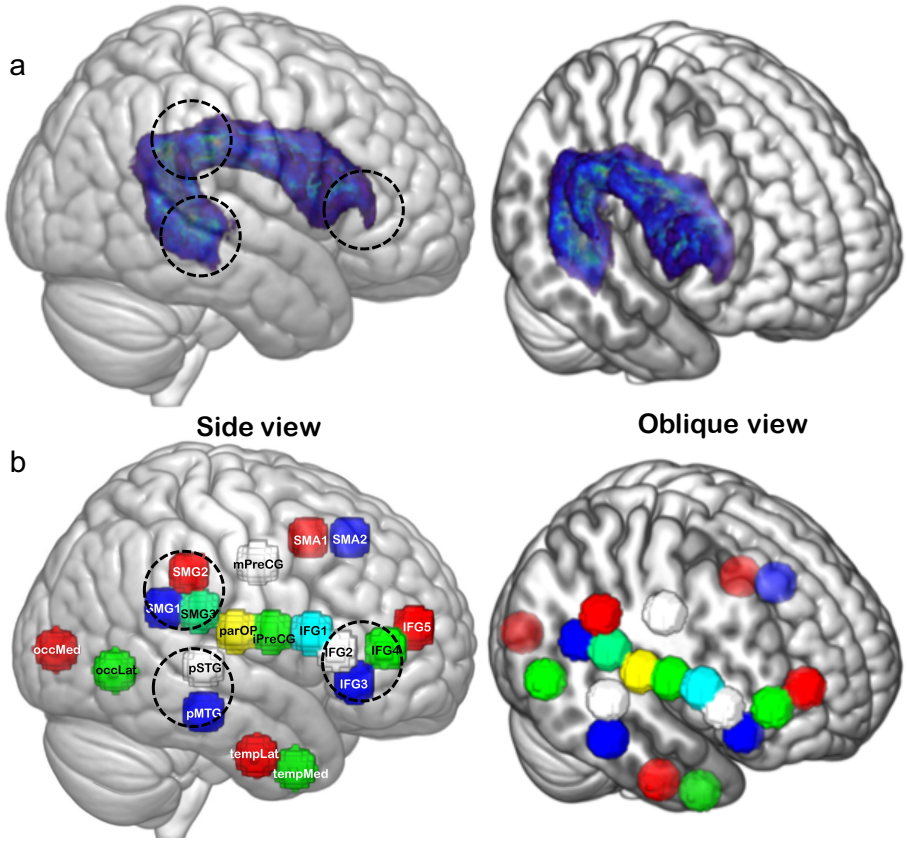

Figure 2: (a) Representative image of the right hemispheric Arcuate Fasciculus tract shown in standard brain and the canonical location of three nodal regions of the AF network targeted by brain-stimulation. (b) Anatomically identified spherical ROIs representing cortical regions within and outside of the AF network (shown here for the right hemisphere, but mirror left hemisphere regions were identified as well). Side and oblique views are shown to demonstrate the mesial location of AF tract and ROIs. Three dotted circles represent nodal end points of AF tract where the anodal electrodes for ME montage are placed.

Pre-processing dataset

fMRI to feature vector calculation

We used a graph theoretical network analysis toolbox called GRETNA for preprocessing and calculation of DFC matrices[15]. First, DICOM images were converted to 3D (T1w images) and 4D (rs-fMRI) NIfTI (Neuroimaging Informatics Technology Initiative) datasets using dcm2niix[16]. Since the BOLD signal shows some T1-saturation effects in the first volume acquisitions, we excluded the first minute of the rs-fMRI data from the analysis. This was followed by preprocessing steps such as slice timing correction, realignment of rs-fMRI images, spatial normalization to the standard MNI space using the dartel segmentation method, smoothing, detrending fMRI data, and regressing out covariates such as white matter signal, CSF signal and head motion. The regression was carried out using a published strategy[17]. After the DFC matrices were generated, we used a custom program developed in MATLAB to get the feature vector for each subject and each set of ROIs to be used in the ML part.

Machine learning approach

Dataset preparation

The prepared dataset is from 33 participants contributing 20 no-stimulation sessions, 25 single-electrode sessions, and 24 multi-electrode sessions. Not all participants were able to participate in all three sessions. The machine learning models may effectively memorize the training dataset if the dataset is small[18]. It means that the generalized learning of mapping from inputs to outputs is avoided, and the models will learn the specific inputs and the associated output

results[19]. One of the approaches used in such a case is to add noise to the training dataset, which improves the generalization error and the structure of the mapping problem[20]. Here, the generalization error refers to the measure of how accurately the algorithm can classify the categories for any data that is not provided in the training data. Due to the low number of sessions across all subjects for the three categories, classification models were either performing on the lower side or getting overfitted. To mitigate this problem, we used the classical method of increasing the dataset through upsampling along with the previously mentioned method of adding noise into the training data[21,22]. Possible outliers in the feature vectors that could create bias were removed using DBSCAN clustering[23]. This was followed by randomizing the entire set of vectors. We divided the prepared dataset into two different parts: 80% of the entire dataset was used for the training of the models and the remaining 20% was used for validation and testing of the model.

