Decoding Movements from EEG with a Convolutional Recurrent Network

Authors: Tai Nguyen, Liangyu Tao, Cherelle Connor

1. Introduction

Electroencephalography (EEG) is a prevalent method for the detection of electrical activity within the brain. It provides researchers with a quick, cost-effective, and non-invasive method of brain mapping and neuroimaging. EEG is often used for identification and analysis of brain functionality. Interpretation of the signals can provide valuable insight into the areas of the cortex responsible for processing specific information at a given time.

Due to its accessibility and noninvasive nature, EEG is often a popular choice when developing brain-computer interfaces. EEG based brain-computer interfaces (BCIs) provide a feasible method for the analysis of neural states using offline or real-time online electrical potentials [1]. Furthermore, BCIs can be used to control assistive or rehabilitative technologies like neuro-prosthetics, potentially changing the lives of individuals with spinal cord injuries or missing limbs. In addition, these interfaces can refine user control in health and entertainment devices [2]. These interfaces capitalize on the EEG signals elicited during motor execution or movement intention, by using sensorimotor rhythms collected from the motor strip of the brain, typically occurring in the central, frontal, and parietal areas of the brain neocortex [3, 4]. These rhythms are most commonly associated with hand and arm movements and occur in both the mu (8-13 Hz) and beta (13-24 Hz) frequency bands [5]. Accurate classification of the rhythms within these bands is key to determine the user's intended movements. Therefore, BCIs developed for motor movement typically used these frequency bands as features for classification [6].

Motor imagery BCIs operate by having the user imagine or conduct a simple sensory-motor task on the right and left side of their body [6]. However, MI is notoriously difficult to use for several reasons. Most commonly, noisy signals can make the data difficult to interpret; the repeatability of the trials can be tiresome or monotonous to participants, which can lead to a decrease in performance the potential for variation between different subjects is high. Due to the difficulty in MI BCIs, it is often beneficial to develop classifiers using motor execution (ME) before implementing the network architecture in the context of MI.

There are two major complications that occur when dealing with BCI development: EEG denoising and trial rejection. EEG denoising is an integral part of the assessment and analysis of brain functionality. Properly denoising EEG signals can often be difficult as the distance between the electrodes and neurons creates a reduction in amplitude that can distort the signal, increasing its signal-to-noise ratio (SNR) [7, 8]. The SNR describes the difference between the targeted signal and the noise

floor [9]. The signals that comprise the noise can originate both externally and internally of the participant, in the form of EEG artifacts. Artifacts can be caused by improperly connected electrodes, eye blinks, and muscle activities, and also by involuntary movements such as breathing and heartbeats. Although difficult to determine, these artifacts can be rejected using preliminary statistical analysis like bandpass filters or independent component analysis (ICA) [10]. Bandpass filters can be used to attenuate frequencies outside the targeted range. Independent component analysis is a method frequently used for the separation of artifacts from EEG signals by separating sources that are identified as linearly independent of one another [11]. It is important to correctly reject these artifacts as they can contaminate the signal, resulting in the occlusion of weaker signals or the loss of important spatial information, leading to overall inaccurate statistical analysis [12].

Once the signal is denoised there are several ways to classify it. Methods such as linear discriminant analysis and support vector machines have been traditionally held as the standard for BCI evaluation and classification [13]. However, it is possible that a lot of information can be lost without the use of more specialized extraction and classification techniques. This can be especially detrimental to the acquisition of accurate classification of the signal for brain-computer interfaces. While these linear methods do not provide a high level of accuracy required by BCI, they are quite valuable in gaining insight into EEG signals and the neural source space during BCI tasks [14].

With larger datasets and more powerful computer hardware becoming widely available, the use of machine learning for the analysis and understanding of EEG has increased [15]. Using deep learning, advancements have been made in several brain-computer interface fields like object recognition, auto spelling, motivation decoding amongst others and there is growing evidence that neural networks provide an efficient and accurate method for classification of EEG patterns. Convolutional neural networks (CNN) and recurrent neural networks (RNNs) are quickly rising in popularity in EEG decoding architectures [15, 16]. Recently, it has been shown that a combination of CNN and RNNs together achieves much higher performance than using just using CNNs or RNNs. The gold standard of this type of hybrid model has shown that a recurrent-convolutional network trained on spectral topography mappings of EEG time series data on a simple working memory task performed much better than many common classifiers including support vector machines, deep belief networks, and logistic regression [17]. However, a later study of movement classification found that the network achieved much lower accuracy when used in the context of movement classification [18]. This may be because the complex spectral feature extraction steps include data compression over a large sampling period while movement intention tasks occur on shorter timescales. The authors instead proposed both a cascade and a parallel convolutional recurrent network architecture that is trained on the progression of spatial patterns of raw EEG activity over short timescale (26 ms). The recurrent part of the architecture was realized using long-short term memory networks (LSTMs). Besides the recurrent feedback loop, LSTM stands out for a special memory unit, enabling models to study long-term temporal dependency and keep it from vanishing or exploding. These features work well in time-domain and frequency-domain EEG data classification [19]. Both the cascade and parallel networks were

able to achieve 98% accuracy for the standard PhysioNet EEG Dataset [20] involving a set of four actions involving anatomically disparate regions. However, it is unknown if this network can achieve similar levels of accuracy for a harder dataset of movements performed by the same anatomical appendage. To tackle this, we present in this study the application of the cascade convolutional-recurrent network on a more difficult dataset involving similar movements using a single-arm [14].

