COHESION PREDICTION BASED ON EEG DATA

By Xueqing Li & Mariia Dolmatova

1. INTRODUCTION

In a social drumming experiment, 50 pairs of participants were recruited to complete 5 drumming sessions, each lasting 4 minutes with varying tasks.

We aim to design a machine learning model to classify high or low cohesion within dyads based on their EEG synchrony. Previous empirical research has demonstrated that neural activity synchrony detected by EEG can predict team performance (Reinero, D. A., Dikker, S., & Van Bavel, J. J., 2021). This inter-brain synchrony has also been shown to have a positive association with affective sharing abilities (Wang, Ren, & Chen, 2024). Moreover, preliminary analyses from the lab suggest that EEG synchrony predicts interpersonal cohesion within dyads (though these findings are unpublished, they provide a basis for this investigation).

Through our approach, we could uncover which EEG features (e.g., frequency bands or regions) contribute most significantly to the predictions. Aligning these findings with prior studies could provide more detailed insights into neural activity synchrony and its relationship with interpersonal cohesion. Ultimately, our work could inform future research directions by advancing our understanding of the mechanisms of synchrony and its applications in team performance, social bonding, and affective interactions.

2. METHODS

2.1 EEG synchrony

We calculate the EEG synchrony based on the correlation of the individual's EGG data with each other in the pair, with sampling rate = 500 Hz. For this project, we focused on the EEG data of participants from the last Freestyle drumming session, which represents the most natural and uninfluenced interaction as it was purely freestyle and involved synchronized drumming without external instructions, making it the best approximation of participants' spontaneous and natural coordination.

2.1.1 EEG preprocessing

The EEG data was pre-processed in MATLAB (vers. 2024b).

The reasons we used MATLAB are as follows: Firstly, MATLAB is a well-established set of tools specifically tailored for signal and EEG processing. Secondly, MATLAB is user-friendly, offering separate windows for managing variables and quick access to their contents with a single click. With 86 signals to process, numerous meta-variables to create, and frequent data reshaping at every step, it helped us massively to track progress and ensure accuracy. Additionally, we can create independent programs in MATLAB based on the variables from a primary program, which enables us to test different approaches and experiment. Finally, the lab data was already in .mat files. While we could have used MATLAB to generate the initial raw table and export it to a Python environment, the benefits mentioned above made MATLAB the better choice.

We inspected the data of each participant. There were 6 pairs' data missing, and 1 sample ended much earlier than the others, resulting in much less data points than others. Therefore, we ended up using 43 pairs of participants' data. The data was passed the signals through [0, 45 Hz] band before Artifact detection. Then we performed Amplitude detection (threshold = 30.000 points), Gradient artifact detection (threshold = 75) for getting rid of extremely rapid signals.

Five brain waves (δ, θ, α, β and γ) were separated with the cleaned signals. The edges of the Hz values for each band were normalised according to the Niquest frequency of our sampling rate (500 Hz). After that, the Butterworth filter was constructed to ensure the flattest possible response to the changes in the signal for avoiding additional peaks. Each signal and each electrode recording were processed individually. It results in an array of dimensions 86(participants) x 5(waves) by 8(electrodes).

We decided to separate the data by certain window frames to enlarge the size for a better model performance, and to allow us to explore if there is a certain trend in correlation data happening in specific time frames. Separation was performed relying on the sampling rate per second, the last part was usually 500 points smaller than others, since we had 119520 points in total. The time windows were: 2 seconds, 4 seconds, 24 seconds, 48 seconds; resulting in the signal being separated by 120, 60, 10, and 5 parts respectively.

For exploration, we did a PCA analysis with our biggest dataset (120 parts) and the smallest dataset (no dissection) to find out what is the principal components that contributes the distinguishment the most. We found out in the biggest dataset that Gamma band is the most important band, Electrode 7 (C4) and 5 (CZ) also contributes more than others (see Fig.1). In the smallest dataset, Gamma band also shows high influence and electrode 8 (F4) contributes the most (see Fig.2).

