Classification of Emotions Based on EEG Data Using Connectivity Features

Anja Popović

89201065@student.upr.si

University of Primorska
The Faculty of Mathematics,
Natural Sciences and Information Technologies
Glagoljaška ulica 8

6000 Koper, Slovenia

Peter Rogelj

peter.rogelj@upr.si

University of Primorska
The Faculty of Mathematics,
Natural Sciences and Information Technologies
Glagoljaška ulica 8

6000 Koper, Slovenia

ABSTRACT

Emotions are a fundamental element of human experience. They influence a person's perception of their environment, behavior, and social interactions. In this study, we tested an innovative emotion classification approach using electroencephalography (EEG). In particular, we investigated the capabilities of brain connectivity analysis methods for recognizing and understanding emotions. We used the Granger causality connectivity metrics, which estimate the directional connectivity between brain regions of individual electrodes. The computed connectivity values for each electrode pair were used as features for classifying emotions. The proposed method was tested on four datasets. Finally, we showed a method for identifying characteristic differences in brain connectivity for different emotions, which can contribute to future neuroscience research.

KEYWORDS

EEG, Emotions, Connectivity matrices, Granger causality, Feature vectors, Neural networks, Classification

1 INTRODUCTION

The ability to accurately classify emotional states from electroencephalography (EEG) data has significant implications for mental health monitoring, and adaptive human-computer interfaces [3]. However, the complex and dynamic nature of emotional responses poses a challenge for conventional classification methods. Over the past two decades, emotion processing using EEG has gained significant attention [16]. EEG is a method that measures brain activity using the electrical potentials of electrodes attached to the scalp [13]. EEG system electronics record electrode potentials, which influence the voltage between different electrodes, providing information on neural activity [3]. To learn more about this topic, we recommend reading the book Analyzing Neural Time Series Data: Theory and Practice (Issues in Clinical and Cognitive Neuropsychology) [4]. The motivating force behind this research stems from the desire to improve emotion classification and its explainability and apply them to practical settings. For example, understanding how viewers emotionally respond to advertisements can help craft more effective marketing strategies. By using EEG and eye movement technology, researchers can analyze subjects' brain activity and eye movements while viewing advertisements, thereby assessing the effectiveness of specific content [15]. The present research focuses on connectivity features within the brain, which, when plotted, visually present the strength and direction of connectivity between brain regions. This approach helps to elucidate the dynamics of the brain neural network involved in emotional processing and provides an estimation of how much influence activity at one electrode has on activity in the region of another electrode [8]. By examining connectivity features, we aim to develop a method that can precisely recognize different emotional states.

2 METHODS

2.1 MATLAB and EEGLAB

For this study, we utilized MATLAB and EEGLAB. MATLAB is a high-level programming language and inter- active environment that is widely used in engineering, scientific research, and applied mathematics. Developed by MathWorks, MATLAB is designed for numerical computation, data analysis, algorithm development, and visualization. EEGLAB is an open-source MATLAB toolbox designed for the analysis of electroencephalography (EEG) data. It provides a comprehensive environment for processing and visualizing EEG data, data importing, and advanced analysis techniques such as independent component analysis (ICA).

2.2 Connectivity Measurements and Granger Causality

In neuroscience, connectivity measurement involves techniques that determine how different brain regions communicate and interact. In our study, we employed Granger causality to compute connectivity matrices that depict the directional interactions between signals at different EEG electrodes, providing insights into the dynamic connectivity of brain regions. Granger causality is based on the principle that if a time series X can improve the prediction of some other signal Y in comparison to predicting signal Y from its own historical data only, then there must be an information flow from the data sources of X to the observed processes in Y [7]. However, it is important to clarify that this does not imply a causal relationship in which X directly influences Y; rather, it indicates that sources of activity in X are also involved in affecting Y. To compute GC, two Vector Autoregressive (VAR) models are employed: Univariate VAR predicts the target time series using only its own past values, whereas the bivariate VAR uses the past values of both the target and potentially causal time series [11]. The reduction of error when using the bivariate model in comparison to the univariate model indicates the level of the electrodes' causal relationship. GC is computed for each electrode pair in both directions.

2.3 Machine Learning and Classification Learner Tool

Machine learning is a field of study focused on developing statistical algorithms that can learn patterns from data and make predictions or decisions based on such learning [6]. Central to machine learning is the concept of features, which are measurable properties or characteristics extracted from samples relevant to the target task. In the context of this study, feature vectors represent structured sets of features derived from the data samples. These feature vectors serve as inputs to the classification process, where the goal is to enable the classification of previously unseen data into predefined classes. We have employed several classification models provided by the Matlab classification learner tool, all with their default parameters. For the validation scheme, we used 5-fold cross-validation; that is, we divided the dataset into 5 equal-sized folds.