Classification algorithm

The machine learning (ML) models random forest, k neighbors' classifier, naive bayes, decision tree classifier, gradient boosting classifier were used on the feature vectors from the DFC matrices. The objective of the ML models is to classify the DFC matrices into three categories: NS, SE, and ME[24].

Ten-fold cross-validation was deployed while training the models. The use of multiple models has helped us compare performance across different models along with a varying number of volumes.

Independent validation

To validate the results of the classifier and to assess the performance of the classification model, we focused on three parameters: Receiver Operating Characteristic (ROC) score, Matthews Correlation Coefficient (MCC), and Accuracy (based upon the test dataset). The ROC score shows the relationship between sensitivity and specificity for every possible cut-off when performing a combination of predictions. The MCC score calculates the difference between the actual values and predicted values of the classification which is equivalent to the chi-square statistics for a 2×2 matrix table[25]. While evaluating the models using different parameters, we have also considered each model's sensitivity and specificity. Accuracy was calculated using the ratio of true positives and true negatives against all predictions.

Model interpretation using Shapley Additive Explanation (SHAP)

There are various methods to estimate the importance of the attributes. In tree-based models like RF and decision tree, attribute importance is pre-assigned which greatly influences the output. The contribution of an attribute is completely different from its importance towards the output. The importance is based on the quantitative approach that determines which attributes significantly drive the model's performance. Whereas the contribution of the attributes helps to understand an intuitive explanation for the output, beyond just identifying what attributes strongly affect the classification performance.

In this study, we have used the Gini index to interpret our model. We can determine how much a particular attribute contributes toward the average decrease in the error of the model classification. Each value from the feature vector represents a correlation between ROI pairs and Gini index and helps us identify the importance of each ROI pair or an attribute. However, this method is insufficient in explaining how individual attributes affect the prediction.

Therefore, we also used the SHAP method to determine quantitatively how each attribute contributes to the model's performance[26,27]. The SHAP uses the game theory concept to calculate the contribution of each of the attributes combined with the prediction model and explanation model using the various methods. Using the SHAP method we can determine the contribution of each attribute to the model classification.

