Classification of fNIRS Finger Tapping Data with Multi-Labeling and Deep Learning

Natalie M. Sommer, *Member, IEEE*, Burak Kakillioglu, *Student Member, IEEE*, Trevor Grant, Senem Velipasalar, *Senior Member, IEEE*, Leanne Hirshfield

***Abstract*— Studying the relationship between the brain and finger tapping motions can contribute towards an improved understanding of neuromuscular impairment. Furthermore, acquiring brain data signals non-intrusively during finger tapping exercises, and building a robust classification model can aid in the field of human computer interaction. In this paper, we present a promising approach for spatially descriptive multi-labeling of spatiotemporal functional Near Infrared Spectroscopy (fNIRS) data to autonomously detect different finger tapping levels in different regions of the brain simultaneously. Our multi-class multi-labeling technique assigns labels to the left and right index fingers, and a given label describes one of three different finger tapping frequencies (rest, 80bpm, and 120bpm) to be monitored in the corresponding contralateral spatial location in the brain's motor cortex. We train a CNN/LSTM-based network to classify the aforementioned finger tapping levels spatially and simultaneously. The evaluation, based on simultaneous multi-label predictions for two brain regions, is performed with a metric commonly used in multi-labeling, Hamming Loss, along with confusion matrix-based measurements. Promising testing results are obtained with an average Hamming Loss of 0.185, average F-Score of 0.81, and average Accuracy of 0.81. Moreover, we explain our model and novel multi-labeling approach by generating Shapley Additive Explanation values and plotting them on an image-like background, which represents the fNIRS channel layout used as data input. Shapley values help to add interpretability to our deep learning model and by confirming expected results, offer a pathway to the future development of complex deep learning models that attempt to predict social-cognitive-affective states.**

***Index Terms*— CNN, finger tapping, fNIRS, Hamming Loss, LSTM, machine learning, multi-labeling, spatially descriptive, spatiotemporal**

I. Introduction

Research in Human Computer Interaction (HCI) and Brain-Computer Interfacing (BCI) continues to make strides in the advancement of machines as an aid to people with disabilities and for rehabilitation. Gathering brain data through non-intrusive mediums along with finding robust ways of interpreting the data have been an integral part of this research. For instance, through image processing, a nose tracking cursor and an eye gazing interface give people with impaired motor function the ability to communicate with others [1], [2]. In another work, a classification model was presented based on the use of Electroencephalogram (EEG) data to distinguish between mental counting and wrist rotations [3]. EEG data was also used by León [4] to classify combinations of hand and feet movement imagery for robotic arm control. Other examples include utilizing a hybrid set of data, such as a combination of EEG and Functional Near Infrared Spectroscopy (fNIRS), to detect different types of motor imagery. Such experiments included classifying the motor imagery related to the force and speed of right hand clenching [6] and motor imagery of left and right hand grasping [7]. Another area of interest in HCI/BCI is to explore the separate ability of fNIRS signals to represent movement and imagery. Three right foot soccer playing motion imageries (i.e. passing, stopping, shooting) were classified by Li et al. [5] with an average accuracy of approximately 79%. Right thumb and little finger physical tapping were distinguished with a validation accuracy of 97.17% by Woo et al. [8]. All the aforementioned studies offer ways of helping those who have compromised motor abilities by providing possible means of communication with the outside world. Moreover, increasing our knowledge of the brain and how to best capture and classify its signals gives hope to those who suffer from locked-in syndrome.

Functional magnetic resonance imaging (fMRI) represents the gold standard for brain measurement in cases where it is possible to place participants in the fMRI magnet with minimal motion permitted. Although fNIRS cannot measure deep brain structures like fMRI, it can take comparable measurements of hemodynamic responses across the brain cortex, and it can do so in naturalistic real-world environments due to portability, ease of set up, and decreased sensitivity to a subject's motion [9]. Thus, fNIRS could be a suitable device for BCI applications, where target users can wear the non-invasive device in their naturalistic environments. The motion

"The information, data, or work presented herein was funded in part by National Science Foundation (NSF) under Grant 1739748, Grant 1816732 and by the Advanced Research Projects Agency-Energy (ARPA-E), U.S. Department of Energy, under Award Number DE-AR0000940. The views and opinions of authors expressed herein do not necessarily state or reflect those of the United States Government or any agency thereof. "

Natalie M. Sommer, Burak Kakillioglu and Senem Velipasalar are with the Department of Electrical Engineering and Computer Science, Syracuse University, Syracuse, NY, 13244 USA (e-mail: {nmsommer, bkakilli, svelipas}@syr.edu).

Trevor Grant and Leanne Hirshfield are with the Institute of Cognitive Science, University of Colorado, Boulder, CO, 80309 USA (e-mail: {trevor.grant, leanne.hirshfield}@colorado.edu).

sensitivity of fNIRS sensors is also smaller than that of EEG sensors and fNIRS can provide greater spatial resolution [10]. Although its acquisition of neural information is restricted to 1 cm below the surface of the brain, fNIRS has been a popular way of collecting brain data signals as it is capable of capturing hemodynamic information in a non-restrictive and practical way while providing ample spatial and temporal resolution.

The classification of finger tapping activities with the use of fNIRS data is a popular area of research [10], [16]. Finding a suitable interpretation of data collected by fNIRS sensors as it relates to a subject's physical or mental activity is key to making the data useful for HCI/BCI. In this paper, we propose an approach that aims at capturing the spatiotemporal nature of the fNIRS data with a robust deep learning algorithm, which combines a Convolutional Neural Network (CNN) and Long Short Term Memory (LSTM). Moreover, assigning labels to the data, which represent the spatial nature of the fNIRS probe configuration can yield information about the location of activation. Therefore, by handling the readings acquired by the fNIRS probe channels similar to video frames, we can apply our novel spatially descriptive multi-label/multi-class deep learning classifier, which we first introduced to detect different activity levels in various regions of interest simultaneously in video data [17]. In this case, finger tapping data would be formatted to represent two regions of interest in the brain, namely, the left and right motor cortices. Channel readings, based on the corresponding probe locations, would be formatted to represent these two sides of the brain.

The data used in this study was collected at a University in the Western United States. The dataset contains 51 sessions of index finger tappings from the right and left hands at three different frequencies from 12 participants. Each session included sequences of both index fingers tapping at the same frequencies, single index finger tappings and rest periods. The different levels of finger tapping frequency (rest, 80bpm, and 120bpm) lend themselves well to a multi-class labeling schema.

Then, applying our novel spatial multi-labeling technique of designating two labels to the two sides of the brain, each of which is assigned a finger tapping frequency level, classification would be based on a multi-label/multi-class schema. This is a novel approach to multi-labeling since in most studies, binary labels are used for the presence (1) or absence (0) of each class in a multi-label [18], [19]. Moreover, in these cases, the multi-label does not provide spatial information the way that our schema does.

We propose a unique and promising approach to classifying various frequency levels of index finger tapping *simultaneously* with the use of a multi-label/multi-class labeling schema. Our novel multi-labeling approach provides a concurrent detection of different levels of brain activation in the two sides of the brain.The spatiotemporal nature of the fNIRS signal acquired during finger tapping trials is captured during the training and validation of a Convolutional LSTM model. Different from our previous work [17], which focuses on video data, in this work (i) the labels are assigned to fNIRS data formatted to represent the probe configuration for the two sides of the brain; (ii) we employ and generate Shapley Additive Explanation values [56] to help explain the spatial characteristics of our Convolutional LSTM model; (iii) we plot the Shapley values on an image-like background, which represents the fNIRS channel layout used as data input. Applying Shapley values is notable as it can be used to view the structure of deep learning models (and these deep learning 'black boxes' can often be difficult to interpret) and to show the regions of the brain that were most important for different model predictions. Finger tapping has been heavily studied in the brain measurement domain, and there are known brain activations in the contralateral primary motor cortices, based on which finger is being tapped (left finger activates right primary motor cortex, and vice versa). We use our Shapley values to show that the most predictive channels in our deep learning model align with known spatial characteristics of the brain during finger tapping. By using our model explainability techniques on a benchmark finger tapping task, we demonstrate the potential of applying this approach to more complex classification tasks (e.g. classifying types of workload or emotional states from fNIRS data), where the use of Shapley values can help to explain the model, while also having the potential to add to the field of cognitive neuroscience, whereby complex interactions between interconnected brain regions could be identified with an explainability technique.

The organization of this paper is as follows: We discuss the related work in Section II. We then describe our network model and illustrate the formatting of our dataset, and labeling structure in Section III. We present our experimental results along with further analysis of our proposed method with Shapley Additive Explanations (SHAP) values in Section IV. We then conclude the paper in Section V.

II. Related Work

Finger tapping abilities have been studied as a way of determining the progression of Parkinson's Disease [22], [23]. Similarly, other neuromuscular disabilities caused by cerebral palsy and stroke can be better understood with finger tapping exercises [24], [25]. This means that studying the relationship between the brain and finger tapping motions can contribute towards an improved understanding of neuromuscular impairment. Furthermore, by acquiring brain data signals non-intrusively during finger tapping exercises and building a robust classification model, can aid in the fields of HCI and BCI for people with compromised motor function. Training BCI applications on real, or imagined, finger tapping motions has been heavily studied in part because finger tapping has been found to result in consistent patterns of activation in brain areas involved in motor function. Specifically, right finger tapping shows reliable activation in the left primary motor cortex, and vice versa. Tapping both fingers simultaneously will result in both left and right primary cortices being activated (amongst a host of other regions that are implicated in the execution of motor function) [4], [57].

