Causality analysis of EEG signals using Granger causality and implementation of the Graph Neural Network

Bioengineering for Neurosciences project exam

Greta Feregotto, Erica Li Calzi, Francesco Fantechi

A.A. 2022-2023

Relatori: Antonio Lanatá
Correlatori: Pietro Tarchi, Lorenzo Frassineti, Federico Calá

UNIVERSITA' DEGLI STUDI DI FIRENZE
Facolta di Ingegneria
Corso di Laurea Magistrale in Ingegneria Biomedica

1. Abstract

The explosion of user-generated, untagged multimedia data in recent years, generates a strong need for efficient search and retrieval of this data.
Recently, the study of emotion recognition has received increasing attentions by the rapid development of noninvasive sensor technologies, machine learning algorithms and compute capability of computers.
For this reason, EEG signals from 26 subjects were analyzed to get a better insight on connectivity analysis between different areas of the brain, while subjects were evaluating a series of real or fake images. Three types of images (positive, negative, and neutral) were utilized to elicit specific reactions of subjects.
The power spectrum density (PSD) features of EEG were extracted by time-frequency analysis, and the relative features, along with Granger causality (GC) matrix, were selected as inputs of a Graph-Embedded Convolutional Neural Network (GCNN).
The GC matrix led to understanding that different brain areas were communicating while the subjects were performing the task.
The GCNN implemented is useful to understand whether the subjects are able to distinguish a real image from a fake one. The lack of enough data has led to results that are not so useful for drawing conclusions, but this study can be a good starting point for inspiration for future works.

2. Introduction

Emotion recognition allows machines to understand the emotional states of humans to make more humanized decisions, a process that plays an important role in the research fields of artificial intelligence and human-machine interaction.
Human emotion states directly affect behavioral and physiological signals, which are the two most used methods to measure different frames of mind. Although behavioral signals, such as facial expression and speech, are commonly used for affective computing, physiological signals have incomparable advantages over behavioral signals, since they are spontaneous with their emotions and difficult to be disguised. Among physiological signals, electroencephalogram (EEG) is commonly used to identify various emotion states, due to higher classification accuracy. EEG signals are obtained directly from the cerebral cortex, and thus directly reflect changes in human emotions. As a kind of noninvasive signal, EEG provides an effective and comprehensive way to measure the brain's electrical activity, which has millisecond time resolution and contains richer information related to emotions. Therefore, emotion recognition using EEG signals has recently attracted more attention.
In this work EEG signals have been used to analyze brain activation and connectivity between different brain regions in response to optical stimuli for the recognition of real and deep fakes images.
Informally, deep fakes can be defined as realistic digital media (images, videos, or audio tracks) depicting untruthful content, obtained either by manipulating pristine material or generated from scratch. The attribute deep refers to the use of algorithms based on deep learning, a subfield of modern artificial intelligence (AI), which pushed the boundaries for many applications, including media data manipulation and generation.

The remainder of this paper is organized as follows:

- Section 3 introduces the related work and the state of the art about EEG analysis for connectivity among different cerebral areas.
- In Section 4 are briefly described the dataset used to generate images

shown to the subjects and the experiment assessment for the acquisition of EEG signals.

- Section 5 explains preprocessing methods for feature extraction from EEG channels.
- The proposed neural network to analyze connectivity features and causality is explained in Section 6.
- Finally, in Section 7, experimental results and conclusions are presented.