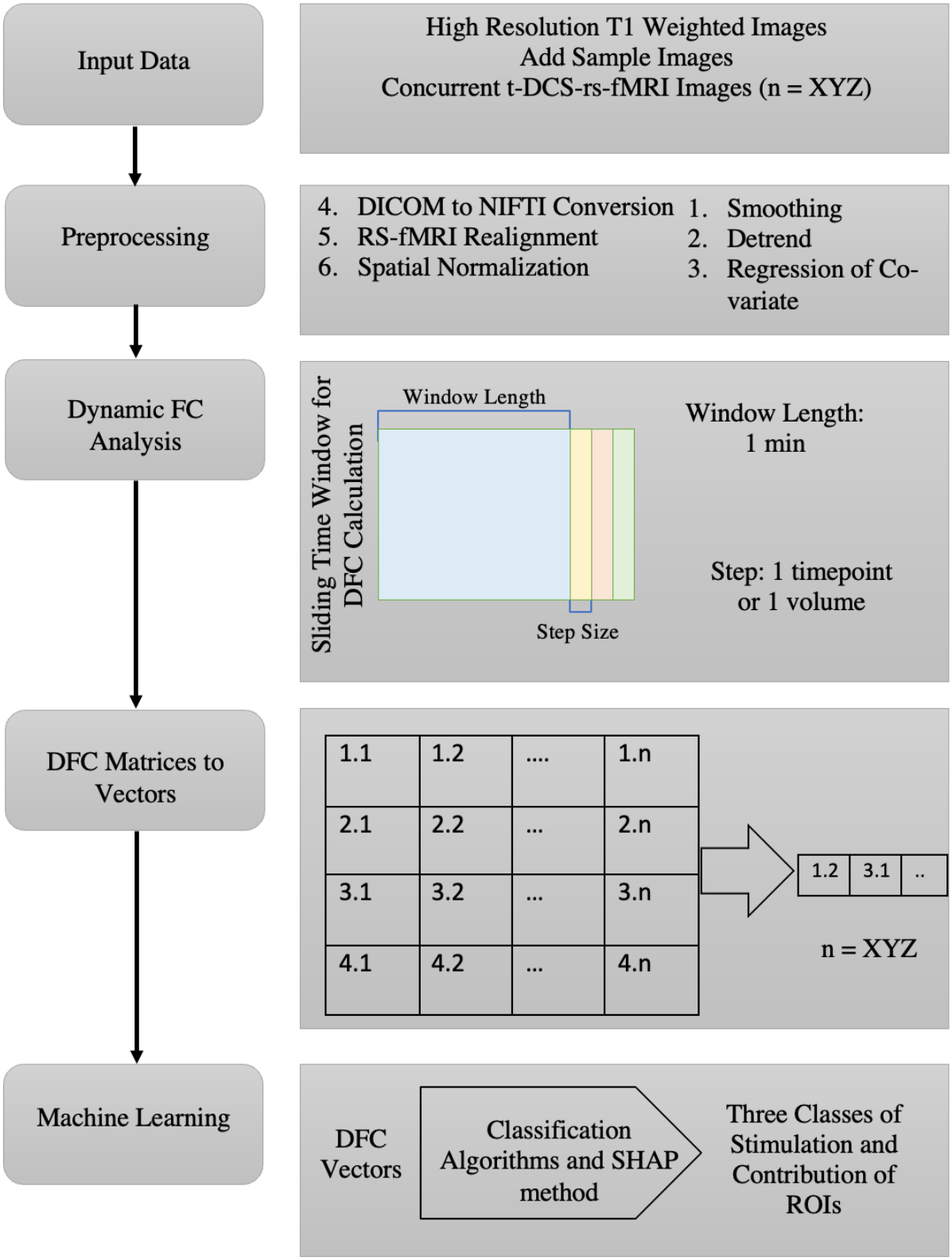

Figure 3: Data preprocessing flow-chart

Proof-of-concept feedback system

A classification model described above will be tested to identify the ML model capable of identifying the effects of ME and SE stimulation on the fMRI data. Additionally, SHAP method and Gini index interpretation will be used to identify the top ROI pairs that contribute strongly towards this classification.

This was done as follows: whenever a new fMRI dataset was presented to the system, it was processed to calculate the DFC matrices using GRETNA software which would then be converted to the feature vectors and fed as input to the classification model. The pre-trained classification model would inform if the fMRI data showed responses that correspond to the effects generated with ME and SE stimulation. Additionally, top contributing ROI pairs can be compared against the top contributing pairs generated while training the model.

This proof-of-concept feedback system is designed to help in the evaluation of unknown (newly acquired) fMRI data using classification model and corresponding interpretation indices which can inform about the stimulation and similarity of the top contributing ROI pairs with the ROI pairs that strongly influence the classification in validation. Using this information, stimulation parameters can be adjusted to further invoke the desired responses in fMRI.

Results

Evaluation of classification performance

ROC and MCC score

Table - 1: ROC and MCC across different ROIs for the classification shown by various models

Model	No. of ROIs - 112		No. of ROIs - 38		No. of ROIs - 26		No. of ROIs - 22		No. of ROIs - 16	
	ROC	MCC	ROC	MCC	ROC	MCC	ROC	MCC	ROC	MCC
K Neighbors Classifier	0.6822	0.6402	0.9055	0.8593	0.9099	0.8697	0.9247	0.8459	0.8898	0.8213
Random Forest Classifier	0.7321	0.6789	0.8803	0.8207	0.8889	0.8124	0.9012	0.8411	0.8646	0.7984
Logistic Regression	0.7091	0.6211	0.8519	0.7794	0.8697	0.7985	0.8899	0.7456	0.7836	0.6732
Decision Tree Classifier	0.633	0.5154	0.7553	0.6338	0.7375	0.6741	0.8838	0.7982	0.7533	0.6173
Naive Bayes	0.654	0.4879	0.72	0.5856	0.7356	0.6569	0.7099	0.6428	0.5596	0.3215

The ROC score has been generated using the true positive rates and false positive rates of the models. The MCC score is used as a measure of the quality of binary classifications and has been generated using the formula based on true positives, false positives, true negatives, and false negatives. In table 1, we represent the ROC and MCC scores for each of the models. A score of 1.0 for the ROC indicates the model performs ideally. In the case of the k-nearest neighbor (knn) classifier, the ROC was consistently 0.90 (except when 112 ROIs or 16 ROIs were used) or higher when compared to the other models, this displayed the best performance. Although the Random Forest Classifier came close to the values of the ROC curve it was consistently lower compared to knn. A score of 1.0 for the MCC indicates the perfect agreement between the prediction and observation. MCC values for the knn classifier were close to 0.85 when 38, 26 and 22 ROIs were used to calculate feature vectors. Similar to ROC results, MOC values with random forest classifier were close to the knn performance but were slightly lower. Logistic regression, decision tree, and

naïve Bayes classifiers' performance was inferior to the knn or random forest classifiers irrespective of the number of ROIs used to generate feature vectors.

Classification accuracy

In table 2, we show the overall classification accuracy of different models with varying sets of ROIs. It is observed that the overall accuracy improved when the number of ROIs were reduced from 112 to lower numbers, and it was highest [0.9247] when using the dataset generated using 22 ROIs and k-nearest neighbor (knn) classifier.

Table - 2: Accuracy across different ROIs for the classification shown by various models

Model	No. of ROIs - 112	No. of ROIs - 38	No. of ROIs - 26	No. of ROIs - 22	No. of ROIs - 16
K Neighbors Classifier	0.683	0.9056	0.9099	0.9247	0.8833
Random Forest Classifier	0.731	0.88	0.8889	0.9012	0.8689
Logistic Regression	0.71	0.8518	0.8697	0.8899	0.7869
Decision Tree Classifier	0.63	0.7556	0.7375	0.8838	0.75
Naive Bayes	0.6544	0.722	0.7356	0.7099	0.5574