2.4 Neural Networks and Neural Network Classification

Artificial neural networks are machine learning models that make decisions like the human brain, utilizing layers of interconnected neurons to model and recognize complex patterns in data [2]. They can handle various tasks, including regression, clustering, and classification. For neural network classification, we designed our own artificial neural network model using the Deep Network Designer in MATLAB, which provides an intuitive and interactive user interface for creating deep learning networks. To better understand the decision process which is the internal mechanism by which the network arrives at a specific classification decision, we used gradient-based saliency maps. A saliency map highlights the most important features of the input data that influence the model's output [12]. It essentially shows where the model is "looking" or focusing on its attention. A gradient map, in particular, is a specific type of saliency map that uses gradients to determine the importance of the input features. By computing the gradient of the output with respect to the input features, we can identify which input features have the most significant impact on the output [12].

2.5 Statistical Analysis of Feature Importance

In addition to neural networks, statistical analysis plays a crucial role in model interpretation. One common feature selection method is the minimum redundancy maximum relevance (MRMR) method. MRMR is used to select features that are highly relevant to the target variable while ensuring that the selected features are minimally redundant with each other [14]. This method helps identify features that have the most influence on the model's predictions, thereby providing insights into the decision-making process of the model.

3 DATASETS

In this study, we used four different datasets. The data were obtained from several sources to ensure a comprehensive representation of neural activity across various contexts. In the following section, we will explain how the experiments were conducted for each dataset. The datasets were organized based on their chronological order within the research timeline rather than their effectiveness or the significance of the results.

(1) **Dataset 1: EMOTIONS 1** - The data was obtained from 21 participants. They were presented with pictures in 20 repeated blocks. Each block included carefully selected pictures that induced four different emotions ('veselje' (happiness), 'strah' (fear), 'gnus' (disgust), and 'nevtralno' (neutral)), presented in a specific sequence. To minimize the impact of the sequence on brain activity, the order of categories was altered for every 10 participants. Pictures within each category were presented for 1 s, with 2-5 s pauses, and a fixation cross was displayed in the center of the screen. The number of EEG signals was 32, with a duration of a few seconds. (Source: UL MF [1])

(2) **Dataset 2: EEG MOTOR MOVEMENT/IMAGERY DATASET** - The data was obtained from 109 volunteers who performed different motor and imaging tasks, and 64-channel EEG data was recorded using the BCI2000 system. Each subject performed two 1-minute baseline runs (one with eyes open, one with eyes closed), in addition to the primary runs where motor and imaginary movements were analyzed. The number of EEG signals was 64, with a duration of a few seconds. (Source: BCI2000 [10])

(3) **Dataset 3: EMOTIONS 2** - Note that the data from this dataset came from the same source as the first dataset. This experiment was conducted as a follow-up to the initial study, with some modifications. The differences are that it was conducted on 50 participants and the display time of each image was extended by 5 seconds. The number of EEG signals was 32, with a duration of a few seconds. (Source: UL MF [1])

(4) **Dataset 4: SEED** - Fifteen subjects participated in the experiment, where they were shown 15 Chinese movie clips that caused positive, neutral, or negative emotions. The number of EEG signals was 62, with a duration of approximately 4 minutes. (Source: SEED [17][5]).

4 EXPERIMENTS AND RESULTS

The data required preprocessing to ensure that they were suitable for further computation. Preprocessing included steps such as filtering to remove noise and artifacts. We applied a bandpass filter to each dataset to remove frequencies below the low-pass threshold and above the high-pass threshold. This stage is crucial because it removes low-frequency drift and high-frequency noise, which can obscure the EEG signal of interest. After filtering, we performed resampling. It is important to note that the sampling frequency should be at least twice as high as the highest frequency present in the signals. This is known as a Nyquist-Shannon sampling theorem [9].

The following step was the epoching of the data. This process splits the data into several epochs, which are defined based on reference time points obtained from recording events such as the presentation of visual stimuli, auditory cues, and motor actions. Next, we calculated the connectivity matrices. For each epoch, the connectivity matrix was obtained using the Granger causality method. For each pair of brain regions, a mathematical model was used to determine the extent of this predictive relationship, resulting in a value in the connectivity matrix. Before classification, we transformed the obtained matrices into feature vectors. The transformation involved flattening or reshaping the matrix into a one-dimensional array, where each element of the array represents a specific feature. The last column in the obtained table contained the label for each epoch. Afterward, the data were tested using various classifiers. Finally, we used statistical and neural network analysis to investigate the patterns of brain function that enable the classification of emotions.

For each dataset, we applied the necessary methods and after the classification of the data we obtained the following results:

(1) **Dataset: EMOTIONS 1** - The highest accuracy was 26.2% obtained by Narrow Neural Network classifier across four classes. Therefore, the classification failed, as the obtained results are at the level of randomness.