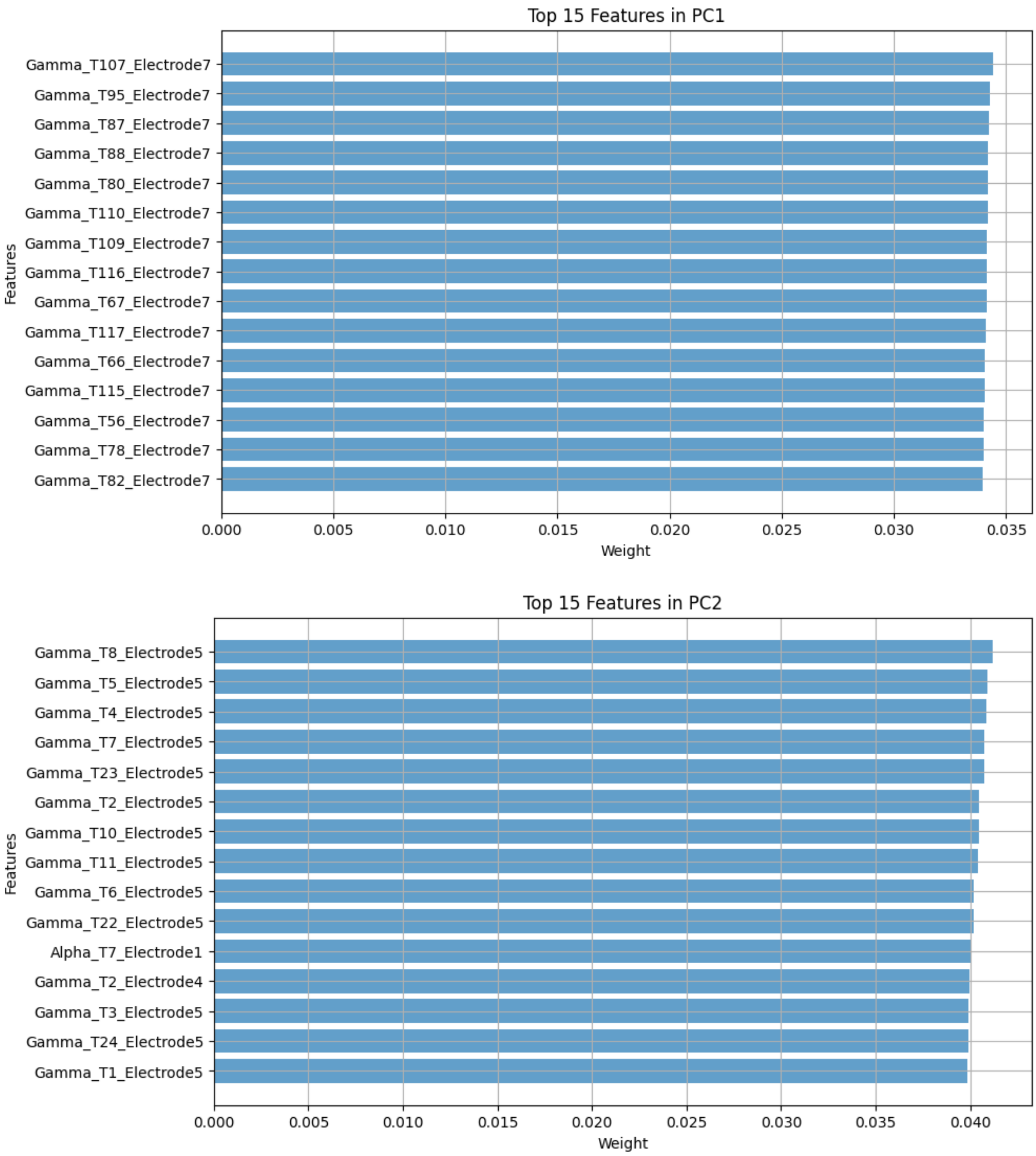

Fig.1 The PCA analysis in dataset dissected by every 2 seconds.

