Trial-based Classification of Haptic Tasks Based on EEG Data

Haneen Alsuradi[1] and Mohamad Eid[2]

Abstract— **With the increasing popularity of neural imaging techniques such as electroencephalography (EEG), developing quantitative measures to characterize haptic interactions is becoming a reality. Meanwhile, machine learning is a promising approach for trial-based EEG data analysis. This work presents a model that can distinguish between passive and active kinesthetic interactions based on a single trial EEG data. An interactive task that involves hitting a ball using a racket is developed under passive and active kinesthetic settings using a haptic device and a computer screen. Temporal and frequency domain features are extracted from the motor and somatosensory cortices, and a proposed 2-D CNN model is trained on data extracted from 19 participants. The model achieves a mean accuracy of 84.56%, 93.96%, and 95.89% across 5-fold validation when using one, four, or six electrodes, respectively. The model mechanism is assessed using an explainable machine learning algorithm, LIME, which shows that the model utilizes sensible features from a neuroscience perspective towards its prediction. This work paves the way for a better understanding of the neural mechanisms associated with kinesthetic haptic interaction, which proves helpful in many applications such as motor rehabilitation and brain-computer interactions, in addition to modeling the haptic quality of experience objectively.**

I. Introduction

Haptic technologies have paved the way for making touch part of the information flow between the user and the computer. Haptic interaction involves bidirectional communication of cutaneous (such as contact, pressure, and vibration) and kinesthetic (force or motion) sensations. Two types of haptic exploration modes can be distinguished: active and passive [1]. In active interaction, the user moves their body to initiate the haptic interaction, whereas passive interaction does not involve any body movement. Active exploration is predominantly exploited in applications dealing with object manipulation and grasping [2], whereas passive exploration is preferred in training and education [3].

Recently, there has been an increased interest in studying touch perception using neuroimaging techniques, and in particular, electroencephalography (EEG) [4]. Compared to other neural imaging techniques such as fMRI, EEG is easy to use, portable, affordable, and provides superior temporal resolution [5]. In addition to unveiling information about how touch information is encoded in the brain, EEG data can train Machine Learning (ML) models to allow for an automatic interpretation of neural activities associated with physical interaction and thus quantify the perceptual haptic experience [6][7].

Developing a model to classify neural activities associated with passive and active haptic interaction is crucial for several application such as rehabilitative and therapeutic interventions designed to mitigate sensorimotor impairments [8], human-robot interaction [9], brain-computer interaction [10], and quantitative evaluation of the quality of haptic experience [4]. In this paper, we present a Convolution Neural Network (CNN) model to identify passive and active haptic tasks using a single trial of EEG data.

II. Related Work

Neural mechanisms associated with active and passive cutaneous interaction have been long and extensively studied in the literature [11]. Most studies focused on the interaction with textured surfaces. For instance, a device is introduced in [12] to provide dynamic passive stimulation that mimics the movement of sliding an object against a participant's finger. Results demonstrated a bilateral desynchronization in the alpha band throughout the passive stimulation. Another study showed a significant relationship between the power of the beta band and the discrimination between soft and rough textures [13]. Interestingly, a recent study showed that both active and passive cutaneous interactions, though physically different, activate similar cortical areas in the brain (contralateral central and parietal areas) [14].

Meanwhile, there has been a rising interest in studying the cortical activity associated with passive and active kinesthetic interactions using EEG. It is known that active kinesthetic interaction can evoke greater brain activation due to the activation of both the motor and somatosensory cortices as compared to passive kinesthetic interaction that involves only the somatosensory cortex [15]. A feasibility study to examine brain responses to kinesthetic interaction in a 3D virtual environment revealed variations in the peak magnitudes and latencies of the event-related potential (ERP) responses [16]. The kinesthetic interaction involved force and vibrotactile feedback while flying a virtual drone.

A few attempts were made to classify kinesthetic interaction using EEG and machine learning. For instance, two types of active haptic tasks, namely catch and touch, are classified based on EEG data using a three-layer neural network [6]. Results showed that a small number of electrodes (C3, C4, P3, and P4) provided the highest classification accuracy. Another study utilized a deep convolutional neural network model to classify the kinesthetic motor imagery task of walking [17].