The classification of brain signals, which show greatest activation during finger tapping motions, can also lead to establishing a brain mapping based on sensor locations [27] thereby assisting in motor skill rehabilitation [26]. Moreover, the study

of finger tapping motor imagery is instrumental in the fields of HCI/BCI [30], [31]. Neural correlates exist between imagined motor movements and real motor movements [68], [69]. Also, neural correlates of different finger tapping frequencies can be detected through hemodynamic responses [70]. Therefore, being able to detect areas of brain activation during physical finger tappings at different frequencies can impart a better understanding of brain activity during imagined finger tappings which could offer a means of communication through BCI for locked-in patients. Moreover, classifying brain activation patterns during finger tapping activities can help measure improvements gained through BCI stroke rehabilitation [26] and assess the progression of Parkinson's Disease [70].

Since hemodynamic response is correlated to finger tapping imagery and execution [28], [32], the study of fNIRS signals evoked during this type of activity is fitting. In general, research based on fNIRS signals has gained popularity over the last several years [33], [34]. Using optical wavelengths between 650 nm and 1000 nm produced by the emitter probes, the detector probes detect the reflected light due to changes in oxygenated (OXY) and deoxygenated (DEOXY) hemoglobin concentrations at the cerebral cortex. The changes in hemoglobin oxygenation are a result of neural activity elicited by stimuli.

The study of neural activation due to finger tapping using fNIRS signals has shown interesting results. Bak et al. [11] performed a ternary classification of left index finger, right index finger tapping, and foot tapping with an SVM model, and obtained an average accuracy of about 70.4%. Woo et al. [8] used a deep convolutional generative neural network to augment their data, and trained a CNN model, which produced a validation accuracy of 97.17% for thumb tapping and little finger tapping classifications. Nazeer et al. [12] used vector-based phase analysis features and a Linear Discriminant Analysis (LDA) model to distinguish between left index finger tapping, right finger tapping, and rest. The results, based on a sample size of seven, were 85.4 +/- 1.4% for this three-class distinction.

It is difficult to compare the results of the aforementioned studies since different models were used, and different finger tapping exercises were conducted. All were based on fNIRS data, and reported robust classification accuracies. However, the disadvantage of using shallow learning (i.e. SVM, LDA) is the need to choose features that best fit the nature of the data. Nazeer et al. [12] reported robust classifications results with LDA with a small sample size of seven subjects. Woo et al. [8] reported their robust results based on a sample size of 11 subjects and with a deep learning CNN model. The latter study along with research by Trakoolwilaiwan et al. [36] and Wickramaratne at al. [37] have demonstrated that CNN-based algorithms are suitable for learning patterns in raw fNIRS data, thereby eliminating the need for generating handcrafted features. Additionally, recent research has been conducted to support the use of LSTM networks in the classification of fNIRS data [38], [39]. Although the data was not finger tapping related, improvements in classification results were shown as the algorithm was able to capture the temporal characteristics of the data.

The monitoring of simultaneous activity based on spatiotemporal data has been studied for a variety of applications. For example, a Convolutional LSTM was used to predict the simultaneous demands of different modes of transportation in an urban environment [21]. Also, using magnetoencephalography (MEG) and SVM, four types of simultaneous bilateral hand movements were classified with average accuracies of 75% and 70% for physical and imagined movements, respectively [20]. Classification of concurrent events offers a snapshot view of the output states through time providing additional information related to the correlation of these actions. For example, this provides a means of exploring the neural correlates of finger tapping performed by fingers on different hands. Our interest in detecting simultaneous activity in spatiotemporal data necessitates the use of multi-labels. Multi-labeling has been used in video and other spatiotemporal data classification [40], [41]. However, most multi-labeling schemas in current research are based on selecting one of more descriptors (classes) and then referring to them as a multi-label. For example, the multi-labels in the YouTube-8M database [42] are annotations, which describe the contents of the video. Similarly, in combating noise pollution, environmental sounds, which include simultaneous sounds from different sources, are tagged, thereby forming the multi-labels [43]. In the medical field, multi-labels have been used in classifying motor execution and imagery. For example, Olsson et al. [44] classified compound hand movements based on high density surface electromyography (HD-sEMG) recordings using a series of labels that describe the basic movements (i.e.: individual finger movements) needed to attain the final compound movement (i.e.: fist). Therefore, the basic movements defined the individual labels, which were used to build the multi-labels. León [4] used multi-labels to represent different combinations of hand and feet motor imagery captured by EEG data. One of the labels would be assigned a '1' based on the motion(s) detected, and all other labels would then receive a '0'. The previously mentioned examples utilize a pool of classes to choose from in building multi-labels. To the best of our knowledge, the spatially descriptive multi-label/multi-class approach that we employ in our research has not been attempted. In our case, the labels represent spatial regions of interest. Each region of interest gets assigned one of many classes depending on the activity level in the area that is monitored.

Ensuring that labels are balanced is important to avoid classification skew, which can impact results. Multi-labeling can magnify the problem of label imbalance [40]. Data augmentation is a technique, which increases the number of samples, and helps balance labels [8], [45]. For example, when classifying different levels of activity in surveillance videos [17], we augmented the video data by creating additional videos through rotation of frame quadrants to increase samples of minority labels. Additionally, another approach to label and class balancing is to use resampling techniques, such as Random Undersampling (RUS) and Synthetic Minority Oversampling Technique (SMOTE). SMOTENN [47], [59], an interesting balancing technique, combines both oversampling and undersampling. More specifically, it combines SMOTE

oversampling (for the minority classes) with Edited Nearest Neighbors undersampling (for the majority classes). This balancing technique is utilized in our proposed method.

III. Proposed Method

By considering the two sides of the brain as the two regions of interest, we use the fNIRS probe layout shown in Fig. 1. Paying attention to the channels in between the sources (in red) and detectors (in blue), the OXY and DEOXY data channels in our fNIRS readings were separated into the two sections of interest, namely, left primary motor cortex and right primary motor cortex. Then, the channels in each respective region were rearranged to reflect the proximity of detectors to the sources. The final configuration is shown in Table I. The two distinct spatial regions each correspond to a label assigned to an index finger as part of our multi-label. With this set-up, we are able to determine the types of activation taking place in each side of the brain simultaneously during index finger tappings.

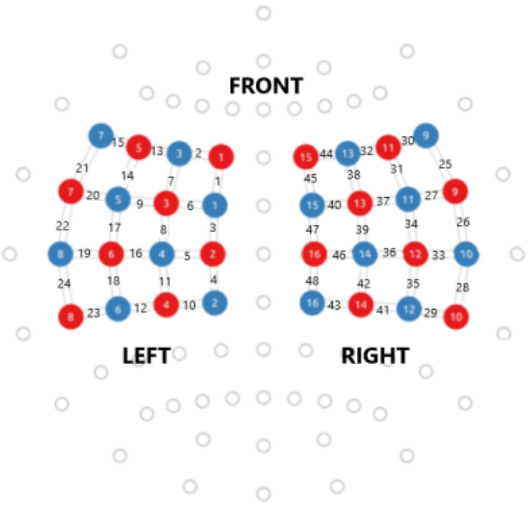

Fig. 1. fNIRS probe layout, with regions on the left and right primary motor cortex covered

LEFT MOTOR CORTEX						RIGHT MOTOR CORTEX					
CH_1	CH_14	CH_20	CH_6	CH_11	CH_24	CH_45	CH_31	CH_27	CH_40	CH_42	CH_28
CH_2	CH_15	CH_17	CH_3	CH_16	CH_23	CH_44	CH_30	CH_34	CH_47	CH_36	CH_29
CH_7	CH_21	CH_9	CH_4	CH_18	CH_12	CH_38	CH_25	CH_37	CH_48	CH_35	CH_41
CH_13	CH_22	CH_8	CH_5	CH_19	CH_10	CH_32	CH_26	CH_39	CH_46	CH_33	CH_43

TABLE I: Channel order format to reflect probe configuration in the left and right regions of interest

In our labeling schema, each spatial region corresponds to an index finger which gets assigned one of three possible finger tapping frequencies: rest, 80bpm, and 120bpm. Our spatially descriptive multi-labels enable the recognition of concurrent finger tappings by the right and left index fingers simultaneously, and the identification of different rates of tapping through a *multi-class* descriptor for each label. Our labeling schema, like most multi-labeling schemas, is more prone to label imbalance. That is why we chose SMOTENN [47], [59], which uses hybrid oversampling/undersampling, to balance our labels. The effect of SMOTENN on the class sample distributions for our two labels is shown in Table II.

Our network model, dataset and the labeling structure are described in detail in the following subsections.