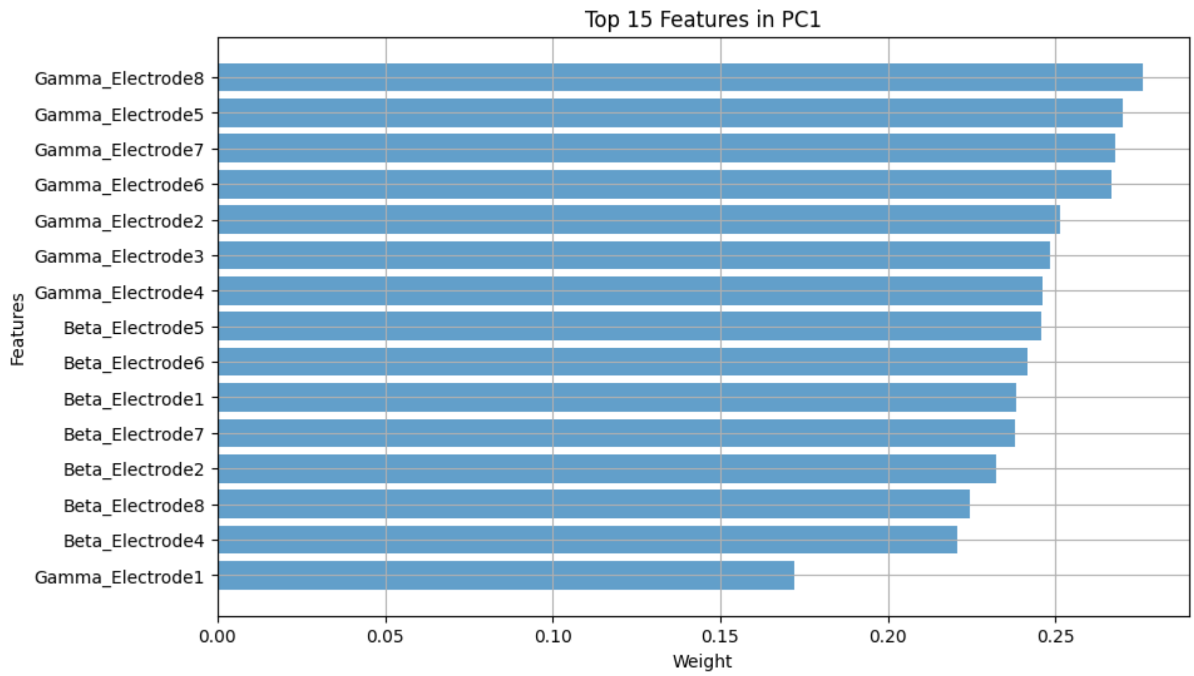

Fig.2 The PCA analysis in the dataset without dissection.

We further processed the data from MATLAB in python (vers. 3.11.11) to reshape it into a suitable format as input.

2.2 Cohesion scores and preprocessing

After each session, participants completed a cohesion questionnaire in 1 to 6 Likert scale to evaluate their perceived cohesion with their partner. The cohesion score for each pair was calculated as the average of both participants' individual scores within the dyad. There is no predetermined cut off of the measurement we used (Podsakoff & MacKenzie, 1994), to classify the cohesion score into binary labels as low or high cohesive, we decided to set a threshold of 4.5, which reflects the expectation that, by the final session of the experiment, participants had already established some level of cohesion due to their prior interactions and practice. It as well reflects a more balanced number of labels (Fig.3 & Fig.4), while not sacrificing the objectiveness too much. Thus, dyads whose cohesion score lower or equal to 4.5 is classified as low cohesive, while dyads whose cohesion score higher than 4.5 is classified as high cohesive. We processed the row cohesion data of each participant in python (vers. 3.11.11) to obtain the binary labels for our model.