(2) **EEG MOTOR MOVEMENT/IMAGERY DATASET** - The highest accuracy was 90.0% obtained by the Wide Neural Network classifier across two classes. The successful results proved that our method was correct.

(3) **Dataset: EMOTIONS 2** - The highest accuracy was 27.3% obtained by SVM classifier across four classes. Similarly, as with the first dataset, the classification was not successful.

(4) **Dataset: SEED** - results obtained for this dataset across three classes are in the following table:

Classifier	Results
Wide Neural Network	73.3%
Medium Neural Network	71.6%
Narrow Neural Network	71.0%
SVM	69.5%
Trilayered Neural Network	68.0%
Bagged Trees	67.9%
Bilayered Neural Network	67.4%
KNN	60.7%
Subspace KNN	60.7%
Fine Tree	50.7%
RUSBoosted Trees	49.6%
Kernel Naive Bayes	47.7%
Efficient Logistic Regression	45.5%
Quadratic Discriminant	33.3%

Table 1: Results of Different Classifiers - Dataset: SEED

From Table 4, we can observe that the model was able to successfully classify emotional states in dataset 4. It can also be noticed that neural network algorithms performed the best. Therefore, the purpose of further analysis was to determine whether such analysis could aid in understanding brain processes by identifying which connectivity pairs were significant for a certain emotion type. We performed the analysis of feature importance on dataset 4 using two different approaches: statistical and neural networks-based.

4.1 Statistical Analysis

For the statistical approach, we used the `fscmrmr` function available in MATLAB, which ranks features for classification using the minimum redundancy maximum relevance (MRMR) algorithm. The outputs of the `fscmrmr` function were used to create a matrix representing the importance scores of each feature, which could be further analyzed or visualized. To visualize the relationships between different features or electrodes in the dataset, we plotted an EEG connectivity map.

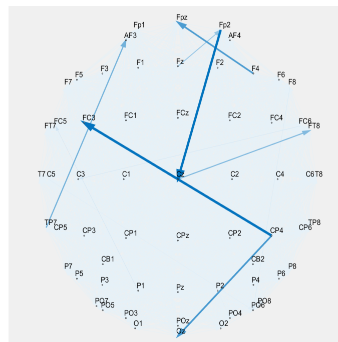

Figure 1: Feature Connection Map
The figure shows the importance of a feature (connectivity pair) as assessed by MRMR. Darker arrows indicate stronger connections.

The arrows represent the connections between electrodes, with the thickness or color indicating the importance or strength of the connection. This visualization can provide insights into the relationships between different features and help interpret the importance scores obtained from the MRMR function analysis. Therefore, we showed that using the statistical approach, we can visualize the relationship between electrodes. Because the features were obtained using Granger causality, we gained detailed insight into the direction of their influences.

4.2 Neural Network Analysis

In addition to classification performed by the Classification learner, we implemented a neural network classifier, which can be explained by plotting saliency maps. This enabled us to make a comparison between identifying feature importance statistically and as an explanation of classification models. We used a neural network model designed in the Deep Network Designer in MATLAB. After the training of the model, we plotted confusion matrices to evaluate the model's performance on both the test and training data. The confusion matrix for testing data demonstrated a validation accuracy of 72.4%. With an accuracy of 99.2%, the confusion matrix for training data achieved nearly perfect classification performance with very few misclassifications. The significant inconsistency in performance between training and testing suggests that the model overfitted to the training set, failing to generalize well to new data. Nevertheless, the findings are still congruent with the primary goal of our study, which was to demonstrate that using EEG technology allowed us to infer the direction of influences between electrodes, thereby providing deeper insight into emotion processing in the human brain. It is notable that the accuracy achieved using our model (72.4%) is comparable to the highest accuracy obtained with the classification learner tool (73.3%).

Finally, by completing neural network analysis, we once again showed that the importance of connectivity features for classification by the neural network can be visualized by plotting saliency map data as an EEG connectivity map.

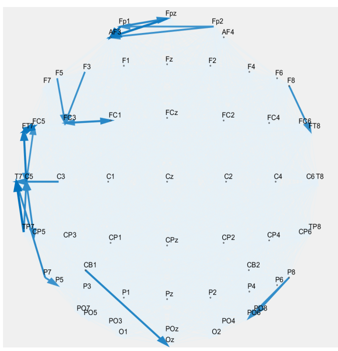

Figure 2: Feature connection map for average gradient map
The most significant connections in classification are indicated by dark blue arrows.