Before SMOTENN	Label #1	Label #2
Class '0'	45.9%	47.1%
Class '1'	27.2%	26.3%
Class '2'	26.8%	26.6%
After SMOTENN		
Class '0'	34.5%	33.8%
Class '1'	32.9%	32.5%
Class '2'	32.6%	33.7%

TABLE II: Class distribution before and after label balancing with SMOTENN

A. Network Model

Since the fNIRS signals acquired during index finger tappings are spatiotemporal, we chose to base our network model on a Convolutional LSTM to be able to detect the spatial and temporal properties of the data. The structure of our network model is shown in Fig. 2. The fNIRS input data is formatted into two regions of interest and labeled with our multi-labels. The network contains two 2D convolutional layers, with the first followed by a max pooling layer. Then, following a dropout layer, there is an LSTM layer and a dense layer, which leads to a final output layer. An adaptive learning rate, which decreased by a factor of 2 to a minimum of 10^{-6}, was used along with binary cross-entropy loss and $rmsprop$ optimization. Since the class values in our multi-labels are one-hot encoded and the multi-output layer uses sigmoid activation, binary cross-entropy was the suitable choice. The adaptive learning rate and early stopping monitor validation loss so that adjustments are made to the learning rate if the validation loss does not decrease after two epochs and consequently, training automatically stops if the validation loss has stopped decreasing after three epochs. After training, the model was tested on *unseen* fNIRS finger tapping sessions formatted the same way as our training/validation data (detailed next).

As illustrated in Fig. 2, the input to our proposed network model is formatted into two sections, wherein the left and right include the channels, which capture the changes in hemoglobin concentrations for the left and right motor cortices for each participant. Specifically, the channels for each region of interest were grouped into chunks of 50 samples with a sliding window of 10 samples. Batches of these groups were then used to train and validate our model.

B. Dataset

Our fNIRS data was collected with a NIRx NirsSport2 device at a sampling rate of 10.2 Hz. We used the standard NIRX montage that covers the right and left primary motor cortices. Channel readings of changes in light intensity corresponding to changes in OXY and DEOXY hemoglobin concentrations at the cerebral cortex were recorded by the Aurora fNIRS software. Data was bandpass filtered from 0.01 Hz and 0.5 Hz to remove noise. The modified Beer-Lambert Law was applied to convert the light intensities into data representative of relative change in OXY and DEOXY hemoglobin. Z-score normalization was applied to each channel. For each sample (OXY and DEOXY), the channels belonging to the left motor

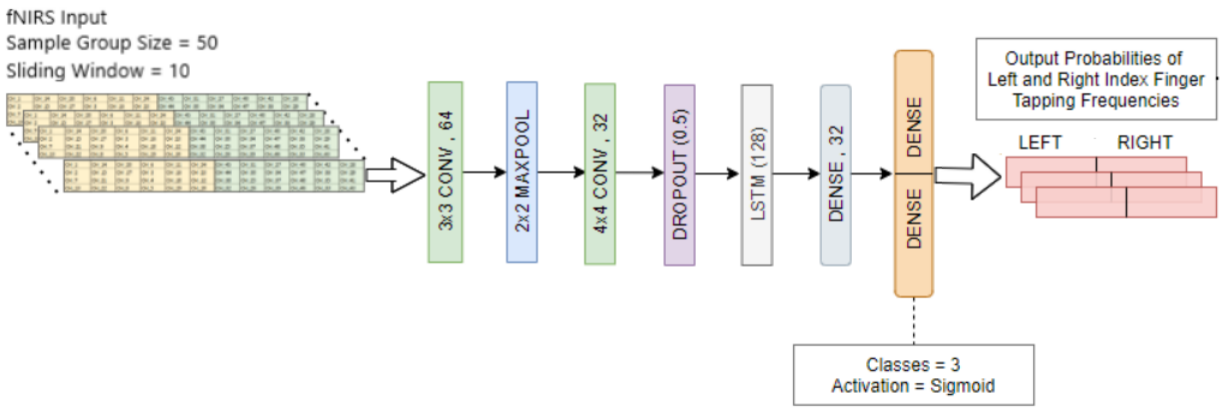

Fig. 2. Proposed deep learning model structure

cortex were rearranged and reshaped to correspond to the probe layout. Subsequently, after the channels for the right motor cortex were formatted accordingly, the two sets of data were concatenated as illustrated in Table I. The final format of the data files included a layer of 4x12 channel-ordered OXY data followed by 4x12 channel-ordered DEOXY data.

Each session included sequences of the following finger tapping combinations: both index fingers at rest; one index finger at rest and the other tapping at either 80bpm or 120bpm, both index fingers tapping at 80bpm, and both index fingers tapping at 120bpm. Due to the complexity of tapping index fingers at two different frequencies simultaneously, the combinations, namely $[80bpm, 120bpm]$ and $[120bpm, 80bpm]$, were not included in the trials. Therefore, there were seven possible multi-labels. Using two labels to spatially represent the left and right motor cortices, one of three classes characterizing the tapping frequency was assigned to each label. The SMOTENN [47], [59] label balancing algorithm was applied to the training and validation data before samples were grouped in sets of 50 with a sliding window of 10 samples. During the grouping process, the label of the middle sample in each group at each location (region of interest), i.e. the label of the 25th sample in this case, was used as part of the multi-label. With a total of 51 finger tapping data acquisition sessions from 12 participants, we initially tested our design by setting aside 10% of the data sessions for testing and trained/validated a model on the remaining part of the data using a random 80% training / 20% validation split. This process was repeated several times. The testing sessions were chosen so that each participant appeared at least twice in the total of all testing trial runs. Likewise, we also conducted a leave-one-subject-out 11-fold cross-validation with a similar process. For each run, a participant's session was set aside for testing. Remaining sessions, not including the subject's other trials, were partitioned in a random 80% training / 20% validation split. One participant was excluded from testing due to an imbalance in the distribution of the number of trial runs.

C. Proposed Multi-Labeling Structure

Given two spatial regions of interest, namely, the left and right motor cortices, we established a novel approach to multi-labeling. Due to the contralateral relationship between an index finger's tapping motion and motor cortex activation [58], the left index finger was assigned a label which represents the right motor cortex. The right index finger was assigned a label that represents the left motor cortex. The three tapping frequencies in our data were rest, represented with a '0', 80bpm, represented with a '1', and 120bpm, represented with a '2'. The definition for each type of finger tapping frequency that was used in our labeling schema is shown in Table III.

LEVEL OF TAPPING	DEFINITION
0	rest (No tapping)
1	Index finger tapping at 80bpm
2	Index finger tapping at 120bpm

TABLE III: Three different types of finger tapping frequencies are to be detected in each of the two sides of the brain

As we rearranged and reshaped our fNIRS data into our two regions of interest for each input file (for both OXY and DEOXY layers), we combined the corresponding labels (left index and right index finger) for our multi-labels. For example, suppose that for a particular sample of finger tapping data, the right and left regions of the brain are labeled to have left index finger tapping at 120bpm (level of tapping = '2'); and right index finger tapping is at rest (level of tapping = '0'), respectively. Then, the multi-label of $[120bpm(2), rest(0)]$ would be assigned to the pertinent sample of fNIRS data as shown in Table IV.

LABEL #1 Left Index Finger Level of Tapping	LABEL #2 Right Index Finger Level of Tapping
2	**0**

TABLE IV: Example of our spatially and tapping-level descriptive multi-label schema

The finger tapping combinations in our fNIRS database resulted in seven possible multi-labels, namely, $[rest, rest]$, $[rest, 80bpm]$, $[rest, 120bpm]$, $[80bpm, rest]$, $[120bpm, rest]$, $[80bpm, 80bpm]$, and $[120bpm, 120bpm]$. For each cross-validation run, testing was performed on the five finger tapping sessions that were not in the training and validation set. After

testing, predictions were made, and each type of finger tapping level was assigned a probability for the respective region of interest. The class of the finger tapping frequency (rest(0), 80bpm(1), or 120bpm(2)) with the highest probability was then chosen as part of the predicted multi-label.

D. Evaluation Criteria

We report our results based on the micro metrics of F-Score and Accuracy. Additionally, we use the measure of Hamming Loss, which is a common metric used in multi-label research, and offers an overall look at a classifier's prediction error [48]. It is an instance-based metric, since it is based on the entire multi-label prediction for each time period of testing data. This metric is often used in BCI and HCI applications. For example, Kalansooriya et al. [46], reported Hamming Loss scores between 0.198 and 0.311 when comparing classifiers to detect emotion from EEG signals retrieved during video game playing. Devlaminck et al. [60] sought to minimize Hamming Loss when comparing classification methods in a BCI application. Reported values between 0.17 and 0.57 were correlated with changes in accuracy measurements. Therefore, to provide a complete picture of our model's performance , we also report micro-averaged F-score and Accuracy, label-based metrics, to consider the label assigned to each region of interest in the testing results separately [48], [49], [50].

i) Hamming Loss. For each instance, we compared the group of two predicted classes of finger tapping in our multi-label with the ground truth label assignments. The average Hamming Loss for each testing finger tapping trial was calculated as shown in Eq. (1), where S represents the number of seconds in a testing finger tapping trial, and $p_{i,j}$ and $g_{i,j}$ indicate the predicted level of tapping and the ground truth level of tapping, respectively. Therefore, for each label within a multi-label, a mismatch is assigned a 1. These values are then added and averaged over the product of the number of labels (2) in a multi-label and the time span of the recording in seconds (S).

$$\frac{1}{2S}\sum_{i=1}^{S}\sum_{j=1}^{2}[if(p_{i,j}=g_{i,j},0,1)] \tag{1}$$

ii) Micro-Averaged F-Score and Accuracy. In order to provide a complete set of performance criteria, we also consider label-based (i.e. single side of the brain) metrics, such as Micro-averaged F-Score and Accuracy.