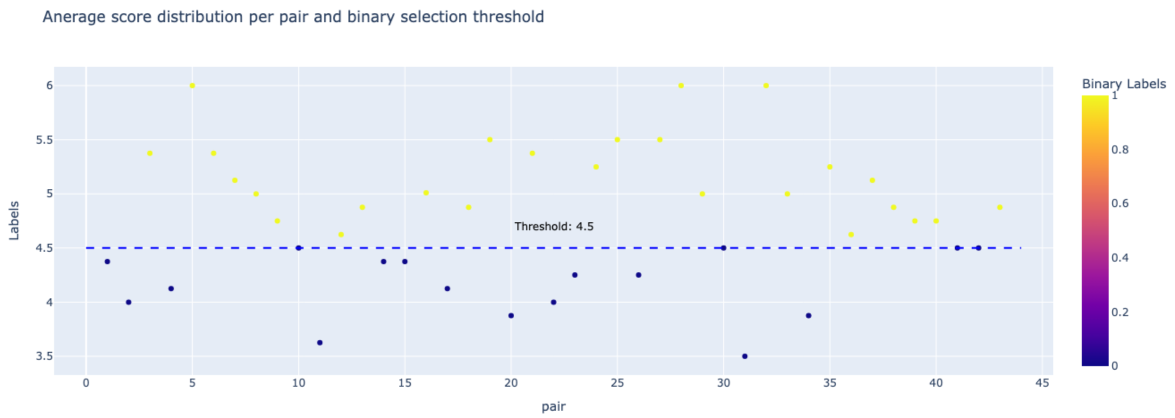

Fig.3 Label distribution (1)

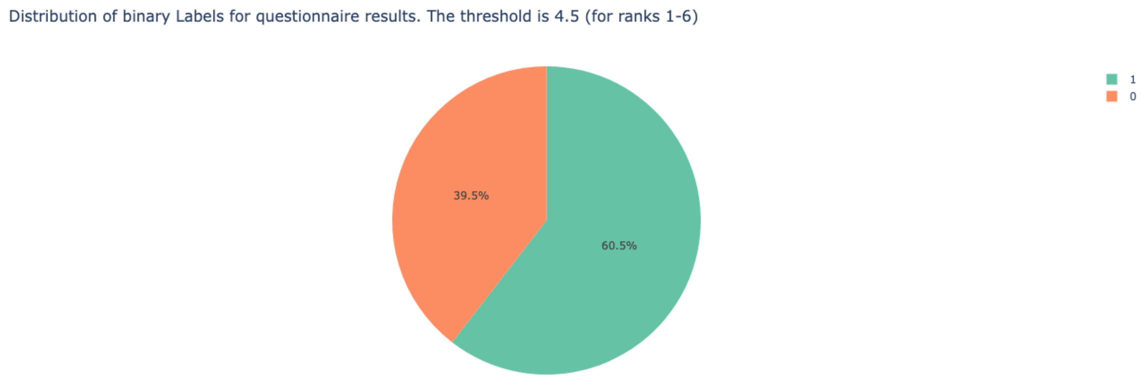

Fig.4 Label distribution (2)

2.3 Machine learning model

In our project, we decided to compare two models: 1) SVM Model and 2) CNN Model. SVM handles non-linearity, reduces overfitting with regularization parameters. Our CNN model is primarily serving for comparison against the SVM. We understand that a CNN may not be the best choice for our dataset, as CNN's many parameters can easily lead to overfitting. Our features involve time, frequency bands, and electrodes, which are intricately interwoven, it is hard to separate it into related dimensions.

2.3.1. SVM Model details

We used SVM Model for all 5 datasets we have. We used GridsearchCV to determine the final parameters and combined the 5-fold cross validation, the kernel are chosen between rbf, linear and polynomial, and F1 score is choose for evaluating model performance. Considering the unbalanced labels, we emphasized the weight of label “0” by balancing the class weight.