3. Realated works

In this section, are first introduced the related works about the state of the art of EEG analysis for connectivity among different cerebral areas.
As described in papers [2,3], there are three types of brain connectivity: structural connectivity that characterizes the brain regions anatomical connectivity; functional connectivity is described as the relations and interactions between separated brain areas, it only contains the statistical dependencies between EEG signals, and cannot provide any causal or directional information; effective connectivity refers to direct influence of one neural system on others and can further measure the causal relationship between EEG signals. Therefore functional and effective connectivity are the most studied to estimate human brain areas connectivity.
There are various techniques to evaluate functional connectivity such as Correlation, Phase Synchronization (PS) and Coherence; in addition further studies have found out that the MI provided excellent information for emotion classification [2].
The analysis of the effective connectivity would be useful instead for the discriminative EEG feature extraction to improve the performance of emotion recognition and considering the non-linearity of the brain system. Quite recently, De Castro Martins et al. [4] investigated the existence of a causal connection, the magnitude and direction of causal propagation, and the most appropriate time lags to determine causality. They posit that investigating connectivity between specific region pairs that may form part of larger networks should still shed light on candidate mechanisms. In particular, De Castro Martins et al. studied pairwise "dynamic" connectivity between ROIs using a nonlinear data-driven approach known as Convergent Cross Mapping (CCM), a novel method to study the coupling between time series. It is a type of empirical dynamic modeling (EDM), which are non-parametric frameworks for modeling nonlinear dynamic systems. EDMs are an alternative and highly flexible approach to use explicit equations since these equations can be impractical when the exact mechanisms are unknown or too complex to be characterized with existing datasets [4].
As studied in papers [5,6], another method employed to estimate the strength of coupling, causal configurations and the information transfer between brain regions is also the Transfer Entropy (TE). The TE considers this communication without assuming a specific relationship between signals, resulting more applicable to nonlinear systems. Further it results useful for the classification of emotions, quantifying the influential effect between the channels [6]. However, these estimators could lead to high level of uncertainty due to a significant degree of freedom, and they might require large amounts of data to obtain reasonable estimates [7].
Among the many choices for understanding relationships between time series, Granger causality is a commonly used framework for time series structure discovery that quantifies whether the past of one time series helps in predicting the future values of another time series. It is based on two major principles: (i) the cause happens

prior to the effect, and (ii) the cause makes notable changes in the effect. Generally speaking, granger causality occurs if and only if the predicted value of Y based on the past values of X and Y are better than predictions based on the past values of Y alone. GC also allows the study of an entire system of time series to uncover networks of causal interactions.

The only negative aspect of GC is that it's only capable of reconstructing linear interactions, while inter-regional couplings often show nonlinear causal relationships [5,6]. A positive aspect is that it can inhibit the noise connection between EEG channels and also make the selected channel connection more suitable for the recognition law of the human brain. GC was employed by Raha et al. [8] to identify the consecutive emotional states which are more likely to be Granger causal in the human brain by sampling the entire signal in order to determine whether the previous state is granger causal to the current one.

Granger causality is known to work well on stationary signals, but it is well established that EEG signals exhibit non-stationary nature. In order to ensure that there are stationary signals, it's necessary to decompose each channel time series into 30 data frames with overlapping windows [6]. Even Wang et al. use Granger causality analysis (GCA) to capture the causal cortical network dynamics during dynamic or static facial expressions of emotion [9].mind. Indeed neural networks are capable of representing complex, non-linear, and

Once used the methods above to find connectivity relationship among different brain areas, neural networks were employed in the emotion recognition field classifying the frame of non-additive interactions between inputs and outputs via a series of transformations unlike the use of multivariate GC alone working with EEG signals that lose the spatial information and just provide a rough way to describe brain activities.

As can be seen in [1] Bashivan et al. proposed a method to interpolate the EEG discrete energy features extracted from different electrodes into images to measure the energy changes over the scalp. To build the connections among different regions in the scalp for EEG emotion recognition, they proposed a Convolutional Neural Network (CNN). Since it is quite difficult for CNN to capture functional dependencies among different regions, the graph theory deserved to be introduced. The graph CNN (GCNN) provides an effective way to model the relationship between different nodes and the hierarchic structure to extract the deep features in graph domain.

Another efficient procedure is the GC-CNN, the mixture of GCNN and GC analysis to guide the construction of the adjacency matrix. In this way, the obtained adjacency matrix will be more suitable for the recognition law of the human brain, because it is calculated through the causal relationship among EEG signals [10].

The GC method, combined with neural networks, could be used to deal with the problem of nonlinearity and varying-length signal delays. Li et al. hypothesized that the emotional cortex interacts with the motor cortex during the mutual regulation of emotion and movement. They proposed a brain connectivity analysis method based on electroencephalogram signal processing: bidirectional long short-term memory Granger causality (bi-LSTM-GC) [3].