None of the previous studies classify the type of kinesthetic task on the basis of being passive or active with single-trial EEG data. This paper aims to develop a CNN model to

[1]Haneen Alsuradi is affiliated with New York University, Tandon School of Engineering, NY, USA. [2]Mohamad Eid is affiliated with Engineering Division, New York University Abu Dhabi, Saadiyat Island, Abu Dhabi, 129188, United Arab Emirates. Email: {haneen, mohamad.eid}@nyu.edu.

classify kinesthetic interaction as passive or active based on single-trial EEG data. The active task involves hitting a tennis ball with a racket, whereas the ball falls off and collides with the racket in the passive task. Furthermore, explainable ML is utilized to identify the most influential EEG features for classifying a kinesthetic interaction as passive or active.

III. Experimental Design

A. Apparatus and tasks

Fig. 1 depicts the recording environment and the apparatus used in the study. Participants were asked to sit comfortably on a chair, approximately one meter away from a computer screen, and to hold the stylus of the haptic device (Geomagic Touch [1], 3D systems, United States) with their right hand while resting their arms on the table. The game was developed using Unity game engine version 2018.4.5f1 and Openhaptics Unity toolkit. The aim of the task is to bounce a tennis ball using a racket shown on the screen and controlled by the haptic device. During the passive task, participants are asked to press the button on the stylus and passively hold the racket waiting for the ball to fall off and collide with the racket. However, in the active task, participants are asked to actively move the racket up towards the ball. Force feedback is delivered when the ball bounces off the racket's surface regardless of the task type. The experimental task sequence for a single trial is illustrated in Fig. 2 for both the passive and the active tasks. A total of 10 runs were conducted divided equally between the active and the passive task and presented in a counterbalanced fashion. Each run consisted of 10 trials of the ball bouncing task. Thus, in total, we collected 950 trials for the passive task and another 950 trials for the active task.

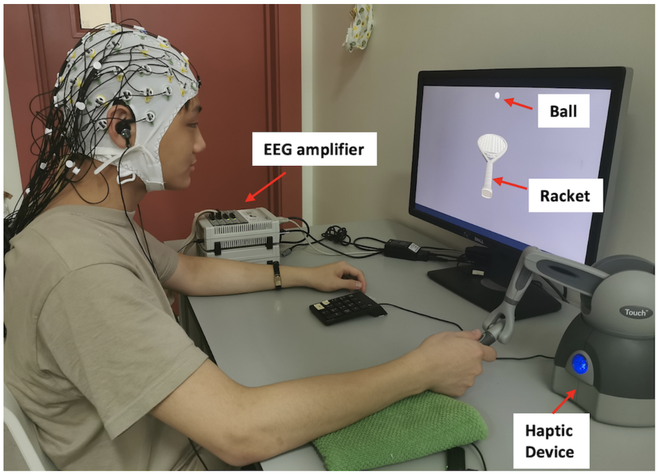

Fig. 1: Apparatus and experimental setup

B. Participants

Nineteen right-handed and healthy subjects (10 females, 9 males) aged 18 to 40 years with no reported traumatic brain injuries, neural abnormalities, and/or muscle atrophy were recruited in this study. The exclusion criteria include participants below the age of 18 or left-handed individuals. The study was approved by New York University Abu Dhabi Institutional Review Board (IRB: #HRPP-2019-120) and was conducted per the Declaration of Helsinki, following its guidelines and regulations. Written informed consent was obtained from all participants after being informed about the study's purpose and procedure.