2.3.2 CNN Model details

We only used CNN Model for the dataset which has been dissected into 120 segments (split every 2 second) as CNN model is mainly for comparison with the SVM model.
We chose 1D CNN as we don't genuinely have two distinct features suitable for a 2D approach nor do we have a large enough dataset. We also use the sample norm, 5-fold Cross-Validation, dropout, early stopping, max pooling and batch normalization. Our hyper parameters are typical for a smaller dataset. There are 2 convolution layers.
In our hypothesis, we believe SVM Model has a better performance than CNN Model as SVM handles better for small datasets. Our criterion is BCE With Logits Loss for balancing the labels.

3. RESULT

3.1 SVM Model

We tested all five datasets with SVM model, the F1 score is not varied significantly across different dataset (Fig.5 and Fig.6), rbf and Polynomial are the selective kernel. The best performance combination by F1 score is the 5 parts dataset (0.77). We checked the individual performance with different dataset, for example, Figure 7 shows the different process of the model with the smallest dataset. There is not much difference between datasets in different sizes and no linear correlation between input size and performance.

	Dataset	K Value	C	Gamma	Kernel	Best F1 Score
0	reshaped_correlations.csv	5	1.0	10.0	rbf	0.753407
3	reshaped_correlations5.csv	all	0.1	0.1	poly	0.770256
1	reshaped_correlations10.csv	5	1.0	10.0	rbf	0.753407
4	reshaped_correlations60.csv	5	1.0	10.0	rbf	0.753407
2	reshaped_correlations120.csv	40	0.1	0.1	poly	0.756703

Fig.5 SVM model result of all 5 dataset.

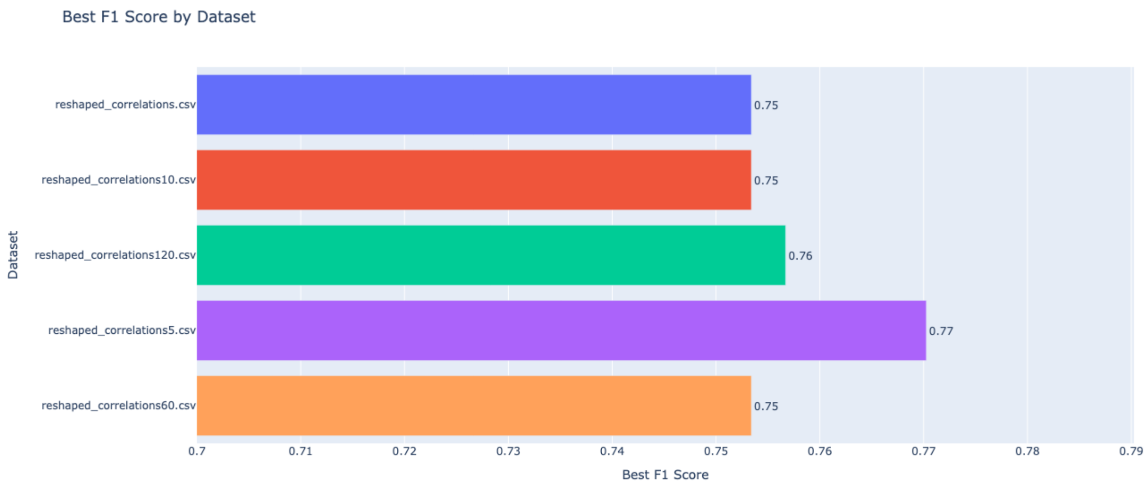

Fig.6 Visualization of the performance between 5 datasets.