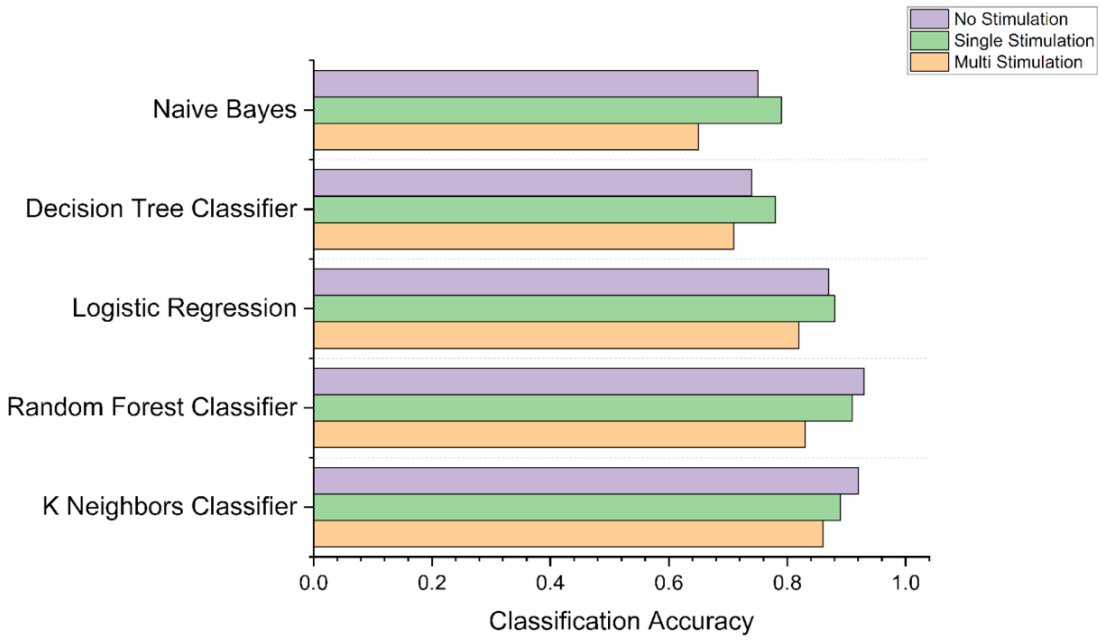

Figure 4: Accuracy of models to classify different stimulations

Figure 4 shows the ability of top 5 performing models to classify the three types of stimulations with feature vectors extracted from fMRI data with 22 ROIs. For each model, bar graphs represent classification accuracy for each class. It can be observed that in some cases, the knn classifier is not the one which classifies with highest accuracy. However, it is our best performing model because when taken overall, it shows highest accuracy. Figure 4 also shows that the

random forest model classifies the NS and SE with the highest accuracy values while the knn model classifies the ME category with the highest accuracy.

Analysis of the contributions and importance of attributes

For further analyses, we have selected the 22 ROI and 26 ROI sets because these sets showed the best performance in machine learning predictions and took comparatively less computational time.

In figure 5, the importance plot shows the model's ability to classify based on the Gini index for each attribute using 22 ROIs and 26 ROIs, respectively. The top ROI pair from the data with 26 ROIs has the Gini index of 0.174, and subsequently, the tenth most important pair has the Gini index of 0.031. The top ROI pair from the data with 22 ROIs has the Gini index of 0.246, and subsequently, the tenth most important pair has the Gini index of 0.019. Although the sum of the Gini index for all pairs is equal to 1, the top 5 ROI pairs in the 26 ROIs and 22 ROIs contribute more than 50% towards it.

The importance doesn't tell us about the contribution of each attribute towards a specific output. The SHAP summary plot's introduction helps in performing an in-depth analysis. From the figure 5b, it is observed that the top pair from the data with 26 ROIs has the highest contribution of 0.086 towards the model's output of classifying a vector into ME. In contrast, the second ROI pair has the highest contribution of 0.081 towards the model's output of classifying a vector into not ME. Similarly, figure 5d represents the SHAP values for pairs of 22 ROIs. It is observed that the top pair has the highest contribution of 0.118 towards the model's output of classifying a vector into ME. In contrast, the fourth and fifth ROI pairs have the highest contribution of 0.055 towards the model's output of classifying a vector into not ME. This shows how each ROI pair can affect a specific output by the model.