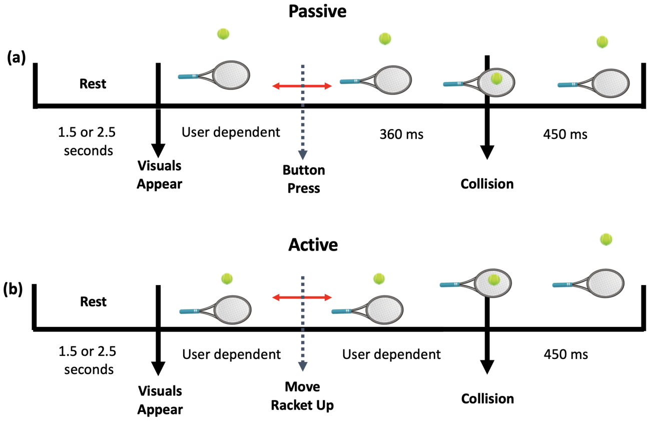

Fig. 2: Schematic representation of the experimental task

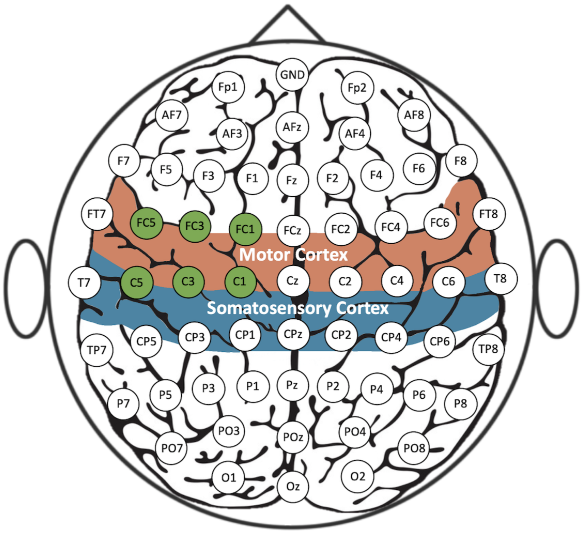

Fig. 3: Positions of the selected electrodes. Electrodes lie above the motor and the somatosensory cortices.

C. EEG data pre-processing

EEG data were recorded at a 1 kHz sampling rate using an EEG amplifier and a 64 Ag/AgCL based electrode set (BrainAmps Standard[2], Brain Products, Germany). Four channels at the EEG cap circumference were excluded (FT9, FT10, TP9, and TP10). A 0.1–85 Hz bandpass filter and a 50 Hz notch filter were applied to the data, followed by applying the Artifact Subspace Reconstruction (ASR) [18] method to remove high-amplitude artifacts; the following

[1] https://www.3dsystems.com/haptics-devices/touch

[2] https://www.brainproducts.com/productdetails.php?id=74

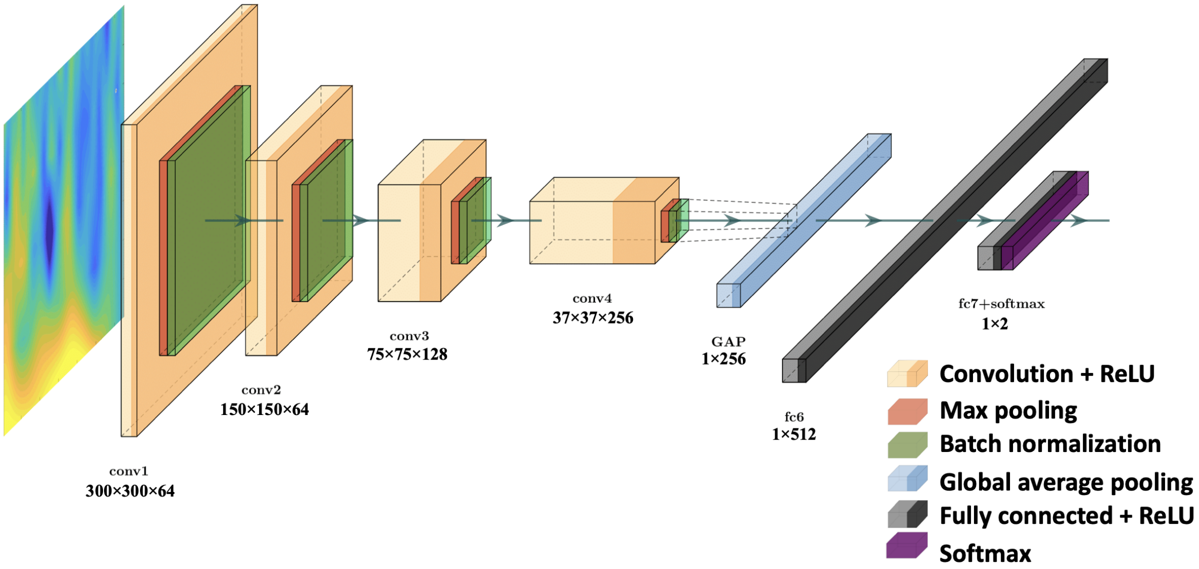

Fig. 4: Proposed 2-D-CNN architecture

parameters were used (argflatline=10, arghighpass=[0.025 0.075], argchannel= 0.8, argnoisy=4, argburst=20, argwindow='off'). Channels were then re-referenced using the Common Average Referencing (CAR) method while restoring the online reference channel (FCz) to the data set. Since the task is asynchronous (the timing is user-dependent), the data is epoched such that it includes 500ms before and 800 ms after the collision point. EEG data were transformed to the time-frequency domain via Morlet Wavelet transformation [19]. Frequencies are averaged, yielding five frequency bins, corresponding to the five following frequency bands: delta (1–4 Hz), theta (4–9 Hz), alpha (9–13 Hz), beta (13–30 Hz), and gamma (30–80 Hz). Each frequency bin had a baseline correction using the interval between 1000 ms to 800 ms before the collision. Contaminated trials were rejected, yielding 944 trials for the passive task and 928 trials for the active task.