5 CONCLUSION

In this study, we demonstrated that emotions can be classified successfully based on EEG data using connectivity matrices. After testing the proposed method on four different datasets, we found that the highest classification accuracy was achieved with the fourth dataset 73.3%, while the lowest accuracy was observed with the first dataset 26.2%. The fourth dataset was obtained in response to movies rather than pictures, which took longer. This may explain why images did not influence distinguishable connectivity patterns. Additionally, we found that neural network classifiers generally performed best on the fourth dataset. After conducting further analysis on the fourth dataset using both statistical methods and neural networks, we were able to visualize which features had the greatest influence and determine the direction of their impact. We extracted explanations of which connectivity pairs are more important for distinguishing emotions, providing valuable insights that could be used in neuroscientific studies. The ability to classify emotions accurately from EEG signals and visualize the direction of feature influences opens doors for various practical applications, such as real-time emotion recognition systems and adaptive human-computer interfaces. Moreover, this finding leaves space for improvements in emotion classification by exploring alternative methods, such as the complex Pearson correlation coefficient (CPCC). Although this method may provide robust results, when combined with the Granger causality method crucial for its ability to reveal the directional influence between pairs of electrodes, as demonstrated in our study, it could significantly enhance our understanding of neural dynamics in emotion processing.

ACKNOWLEDGMENTS

We would also like to extend our gratitude to the Institute of Pharmacology and Experimental Toxicology, Faculty of Medicine, University of Ljubljana, for generously providing us with the EMOTIONS 1 and EMOTIONS 2 datasets.

REFERENCES

[1] [n. d.]. *Passive viewing of emotion-evoking pictures: Multimodal application of functional near-infrared spectroscopy and electroencephalography.* Master's thesis. Institute of Pharmacology and Experimental Toxicology, Faculty of Medicine, University of Ljubljana. Data collected and pre-processed within the framework of the Master's thesis.
[2] Charu C Aggarwal et al. 2018. *Neural networks and deep learning.* Vol. 10. Springer.
[3] K. Blinowska and P. Durka. 2006. Electroencephalography (EEG). (2006). https://doi.org/10.1002/9780471740360.ebs0418
[4] Mike X Cohen. 2014. *Analyzing neural time series data: theory and practice.* MIT press.
[5] Ruo-Nan Duan, Jia-Yi Zhu, and Bao-Liang Lu. 2013. Differential entropy feature for EEG-based emotion classification. In *6th International IEEE/EMBS Conference on Neural Engineering (NER).* IEEE, 81–84.
[6] Issam El Naqa and Martin J Murphy. 2015. *What is machine learning?* Springer.
[7] Katerina Hlaváčková-Schindler, Milan Paluš, Martin Vejmelka, and Joydeep Bhattacharya. 2007. Causality detection based on information-theoretic approaches in time series analysis. *Physics Reports* 441, 1 (2007), 1–46.
[8] E. W. Lang, A. M. Tomé, I. R. Keck, J. M. Górriz-Sáez, and C. G. Puntonet. 2012. Brain Connectivity Analysis: A short survey. *Computational Intelligence and Neuroscience* 2012 (2012), 1–21. https://doi.org/10.1155/2012/412512
[9] Emiel Por, Maaike van Kooten, and Vanja Sarkovic. 2019. Nyquist–Shannon sampling theorem. *Leiden University* 1, 1 (2019), 5.
[10] Gerwin Schalk, Dennis McFarland, Thilo Hinterberger, Niels Birbaumer, and Jonathan Wolpaw. 2004. BCI2000: A General-Purpose Brain-Computer Interface (BCI) System. *IEEE Transactions on Biomedical Engineering* 51, 6 (2004), 1034–1043. https://doi.org/10.1109/TBME.2004.827072
[11] Anil Seth. 2007. Granger causality. *Scholarpedia* 2, 7 (2007), 1667.
[12] Karen Simonyan, Andrea Vedaldi, and Andrew Zisserman. 2013. Deep inside convolutional networks: Visualising image classification models and saliency maps. *arXiv preprint arXiv: 1312.6034* (2013). https://doi.org/10.48550/arxiv.1312.6034
[13] Michal Teplan et al. 2002. Fundamentals of EEG measurement. *Measurement science review* 2, 2 (2002), 1–11.
[14] The MathWorks Inc. 2022. MATLAB. Version R2022a. https://www.mathworks.com/
[15] L. Wang. 2019. Test and evaluation of advertising effect based on EEG and eye tracker. *Translational Neuroscience* 10, 1 (2019), 14–18. https://doi.org/10.1515/tnsci-2019-0003
[16] X. Wang, D. Nie, and B. Lu. 2014. Emotional state classification from EEG data using machine learning approach. *Neurocomputing* 129 (2014), 94–106.
[17] W. Zheng, W. Liu, Y. Lu, B. Lu, and A. Cichocki. 2018. EmotionMeter: A Multimodal Framework for Recognizing Human Emotions. *IEEE Transactions on Cybernetics* (2018), 1–13. https://doi.org/10.1109/TCYB.2018.2797176