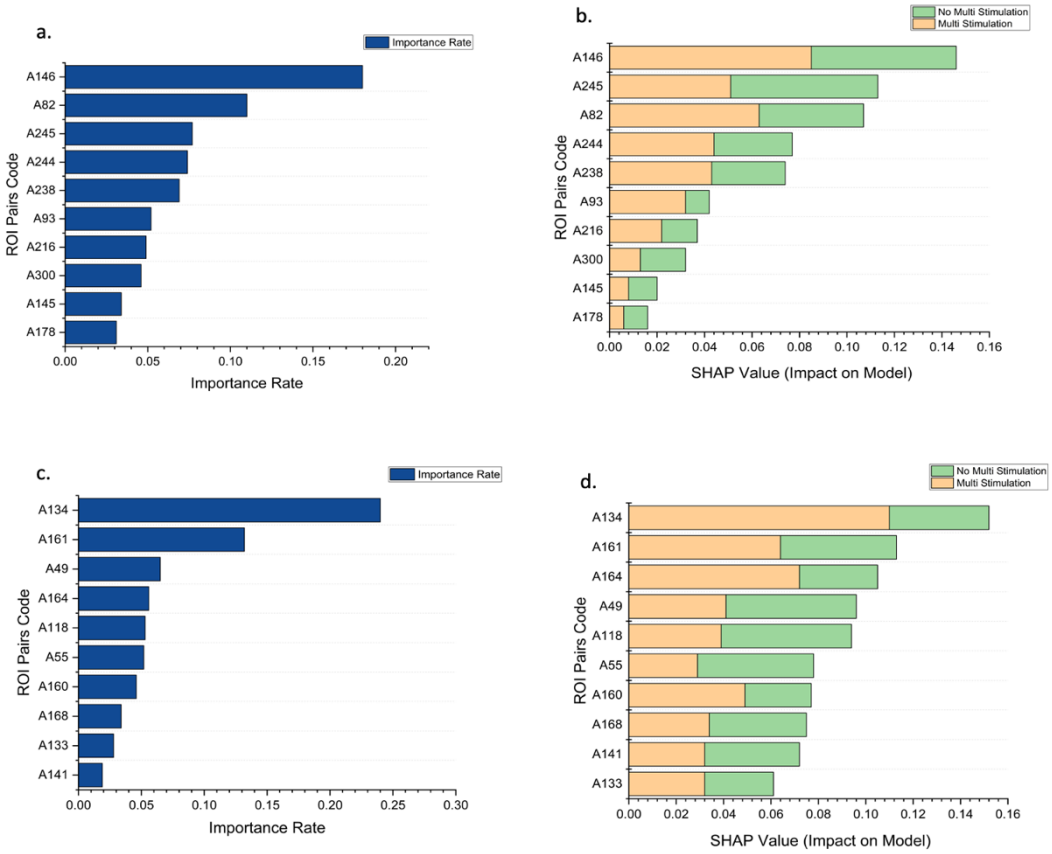

Figure - 5: Importance rate and contribution of ROI pairs towards classifying ME stimulation using Gini Index and SHAP value: (a & b) 26 ROIs and (c & d) 22 ROIs.

Time Taken for computation for processing of ROIs

We used a computer with 64 gigabytes of RAM, NVIDIA GeForce RTX 2080 of 4 gigabytes, and AMD Ryzen 9 5900X CPU with UBUNTU 18.04 operating system. The GRETNA program was installed on MATLAB 2021b (MathWorks, Natick, MA) which was used to carry out the entire processing pipeline including calculation of the DFC matrices. Time required for preprocessing steps was 11 minutes and 26 seconds and it did not differ with number of ROIs. In contrast, the calculation of DFCs showed exponential decrease in time required with a decrease in the number of ROIs. In figure 6, we see that the time taken for the computation of DFC calculation for different numbers of ROIs has inverse exponential relation with a decrease in the number of ROIs. The maximum time taken to calculate the DFC is 6minutes and 18 seconds with 112 ROIs whereas it came down to just 27 seconds when calculated for 16 ROIs.

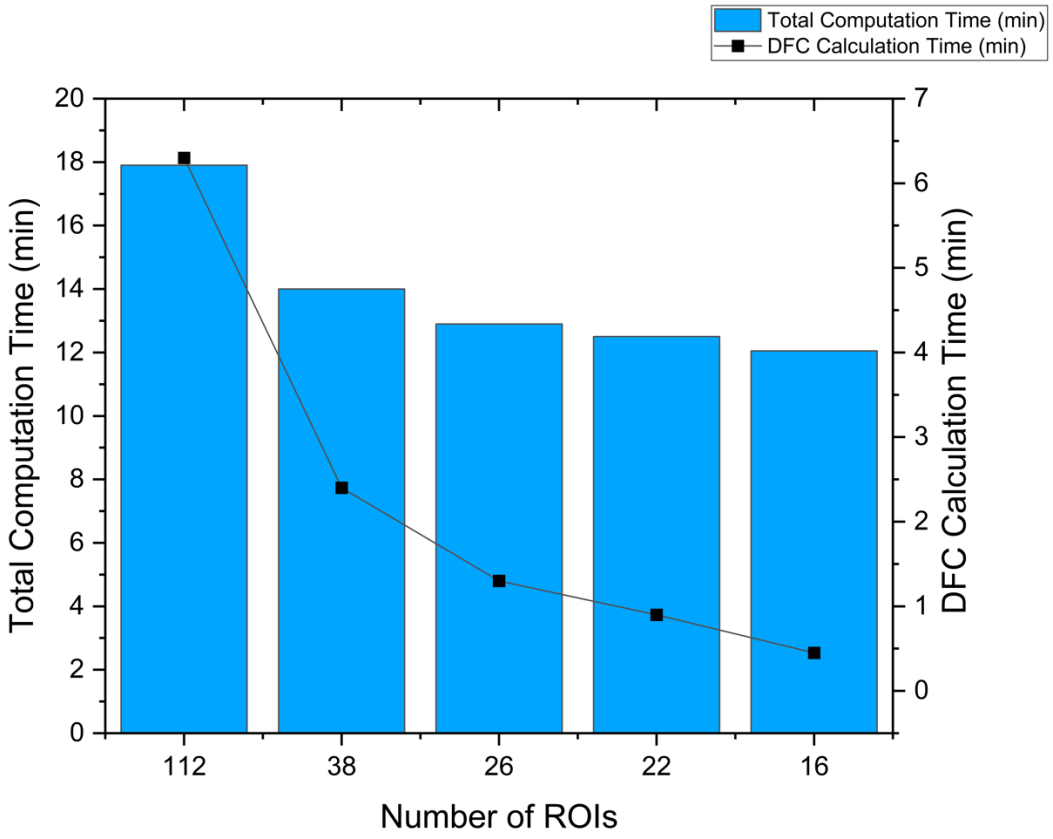

Figure - 6: Time taken for the processing to prediction varying number of ROIs

Evaluating the accuracy of proof-of-concept feedback system

In this study, we used our best performing pre-trained model (knn) to predict each newly acquired datasets for the classification of simulation and to identify the importance and contribution of several ROI pairs.