Three paradigms were included in the simulation test. These included linear simulations (model A), nonlinear simulations with varying lag lengths (model B), and bidirectional nonlinear simulations with varying lag lengths (model C). In the test result from model A, RNN-GC gave the most accurate results. All the dependencies presupposed in model A were correctly captured, with no false alarms. Dependencies were successfully detected in some cases using bi-LSTM-GC, but invalid detections also occurred. In the test result from model B, RNN-GC and bi-LSTM-GC captured all dependencies without any false detections. Using NN-GC, all dependencies were detected, but the detections were obscured by noise from false detections. Moreover, invalid detections occurred in the tests of signals with long time delays. In the test result from model C, bi-LSTM-GC exhibited the best accuracy, with all the dependencies being correctly captured and without false detections. RNN-GC achieved the detection of nonlinear and long delays well, but it could not deal with the detections of forward dependency [3].

4. Dataset and experiment

60 black and white images of caucasian human faces (see figure 1) have been taken from the CK+ dataset. Of these 60 images, 30 were of men and 30 of women and each group of 30 images was equally labeled into three frames of mind: positive (happy), negative (negative men were in turn subdivided into sadness, anger, disgust and fear and negative women into sadness and anger) and neutral. These images have been preprocessed with a black padding and then mixed according to frames of mind through the software Playform to create artificial and counterfeit images of no-existing people.

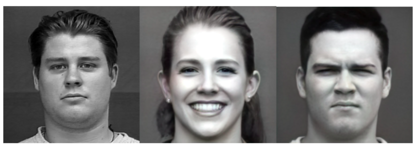


Figure 1: samples of black and white images of caucasian human faces from CK+ dataset

A fraction of the true 60 images and the new 30 fake ones have been then shown to 26 participants (13 men and 13 women) aged between 28 and 19. The experiment was divided in 3 sections: in the first and second, one image at a time was presented to the subjects (for a total of 30 images shown for each of these two phases to each subject), who had to decide whether it was real or artificial. In the third, instead, two images were presented at the same time (for a total of 24 couples of images shown to each subject) and the subjects had to decide which of the two was real (see figure 2). Not necessarily the two images presented contained a true one.

Figure 2: sample of the third phase of the experiment (fake figure on the right and true figure on the left)

Each subject was equipped with a cap for the extraction of the scalp EEG signal, two electrodes on the chest for the extraction of Heart Rate Variability (HRV), a finger clamp to measure sweating and Tobii glasses calibrated for eye tracking. The participant was made to sit comfortably in front of a computer to limit movement artifacts and an initial relaxation phase was performed to calibrate the instruments and assume the baseline.
Then the first and second phase of the experiment started showing each image for 10 seconds through a handmade python slider program. At the end of this time the subject could take all the time necessary to make his decision which he indicated with the keys on the keyboard: $m = true$, $z = artifact$. For the third phase subjects chose which face between the double presented was real: if it was the one on the left they had to press z or m if it was the one on the right.
The program used for the images slider was created with python and for each phase

of the experiment it saved the choices made by the current participant, the real response and the time taken in a txt file. For our subsequent experiments we focused on the analysis of the extracted EEG signal.

5. EEG preprocessing

The entire project's code written in Matlab and Python language has been versioned and is available on the GitHub link: https://github.com/francesco-ftk/EEG_GNN.git

11 subjects whose signals were too noisy were excluded from the analysis of the EEG signals. The subjects included in further analysis were those with $index = [2, 3, 6, 9, 11, 12, 13, 15, 17, 18, 19, 22, 23, 24, 26]$. The EEG signals has been extracted and to achieve this purpose the *trigger.m* script has been modified by adding the *eval* function as follows: eval(['EEG_subject_' num2str(index) '=mat_eeg;']). In this way, 15 matrices were obtained in output from *trigger.m* script in which the size of the rows (on average 440000) varies according to the number of samples acquired for each subject and the number of columns (18) corresponds to the EEG channels.

For the preprocessing the following steps have been implemented:

- The EEG signals have been plotted to see if the signal had trouble acquiring.