IV. Deep Learning Classifier

A. Feature extraction

As the EEG data are high dimensional, it is crucial to well-prepare the data before feeding it to a classifier. In particular, the epoched EEG data have three dimensions: electrodes, time, and frequency bands. Since this classification task is based on differentiating a motor movement neural activation (active task) from a motor stillness neural activation (passive task), the cortical regions of interest (ROI) are identified as the contralateral motor and somatosensory cortices [15]. Time-frequency heatmaps are generated from the six electrodes highlighted in Fig. 3 and used to train a 2-D convolutional neural network (CNN). A single heatmap image represents a single trial, which has time as the x-axis (500 ms before the collision up to 800 ms after the collision), frequency band as the y-axis (1: delta, 2: theta, 3: alpha, 4: beta, 5: gamma) and the power as the pixel values. The model was trained based on three scenarios: using one electrode (C1), using four electrodes (C1, C3, FC1, and FC3), and using six electrodes (C1, C3, C5, FC1, FC3, and FC5). Heatmaps are averaged in scenarios where more than one electrode is used. A single heatmap image is sized as 300 pixels×300 pixels×3 RGB channels. We use colored RGB images instead of the power coefficients of each time-frequency data point to visualize the features that contribute towards the classification decision at the explainability step later on. Particularly, we use the classical colormap transformation, 'jet', commonly used in the neuroscience community which offers a good contrast between high and low power coefficients.

B. Model and training

The proposed 2-D CNN is composed of 4 convolutional layers that reduce in size laterally and increase in depth, as shown in Fig. 4. The first two convolutional layers have 64 filters and 8×8 kernel size and a "same" padding mode. The third and the fourth convolutional layers, on the other hand, have 128 and 256 filters respectively and a 3×3 kernel size. Each convolutional layer was followed by a ReLU activation function, a 2-D max pooling layer with 2×2 pooling window and a batch normalization layer [20]. A global average pooling (GAP) layer is added following the convolutional layers; GAP layers are more native to the convolution process compared to fully-connected layers such that they enforce relationships between the feature map and the target classes [21]. Following the GAP layer, two fully-connected layers were added with ReLU and softmax activations respectively at the output. A dropout layer between the last two fully connected layers is inserted with a dropout value of 0.6 to help in overfitting prevention [22].

The proposed network was trained on 3 data sets: one electrode, four electrodes, and six electrodes heatmap images. Each data set was trained and tested using 5-fold cross-validation, and the mean accuracy is calculated and reported.

Around 15% of the training data (11% of the whole set) is reserved for each fold's validation set. Data is fed in batches of size 16, and the training was run for 50 epochs. A model check-point was set to save the model weights whenever a new higher validation accuracy is achieved at each epoch's end. This ensures we pick the model with the highest validation accuracy instead of the model with the highest training accuracy (typically towards the end of the training process). Adam optimizer [23] is adopted during the training process, and categorical cross-entropy is used as a loss function. We used python programming language along with Keras library for the model implementation.

C. Explainable prediction

The vast majority of literature work reports predictive accuracy as the primary measure to assess a machine learning model's performance [24]. Interpretability, also known as comprehensibility, is another critical factor that is sometimes overlooked. There is an inherent trade-off between accuracy and interpretability due to the usually complex nature of accurate models [25]. Most recently developed accurate models are based on deep learning networks that are usually treated as a black box and are relatively hard to interpret [26]. However, understanding the reasons behind a particular prediction is essential in providing insights into the model's working mechanism and developing trust towards the model in the decision-making process [27].