In each case, we found that the similarity of top 10 ROI pairs is consistently higher than 60% with the averaged classified list of ROI pairs for ME class. The most important note is that in each independent dataset's predictions, we identified that the top 2 contributing ROI pairs (SMA1_R – iPreCG_R and ParOp_R – IFG3_R) that match the corresponding averaged classified list of pairs, as shown in figure 7. The total time taken in this process of prediction of the newly acquired dataset was about 30 minutes including the processing of fMRI images.

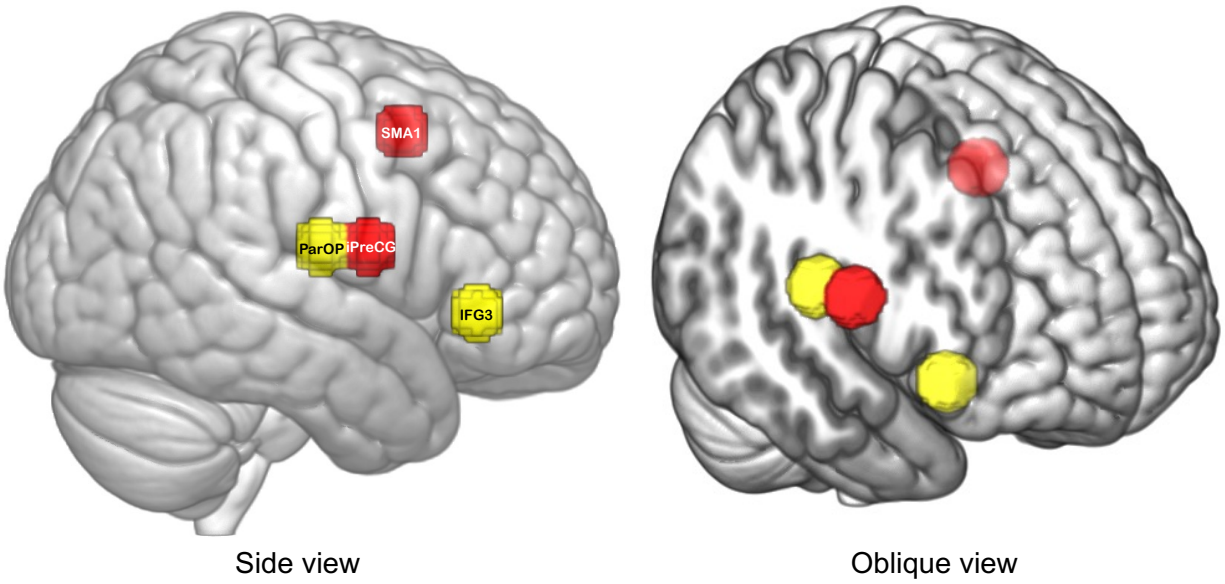

Figure 7: The top 2 ROI pairs contributing towards the classification of multi-electrode stimulation

Discussion

We developed a classification method that used a complex 4-D fMRI data set acquired during concurrent tDCS-MRI experiments to identify whether or not (1) stimulation has been applied and (2) electrode montage has an effect on network engagement. A multi-electrode montage targeting nodal points of access more reliably engages the AF network than a single electrode montage or a no-stimulation condition. Additionally, we have also developed a proof-of-concept feedback system using machine learning methods to determine as quick and efficiently as possible whether or not the ME montage modulates the targeted brain regions or not. We presented a unique and innovative approach to use machine learning applied to 4-D concurrent tDCS-fMRI data to classify tDCS effects based on stimulation montage and to generate a real-time feedback system to verify engagement of the targeted network.

Advancement of the novel machine learning approaches, increase in computational processing speed, and data analysis techniques have fueled the use of artificial intelligence (AI) techniques to solve multi-dimensional data-intensive problems[28]. The use of AI has shown promising results in different healthcare applications including X-ray & CT evaluation, development of antibiotic drugs, cancer classification, biomarker detection, and diagnosis[29]. Only a limited number of research studies have evaluated the use of AI to understand the effects of noninvasive brain stimulation techniques parameters and their topical specificity. Al-Kaysi and colleagues used baseline EEG spectra and mood-cognition evaluation measures from 10 subjects' data to predict if tDCS will positively modulate the mood[30]. Kajimura and colleagues showed that functional MR images acquired after active stimulation differed compared to sham; however, this study suffered from two limitations: there were few participants in each group (N=12) and applying tDCS outside the scanner which means that the study investigated effect of tDCS in different physical conditions than in which it was administered[31]. In other studies, ML methods were used to perform binary classification in complex data, outcome variables and their relationship to the stimulation, and how other baseline factors affect the outcome variables[30,32–34]. All of these studies, used ML methods to either understand/investigate effects of the stimulation in two groups namely 'Stimulation' and 'No-Stimulation.' Our study was more complex than previous studies in two aspects, first, we used concurrent tDCS-fMRI allowing us to record real time effects of stimulation and we applied ML methods to perform classification of three categories including two different stimulation montages. The classification accuracies demonstrate the ability of the ML model to detect if stimulation was applied or not. Additionally, the ML model reliably classifies condition into ME, SE, and NS stimulation categories with the ME category achieving the highest accuracy.