- Band-pass filtering between $1Hz$ and $50Hz$ (EEG frequency band interval) with sampling frequency of $300Hz$ (see inside script *topoplot.m* for the code). In figures 3 and 5 are shown the most significant raw signals (subjects 2, 6) and in figures 4 and 6 are shown the same signals filtered.

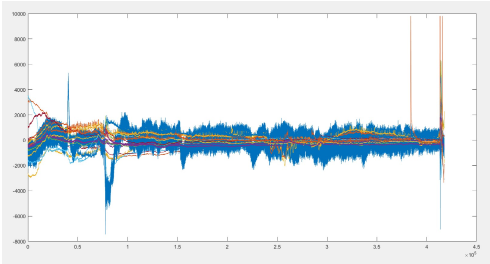


Figure 3: subject 2 EEG signal

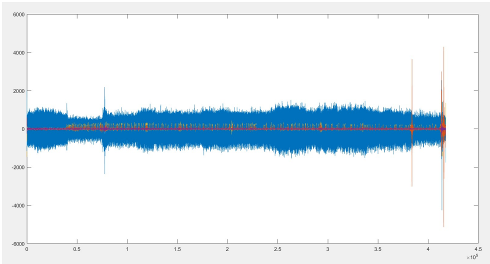


Figure 4: subject 2 EEG filtered signal

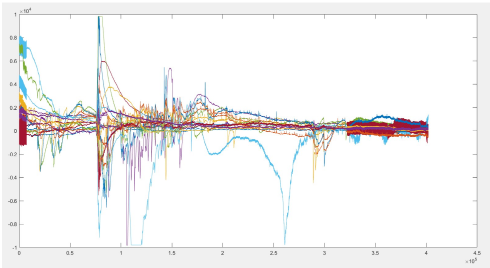


Figure 5: subject 6 EEG signal

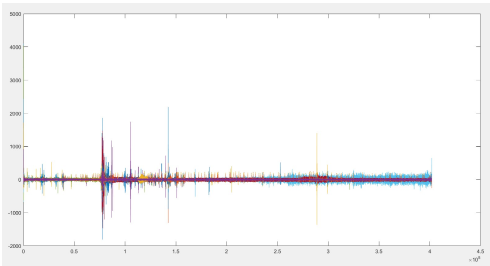


Figure 6: subject 6 EEG filtered signal

- The signals were then subdivided in the 5 typical EEG frequency bands: delta $1-4Hz$, theta $4-8Hz$, alpha $8-14Hz$, beta $14-30Hz$, gamma $30-50Hz$.

- Utilizing the trigger values (from *trigger.m* script), signals have been divided into windows, each window corresponding to the time interval of image visualization.

- For each window the GC matrix has been calculated (see script *GC_Input_matrix.m* for the code). The GC matrix was computed for alpha, beta, gamma and theta frequency bands, not for delta because delta frequency band corresponds to brain inactivity and so it is not useful for the purpose of the project. The Granger Causality matrix has been visualized as can be seen in figure 7 (see script *plot_Granger_Causality_matrix.m* for the code).

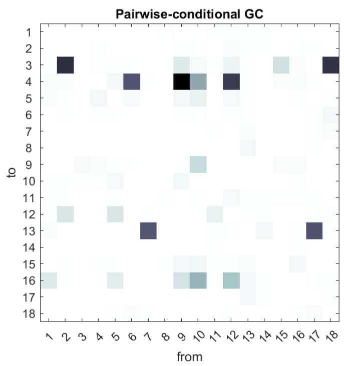


Figure 7: Granger Causality matrix for alpha EEG signal subject 2

- The energy features of the signal were extracted (power spectral density) for the next processing (see *topoplot.m* for the code).

- These values were then normalized and scaled in a bounded range between 0 and 1. Afterwards the mean value for each window, over matrix columns, has been taken (see *main.py* for the code).

For a better understanding of EEG signal distribution over the scalp, Topoplot (a topographic map of an EEG field as a 2-D circular view - looking down at the top of the head - using cointerpolation on a fine cartesian grid) have been plotted, as shown in figure 8.