A few ML interpretability methods are developed in recent years such as SHAP explainer [28], Gradient-weighted Class Activation Mapping (Grad-CAM)[29], and Local Interpretable Model-agnostic Explanations (LIME). LIME is an explanation algorithm that can explain a particular prediction of any classifier by developing a local and interpretable approximation to the original model [30]. In this work, we decided to use LIME as it is well-developed by the community and has its own library written using python programming language [31]. LIME works as follow: suppose we have model f that is applied on an observation (trial) x to give a prediction $f(x)$ where $f(x)$ is to be explained. Let g be a local and an interpretable model that is approximate to f in the neighborhood of x; such that $g \in G$, where G is a group of potentially simple and explainable models (i.e., decision tree, linear model). A complexity measure of g is defined as $\Omega(g)$, and a proximity measure of observation x to another observation z is defined as $\pi_x(z)$ as to define a locality around x. Ultimately, we define $L(f, g, \pi_x)$, which is a measure of how unfaithful model g towards approximating model f in the locality of π_x. Thus, the explanation offered by LIME is obtained by:

$$\xi(x) = \underset{g \in G}{\operatorname{argmin}}(L(f, g, \pi_x) + \Omega(g)) \tag{1}$$

To find an explanation for a single trial $f(x)$, a fake data set is generated in the locality around the trial being explained, along with their corresponding predictions using the model f. A model g is to be fit to the fake data set such that $\Omega(x)$ is minimized. The weights of the model g, which is simple and interpretable, are used to generate an explanation of f in the locality of x.

In this work, we use LIME to produce local explanations on images. Particularly, LIME highlights the regions (called super-pixels) with positive weights towards a specific prediction (passive vs. active task). This provides us with intuition on why the model predicted a particular label for an input heatmap image by highlighting the positively related regions towards the predicted class. In other words, the model basically highlights the time period and frequency band at which the power spectral density is crucial in the classification process.

V. Results and Discussion

A. Kinesthetic Task Classification

Comparing the global average (average over trials and participants) heatmap images of the passive and active task at the the region of interest showed a clear difference in the neural activation before and after the collision point across the different frequency bands. However, examining single trials' heatmap images from the two categories was much more challenging to distinguish. In a real-time setting, the smaller the number of electrodes needed to distinguish between a passive or active kinesthetic action, the more affordable and preferable it is. Thus, the proposed 2-D CNN model was trained and tested on heatmap images generated from a single electrode (C1), four electrodes (C1, C3, FC1, and FC3), and six electrodes (C1, C3, C5, FC1, FC3, and FC5) all from the highlighted ROI. Fig. 5 shows that using data from a single electrode yields a mean accuracy of 84.56% over 5-folds while using four and six electrodes yields an accuracy of 93.96% and 95.89%, respectively.

Interestingly, the boxplot in Fig. 5 further shows that the variability of the accuracy across folds is reduced with increasing the number of electrodes. This can be attributed to the averaging process across electrodes, which helps eliminate noise and strengthen the signal-to-noise ratio (SNR), thus improving the model's robustness. Similarly, the mean precision and recall increase with the increase in the number of electrodes. Fig. 5-b shows the training loss over epochs. The number of epochs it takes the model to reach the corner point varies in a non-linear fashion with the number of electrodes used. The loss reduces slowly with one electrode and rapidly with four and six electrodes. Increasing the number of electrodes beyond four delayed the occurrence of the corner point. Together with the increase in the interquartile range in the precision and recall when moving from four to six electrodes, this could be an indication that a more robust model could be achieved using four electrodes. Additionally, four electrodes demand fewer computational resources and thus is preferable for real-time classification.

B. Explainable Classification

Using LIME method on a couple of the passive and active trials revealed an interesting pattern that is learnt by the classifier. In this study context, where the input of the model is an image, LIME explainer is expected to provide a visual