In addition to model performance, it is unknown if there is any physiological insight that can be garnered by investigating the features that the network looks at to generate classifications. One method to investigate and visualize the features that are most important for model classification is to use saliency maps [21]. Saliency maps are typically used in computer vision to identify the most unique pixel qualities in an image, and they can provide an invaluable tool for revealing meaningful features in the form of EEG spatial distribution [22]. Traditional BCI studies indicate EEG channels in the central and frontal areas are largely mapped to regions such as the primary motor cortex (M1), sensory cortex (S1), premotor cortex (PM), and the posterior parietal cortex (PPC). PM and PPC are known to be involved with planning[23, 24] while M1 is involved with movement execution [25] and S1 receives proprioceptive feedback which is integrated at the PPC [26]. To test if the spatial properties that the convolutional-recurrent network utilizes are consistent with these neuroanatomical and functional studies, we present saliency maps for the 2D data meshes in EEG space where the model gives the highest confidence in each given action and show preliminary efforts into understanding the electrophysiological features that allow the cascade network to obtain high classification accuracy.

2. Methods – Data Processing

In this section, we will outline the dataset, preprocessing steps, and architectures of our neural network. Our methods were based on a combination of the methods described by the authors of the paper in which our data originated [14] and by the paper which our cascade network model is based on [18].

2.1 Data Behavioral paradigm

This study utilizes the dataset set created by Patrick Ofner and his colleagues for the assessment of time-encoded low frequencies in EEG signals [14]. The data were collected from fifteen healthy participants from 61 channels spanning the central, frontal, parietal, and temporal areas of the brain. The data was collected in two separate sessions conducted over the course of one week. The sessions were split into motor execution (ME) and motor imagery (MI). In each session, the participants were seated in a chair with their arm placed in an exoskeleton, with a DT Data Glove (5DT, USA), in a neutral position. Using a combination of audio and visual cues the participants were instructed to execute or imagine a movement class. There were six-movement classes in total and one rest class. The movement classes consisted of elbow flexion/extension, forearm pronation/supination, and hand open/close. Each participant completed 10 runs with 42 trials per run (6 trials per class), which corresponds to 60 trials per movement class per session.

The data was stored in general data format files for each subject, session, and run. The channels labels of the data were

representative of the EEG electrode positions (1-61), the EOG positions (62-64), the data glove sensors (65-83), and the exoskeleton sensors (84-96). The entire dataset can be downloaded at (http://bnci-horizon-2020.eu/database/data-sets).

2.2 Data consolidation and compression

To assess the large (25GB) EEG dataset, the files were consolidated using Python's pickle module. Python's pickle module deconstructs Python objects into databases that contain the necessary information to reconstruct the object into another python file, allowing for the compression of the large EEG datasets into a more manageable database (1.3 GB). With the newly extracted EEG data, stored in the pickle file preprocessing was carried out on the dataset.

2.3 Preprocessing

EEG channel data was first down-sampled from 512 Hz to 128 Hz prior to consolidation into a pickle database. Each trial in the database was then subjected to first-order (linear) baseline subtraction. Then the trials with abnormal joint probability and kurtosis in the EOG channels were rejected. Next, the trials were aligned to movement onset. Finally, the 1D EEG channel data at each time point for each trial was converted to a 2D spatial mesh. The entire process is shown in Figure 1 and individual steps are described in detail below. All preprocessing steps were performed in custom-written Python programs.

2.4 Trial rejection

To detect trials with abnormal EOG channels, we marked artifact trials as those with abnormal EOG activity through three criteria: 1.) EOG channels contain values greater than 200 uV above the mean, 2.) trials that are above 5x the standard deviation of the trial and population joint probability, and 3.) trials that are above 5x the standard deviation of the kurtosis.

2.5 Alignment by movement onset

Movement onset was detected based on kinematic channels located on the exoskeleton and glove. For elbow flexion/extension, we used channels (87 to 89) which corresponded to elbow positions in the x,y, and z space. For forearm pronation/supination, we used channel 94 which was provided as an all-inclusive forearm movement data. For the opening/closing of the hand, we used channels 65 to 79, which correspond to sensors located on the hands. For each kinematic channel, we performed baseline subtraction based on the average voltage of the first half-second of each trial. For elbow flexion/extension, we then summed the elbow channels into a mean channel. For hand opening/closing, we performed PCA on the 15 hand channels and extracted the primary principal component. We obtained the classifier data vector (v1) by calculating the absolute value of the 1D linear embedding of channels and performing a 125 ms 3rd order Savitzky–Golay filter.