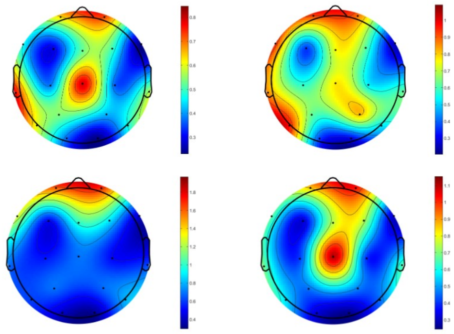


Figure 8: subject 2 topoplot corresponding to four windows

Topoplots (see script *topoplot.m* for the code) have been generated by the matlab library EEGLAB providing EEG signals and electrode position on the scalp.
The EEG sensors for the $DSI - 24$ system are mounted in a lightweight, user-adjustable headset, which positions the sensors at the nominal $Fp1$, $Fp2$, $F7$, $F3$, Fz, $F4$, $F8$, $T3$, $C3$, Cz, $C4$, $T4$, $T5$, $P3$, $P4$, $T6$, $O1$ and $O2$ positions of the International 10/20 System. The $DSI-24$ system is available with options to measure from either the mastoid locations ($M1$, $M2$) or earlobes ($A1$, $A2$). A reference sensor (Common Mode Follower, CMF) is placed at the nominal Pz position.

6. Neural network

In this section the steps followed to implement the Neural Network are explained. In order to check the existence of a correlation between data and brain areas the aim of the Network is to learn a pattern in EEG signals to assert if a subject was able to distinguish a fake image from a real one.
To summarize, for each subject there were 240 input data of size 18×3000 (since each subject had viewed 60 images and the frequency bands taken into account were 4), for each of which it was calculated the

Granger Causality matrix $[18 \times 18]$ and the mean energy value for column $[18 \times 1]$, so, having 15 subjects, in total 3600 pair of datas (input data and GC matrixes) were obtained.
For the Neural Network the Python library Pytorch have been used.
The implemented Neural Network (NN) is divided into two sections:

- A Graph Neural Network (GNN) that performs a message passing among the mean values of the nodes of the Graph (where nodes correspond to the electrodes and the existence of edges between nodes is defined by a threshold function on GC matrix).
- A Multilayer Perceptron (MLP) that, according to the result of the first section, gives a probability score for each class (true or false).

The GNN was formed by 18 nodes (the electrodes) and the initial adjacency matrix was constructed with a threshold of GC matrix, for each frequency band.
This network perform the message passing (see figure 9) with this formula:

$$H^{(l+1)} = \sigma(\bar{D}^{-1/2}\bar{A}\bar{D}^{-1/2}H^{l}W^{l})$$

where H is the node value, $\bar{D}$ is the Diagonal matrix, $\bar{A}$ is the Adjacency matrix A plus the Identity matrix I (so that each node sends its own message also to itself), W are the GNN weight parameters and σ represents an arbitrary activation function.

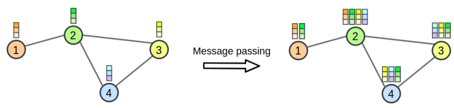


Figure 9: GNN message passing process

Instead of defining a matrix $\bar{D}$, the summed messages can be simply divided by the number of neighbors afterward. Additionally, the weight matrix has been replaced with a linear layer, which additionally has allowed to add a bias (see code in *main.py*).
In a GNN, feature exchange between nodes beyond its neighbors have to be allowed.
This can be achieved by applying multiple GCN layers, which give the final layout of a GNN. The GNN can be build up by a sequence of GCN layers and non-linearities such as ReLU (see figure 10).