Since our labels can be assigned one of three classes (0,1, or 2) to represent three levels of finger tapping frequency, it is best to perform a micro-average of these parameters. For instance, instead of taking an average of the calculated Accuracy for each level of finger tapping, we determine the overall Accuracy for all levels of tapping at once. The calculations of Micro-Averaged F-score (μA_Fscore), and Micro-Averaged Accuracy (μA_Accuracy) are shown in Eq. (2), and (3), respectively.

$$\mu A\_\text{FScore} = \frac{\sum_{i=0}^{2}(2*TP)_i}{\sum_{i=0}^{2}((2*TP)_i+(FP)_i+(FN)_i)} \tag{2}$$

$$\mu A\_\text{Accuracy} = \frac{\sum_{i=0}^{2}((TP)_i+(TN)_i)}{\sum_{i=0}^{2}((TP)_i+(TN)_i+(FP)_i+(FN)_i)} \tag{3}$$

IV. EXPERIMENTAL RESULTS

Our goal was to detect different levels of finger tapping in two different spatial sides of the brain simultaneously. Through our novel multi-labeling schema and deep learning-based algorithm, we built a model to be able to perform automated classification of spatiotemporal finger tapping data and determine the simultaneous types of tapping taking place by both index fingers. Initially, we tested our design by setting aside 10% of the 51 data acquisition sessions for testing and trained/validated a model on the remaining part of the data. Since the SMOTENN [47], [59] algorithm was applied to this data, the number of instances for different classes, was void of skew and imbalance. After balancing, approximately 200 minutes of multi-labeled fNIRS finger tapping data was available for training and validation using a random 80% training / 20% validation split. Optimal results have been obtained with a window size of 5 seconds and a 1 second sliding window. During training and validation, we monitored training to make sure that validation loss was always smaller than training loss to prevent any possible overfitting. Once the validation loss did not improve for three epochs with an adaptable learning rate, training/validation would end. An overall validation accuracy of 99.52% for both labels supports the validity of our classification schema.

A summary of training and validation metrics is presented in Table V. Average validation losses were always smaller than Average training losses.

Run #	No. of Epochs	Average Training Loss	Training Accuracy Label #1	Training Accuracy Label #2	Average Validation Loss	Validation Accuracy Label #1	Validation Accuracy Label #2
1	27	0.0351	0.9940	0.9942	0.0178	0.9948	0.9947
2	35	0.0214	0.9969	0.9969	0.0064	0.9962	0.9948
3	43	0.0221	0.9963	0.9965	0.0063	0.9941	0.9948
4	19	0.0644	0.9894	0.9887	0.0593	0.9919	0.9932
5	34	0.0197	0.9975	0.9975	0.0106	0.9974	0.9979
6	30	0.0327	0.9941	0.9944	0.0198	0.9953	0.9949
7	39	0.0250	0.9965	0.9965	0.0114	0.9967	0.9959

TABLE V: Training and Validation metrics for 7 cross-validation runs

After training and validation, our classifier is then tested on *unseen* similarly configured finger tapping session files. Results were compared to the database's ground truth label annotations. We determined the overall Hamming Loss using Eq. (1) by calculating the average for all groups of two labels for the *unseen* testing finger tapping sessions. We repeated this process seven times and our results are shown in Table VI. These values indicate that our classifier is capable of making correct predictions for both sides of the brain on the testing finger tapping data an average of about 80% of

the time. Additional metrics, namely Micro-Averaged F-score and Accuracy values, were also calculated for each region of interest (i.e. side of the brain), and the results are presented in Table VII.

Cross-Validation Run	1	2	3	4	5	6	7
Average Hamming Loss	**0.185**	**0.209**	**0.235**	**0.217**	**0.161**	**0.211**	**0.188**

TABLE VI: Average Hamming Loss of fNIRS finger tapping sessions set aside for testing for 7 cross-validation runs

Cross - Validation	**Micro-Averaged F-Score**		**Micro-Averaged Accuracy**	
Run #	Label #1	Label #2	Label #1	Label #2
1	0.819	0.823	0.810	0.818
2	0.803	0.787	0.790	0.786
3	0.775	0.748	0.750	0.752
4	0.793	0.819	0.804	0.812
5	0.838	0.839	0.829	0.841
6	0.760	0.783	0.765	0.759
7	0.827	0.803	0.814	0.811
Average	**0.802**	**0.800**	**0.796**	**0.797**

TABLE VII: Right (Label #1) and Left (Label #2) sides of the brain average label metrics for 7 cross-validation runs

This type of testing provided a general representation of our classifier's ability to classify three different finger tapping frequencies from the two sides of the motor cortex simultaneously. In order to further support our approach, we performed additional testing with a leave-one-subject-out 11-fold cross-validation. For each run, a participant's session was set aside for testing. The remaining sessions, not including other trials performed by the subject, were partitioned in a random 80% training and 20% validation split. One participant was excluded from testing due to an imbalance in the distribution of the number of trial runs. Average Hamming Loss results are shown in Table VIII. This metric displays our classifier's ability of making correct predictions for both sides of the brain on the testing finger tapping data an average of about 83% of the time. Also, Micro-Averaged F-score and Accuracy values, were also calculated for each region of interest (i.e. side of the brain), and the results are presented in Table IX. A discussion of all results is provided in Sec. IV-B.

Subject	1	2	3	4	5	6	7	8	9	10	11
Average Hamming Loss	**0.176**	**0.159**	**0.158**	**0.109**	**0.164**	**0.188**	**0.210**	**0.181**	**0.182**	**0.175**	**0.144**

TABLE VIII: Average Hamming Loss of 1-vs-all cross-validations for 11 participants

A. Visualizing our Proposed Method with SHAP

Motivated by our promising Hamming Loss, Micro-Averaged F-Score and Micro-Averaged Accuracy values, our next goal was to gain further understanding of our network model with Shapley Additive Explanations (SHAP) values. SHAP, with its basis in game theory [56], is a way of illustrating model interpretation by assigning impact values on learned features (in the case of deep learning) as they relate to a model's predictions [51], [52]. Using the deep explainer, which is specialized for neural network models, we generated the SHAP values to visualize how our model handles the channel readings when making predictions. The SHAP values reflect a channel's marginal contribution to the output class predictions.

Cross - Validation	**Micro-Averaged F-Score**		**Micro-Averaged Accuracy**	
Subject #	Label #1	Label #2	Label #1	Label #2
1	0.824	0.735	0.822	0.723
2	0.824	0.853	0.834	0.846
3	0.868	0.809	0.844	0.817
4	0.956	0.824	0.933	0.814
5	0.810	0.862	0.809	0.867
6	0.809	0.838	0.813	0.838
7	0.779	0.794	0.812	0.799
8	0.797	0.75	0.825	0.764
9	0.779	0.853	0.821	0.818
10	0.825	0.825	0.827	0.844
11	0.864	0.848	0.876	0.855
Average	**0.830**	**0.817**	**0.838**	**0.817**

TABLE IX: Right (Label #1) and Left (Label #2) sides of the brain label metrics for leave-one-subject-out 11-fold cross-validations

SHAP values are determined based on our deep learning model and testing data. More specifically, in our CNN/LSTM model, an instance (window size = 5 seconds) is represented by the continuous changes in OXY and DEOXY hemoglobin measurements, collected across 48 channels over the Left and Right motor cortices. Thus, when SHAP evaluates a given test instance, it is looking at the value of each of the 96 features/inputs in that instance in the channels shown in Fig. 1. Twenty-four channels are over the left motor cortex and 24 channels are over the right motor cortex. If we consider Fig. 3, we have taken our pretrained CNN/LSTM model, along with each sample representing the condition of [rest(0), rest(0)] and used SHAP to see how the pre-trained model evaluates the OXY and DEOXY features in each instance to ultimately make the final prediction of [rest(0), rest(0)]. We repeat this process for every single [rest(0), rest(0)] 5 second long instance, treating it as a test instance, and using SHAP to evaluate its features. The resulting figure displays the average of the SHAP values for all instances of the target condition. In Fig. 3, the values for '0'(rest) that have the highest impact are shown in red and represent the channels on the right side (top right) and also left side (bottom left) of the layout background from Table I.

A large body of research has found that real, or imagined, finger tapping has reliably been found to activate the primary motor cortices. Tapping the right finger activates the left

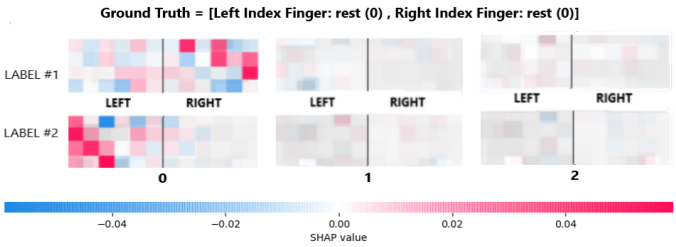

Fig. 3. SHAP values for multi-label [rest(0), rest(0)]

primary motor cortex, and vice versa for the left finger. Tapping of both fingers has been found then to activate both regions simultaneously [4], [57].