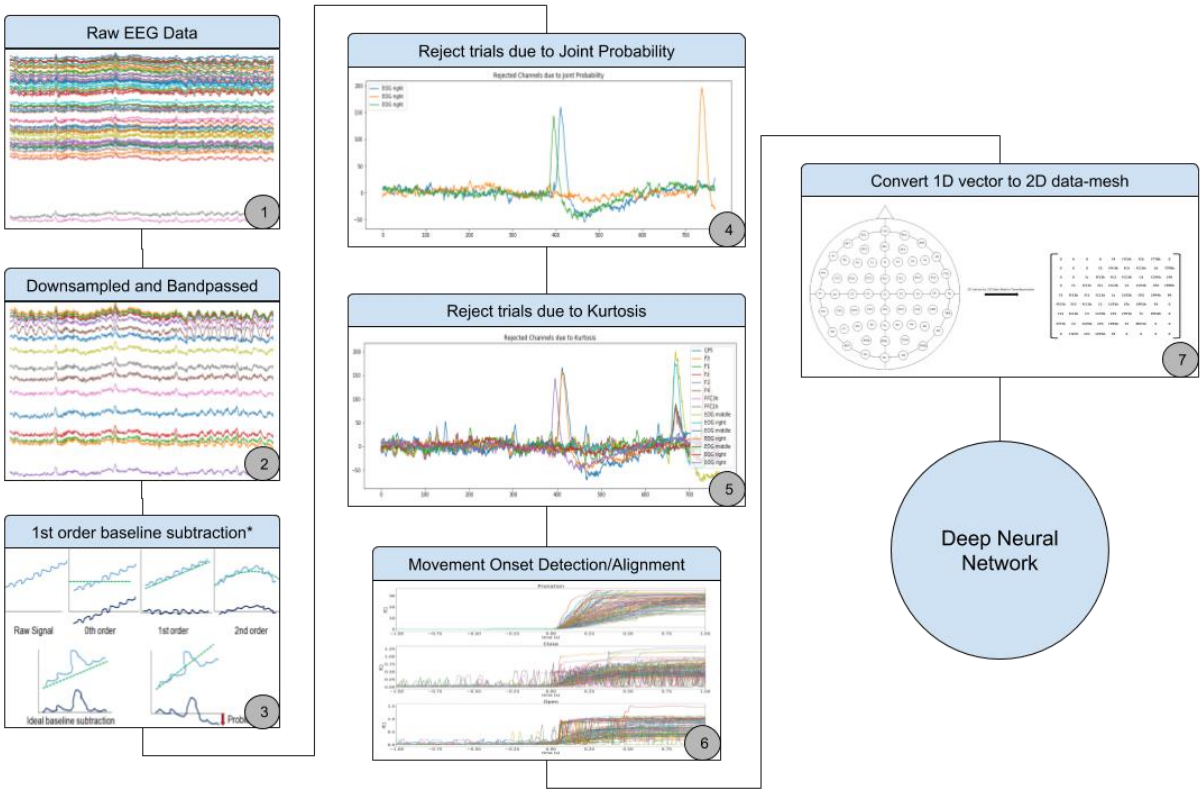

Figure 1: Overview of data processing pipeline

Movement onset was calculated as the last instance before v1 reached 1/3 of the maximum value where v1 and the derivative of v1 are below respective thresholds. This was done to counter noisy data or tremors before movement onset. Thresholds were determined empirically for different movement types. The EEG data were then aligned by movement onset and trimmed to only contain data from 1 second before onset to 1 second after onset.

2.6 Converting from 1D EEG vector to 2D mesh

The 1D EEG data vectors were converted to a 2D EEG data mesh using the spatial information of the electrode distribution of the acquisition system. This is done by first initializing an empty 9x9 mesh map. That map is filled with the row and column positions of each electrode. The mesh is then populated with the electrode coordinates, creating the 2D signal map that preserves relative spatial positioning.

3. Methods - Neural Network

3.1 Cascade model architecture

The deep neural network models we attempted to evaluate came from Zhang et. al we only conducted an extensive examination of the cascade model.

This model features a combination of

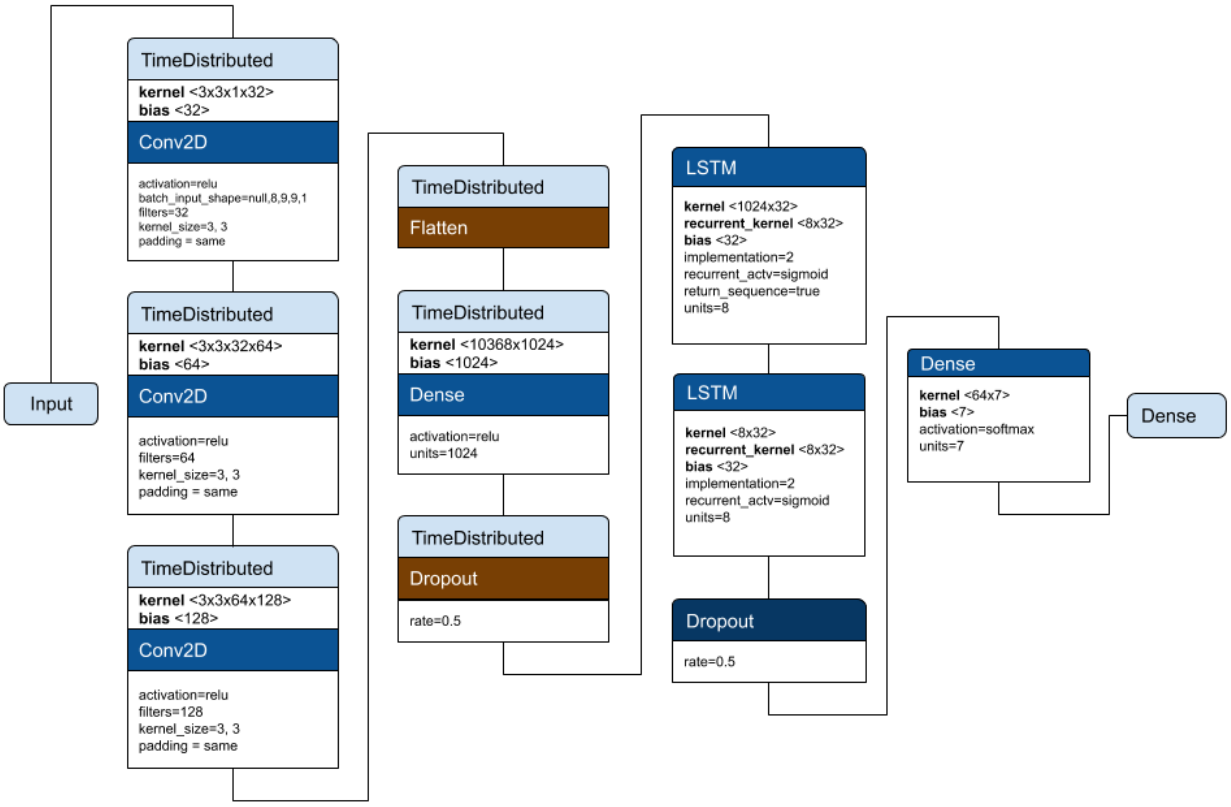

Figure 2: Network architecture of the cascading convolutional recurrent neural network