Our secondary goal was to develop a real-time feedback system that can use the shortest fMRI scan to evaluate the targeted nature of the stimulation. To achieve this goal, we first trained a model to classify the online tDCS-fMRI data into the three stimulation categories. To achieve near real-time feedback, the processing time must be reduced. Different sets of ROIs were evaluated for computing efficiency and accuracy. Different ROI sets allowed us to consider identify the number of ROIs that provide highest accuracy. As shown in the results, ROI sets with 22 ROIs

perform classification at overall accuracy close to 90% and time required for feature calculation close to 1 minute compared to 6 minutes required when using 112 ROI annotated over the whole brain[11–14]. In addition to identifying how many ROIs would be optimal, we also investigated which classification model would provide the best possible accuracy, ROC, and MOC scores. Results suggest that KNN model performs at the highest accuracy of 92% closely followed by the random forest model with the accuracy of 90%.

The proof-of-concept feedback system used DFC between ROI pairs as an individual feature; hundreds of such features were used as input for the ML method. However, identifying the anatomically relevant ROIs not only improved the accuracy, reduced time required, but also improved the possibility of using ML method as a feedback system by allowing us to identify top contributing features. The Gini index and SHAP method were used to calculate the top contributing features that are strongly affecting the classification. Incidentally, top 5 out of 231 features contribute close to 50% towards the classification, highlighting the importance of the top contributing features. The feedback system developed can be used as a test module that will assist in confirming the stimulation parameters applied to an individual's brain are suitable to invoke desired response by recording fMRI scan for less than 4 minutes and applying concurrent stimulation for about 2 minutes.

Testing of the independent fMRI dataset with the feedback system showed that the top DFC of the two ROI pairs contributing towards the classification into stimulation type highly matches with the top two contributing features generated from the group analysis. The top two ROI pairs that are identified as strong contributors to the ME classification were 'SMA1_R – iPreCG_R' and 'ParOp_R – IFG3_R.' (see Figure 7 for details of the anatomical location of these spherical ROIs). These ROI pairs represent nodal regions and relay regions of the right AF network, connected via the AF tract. The ME montage is designed to stimulate multiple nodal regions of the right AF network with the goal to increase connectivity between them as well as other related network regions. The results of our classification analysis suggest that it is the change in connectivity between regions belonging to this network that enhances the classification output. Further, 60% of the top 10 features that contribute towards the classification are common across subjects. The number 60% is not just more than a random chance, but it is a significantly stronger measure of commonalities in the response across subjects as these pairs would contribute more than 50% towards classification of the subject to a particular stimulation type.

There are certain limitations where the ML approach can be improved. The first one being the time required for preprocessing, we are exploring if improving computing power and using high speed memory devices will reduce the time required, additionally, we are also exploring the possibility of performing these operations over high performing computing machines to harness their processing speed. This would allow us to shorten the period from concurrent tDCS-fMRI acquisition to the feedback on targeted engagement to the experimenter. Second, the current proof-of-concept feedback system could be further optimized by the use of an MR-compatible multiplexed electrode system that can change where the current is applied to achieve network effects using a pre-mounted array of electrodes making it less likely that a subject would have to be moved out of and back into the scanner again and adjusting location of the electrodes.

In conclusion, machine learning was developed to identify the best electrode montage to show the most engagement of a targeted brain network spanning multiple lobes of the brain with high accuracy. ROI pairs whose DFC values contribute strongly towards the multi-class classification were identified in the order of importance. The top 5 out of 231 ROI pairs have a cumulative importance of 50%. Furthermore, a prototype of a machine learning real-time feedback system was introduced which would only require 1 minute to extract the DFC calculation-parameters and 1 minute for ML calculations.

Reference

1. Nitsche, M. A. & Paulus, W. Sustained excitability elevations induced by transcranial DC motor cortex stimulation in humans. *Neurology* **57**, (2001).
2. Vines, B. W., Nair, D. G. & Schlaug, G. Contralateral and ipsilateral motor effects after transcranial direct current stimulation. *Neuroreport* **17**, (2006).
3. Zheng, X., Alsop, D. C. & Schlaug, G. Effects of transcranial direct current stimulation (tDCS) on human regional cerebral blood flow. *Neuroimage* **58**, (2011).