When comparing these SHAP values to the channel layout in Table I, we can locate the specific channels with the largest positive impact. This is shown in Fig. 4. Therefore, this allows us to visualize the channels with the highest positive prediction impact on the probe layout as shown in Fig. 5.

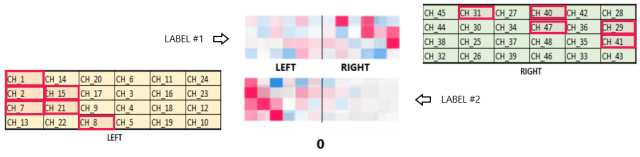

Fig. 4. Mapping SHAP values to specific channels for multi-label [rest(0), rest(0)]

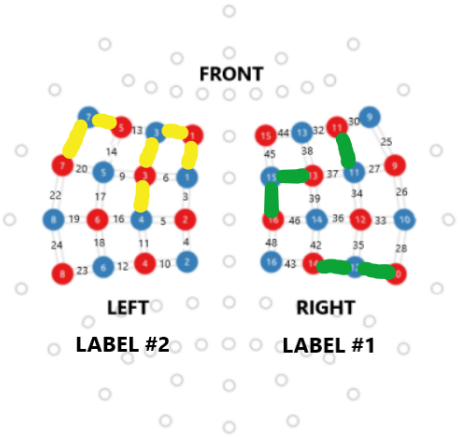

Fig. 5. Highlighting channels with highest positive SHAP values on probe layout for multi-label [rest(0), rest(0)]

SHAP values, based on the other finger tapping combinations were similarly generated. Subsequently, the SHAP values for the multi-labels [rest(0), 80bpm(1)] and [rest(0), 120bpm(2)] are shown in Fig. 6 and Fig. 7, respectively. In Fig. 6, the multi-label [rest(0), 80bpm(1)] represents the simultaneous action of the right index finger tapping at 80bpm while the left index finger is at rest. Similarly, in Fig. 7, the multi-label [rest(0), 120bpm(2)] represents the left index finger at rest while the right index finger is tapping at 120bpm. In both results, as expected, a greater number of positive SHAP values are shown in the right side channels for a prediction of '0' for Label #1. The second label shows the brightest positive SHAP values in the left side channels for predictions of '1' and '2' for multi-labels [rest(0), 80bpm(1)] and [rest(0), 120bpm(2)], respectively. It is interesting to note the difference in scale between Fig. 6 and Fig. 7, which seems to indicate that the activation levels for the simultaneous detection of an index finger at rest while the other one is tapping at 80bpm is lower than when the tapping is at a higher rate, 120bpm.

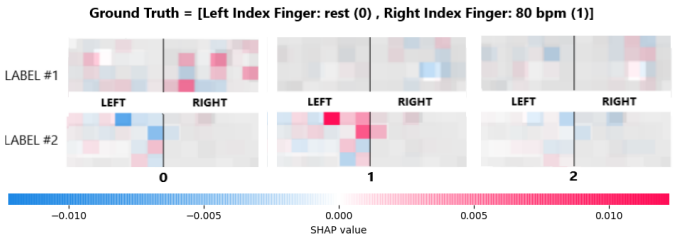

Fig. 6. SHAP values for multi-Label [rest(0), 80bpm(1)]

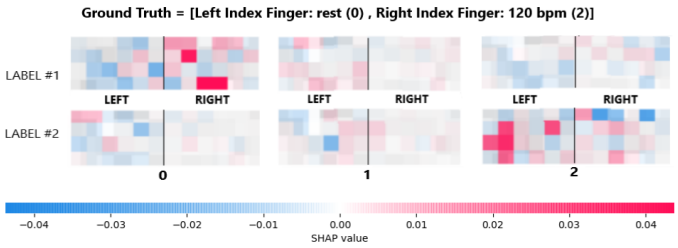

Fig. 7. SHAP values for multi-label [rest(0), 120bpm(2)]

A similar difference in scale persists when comparing the SHAP values for the multi-labels [80bpm(1), rest(0)] and [120bpm(2), rest(0)] as shown in Fig. 8 and Fig. 9, respectively. In Fig. 8, the left finger is tapping at 80bpm while the right finger is at rest. In Fig. 9, the left finger is tapping at the higher rate, 120bpm, and the right finger is at rest. Corresponding channels from the two sides of the motor cortices show the brightest SHAP values, as expected.

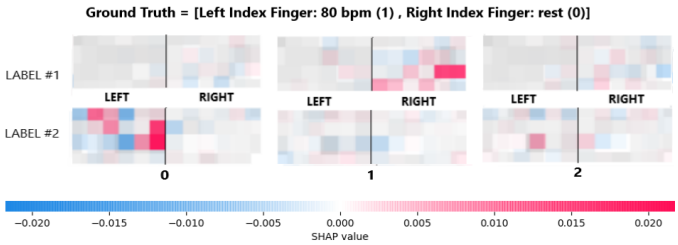

Fig. 8. SHAP values for multi-label [80bpm(1), rest(0)]

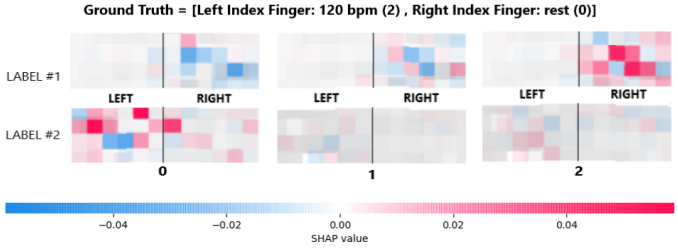

Fig. 9. SHAP values for multi-label [120bpm(2), rest(0)]

Finally, the SHAP values for the multi-labels, which represent both index fingers tapping at the same rate, namely [80bpm(1), 80bpm(1)] and [120bpm(2), 120bpm(2)], are shown in Fig. 10 and Fig. 11, respectively. As anticipated, simultaneous activation is seen in the channels representing both sides of the motor cortices for the same classes in the two figures.

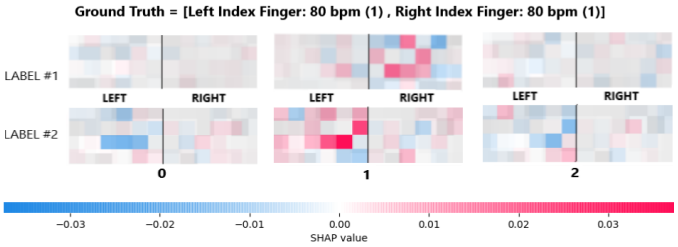

Fig. 10. SHAP values for multi-label [80bpm(1), 80bpm(1)]

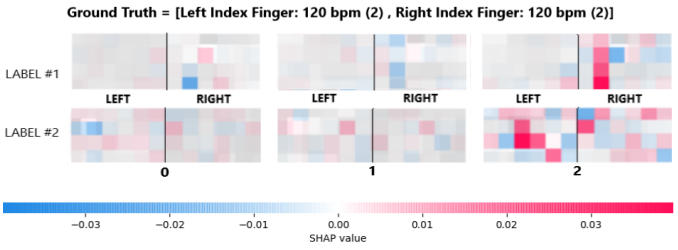

Fig. 11. SHAP values for multi-label [120bpm(2), 120bpm(2)]

Overall, the SHAP values support the well-established contralateral correlation between the sides of motor cortex activation and finger tapping hand [58]. Also, higher levels of activation are exhibited when one finger is at rest and the other finger is tapping at a class level of '2' as compared to a class level of '1' as seen in Fig. 6 through Fig. 9. The smaller scale for the resulting 80bpm SHAP values (Fig. 6 & Fig. 8) as compared to the 120bpm (Fig. 7 & Fig. 9) supports this. Therefore, this study reinforces the use of SHAP values as a way of understanding a deep learning 'black box' model. The locations of the brightest positive SHAP values provide an affirmative backing of León's [4] statement that: "During the preparation of a motor task certain cortical and subcortical regions are activated; contrarily, in the course of motor imageries most of the activity is found within the primary motor cortex over the corresponding contralateral hemisphere."

By showing that SHAP produces expected results on finger tapping data, where we have known a priori hypotheses about the patterns of activation expected during different finger tapping conditions, we propose that SHAP can and should be further explored on more complex deep learning models that attempt to predict a variety of social-cognitive-affective states (e.g., workload, attention, trust, frustration). With a growing body of literature using machine learning to predict these mental states from brain data, there is a growing need to evaluate and interpret these models [61], and SHAP can provide one additional means of interpretation. Furthermore, with new wearable biotechnology developments, we are seeing more models that are trained from data collected in operational settings [62] - [67]. These models may make predictions of complex social-cognitive-affective states and SHAP can be used on those models to add to the neuroscience of human information processing, shedding light on the neural correlates and interconnected brain regions underlying these highly complex states.

B. Further Discussion of Results

We have presented all seven of our multi-labels using SHAP values, and were able to demonstrate that the positive impact values (in red) were located in corresponding sides of the brain. For example, the SHAP values for the multi-label [120bpm(2), rest(0)] are showing the highest activation on the right motor cortex for class '2' and left motor cortex for class '0'. These map correctly to the left index finger tapping label of '2' and the right index finger tapping label of '0'. This supports our model's weight assignments for the learned features after training and validation.