2018 paper: "Cascade and Parallel Convolutional Recurrent Neural Networks on EEG-based Intention Recognition for Brain-Computer Interface". We intended on gauging the performance of both the cascade and the parallel model demonstrated in Zhang's work because the author claimed that his models can work on raw EEG data. However, due to the paper's code being outdated, and various technical difficulties during the process of converting the Ofner's dataset to a version which is compatible with the Zhang's models, Convolutional Neural Networks (CNNs) and a Long-Short Term Memory (LSTM) Network. The word "cascade" means that the outputs of the CNNs are fed into the LSTM network, whose output is used for classification. Each CNN has a topology described by 3 x 2D layers with the same kernel size of 3 by 3 (with zero-padding to make sure the feature maps have the same size as the input data images) but different number of feature maps: 32, 64, and 128 (Equation 2). The output of the 3 convolutional layers is fed into a fully-

connected layer of size 1 by 1024. The number of CNNs in the architecture is determined by the number of 2D data mesh (called frames) in one segment. Since the model we built has 8 frames per segment, each segment has 4 overlapping frames, the CNNs' outputs are 8 sequences of 1 by 1024 (Equation 3). This entire output will describe the spatial feature for the entire segment. On the other hand, the LSTM we built network has 2 layers, each layer with 8 (the number of frames in a segment) unit. The output time sequence of each unit in the first layer is fed into the corresponding unit in the second layer as an external input. The output of the LSTM network will be the output of the last unit in the second layer (with respect to the frame in the segment) (Equation 4). This result represents both the spatial and the temporal features of the entire segment. To do classification from this result, we decrease its dimensionality by first feeding it into one fully-connected layer that has 64 neurons (Equation 5). Finally, the output from this layer is connected to another fully-connected layer with 7 neurons (7 classes) with a soft-max activation function for categorical classification (Equation 6 & 7). The final model was written using Keras.

$$Seg_j = (frame_i, frame_{i+1}, \dots, frame_{i+7}) \quad \text{eq.1}$$

$$f_j = CNN_j(Seg_j), f_j \in \mathbb{R}^{9x9x128} \quad \text{eq.2}$$

$$F_j = Dense\left(Flat(f_j)\right), F_j \in \mathbb{R}^{1024} \quad \text{eq.3}$$

$$H_{lstm} = LSTM(F), H_{lstm} \in \mathbb{R}^{8} \quad \text{eq.4}$$

$$H' = Dense(H_{lstm}), H' \in \mathbb{R}^{64} \quad \text{eq.5}$$

$$H = Dense(H'), H \in \mathbb{R}^{7} \quad \text{eq.6}$$

$$Cascade = Softmax(H) \quad \text{eq.7}$$

3.2 Training

To get the data ready for training, we first augmented the dataset such that: for every trial in each class, the number of frames N is expanded into M, where M is the final number of frames after applying a sliding window of size = 8 with overlap = 4. Then, the frames for all trials in a class are accumulated into a big array of size P inputs. Each input having a size of 8x9x9x1 (window x size of 2D EEG x 1).

The training was carried out by optimizing the categorical cross-entropy loss function. We trained the network using the Adam optimizer [27] with a learning rate of 10e-4 and a batch size of 64. The model was trained on 64% of the dataset, validated on 16% of the dataset, and tested on the remaining 20%. The large number of parameters in the model made it susceptible to overfitting. Therefore, we added a dropout rate of 0.5 at the end of the convolutional portion of the model and at the output layer of the LSTM. In addition, we applied early stopping my monitoring the model's performance over the validation set and saved the model that achieved the highest validation accuracy. Training was performed on a Google Colabs GPU and takes on average 160s per epoch. We found that the model converged in 50 epochs (Figure 3A).

4. Results

4.1 Single arm movements can be well classified using the cascade CNN-RNN framework

To determine if movements performed by a single arm of individuals can be decoded by EEG activity, we trained a neural network classifier on all six movement and one rest class for the ME dataset. We obtained with 85% validation accuracy, 99% training accuracy and 85% testing accuracies with little deviation (data not shown) in accuracy across

multiple trained models (n=3). The training loss and accuracy of the model used for the rest of the study is shown in Figure 3 A and B.

To investigate whether movement is more likely to be misclassified to more similar movements (i.e. hand open to hand close), as opposed to more disparate movements (i.e. hand open to forearm pronation), we computed a confusion matrix based on the testing dataset (Figure 3C). Surprisingly, we found that movement similarity did not cause higher misclassifications. Testing set the aggregation of movement classes against rest for movement-vs-rest.