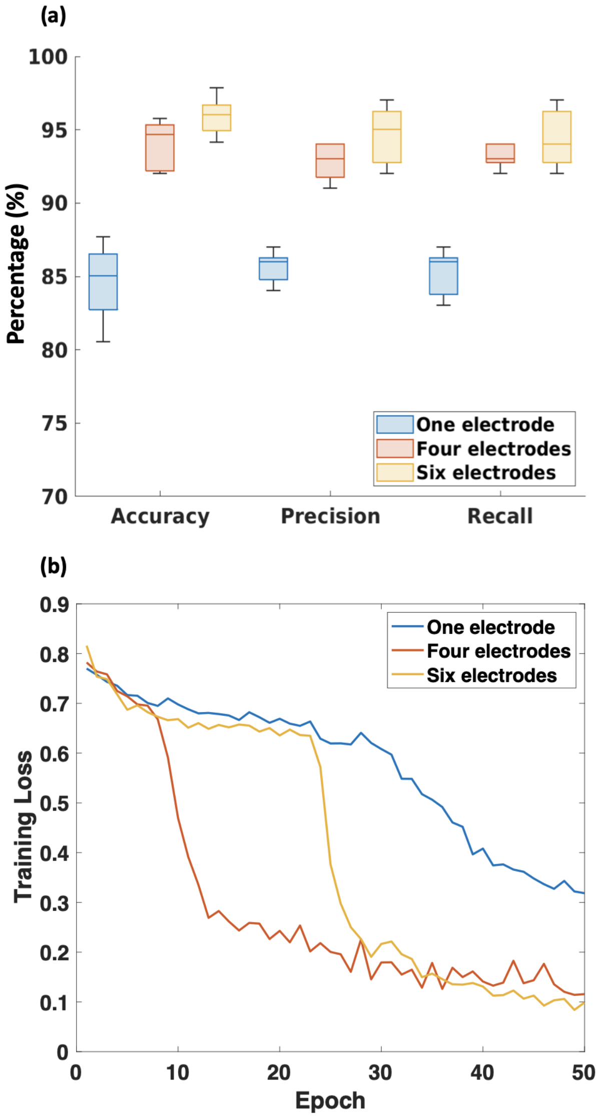

Fig. 5: (a) Box plot of the model metrics based on the utilized electrodes (b) Training loss over epochs based on the utilized electrodes

representation (e.g., patches of the input image) that provides a qualitative understanding of the relationship between that patch and the model's prediction. Fig. 6 shows two explained examples of correctly classified passive and active trials. The model classifies the passive trial as passive with a probability of 85.79% and the active trial as active with a probability of 99.97%. LIME highlights the parts of the image that were most influential in the model's decision. It can be observed in Fig. 6-b that a strong desynchronization concentrated in the alpha and beta bands (bands 3 and 4) was the most influential feature in classifying the trial as active. On the other hand, a synchronization in the delta and theta bands (bands 2 and 3) in the passive trial was the most influential feature in classifying the trial as passive, as shown in Fig. 6-a.

From a neuroscience point of view, EEG research has shown that oscillations in the theta, alpha, and beta bands are all involved in various processes related to self-motion perception and motor functions [32]. For example, it has been repeatedly reported that there are at least two types of *mu* rhythms in the alpha frequency band exhibited before and during movement [33]. The first is lower-frequency (8–10 Hz) and widespread across the cortex, which is believed to be movement-type unspecific that serves general motor attention purposes. The other is higher-frequency and movement-type specific localized in the motor-somatosensory cortex. When both alpha and beta desynchronization are coupled, they are typically associated with multisensory body movements or in coordinating visual processing and physical motion together. Both of the described neural phenomena can be observed in Fig. 6-b. On the other hand, a kinesthetic-based passive task exhibits a different neural activation. At the collision point and while the subject is passively holding the racket, an apparent synchronization is observed post the force feedback, mainly towards the lower frequency bands (theta band). Theta synchronization in the proximity of the mid-frontal cortex is documented to play an essential role in multisensory divided attention [34]. In the passive kinesthetic task, subjects are attentive to the visual stimuli displayed on the screen as well as to the force feedback delivered upon the ball collision. Despite the same is true for the active kinesthetic task, *mu* rhythm is predominant in the active task due to the motor movement, which could impact the intensity of theta stimulation upon the force feedback delivery. It can be concluded from this discussion that the model in use could be potentially reliable, as the highlighted regions in the heatmap images deem important in recognizing passive and active kinesthetic tasks.

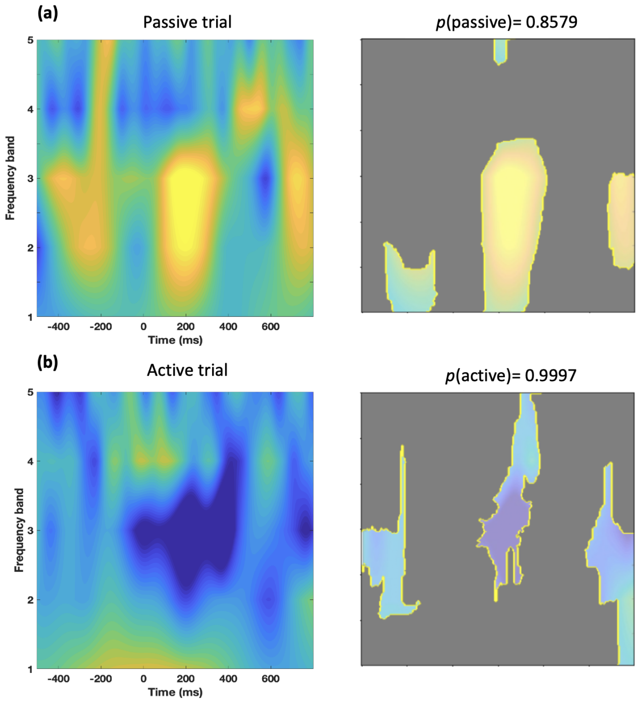

Fig. 6: An example of a local explanation for a passive and an active trials using LIME