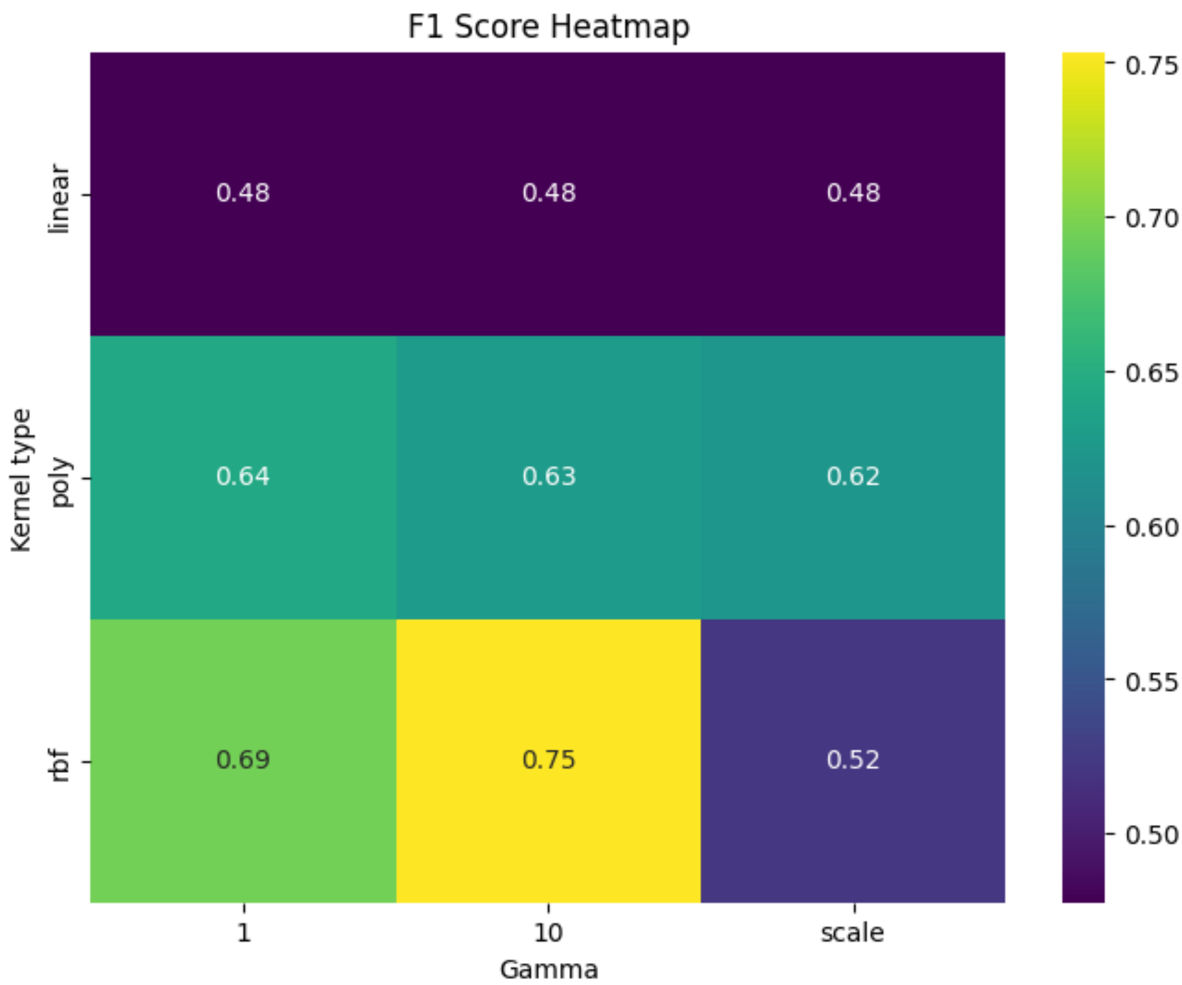

Fig.7 (1) detailed selection process with the smallest dataset.