Neural activity near movement onset is traditionally known to yield higher classification accuracy than activity from further in the past and future. This because the changed in motor-related cortical potentials (MRCPs) are often largest during the period closes to movement onset. It was found in the Olfner paper that the highest classification accuracies occur between 0.13 s to 0.25s after movement onset for movement-vs-movement

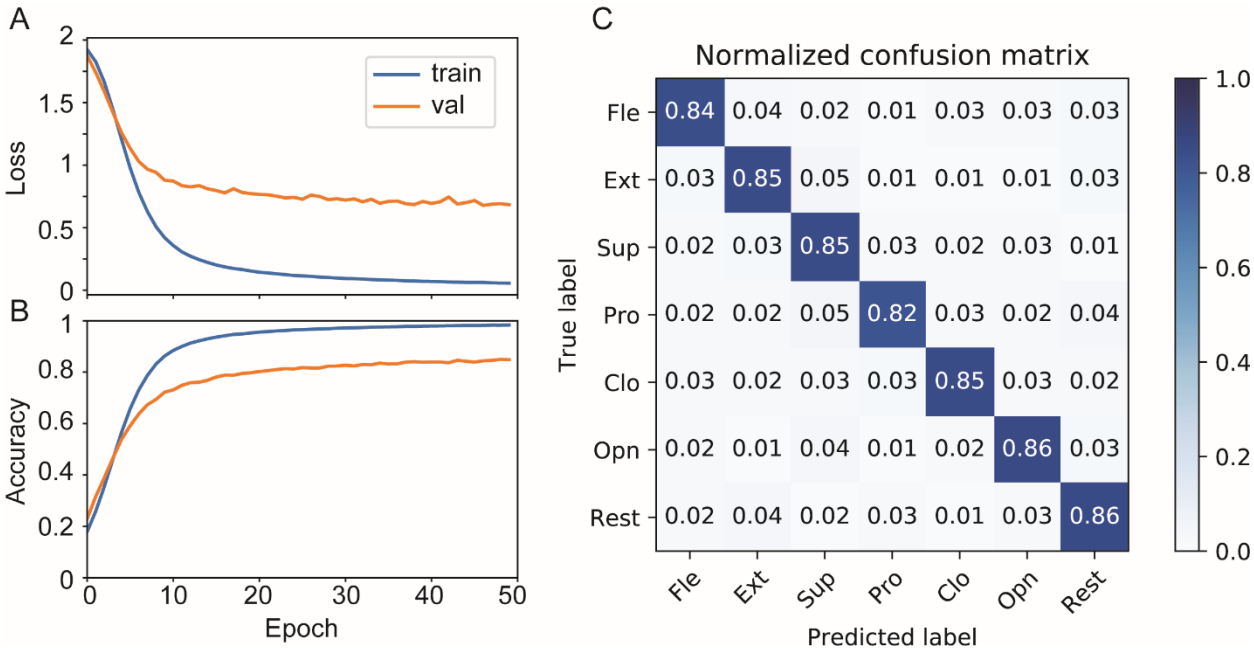

Figure 3: Model's performance. A) Training and validation loss. B) Training and validation accuracy C) Confusion matrix based on testing accuracy

accuracies for each class are consistently high at 85% except for pronation, which is lower at 82%. This is much higher than the peak of 55% accuracy for movement-vs-movement classifications and comparable to the peak of 87% movement-vs-rest accuracy obtained using shrinkage linear discriminant analysis (sLDA) classifiers as reported in the 2017 Olfner paper [14]. It is important to note that the baseline model as described in the Olfner paper is based on individual pairwise classifiers for movement-vs-movement and (Figure 4A) and between -0.13s before and 0.19s following movement onset for movement-vs-rest classification (Figure 4B) [14]. The broader range for movement-vs-rest may be due to the absence of a movement indicator to align rest classes. To determine if the cascade model follow such limitations, we calculated the mean accuracy of each time bin for our entire dataset. Surprisingly, we found that the accuracy was consistently high even at 1 second away from movement onset (Figure 4C). This suggests that the model was

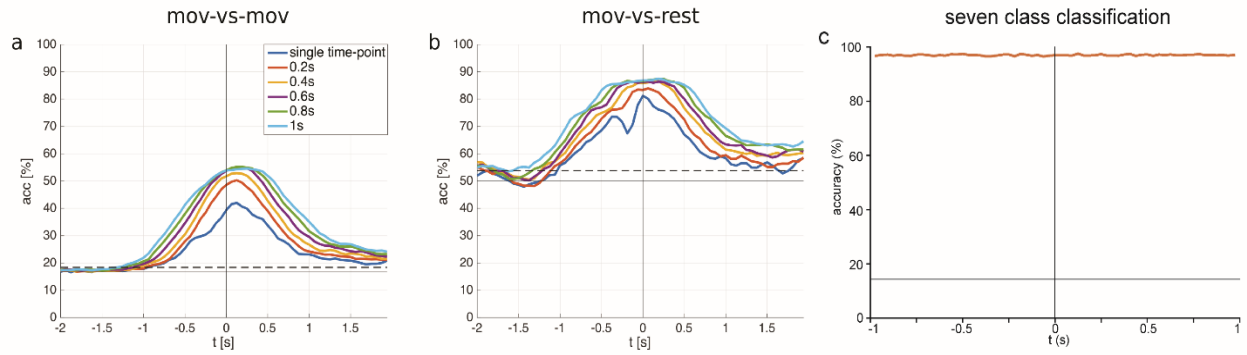

Figure 4: A) and B) Ofner's model's performance with respect to different time window. C) All 7 class classification accuracy for the cascading CNN+LSTM model

able to take advantage of smaller, more nuanced changes in cortical potential that is otherwise noninformative in traditional classifiers. Taken together, the capability to obtain high accuracies for similar movement classes performed on a single-arm across multiple individuals and the ability to predict movements far in the past and future of movement onset shows that the cascade model is a good model for using in movement decoding.