Our Hamming Loss results show that on average, for the first and second testing trials, correct predictions were made 80% and 83% of the time, respectively. The corresponding micro-averaged F-score and accuracy metrics were also close to these results in both cases. These outcomes are better than the accuracy obtained by Bak et al. [11] who obtained an average accuracy of about 70.4% when classifying fNIRS data based on left index finger tapping, right index finger tapping, and foot tapping with an SVM model. Their model was meant to distinguish between the three types of tapping separately and not simultaneously, as in our approach with a multi-label. Moreover, classification was binary and not multi-class. Another study as reported by Woo et al. [8] showed that augmentation followed by training a CNN model resulted in a validation accuracy of 97.17% for thumb tapping and little finger tapping. Once again, a binary classifier was used to distinguish the tappings of two different fingers but, in this case, the fingers were on the same hand. Although these results are based on a simpler classification than ours, for a more commensurate comparison, we tested the classification network employed by Woo et al. [8], AlexNet, on our dataset. We tested the AlexNet algorithm using the training/validation data from our first set of testing trials. Resulting validation metrics (AlexNet) are presented next to the ones extracted from Table V (Our Network) in Table X. The resulting overall average AlexNet validation accuracy of 87.8% falls short of our network's results, thereby supporting our network model as presented in Fig. 2. AlexNet does not include an LSTM layer, which is important to capture label dependency in our multi-labels. This is supported with our network's overall average validation accuracy of 99.52% presented in Tables V and X.

Nazeer et al. [12] used a sample size of seven subjects in support of classification accuracy results of 85.4 +/- 1.4% when distinguishing between left finger tapping, right finger tapping, and rest. This result, although based on a three-class classification like ours, does not include different tapping frequencies and multi-labels. Our approach goes one step further where discerning between left index finger tapping and

Run #	Average Validation Loss (AlexNet)	Validation Accuracy Label #1 (AlexNet)	Validation Accuracy Label #2 (AlexNet)	Average Validation Loss (Our Network)	Validation Accuracy Label #1 (Our Network)	Validation Accuracy Label #2 (Our Network)
1	0.2675	0.8626	0.8821	0.0178	0.9948	0.9947
2	0.2302	0.8954	0.8872	0.0064	0.9962	0.9948
3	0.2579	0.8458	0.8543	0.0063	0.9941	0.9948
4	0.1911	0.8876	0.8623	0.0593	0.9919	0.9932
5	0.1355	0.8974	0.8743	0.0106	0.9974	0.9979
6	0.1901	0.8907	0.8767	0.0198	0.9953	0.9949
7	0.1080	0.8750	0.8951	0.0114	0.9967	0.9959

TABLE X: A comparison of validation metrics for 7 cross-validation runs based on AlexNet and our network

right finger tapping is embedded in our multi-labeling schema allowing us to be able to successfully differentiate between three different tapping frequencies simultaneously for each time step. It eliminates the need to distinguish between right and left finger tapping while providing the ability to classify more complex information.

Our proposed approach introduces a novel spatially descriptive multi-labeling schema, which uses two labels to represent the two sides of the brain. As both labels are classified *simultaneously*, one of three finger tapping frequencies can be assigned to each label. As our SHAP values illustrate, our model is able to learn the correct features of the fNIRS data, which contribute to the class predictions of our multi-labels. Finally, our multi-labeling classification was able to benefit from the use of the LSTM, known for its ability to detect label dependencies [53]–[55].

V. Conclusion

Providing a means of communicating with the outside world for those who have compromised motor function has played an important role in HCI/BCI research. Gaining a better understanding of how to capture the information in fNIRS brain signals to autonomously detect different levels of simultaneous index finger tapping frequencies can contribute to this endeavor. For example, being able classify physical finger tappings at different frequencies can impart a better understanding of brain activity during imagined finger tappings, offering a path of communication through BCI for locked-in patients. Also, understanding brain activation patterns during finger tapping activities can assist in stroke rehabilitation and assess the severity of Parkinson's Disease. We have presented a promising approach to multi-labeling spatiotemporal data to detect different classes in two regions of interest *concurrently*. In this case, we have applied our approach to detect three different levels (rest, 80bpm, and 120bpm) of finger tapping for both index fingers at the same time. The spatial aspect of our fNIRS data, formatted to reflect probe layout, is captured with CNNs and the temporal one is captured with an LSTM. Our novel multi-labeling technique enables us to classify activity on both sides of the brain simultaneously with our network. Our network's SHAP values support its ability to choose appropriate spatial features (which importantly aligns with the known spatial characteristics of finger tapping on the primary motor cortices) when making predictions. By using Shapley's model explainability technique on these benchmark finger tapping tasks, we demonstrate the potential of applying this approach to the more complex neuroscience classification of human information processing tasks, where the use of Shapley values can help to explain the model, while also having the potential of shedding light on the neural correlates and interconnected brain regions underlying highly complex social-cognitive-affective states.

References

[1] S. Khan, M.Sunny, M.Hossain, E.Hossain, and M.Ahmad, *Nose tracking cursor control for the people with disabilities: An improved HCI*, 2017 3rd International Conference on Electrical Information and Communication Technology (EICT), IEEE; 2017, p. 1–5.

[2] O. Prabhune, and P. Rege, *Speakingeyes: Enabling paralyzed people to communicate*, 2019 IEEE 16th India Council International Conference (INDICON), IEEE; 2019, p. 1–4.

[3] P. Goel, R. Joshi, M. Sur, H. Murthy, *A common spatial pattern approach for classification of mental counting and motor execution EEG*, International Conference on Intelligent Human Computer Interaction, Springer; 2018, p. 25–35.

[4] C. Lindig León. *Multilabel classification of EEG-based combined motor imageries implemented for the 3D control of a robotic arm*, URL: https://tel.archives-ouvertes.fr/tel-01549139, NUMBER: 2017LORR0016, Université de Lorraine, 2017, HAL Id : tel-01549139, version 1.

[5] Y. Li, X. Xiong, Z. Li, and Y. Fu, *Recognition of three different imagined movement of the right foot based on functional near-infrared spectroscopy*, Sheng wu yi xue gong cheng xue za zhi= Journal of biomedical engineering= Shengwu yixue gongchengxue zazhi, Vol. 37, No. 2, 2020, p. 262–270.

[6] X. Yin, B. Xu, C. Jiang, Y. Fu, Z. Wang, H. Li, and G. Shi, *A hybrid BCI based on EEG and fNIRS signals improves the performance of decoding motor imagery of both force and speed of hand clenching*, Journal of neural engineering, Vol. 12, No. 3, IOP Publishing; 2015.

[7] G. Hirsch, M. Dirodi, R. Xu, P. Reitner, and C. Guger, *Online Classification of Motor Imagery Using EEG and fNIRS: A Hybrid Approach with Real Time Human-Computer Interaction*, International Conference on Human-Computer Interaction, Springer; 2020, p. 231–238.

[8] S. Woo, M. Kang, and K. Hong, *Classification of Finger Tapping Tasks using Convolutional Neural Network Based on Augmented Data with Deep Convolutional Generative Adversarial Network*, 2020 8th IEEE RAS/EMBS International Conference for Biomedical Robotics and Biomechatronics (BioRob), IEEE; p. 328–333.

[9] V. Scarapicchia, C. Brown, C. Mayo, and J. Gawryluk, *Functional magnetic resonance imaging and functional near-infrared spectroscopy: insights from combined recording studies*,Frontiers in Human Neuroscience, Vol. 11, Frontiers; 2017, p. 419.

[10] T. Wilcox, and M. Biondi, *fNIRS in the developmental sciences*, Wiley Interdisciplinary Reviews: Cognitive Science, Vol. 6, No. 3, Wiley Online Library; 2015, p. 263–283.

[11] S. Bak, J. Park, J. Shin, and J. Jeong, *Open-access fNIRS dataset for classification of unilateral finger-and foot-tapping*, Electronics, Vol. 8, No. 12, Multidisciplinary Digital Publishing Institute; 2019.

[12] H. Nazeer, N. Naseer, R. Khan, F. Noori, N. Qureshi, U. Khan, and M. Khan, *Enhancing classification accuracy of fNIRS-BCI using features acquired from vector-based phase analysis*, Journal of Neural Engineering, Vol. 17, No. 5, IOP Publishing; 2020, p. 056025.

[13] P. Caçola, N. Getchell, D. Srinivasan, G. Alexandrakis, and H. Liu, *Cortical activity in fine-motor tasks in children with developmental coordination disorder: a preliminary fNIRS study*, International Journal of Developmental Neuroscience, Vol. 65, Elsevier; 2018, p. 83–90.

[14] A. Rahimpour, L. Pollonini, D. Comstock, R. Balasubramaniam, and H. Bortfeld, *Tracking differential activation of primary and supplementary motor cortex across timing tasks: An fNIRS validation study*, Journal of Neuroscience Methods, Vol. 341, Elsevier; 2020, p. 108790.

[15] M. Khan, M. Bhutta, and K. Hong, *Task-Specific stimulation duration for fNIRS brain-computer interface*, IEEE Access, Vol. 8, IEEE; 2020, p. 89093–89105.

[16] A. Zafar, U. Ghafoor, M. Yaqub, and K. Hong, *Initial-dip-based classification for fNIRS-BCI*, Neural Imaging and Sensing 2019, Vol. 10865, International Society for Optics and Photonics; 2019, p. 108651N.