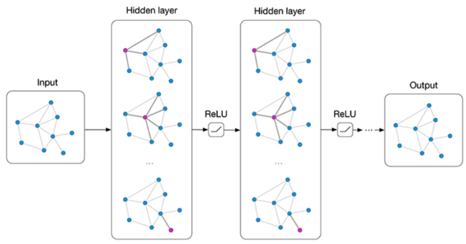
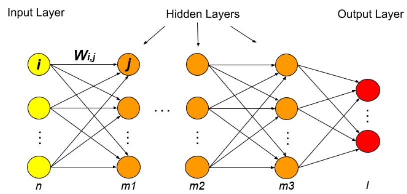


Figure 10: GNN model

Then the result (a $18x1$ Tensor containing the mixed input) has been passed to the MLP (see figure 11) that perform the class prediction.



$$\alpha_j = g(\sum_{i=0}^{N} W_{i,j}\alpha_i)$$

con g funzione di attivazione e α_k output del nodo k-esimo

Figure 11: MLP model

Written with PyTorch library, the Neural Network is defined as follows in figure 12:

```python
input_size = 18
hidden_sizes = [10, 8]
output_size = 2

class MyGCNNModel(nn.Module):
    def __init__(self, c_in, c_out, number_hidden_layer):
        super().__init__()
        self.hidden_layers = number_hidden_layer
        self.projection = nn.Linear(c_in, c_out)
        self.fl1 = nn.Linear(input_size, hidden_sizes[0])
        self.fl2 = nn.Linear(hidden_sizes[0], hidden_sizes[1])
        self.fl3 = nn.Linear(hidden_sizes[1], output_size)

    def layer_GCNN(self, node_features, adj_matrix):
        # Num neighbours = number of incoming edges
        num_neighbours = adj_matrix.sum(dim=-1, keepdims=True)
        node_features = self.projection(node_features)
        node_features = torch.bmm(adj_matrix, node_features)
        node_features = node_features / num_neighbours
        return node_features

    def forward(self, node_features, adj_matrix):
        with torch.no_grad():
            for g in range(0, self.hidden_layers):
                node_features = self.layer_GCNN(node_features, adj_matrix)
                node_features = F.relu(node_features)
            node_features = torch.reshape(node_features, (BATCH_SIZE, 18))
        for j in range(0, BATCH_SIZE):
            for i in range(0, 18):
                if math.isnan(node_features[j][i]):
                    node_features[j][i] = 0.0
        x = F.relu(self.fl1(node_features))
        x = F.relu(self.fl2(x))
        output_net = self.fl3(x)
```

Figure 12: Neural network definition with Pytorch

The network was then trained in batch modality using Tensorboard library of Tensorflow for the cross-validation steps. Labels and ground truth were given as inputs of the Neural Network at the beginning.

On the basis of the NN outputs the loss function (Cross Entropy Loss) could then be used to update the network weights with the optimizer Adam, so that the network could learn and predict the subject answer (see figure 13).

```
trainingSet = CustomDataset(train_input, train_gc, train_label)
trainDataloader = DataLoader(trainingSet, batch_size=BATCH_SIZE, shuffle=True)

criterion = nn.CrossEntropyLoss()
net = MyGCNNModel(1, 1, 3).double()
optimizer = optim.Adam(net.parameters(), weight_decay=1e-5)

writer = SummaryWriter("runs")

for epoch in range(100):  # loop over the train dataset multiple times
    print('Running Epoch: ', epoch)

    # Epoch Train

    for i, data in enumerate(trainDataloader, 0):
        inputs, gcs, labels = data

        optimizer.zero_grad()

        outputs = net(inputs.double().view(BATCH_SIZE, 18, 1), gcs.double())
        loss = criterion(outputs, labels)
        loss.backward()
        optimizer.step()
```

Figure 13: Neural network training steps

7. Results and Conclusion

In this project, EEG signals have been extracted from subjects while they were observing real or artificial face images and at the same time they were trying to distinguish to which category the figures belong.
A Neural Network has been implemented to find a pattern in brain connectivity corresponding to the subject answer.
The results have shown that the proposed method has an accuracy of 53%.
The accuracy obtained is not so useful to draw conclusions, this can be explained by the lack of enough data and by the nature of EEG signal that can be difficult to analyze since it is captured superficially and so very reduced by the scalp's presence.
The consequence of the lack of data can be seen in the figure 14 where the network overfits in the first epochs: the train loss decreases while the validation loss increases.
However this study can be a good starting point for inspiration for future works collecting more data and increasing the number of subjects involved.