4.2 Visualizing the impact of individual EEG channels on model classification

Since the cascade model significantly increases classification accuracy over the baseline sLDA method, it is important to understand what physiological inputs the model uses to make its decision in classifying each class. Traditionally, this idea is realized with great success in convolutional neural networks using saliency maps (citations 1-5). The general formalization in our application is to calculate the gradient of each output class with respect to an input segment of EEG mesh data (S_j). For each class, we used Keras-vis to compute saliency maps for the input segment within the top 10 segments in the testing set that gives the highest confidence in the given class with the highest total absolute activity (Figure 5). Interestingly, the highest saliency generally

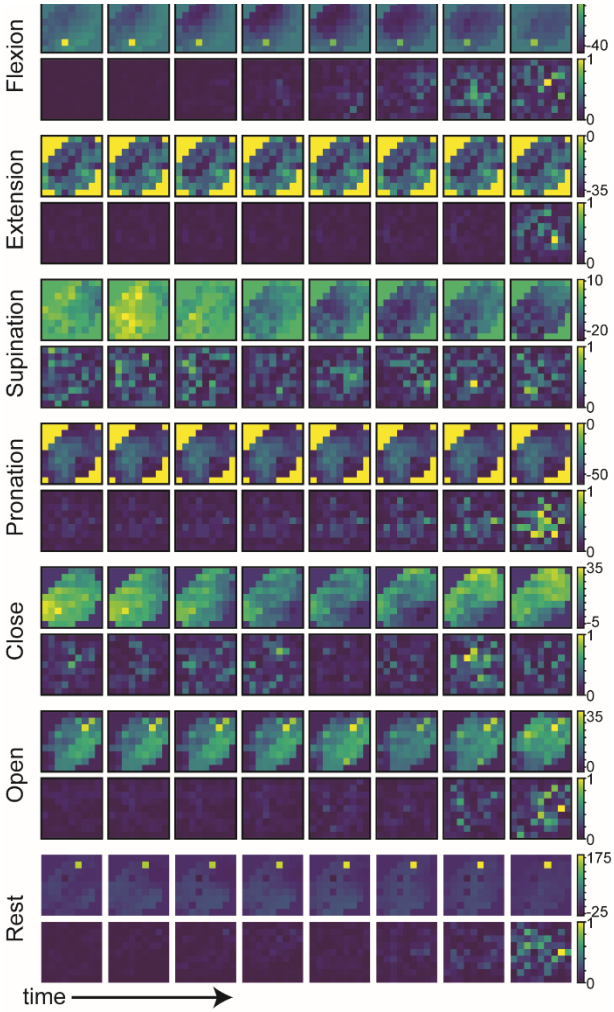

Figure 5: Saliency map for each class. In each class, the top row of images shows the input bin of 8 consecutive time steps. The bottom row shows the corresponding saliency map

occurs within the last two time points across all classes. Furthermore, the channels with the highest activities in the original EEG mesh

space are often not the channels with maximum saliency. This is consistent with the previous observations that the model is looking at smaller, more nuanced changes in EEG activity than in large EEG activity.

Next, we investigated the channels for each class that has saliency values greater than 0.85 (Table 1). We found that most of the significant electrodes are located in the central, frontal-central, and central-parietal regions of the scalp. While some channels such as C2, C3, and FCz have been investigated in the past in regards to BCI applications, many of the other classes are not commonly studied. Investigating the MRCP of the regions, we find that the signals are quite similar in each region. This is likely due to the high correlation between EEG channels and the close spatial proximity of these channels (Figure 6).

Class	Channels
Flexions	C2
Extension	CPP2h
Supination	C3, CCP1h
Pronation	CCP2h, CCP3h, CP1, FCC2h, Pz
Hand close	FC2, FCz
Hand open	CCP1h, CPP4h
Rest	CCP4h

Table 1: Channels with saliency values above 0.85 for each movement class

5. Discussions and improvements

While we have shown the cascading convolutional recurrent neural network was able to give high classification accuracies for a relatively difficult motor execution dataset, there are still multiple avenues of directions that were not investigated due to time constraints. In terms of data, it is unknown if the model will achieve a similarly high level of performance for a motor imagery dataset of the same movement types. In the Olfner paper, the movement-vs-movement classification accuracy dropped from 55% to 27% for the motor imagery dataset. Meanwhile, the movement-vs-rest accuracy dropped from 87% to 73% for the motor imagery dataset [14]. Furthermore, we implemented multiple data preprocessing steps in order to ensure the highest likelihood of modeling success. It would be interesting to systematically remove each type of data preprocessing to investigate whether specific data preprocessing steps are necessary for high model performance. Yet another avenue of investigation is in model architecture and hyper-parameterization. It would be interesting to compare the results of this cascade model to a parallel architecture where the convolutional and recurrent networks run in parallel and their outputs are concatenated before being fed into a SoftMax layer. Finally, more work should be done on visualizing the learned representations of the model. For instance, we can investigate the feature maps of each convolutional layer.