[17] N. Sommer, S. Velipasalar, L. Hirshfield, Y. Lu, and B. Kakillioglu. *Simultaneous and Spatiotemporal Detection of Different Levels of Activity in Multidimensional Data*, IEEE Access Vol. 8, IEEE; 2020, p. 118205–118218.

[18] K. Sozykin, S. Protasov, A. Khan, R. Hussain, and J. Lee, *Multi-label class-imbalanced action recognition in hockey videos via 3d convolutional neural networks*, 2018 19th IEEE/ACIS International Conference on Software Engineering, Artificial Intelligence, Networking and Parallel/Distributed Computing (SNPD), IEEE; 2018, p. 146–151.

[19] M. Cartwright, J. Cramer, A. Mendez, Y. Wang, H. Wu,V. Lostanlen, M. Fuentes, G. Dove, C. Mydlarz, J. Salamon, Justin and others. *SONYC-UST-V2: An Urban Sound Tagging Dataset with Spatiotemporal Context*, arXiv preprint arXiv:2009.05188, 2020.

[20] A. Belkacem, S. Nishio, T. Suzuki, H. Ishiguro, and M. Hirata, Masayuki, *Neuromagnetic decoding of simultaneous bilateral hand movements for multidimensional brain–machine interfaces*, IEEE Transactions on Neural Systems and Rehabilitation Engineering, Vol. 26, No. 6, IEEE; 2018, p. 1301–1310.

[21] J. Ye, L. Sun, B. Du, Y. Fu, X. Tong, H. Xiong, Hui, *Co-prediction of multiple transportation demands based on deep spatio-temporal neural network*, Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining, 2019, p. 305–313.

[22] M. Trager, K. Wilkins, M. Koop, and H. Bronte-Stewart, *A validated measure of rigidity in Parkinson's disease using alternating finger tapping on an engineered keyboard*, Parkinsonism & related disorders, Vol. 81, Elsevier; 2020, p. 161–164.

[23] R. Krupicka, P. Kryze, S. Net'ukova, T. Duspivova, O. Klempir, Z. Szabo, P. Dusek, K. Sonka, J. Rusz, and E. Ruzicka, *Instrumental analysis of finger tapping reveals a novel early biomarker of parkinsonism in idiopathic rapid eye movement sleep behaviour disorder*, Sleep Medicine, Vol. 75, Elsevier; 2020, p. 45–49.

[24] A. Alves-Pinto, S. Ehrlich, G. Cheng, V. Turova, T. Blumenstein, and R. Lampe, *Effects of short-term piano training on measures of finger tapping, somatosensory perception and motor-related brain activity in patients with cerebral palsy*, Neuropsychiatric disease and treatment, Vol. 13, Dove Press; 2017, p. 2705.

[25] J. Birchenall, M. Térémetz, P. Roca, J. Lamy, C. Oppenheim, M. Maier, J. Mas, C. Lamy, J. Baron, and P. Lindberg, *Individual recovery profiles of manual dexterity, and relation to corticospinal lesion load and excitability after stroke–a longitudinal pilot study*, Neurophysiologie Clinique, Vol. 49, No. 2, Elsevier; 2019, p. 149–164.

[26] B. Young, Z. Nigogosyan, L. Walton, J. Song, V. Nair, S. Grogan, M. Tyler, D. Edwards, K. Caldera, J. Sattin, and others, *Changes in functional brain organization and behavioral correlations after rehabilitative therapy using a brain-computer interface*, Frontiers in Neuroengineering, Vol. 7, Frontiers; 2014, p. 26.

[27] M. Khan, and K. Hong, *Active brain area identification using EEG-NIRS signal acquisition*, 2015 International Conference on Automation, Cognitive Science, Optics, Micro Electro-Mechanical System, and Information Technology (ICACOMIT), IEEE; 2015, p. 7–11.

[28] M. Peifer, L. Zhu, and L. Najafizadeh, *Real-Time Classification of Actual vs Imagery Finger Tapping Using fNIRS*, Biomedical Optics, Optical Society of America. 2014, p. BM3A–34.

[29] H. Nazeer, N. Naseer, R. Khan, F. Noori, N. Qureshi, U. Khan, and M. Khan, *Enhancing classification accuracy of fNIRS-BCI using features acquired from vector-based phase analysis*, Journal of Neural Engineering, Vol. 17, No. 5, IOP Publishing; 2020, p. 056025.

[30] S. Zhang, Y. Zheng, D. Wang, L. Wang, J. Ma, J. Zhang, W. Xu, D. Li, and D. Zhang, *Application of a common spatial pattern-based algorithm for an fNIRS-based motor imagery brain-computer interface*, Neuroscience Letters, Vol. 655, Elsevier; 2017, p. 35–40.

[31] R. Sitaram, H. Zhang, C. Guan, M. Thulasidas, Y. Hoshi, A. Ishikawa, K. Shimizu, and N. Birbaumer, *Temporal classification of multichannel near-infrared spectroscopy signals of motor imagery for developing a brain–computer interface*, NeuroImage, Vol. 34, No. 4, Elsevier; 2007, p. 1416–1427.

[32] A. Batula, J. Mark, Y. Kim, and H. Ayaz, *Comparison of brain activation during motor imagery and motor movement using fNIRS*, Computational intelligence and neuroscience, Vol. 2017, Hindawi; 2017.

[33] W. Chen, J. Wagner, N. Heugel, J. Sugar, Y. Lee, L. Conant, M. Malloy, J. Heffernan, B. Quirk, A, Zinos, and others, *Functional near-infrared spectroscopy and its clinical application in the field of neuroscience: advances and future directions*, Frontiers in Neuroscience, Vol. 14, Frontiers; 2020, p. 724.

[34] K. Hong, and M. Yaqub, *Application of functional near-infrared spectroscopy in the healthcare industry: A review*, Journal of Innovative Optical Health Sciences, Vol. 12, No. 6, World Scientific; 2019, p. 1930012.

[35] T. Siddique, and M. Mahmud, *Classification of fNIRS Data Under Uncertainty: A Bayesian Neural Network Approach*, arXiv preprint arXiv:2101.07128, 2021.

[36] T. Trakoolwilaiwan, B. Behboodi, J. Lee, K. Kim, and J. Choi, *Convolutional neural network for high-accuracy functional near-infrared spectroscopy in a brain–computer interface: three-class classification of rest, right-, and left-hand motor execution*, Neurophotonics, Vol. 5, No. 1, International Society for Optics and Photonics; 2017, p. 011008.

[37] S. Wickramaratne, and M. Mahmud, *A Deep Learning Based Ternary Task Classification System Using Gramian Angular Summation Field in fNIRS Neuroimaging Data*, arXiv preprint arXiv:2101.05891, 2021.

[38] U. Asgher, K. Khalil, M. Khan, R. Ahmad, S. Butt, Y. Ayaz, N. Naseer, and S. Nazir, *Enhanced accuracy for multiclass mental workload detection using long short-term memory for brain–computer interface*, Frontiers in Neuroscience, Vol. 14, Frontiers; 2020, p. 584.

[39] L. Xu, Y. Liu, J. Yu, X. Li, X. Yu, H. Cheng, and J. Li, *Characterizing autism spectrum disorder by deep learning spontaneous brain activity from functional near-infrared spectroscopy*, Journal of Neuroscience Methods, Vol. 331, Elsevier; 2020, p. 108538.

[40] F. Herrera, F. Charte, A. Rivera, and M. Del Jesus, *Multilabel classification*, Springer; 2016, p. 17–31.

[41] Y. Wang, D. He, F. Li, X. Long, Z. Zhou, J. Ma, and S. Wen, *Multi-label classification with label graph superimposing*, Proceedings of the AAAI Conference on Artificial Intelligence, Vol. 34, No. 7, 2020, p. 12265–12272.

[42] S. Abu-El-Haija, N. Kothari, J. Lee, P. Natsev, G. Toderici, B. Varadarajan, and S. Vijayanarasimhan, *Youtube-8m: A large-scale video classification* arXiv preprint arXiv:1609.08675, 2016.

[43] M. Cartwright, J. Cramer, A. Mendez, Y. Wang, H. Wu,V. Lostanlen, M. Fuentes, G. Dove, C. Mydlarz, J. Salamon, Justin and others. *SONYC-UST-V2: An Urban Sound Tagging Dataset with Spatiotemporal Context*, arXiv preprint arXiv:2009.05188, 2020.

[44] A. Olsson, P. Sager, E. Andersson, A. Björkman, N. Malešević, and C. Antfolk, *Extraction of multi-labelled movement information from the raw HD-sEMG image with time-domain depth*, Scientific reports, Vol. 9, No. 1, Nature Publishing Group; 2019, p. 1–10.

[45] D. Freer, and G. Yang, *Data augmentation for self-paced motor imagery classification with C-LSTM*, Journal of Neural Engineering, Vol. 17, No. 1, IOP Publishing; 2020, p. 016041.

[46] P. Kalansooriya, GA Ganepola, and TS Thalagala, *Affective gaming in real-time emotion detection and Smart Computing music emotion recognition: Implementation approach with electroencephalogram*, 2020 International Research Conference on Smart Computing and Systems Engineering (SCSE), IEEE; 2020, p. 111–116.

[47] S. Kudugunta, and E. Ferrara, *Deep neural networks for bot detection*, Information Sciences, Vol. 467, Elsevier; 2018, p. 312–322.