VI. Conclusion

This paper demonstrated the use of a 2-D CNN model to classify passive and active kinesthetic interactions using single-trial EEG data. The model performance was compared when using one, four, or six electrodes associated with the motor and somatosensory cortices. The model achieved a mean accuracy of 84.56%, 93.96%, and 95.89% across 5-fold validation when using one, four, or six electrodes, respectively. Although the accuracy from six electrodes data is the highest, the model's performance with four electrodes showed less variability in precision and recall measures and a faster training convergence. We further used an explainable machine learning method, LIME, to assess the classifier's mechanisms in producing its prediction. LIME showed that the model considers important neural features of the input image supported by the neuroscience literature. As for future work, it is essential to compile a more extensive dataset that includes passive and active kinesthetic interactions that vary in task nature, which should help the model learn the kinesthetic interaction's general neural markers regardless of the nature of the task. Furthermore, we plan to develop a neural-based haptic guidance method for rehabilitation therapy. Finally, this study can be considered a step towards building a haptic model capable of objectively describing the passive and active kinesthetic interactions from a neural perspective.

VII. ACKNOWLEDGEMENT

This research is funded by the New York University Abu Dhabi PhD Fellowship Program.

References

[1] J.-L. Rodríguez, R. Velázquez, C. Del-Valle-Soto, S. Gutiérrez, J. Varona, and J. Enríquez-Zarate, "Active and passive haptic perception of shape: Passive haptics can support navigation," *Electronics*, vol. 8, no. 3, p. 355, 2019.

[2] H. Culbertson, S. B. Schorr, and A. M. Okamura, "Haptics: The present and future of artificial touch sensation," *Annual Review of Control, Robotics, and Autonomous Systems*, vol. 1, pp. 385–409, 2018.

[3] W. Barfield, "The use of haptic display technology in education," *Themes in science and technology education*, vol. 2, no. 1-2, pp. 11–30, 2010.

[4] H. Alsuradi, W. Park, and M. Eid, "Eeg-based neurohaptics research: A literature review," *IEEE Access*, vol. 8, pp. 49313–49328, 2020.

[5] A. Lau-Zhu, M. P. Lau, and G. McLoughlin, "Mobile eeg in research on neurodevelopmental disorders: Opportunities and challenges," *Developmental cognitive neuroscience*, vol. 36, p. 100635, 2019.

[6] H. Miura, J. Kimura, N. Matsuda, M. Soga, and H. Taki, "Classification of haptic tasks based on electroencephalogram frequency analysis," *Procedia Computer Science*, vol. 35, pp. 1270–1277, 2014.

[7] H. Alsuradi, C. Pawar, W. Park, and M. Eid, "Detection of tactile feedback on touch-screen devices using eeg data," in *2020 IEEE Haptics Symposium (HAPTICS)*, pp. 775–780, IEEE, 2020.

[8] N. Kang and J. H. Cauraugh, "Force control in chronic stroke," *Neuroscience & Biobehavioral Reviews*, vol. 52, pp. 38–48, 2015.

[9] L. Seminara, P. Gastaldo, S. J. Watt, K. F. Valyear, F. Zuher, and F. Mastrogiovanni, "Active haptic perception in robots: a review," *Frontiers in neurorobotics*, vol. 13, p. 53, 2019.

[10] M. Fleury, G. Lioi, C. Barillot, and A. Lécuyer, "A survey on the use of haptic feedback for brain-computer interfaces and neurofeedback," *Frontiers in Neuroscience*, vol. 14, 2020.

[11] S. J. Lederman, "The perception of surface roughness by active and passive touch," *Bulletin of the Psychonomic Society*, vol. 18, no. 5, pp. 253–255, 1981.

[12] C. Genna, F. Artoni, C. Fanciullacci, C. Chisari, C. M. Oddo, and S. Micera, "Long-latency components of somatosensory evoked potentials during passive tactile perception of gratings," in *2016 38th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, pp. 1648–1651, IEEE, 2016.