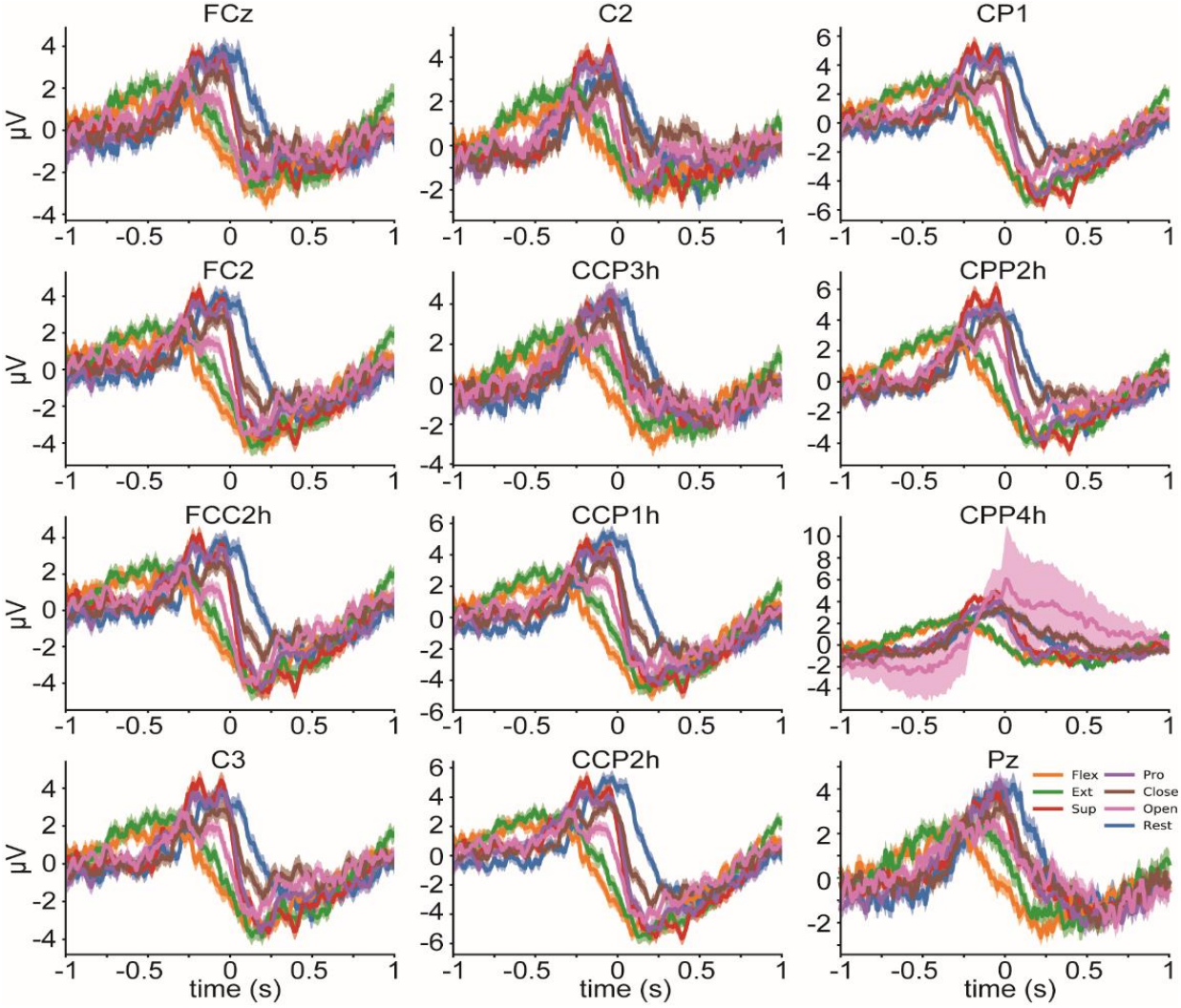

Figure 6: Motor related cortical potentials for each channel with high saliency for all classes

6. References

[1] F. Lotte, "A Tutorial on EEG Signal Processing Techniques for Mental State Recognition in Brain-Computer Interfaces," 10/04 2014, doi: 10.1007/978-1-4471-6584-2_7.

[2] N. Padfield, J. Zabalza, H. Zhao, V. Masero, and J. Ren, "EEG-Based Brain-Computer Interfaces Using Motor-Imagery: Techniques and Challenges," (in eng), *Sensors (Basel),* vol. 19, no. 6, p. 1423, 2019, doi: 10.3390/s19061423.

[3] B. K. Kang, J. S. Kim, S. Ryun, and C. K. Chung, "Prediction of movement intention using connectivity within motor-related network: An electrocorticography study," *PLOS ONE,* vol. 13, no. 1, p. e0191480, 2018, doi: 10.1371/journal.pone.0191480.

[4] I. Toni, D. Thoenissen, and K. Zilles, "Movement Preparation and Motor Intention," *NeuroImage,* vol. 14, no. 1, pp. S110-S117, 2001/07/01/ 2001, doi: https://doi.org/10.1006/nimg.2001.0841.

[5] M. Ahn and S. C. Jun, "Performance variation in motor imagery brain-computer interface: a brief review," (in eng), *J Neurosci Methods,* vol. 243, pp. 103-10, Mar 30 2015, doi: 10.1016/j.jneumeth.2015.01.033.

[6] M. Mousavi, A. Koerner, Q. Zhang, E. Noh, and V. de Sa, "Improving motor imagery BCI with user response to feedback," *Brain-Computer Interfaces,* vol. 4, pp. 1-13, 04/11 2017, doi: 10.1080/2326263X.2017.1303253.

[7] G. Czanner *et al.*, "Measuring the signal-to-noise ratio of a neuron," *Proceedings of the National Academy of Sciences,* vol. 112, no. 23, p. 7141, 2015, doi: 10.1073/pnas.1505545112.

[8] R. Kieser, P. Reynisson, and T. J. Mulligan, "Definition of signal-to-noise ratio and its critical role in split-beam measurements," *ICES Journal of Marine Science,* vol. 62, no. 1, pp. 123-130, 2005, doi: 10.1016/j.icesjms.2004.09.006.