4. Stagg, C. J. & Nitsche, M. A. Physiological basis of transcranial direct current stimulation. *Neuroscientist* **17**, (2011).
5. Shinde, A. B., Lerud, K. D., Munsch, F., Alsop, D. C. & Schlaug, G. Effects of tDCS dose and electrode montage on regional cerebral blood flow and motor behavior. *Neuroimage* **237**, (2021).
6. Splittgerber, M. *et al.* Multichannel anodal tDCS over the left dorsolateral prefrontal cortex in a paediatric population. *Sci Rep* **11**, (2021).
7. Lindenberg, R., Renga, V., Zhu, L. L., Nair, D. & Schlaug, G. Bihemispheric brain stimulation facilitates motor recovery in chronic stroke patients. *Neurology* **75**, (2010).
8. Nitsche, M. A. & Paulus, W. Excitability changes induced in the human motor cortex by weak transcranial direct current stimulation. *Journal of Physiology* **527**, (2000).
9. Fregni, F. *et al.* Transcranial direct current stimulation of the unaffected hemisphere in stroke patients. *Neuroreport* **16**, (2005).
10. Esmaeilpour, Z. *et al.* Methodology for tDCS integration with fMRI. *Human Brain Mapping* **41,** (2020).
11. Makris, N. *et al.* Decreased volume of left and total anterior insular lobule in schizophrenia. *Schizophr Res* **83**, (2006).
12. Frazier, J. A. *et al.* Structural brain magnetic resonance imaging of limbic and thalamic volumes in pediatric bipolar disorder. *American Journal of Psychiatry* **162**, (2005).
13. Desikan, R. S. *et al.* An automated labeling system for subdividing the human cerebral cortex on MRI scans into gyral based regions of interest. *Neuroimage* **31**, (2006).
14. Goldstein, J. M. *et al.* Hypothalamic Abnormalities in Schizophrenia: Sex Effects and Genetic Vulnerability. *Biol Psychiatry* **61**, (2007).
15. Wang, J. *et al.* GRETNA: A graph theoretical network analysis toolbox for imaging connectomics. *Front Hum Neurosci* **9**, (2015).
16. Li, X., Morgan, P. S., Ashburner, J., Smith, J. & Rorden, C. The first step for neuroimaging data analysis: DICOM to NIfTI conversion. *J Neurosci Methods* **264**, (2016).
17. Friston, K. J., Williams, S., Howard, R., Frackowiak, R. S. J. & Turner, R. Movement-related effects in fMRI time-series. *Magn Reson Med* **35**, (1996).
18. Koistinen, P. & Holmstrom, L. Kernel regression and backpropagation training with noise. in *1991 IEEE International Joint Conference on Neural Networks* (1992). doi:10.1109/ijcnn.1991.170429.
19. Skurichina, M., Raudys, Š. & Duin, R. P. W. K-nearest neighbors directed noise injection in multilayer perceptron training. *IEEE Trans Neural Netw* **11**, (2000).
20. An, G. The Effects of Adding Noise during Backpropagation Training on a Generalization Performance. *Neural Comput* **8**, (1996).
21. Lee, Y. J. & Yoon, J. Nonlinear image upsampling method based on radial basis function interpolation. *IEEE Transactions on Image Processing* **19**, (2010).
22. Yap, B. W. *et al.* An application of oversampling, undersampling, bagging and boosting in handling imbalanced datasets. in *Lecture Notes in Electrical Engineering* vol. 285 LNEE (2014).
23. Ester, M., Kriegel, H.-P., Sander, J. & Xu, X. A Density-Based Algorithm for Discovering Clusters in Large Spatial Databases with Noise. in *Proceedings of the 2nd International Conference on Knowledge Discovery and Data Mining* (1996).

24. Ahmad, A. & Quegan, S. Analysis of maximum likelihood classification on multispectral data. *Applied Mathematical Sciences* **6**, (2012).
25. Chicco, D. & Jurman, G. The advantages of the Matthews correlation coefficient (MCC) over F1 score and accuracy in binary classification evaluation. *BMC Genomics* **21**, (2020).
26. Lundberg, S. M. *et al.* Explainable machine-learning predictions for the prevention of hypoxaemia during surgery. *Nat Biomed Eng* **2**, (2018).
27. Lundberg, S. M. & Lee, S. I. A unified approach to interpreting model predictions. in *Advances in Neural Information Processing Systems* vols 2017-December (2017).
28. Jordan, M. I. & Mitchell, T. M. Machine learning: Trends, perspectives, and prospects. *Science* **349**, (2015).
29. Jiang, F. *et al.* Artificial intelligence in healthcare: Past, present and future. *Stroke and Vascular Neurology* **2**, (2017).
30. Al-Kaysi, A. M. *et al.* Predicting tDCS treatment outcomes of patients with major depressive disorder using automated EEG classification. *J Affect Disord* **208**, (2017).
31. Kajimura, S., Kochiyama, T., Abe, N. & Nomura, M. Challenge to unity: Relationship between hemispheric asymmetry of the default mode network and mind wandering. *Cerebral Cortex* **29**, (2019).
32. Albizu, A. *et al.* Machine learning and individual variability in electric field characteristics predict tDCS treatment response. *Brain Stimul* **13**, (2020).
33. Petrovskaya, A. *et al.* Examining the effects of transcranial direct current stimulation on human episodic memory with machine learning. *PLoS One* **15**, (2020).
34. Marotta, N. *et al.* Efficacy of Transcranial Direct Current Stimulation (tDCS) on Balance and Gait in Multiple Sclerosis Patients: A Machine Learning Approach. *J Clin Med* **11**, 3505 (2022).