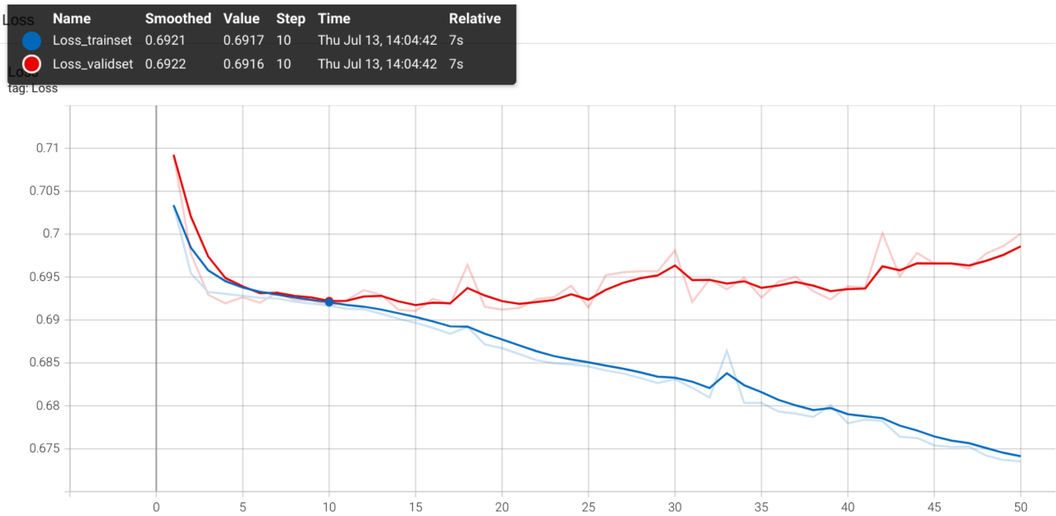


Figure 14: Tensorboard's training and validation loss graph along the epochs

8. References

[1] TENGFEI SONG , WENMING ZHENG , SUYUAN LIU, YUAN ZONG , ZHEN CUI, AND YANG LI. Graph-Embedded Convolutional Neural Network for Image-Based EEG Emotion Recognition, 2022. Emerging Topics in Computing.

[2] Tafreshi, T.F., Daliri, M.R., Mahrad Ghodousi. Functional and effective connectivity based features of EEG signals for object recognition, 2019. Cognitive Neurodynamics 13, 555–566.

[3] Li T., Li G, Xue T and Zhang J. Analyzing Brain Connectivity in the Mutual Regulation of Emotion–Movement Using Bidirectional Granger Causality, 2020. Front. Neurosci. 14:369.

[4] Caio de Castro Martins, Chaminade Thierry, Marc Cavazza. Causal Analysis of Activity in Social Brain Areas During Human-Agent Conversation. Frontiers in Neuroergonomics, Frontiers Media SA, 2022. 3, pp.843005.

[5] Yunyuan Gao, Xiangkun Wang, Thomas Potter, Jianhai Zhang, Yingchun Zhang. Single-trial EEG emotion recognition using Granger Causality/Transfer Entropy analysis, 2020 Journal of Neuroscience Methods.

[6] J. Siva Ramakrishna, Neelam Sinha, Hariharan Ramasangu. Classification of Human Emotions using EEG-based Causal Connectivity Patterns, 2021. IEEE Conference on Computational Intelligence in Bioinformatics and Computational Biology (CIBCB).

[7] Horvàth S., Sultan M. S., Ombao, H. Granger Causality Using Neural Networks, 2022. Cornell University, 1-17.

[8] R. Raha, A. Sengupta, A. Saha. Causality Analysis of Emotional States from EEG Response, 2020. IEEE International Conference for Convergence in Engineering. 410-415.

[9] M.-Y. Wang, Z. Yuan. EEG Decoding of Dynamic Facial Expressions of Emotion: Evidence from SSVEP and Causal Cortical Network Dynamics, 2021. Neuroscience 459, 50–58.

[10] Zhang J., Zhang X., Zhao Q., Improved Graph Convolutional Neural Networks based on Granger Causality Analysis for EEG Emotion Recognition, 2022. International Conference on Computer Engineering and Artificial Intelligence (ICCEAI).