[9] D. M. Goldenholz *et al.*, "Mapping the signal-to-noise-ratios of cortical sources in magnetoencephalography and electroencephalography," (in eng), *Hum Brain Mapp,* vol. 30, no. 4, pp. 1077-86, Apr 2009, doi: 10.1002/hbm.20571.

[10] X. Jiang, G.-B. Bian, and Z. Tian, "Removal of Artifacts from EEG Signals: A Review," (in eng), *Sensors (Basel),* vol. 19, no. 5, p. 987, 2019, doi: 10.3390/s19050987.

[11] N. K. Al-Qazzaz, S. Hamid Bin Mohd Ali, S. A. Ahmad, M. S. Islam, and J. Escudero, "Automatic Artifact Removal in EEG of Normal and Demented Individuals Using ICA-WT during Working Memory Tasks," (in eng), *Sensors (Basel),* vol. 17, no. 6, p. 1326, 2017, doi: 10.3390/s17061326.

[12] S. Kanoga and Y. Mitsukura, "Review of Artifact Rejection Methods for Electroencephalographic Systems," 2017.

[13] M. Iftikhar, S. A. Khan, and A. Hassan, "A Survey of Deep Learning and Traditional Approaches for EEG Signal Processing and Classification," in *2018 IEEE 9th Annual Information Technology, Electronics and Mobile Communication Conference (IEMCON)*, 1-3 Nov. 2018 2018, pp. 395-400, doi: 10.1109/IEMCON.2018.8614893.

[14] P. Ofner, A. Schwarz, J. Pereira, and G. R. Müller-Putz, "Upper limb movements can be decoded from the time-domain of low-frequency EEG," *PLOS ONE,* vol. 12, no. 8, p. e0182578, 2017, doi: 10.1371/journal.pone.0182578.

[15] A. Craik, Y. He, and J. L. Contreras-Vidal, "Deep learning for electroencephalogram (EEG) classification tasks: a review," *Journal of Neural Engineering,* vol. 16, no. 3, p. 031001, 2019/04/09 2019, doi: 10.1088/1741-2552/ab0ab5.

[16] I. E. Shepelev, D. M. Lazurenko, V. N. Kiroy, E. V. Aslanyan, O. M. Bakhtin, and N. R. Minyaeva, "A Novel Neural Network Approach to Creating a Brain–Computer Interface Based on the EEG Patterns of Voluntary Muscle Movements," *Neuroscience and Behavioral Physiology,* vol. 48, no. 9, pp. 1145-1157, 2018/11/01 2018, doi: 10.1007/s11055-018-0679-0.

[17] P. Bashivan, I. Rish, M. Yeasin, and N. Codella, "Learning Representations from EEG with Deep Recurrent-Convolutional Neural Networks," 11/19 2015.

[18] D. Zhang, L. Yao, X. Zhang, S. Wang, W. Chen, and R. Boots, *Cascade and Parallel Convolutional Recurrent Neural Networks on EEG-based Intention Recognition for Brain Computer Interface*. 2017.

[19] G. Zhang, V. Davoodnia, A. Sepas-Moghaddam, Y. Zhang, and A. Etemad, "Classification of Hand Movements from EEG using a Deep Attention-based LSTM Network," *IEEE Sensors Journal,* vol. PP, pp. 1-1, 12/02 2019, doi: 10.1109/JSEN.2019.2956998.

[20] A. L. Goldberger *et al.*, "PhysioBank, PhysioToolkit, and PhysioNet: components of a new research resource for complex physiologic signals," (in eng), *Circulation,* vol. 101, no. 23, pp. E215-20, Jun 13 2000, doi: 10.1161/01.cir.101.23.e215.

[21] K. Simonyan, A. Vedaldi, and A. Zisserman, "Deep Inside Convolutional Networks: Visualising Image Classification Models and Saliency Maps," *preprint,* 12/20 2013.

[22] A. Farahat, C. Reichert, C. M. Sweeney-Reed, and H. Hinrichs, "Convolutional neural networks for decoding of covert attention focus and saliency maps for EEG feature visualization," *Journal of Neural Engineering,* vol. 16, no. 6, p. 066010, 2019/10/23 2019, doi: 10.1088/1741-2552/ab3bb4.

[23] L. H. Snyder, A. P. Batista, and R. A. Andersen, "Coding of intention in the posterior parietal cortex," *Nature,* vol. 386, no. 6621, pp. 167-170, 1997/03/01 1997, doi: 10.1038/386167a0.

[24] A. Riehle and J. Requin, "Monkey primary motor and premotor cortex: single-cell activity related to prior information about direction and extent of an intended movement," (in eng), *J Neurophysiol,* vol. 61, no. 3, pp. 534-49, Mar 1989, doi: 10.1152/jn.1989.61.3.534.

[25] C. Ghez, W. Hening, and J. Gordon, "Organization of voluntary movement," (in eng), *Curr Opin Neurobiol,* vol. 1, no. 4, pp. 664-71, Dec 1991, doi: 10.1016/s0959-4388(05)80046-7.

[26] R. Ackerley and A. Kavounoudias, "The role of tactile afference in shaping motor behaviour and implications for prosthetic innovation," (in eng), *Neuropsychologia,* vol. 79, no. Pt B, pp. 192-205, Dec 2015, doi: 10.1016/j.neuropsychologia.2015.06.024.

[27] D. Kingma and J. Ba, "Adam: A Method for Stochastic Optimization," *International Conference on Learning Representations,* 12/22 2014.