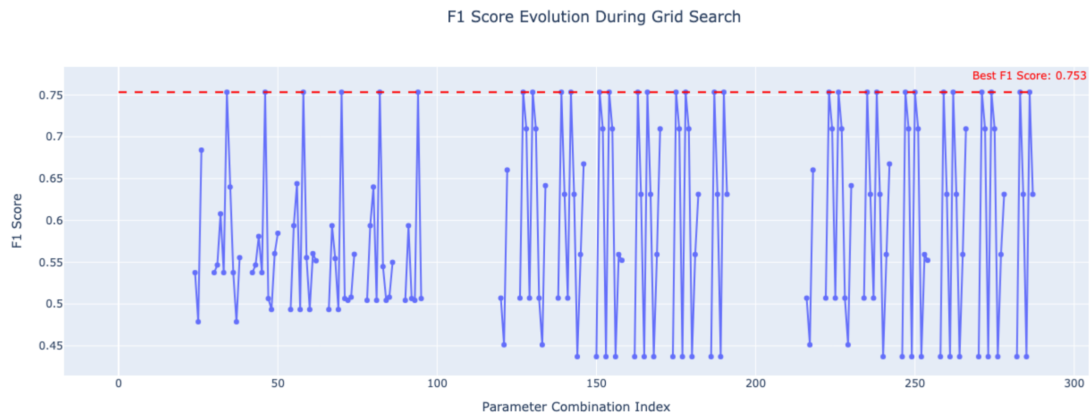

Fig.7 (2) detailed model iteration process with the smallest dataset.

3.2 CNN Model

Below is the outcome of one in the five folds cross-validation session (see Fig.8), the average loss of the five folds is 0.6416, average accuracy is 0.4861 and average F1 score is 0.5463. The outcome of other folds is very similar. From the graph, we can tell the overfitting is a great issue with CNN model. The average performance of CNN is worse than SVM.