[13] S. Eldeeb, J. Ting, D. Erdogmus, D. Weber, and M. Akcakaya, "Eeg-based texture classification during active touch," in *2019 IEEE 29th International Workshop on Machine Learning for Signal Processing (MLSP)*, pp. 1–6, IEEE, 2019.

[14] A. Moungou, E. Vezzoli, C. Lombart, B. Lemaire-Semail, J.-L. Thonnard, and A. Mouraux, "A novel method using eeg to characterize the cortical processes involved in active and passive touch," in *2016 IEEE Haptics Symposium (HAPTICS)*, pp. 205–210, IEEE, 2016.

[15] H. Singh, M. Bauer, W. Chowanski, Y. Sui, D. Atkinson, S. Baurley, M. Fry, J. Evans, and N. Bianchi-Berthouze, "The brain's response to pleasant touch: an eeg investigation of tactile caressing," *Frontiers in human neuroscience*, vol. 8, p. 893, 2014.

[16] S. Tarng, D. Wang, Y. Hu, and F. Merienne, "Towards eeg-based haptic interaction within virtual environments," in *2019 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*, pp. 1179–1180, IEEE, 2019.

[17] A. Craik, A. Kilicarslan, and J. L. Contreras-Vidal, "Classification and transfer learning of eeg during a kinesthetic motor imagery task using deep convolutional neural networks," in *2019 41st Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, pp. 3046–3049, IEEE, 2019.

[18] C. A. E. Kothe and T.-P. Jung, "Artifact removal techniques with signal reconstruction," Apr. 28 2016. US Patent App. 14/895,440.

[19] A. Grossmann and J. Morlet, "Decomposition of hardy functions into square integrable wavelets of constant shape," *SIAM journal on mathematical analysis*, vol. 15, no. 4, pp. 723–736, 1984.

[20] S. Ioffe and C. Szegedy, "Batch normalization: accelerating deep network training by reducing internal covariate shift. arxiv e-prints," 2015.

[21] M. Lin, Q. Chen, and S. Yan, "Network in network," *arXiv preprint arXiv:1312.4400*, 2013.

[22] N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever, and R. Salakhutdinov, "Dropout: a simple way to prevent neural networks from overfitting," *The journal of machine learning research*, vol. 15, no. 1, pp. 1929–1958, 2014.

[23] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," *arXiv preprint arXiv:1412.6980*, 2014.

[24] A. K. Gopalakrishna, T. Ozcelebi, A. Liotta, and J. J. Lukkien, "Relevance as a metric for evaluating machine learning algorithms," in *International Workshop on Machine Learning and Data Mining in Pattern Recognition*, pp. 195–208, Springer, 2013.

[25] A. A. Freitas, "A critical review of multi-objective optimization in data mining: a position paper," *ACM SIGKDD Explorations Newsletter*, vol. 6, no. 2, pp. 77–86, 2004.

[26] G. Alain and Y. Bengio, "Understanding intermediate layers using linear classifier probes," *arXiv preprint arXiv:1610.01644*, 2016.

[27] D. V. Carvalho, E. M. Pereira, and J. S. Cardoso, "Machine learning interpretability: A survey on methods and metrics," *Electronics*, vol. 8, no. 8, p. 832, 2019.

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

[29] R. R. Selvaraju, M. Cogswell, A. Das, R. Vedantam, D. Parikh, and D. Batra, "Grad-cam: Visual explanations from deep networks via gradient-based localization," in *Proceedings of the IEEE international conference on computer vision*, pp. 618–626, 2017.

[30] M. T. Ribeiro, S. Singh, and C. Guestrin, "" why should i trust you?" explaining the predictions of any classifier," in *Proceedings of the 22nd ACM SIGKDD international conference on knowledge discovery and data mining*, pp. 1135–1144, 2016.

[31] M. T. C. Ribeiro, "Lime: Local interpretable model-agnostic explanations."

[32] B. Townsend, J. K. Legere, S. O'Malley, M. v. Mohrenschildt, and J. M. Shedden, "Attention modulates event-related spectral power in multisensory self-motion perception," *NeuroImage*, vol. 191, pp. 68–80, 2019.

[33] G. Pfurtscheller, "Induced oscillations in the alpha band: functional meaning," *Epilepsia*, vol. 44, pp. 2–8, 2003.

[34] A. S. Keller, L. Payne, and R. Sekuler, "Characterizing the roles of alpha and theta oscillations in multisensory attention," *Neuropsychologia*, vol. 99, pp. 48–63, 2017.