[48] R. Pereira, A. Plastino, B. Zadrozny, and L. Merschmann, *Correlation analysis of performance measures for multi-label classification*, Information Processing & Management, Vol. 54, No. 3, Elsevier; 2018, p. 359–369.

[49] G. Tsoumakas, and I. Katakis, *Multi-label classification: An overview*, International Journal of Data Warehousing and Mining (IJDWM) Vol 3, IGI Global; 2007, p. 1–13.

[50] M. Zhang, and Z. Zhou, *A unified view of multi-label performance measures*, Proceedings of the 34th International Conference on Machine Learning Vol 70, 2017, p. 3780–3788.

[51] S. Lundberg, and S. Lee, *A unified approach to interpreting model predictions*, arXiv preprint arXiv:1705.07874, 2017.

[52] Y. Hailemariam, A. Yazdinejad, R. Parizi, G. Srivastava, and A. Dehghantanha, *An Empirical Evaluation of AI Deep Explainable Tools*, 2020 IEEE Globecom Workshops, IEEE; 2020, p. 1–6.

[53] L. Yao, E. Poblenz, D. Dagunts, B. Covington, D. Bernard, and K. Lyman, *Learning to diagnose from scratch by exploiting dependencies among labels*, arXiv preprint arXiv:1710.10501; 2017.

[54] J. Wang, Y. Yang, J. Mao, Z. Huang, C. Huang, and W. Xu, *Cnn-rnn: A unified framework for multi-label image classification*, Proceedings of the IEEE conference on computer vision and pattern recognition, IEEE; 2016, p. 2285–2294.

[55] B. Zhao, X. Li, X. Lu, and Z. Wang, *A CNN–RNN architecture for multi-label weather recognition*, Neurocomputing, Vol. 322, Elsevier; 2018, p. 47–57.

[56] L. Shapley, *A value for n-person games*, Contributions to the Theory of Games, Vol. 2, No. 28, 1953, p. 307–317.

[57] A. Schaefer, R. Kong, E. Gordon, T. Laumann, X. Zuo, A. Holmes, S. Eickhoff, B. Yeo, *Local-global parcellation of the human cerebral cortex from intrinsic functional connectivity MRI*, Cerebral cortex, Vol. 28, No. 9, Oxford University Press; 2018, p. 3095–3114.

[58] C. Horenstein, M. Lowe, K. Koenig, and M. Phillips, *Comparison of unilateral and bilateral complex finger tapping-related activation in*

premotor and primary motor cortex, Human brain mapping, Vol. 30, No. 4, Wiley Online Library; 2009, p. 1397–1412.
[59] https://imbalanced-learn.org/stable/references/index.html. Accessed on 17 February 2021.
[60] D. Devlaminck, W. Waegeman, B. Bauwens, B. Wyns, P. Santens, and G. Otte, *From circular ordinal regression to multilabel classification*, Proceedings of the 2010 Workshop on Preference Learning (European Conference on Machine Learning, ECML), 2010, p. 15.
[61] A. Brouwer, T. Zander, J. Van Erp, J. Korteling, and A. Bronkhorst, *Using neurophysiological signals that reflect cognitive or affective state: six recommendations to avoid common pitfalls*, Frontiers in neuroscience, Vol. 9, Frontiers; 2015, p. 136.
[62] L. Hirshfield, R. Gulotta, S. Hirshfield, S. Hincks, M. Russell, R. Ward, T. Williams, and R. Jacob, *This is your brain on interfaces: enhancing usability testing with functional near-infrared spectroscopy*, Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, 2011, p. 373–382.
[63] M. Lohani, B. Payne, and D. Strayer, *A review of psychophysiological measures to assess cognitive states in real-world driving*, Frontiers in human neuroscience, Vol. 13, Frontiers; 2019, p. 57.
[64] R. McKendrick, B. Feest, A. Harwood, and B. Falcone, *Theories and methods for labeling cognitive workload: Classification and transfer learning*, Frontiers in human neuroscience, Vol. 13, Frontiers; 2019, p. 295.
[65] R. McKendrick, R. Parasuraman, R. Murtza, A. Formwalt, W. Baccus, M. Paczynski, and H. Ayaz, *Into the wild: neuroergonomic differentiation of hand-held and augmented reality wearable displays during outdoor navigation with functional near infrared spectroscopy*, Frontiers in human neuroscience, Vol. 10, Frontiers; 2016, p. 216.
[66] R. Parasuraman, and M. Rizzo, *Neuroergonomics: The brain at work*, Oxford University Press, 2008.
[67] A. Dimoka, *How to conduct a functional magnetic resonance (fMRI) study in social science research*, MIS quarterly, JSTOR; 2012, p. 811–840.
[68] H. Sugata, M. Hirata, T. Yanagisawa, K. Matsushita, S. Yorifuji, and T. Yoshimine, *Common neural correlates of real and imagined movements contributing to the performance of brain–machine interfaces*, Scientific reports, Vol. 6, No. 1, Nature Publishing Group; 2016, p. 1–11.
[69] K. Kilteni, B. Andersson, C. Houborg, and H. Ehrsson, *Motor imagery involves predicting the sensory consequences of the imagined movement*, Nature communications, Vol. 9, No. 1, Nature Publishing Group; 2018, p. 1–9.
[70] C. Wurster, H. Graf, H. Ackermann, K. Groth, J. Kassubek, and A. Riecker, *Neural correlates of rate-dependent finger-tapping in Parkinson's disease*, Brain Structure and Function, Vol. 220, No. 3, Springer; 2015, p. 1637–1648.

Natalie M. Sommer (M'00) received the B.S. degree in electrical engineering and the M.S. degree in electrical engineering from Union College, Schenectady, NY, in 1991. She is currently pursuing the Ph.D. degree with the Department of Electrical Engineering and Computer Science, Syracuse University. Her research interests include deep machine learning methods for activity recognition in video data and fNIRS brain signals.

Burak Kakillioglu received B.S. degree in electrical and electronics engineering from Bilkent University, Ankara, Turkey, in 2015. He is currently pursuing a Ph.D. degree with the Department of Electrical Engineering and Computer Science, Syracuse University, Syracuse, NY, USA. In 2019, he was a research intern with the Computer Vision Technologies Group, SRI International, Princeton, NJ, USA. His research interests include 3D data understanding, computer vision, machine learning, and embedded systems.

Trevor Grant is a graduate student in the Computer Science department at University of Colorado Boulder. His primary focus is on exploring the neurophysiological correlates to different types of mental workload experienced while users interact with virtual environments. Consequently, his work overlaps with various subdomains of cognitive psychology, primarily by using theories of attention, perception and memory to develop models of mental load.

Senem Velipasalar (M'04–SM'14) received the B.S. degree in electrical and electronic engineering from Bogazici University, Istanbul, Turkey, in 1999, the M.S. degree in electrical sciences and computer engineering from Brown University, Providence, RI, USA, in 2001, and the M.A. and Ph.D. degrees in electrical engineering from Princeton University, Princeton, NJ, USA, in 2004 and 2007, respectively. From 2001 to 2005, she was with the Exploratory Computer Vision Group, IBM T. J. Watson Research Center, NY, USA. From 2007 to 2011, she was an Assistant Professor with the Department of Electrical Engineering, University of Nebraska-Lincoln, Lincoln. She is currently a Professor in the Department of Electrical Engineering and Computer Science at Syracuse University.

The focus of her research has been on mobile camera applications, wireless embedded smart cameras, multi-camera tracking and surveillance systems, and automatic event detection from videos. Dr. Velipasalar received a Faculty Early Career Development Award (CAREER) from the National Science Foundation in 2011. She is the recipient of the 2014 Excellence in Graduate Education Faculty Recognition Award. She is the coauthor of the paper, which received the 2017 IEEE Green Communications and Computing Technical Committee Best Journal Paper Award. She received the Best Student Paper Award at the IEEE International Conference on Multimedia and Expo in 2006. She is the recipient of the EPSCoR First Award, IBM Patent Application Award, and Princeton and Brown University Graduate Fellowships. She is a member of the Editorial Board of the IEEE Transactions on Image Processing and Springer Journal of Signal Processing Systems.

Dr.Leanne Hirshfield is an Associate Research Professor at the University of Colorado's Institute of Cognitive Science. Hirshfield directs the research efforts in the System Human Interaction with NIRS and EEG (SHINE) Laboratory. She received her Ph.D.in Computer Science from Tufts University in 2009 having completed her undergraduate and M.S. degrees in Computer Science at Hamilton College and the Colorado School of Mines, respectively. Her specialty is in the areas of Human-Computer Interaction (HCI) and Machine Learning (ML) and she has a wealth of experience designing and evaluating various user interfaces. Dr. Hirshfield's research in the SHINE Lab explores the use of non-invasive brain measurement to passively classify user states in order to enhance usability testing and adaptive system design. Dr. Hirshfield's research focuses on the use of a relatively new, non-invasive brain imaging device called functional near-infrared spectroscopy (fNIRS). The fNIRS device is safe, portable, robust to noise, and it can be implemented wirelessly; making it ideal for research in HCI. While much of her research focuses on fNIRS, Dr. Hirshfield also works extensively with EEG, galvanic skin response sensors, and eyetracking devices in her lab.