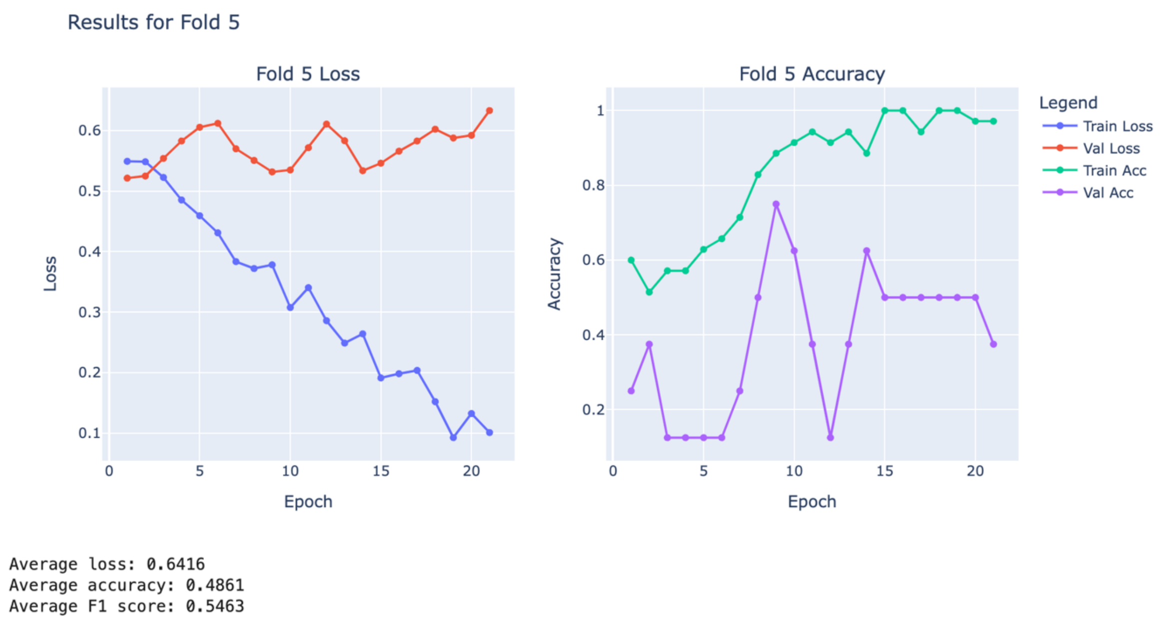

Fig.8 CNN performance

4. DISCUSSION

Our model performance is not high, we believe the result is rooted in our small dataset (43 samples) and unequal distribution of the labels. The dataset is easily causing overfitting, particularly CNN Model, that have more trainable parameters and need more data to learn complex features. Even if we expanded it by dissecting, the correlation between each time window may not very much, regarding the SVM Model result, F1 score is equal to different dataset, it may be due to the overfitting as well. It is hard to practice data augmentation in the current datasets, as the data are numerical data obtained from real-life experiments. The most practical method in the future is to recruit more participants and enhance the dataset size, and to change the method to calculating EEG synchrony and enable the model to learn more. Plus, we focused exclusively on the last session, which could introduce bias into the model. The label process itself is also arbitrary, which may have contributed to the inconsistencies in prediction performance. We also need to re-consider the threshold of the labels, so that we can have a much more balanced data to train the model.

In our PCA analysis, we found that EEG synchrony (correlation) from electrodes C4, CZ, F4 along with gamma-band activity are the most significant components of our input. These findings suggest that these brain areas and frequencies—which are strongly activated during drumming—are closely linked to subjective cohesion, consistent with prior research. Because C4 overlaps with the motor cortex, this highlights the role of motor synchronization in shaping how participants experience subjective cohesion. Additionally, gamma-band activity, which reflects sensorimotor integration (Rieder et al., 2010), further indicates that neural dynamic synchrony is tied to subjective feelings, especially in cooperative contexts (Sinha et al., 2016). The CZ electrodes correspond to the central region of the brain, related to motor and sensorimotor function. F4 responds to anterior cingulate cortex (ACC), and dorsolateral prefrontal cortex (dlPFC). These areas are heavily involved in higher-order cognitive processes and social-emotional functions (Lieberman et al., 2019; Apps et al., 2016).

In future studies, attention should be directed towards the synchrony of specific brain region in the brain and the gamma band activity, along with increasing the sample size and have a more reasonable label set.

REFERENCE

Apps, Matthew A.J., et al. “The Anterior Cingulate Gyrus and Social Cognition: Tracking the Motivation of Others.” Neuron, vol. 90, no. 4, May 2016, pp. 692–707, www.ncbi.nlm.nih.gov/pmc/articles/PMC4885021/, https://doi.org/10.1016/j.neuron.2016.04.018.

Lieberman, Matthew D., et al. “Social, Self, (Situational), and Affective Processes in Medial Prefrontal Cortex (MPFC): Causal, Multivariate, and Reverse Inference Evidence.” Neuroscience & Biobehavioral Reviews, vol. 99, Apr. 2019, pp. 311–328, https://doi.org/10.1016/j.neubiorev.2018.12.021.

N. Sinha, T. Maszczyk, Zhang Wanxuan, J. Tan and J. Dauwels, "EEG hyperscanning study of inter-brain synchrony during cooperative and competitive interaction," 2016 IEEE International Conference on Systems, Man, and Cybernetics (SMC), Budapest, Hungary, 2016, pp. 004813-004818, doi: 10.1109/SMC.2016.7844990.

Podsakoff, P. M., & MacKenzie, S. B. (1994). An examination of the psychometric properties and nomological validity of some revised and reduced substitutes for leadership scales. *Journal of Applied Psychology*, *79*(5), 702–713. https://doi.org/10.1037/0021-9010.79.5.702

Reinero, D. A., Dikker, S., & Van Bavel, J. J. (2021). Inter-brain synchrony in teams predicts collective performance. Social cognitive and affective neuroscience, 16(1-2), 43–57. https://doi.org/10.1093/scan/nsaa135

Rieder, Maria K., et al. “Human Gamma-Band Activity and Behavior.” International Journal of Psychophysiology, vol. 79, no. 1, Jan. 2011, pp. 39–48, https://doi.org/10.1016/j.ijpsycho.2010.08.010. Accessed 28 Nov. 2021.

Wang, D., Ren, Y. & Chen, W. Relationship evolution shapes inter-brain synchrony in affective sharing: The role of self-expansion. Brain Struct Funct 229, 2269–2283 (2024). https://doi.org/10.1007/s00429-024-02841-0