Using EEG Effective Connectivity Based on Granger Causality and Directed Transfer Function for Emotion Recognition

Weisong Wang, Wenjing Sun*
School of Marxism, XinJiang Normal University, WuLuMuQi 830017, Xinjiang, China
School of Marxism, Guangxi Science and Technology Normal University, LaiBin 546199, Guangxi, China

Abstract—Emotion is a complex phenomenon that originates from everyday issues and has significant effects on individual decisions. Electroencephalography (EEG) is one of the widely used tools in examining the neural correlates of emotions. In this research, two concepts of Granger causality and directional transfer function were utilized to analyze EEG data recorded from 36 healthy volunteers in positive, negative and neutral emotional states and determine the effective connectivity between different brain sources (obtained through independent component analysis). Shannon entropy was utilized to sort the brain sources obtained by the ICA method, and average topography helps to add spatial information to the proposed connectivity models. According to the obtained confusion matrix, our method yielded an overall accuracy of 75% in recognizing three emotional states. Positive emotion was recognized with the highest accuracy of 87.96% (precision = 0.78, recall = 0.78 and F1-score = 0.81), followed by neutral (accuracy = 82.41%) and negative (accuracy = 79.63%) emotions. Indeed, our proposed method achieved the highest recognition accuracy for positive emotion. The proposed model in the present study has the ability to identify emotions in a completely personalized way based on neurobiological data. In the future, the proposed approach in the present study can be integrated with machine learning and neural network methods.

Keywords—EEG; effective connectivity; granger causality; directed transfer function; emotion recognition

I. INTRODUCTION

Emotion is a complex phenomenon that originates from everyday issues and has significant effects on individual decisions. The extent of these decisions can affect the personal and social life of the people of society [1]. Emotions are very important in learning and communicating, and their expression plays a big role in human relationships. Emotions can directly affect a person's performance, and therefore, it is important to try to understand their sources and control them [2]. A certain emotion is one of the brain states that is produced by the electrical activity of millions of neurons and is associated with physiological changes in the whole body [3]. When experiencing emotions, the brain's left and right hemispheres are involved in different ways [4]. For example, it has been shown that the left prefrontal area is more involved in dependent emotional reactions and the right prefrontal area is involved in withdrawal reactions [5]. Generally, it is hypothesized that the left hemisphere is dominant in positive emotions, and the right hemisphere is dominant in negative emotions [6].

Electroencephalography (EEG) is one of the widely used tools in examining the neural correlates of emotions [7]. This instrument has many applications in cognitive neuroscience and psychiatry and improves our insight into various behavioral phenomena including depression [8], bipolar disorder [9-17] and hyperactivity [18-22]. Many researchers have applied various machine learning techniques to EEG data to develop an automated EEG-based emotion recognition system [23]. Zhang et al. applied the Empirical Mode Decomposition (EMD) technique to EEGs, calculated the sample entropy from the first four IMFs, and reported an average accuracy of 93.20% in classifying five distinct emotions [24]. Zheng and Lou trained deep belief networks by calculating entropy features from different EEG frequency bands and achieved an accuracy of 86.65% in detecting three negative, neutral and positive emotions [25]. Meng et al. proposed an EEG-based emotion recognition system based on entropy features and cascaded convolutional recurrent neural networks and achieved an accuracy of 94.85% in valence-based classification problems [26]. Hwang et al. utilized convolutional neural networks along with generating topology-keeping differential entropy features to prevent the loss of localized information and achieved an average accuracy of 90% in recognizing three negative, positive and neutral emotions [27]. Nawaz et al. proposed an EEG-based emotion recognition system based on various linear and nonlinear features such as fractal dimension and wavelet energy, feature selection by principle component analysis, and multiple classifiers such as SVM and KNN, and reported an average accuracy of 77.60%, 78.96% and 77.62% in detecting dominance, arousal and valence emotions, respectively [28]. Two recent reviews on EEG-based emotion recognition emphasized the importance of such systems for various scientific fields, such as psychology and cognitive neuroscience, and highlighted the necessity of developing these systems with a special focus on regional brain connectivity [29, 30].

In the last two decades, neuroscience studies have focused on the connections of different areas of the cerebral cortex and how these areas interact when performing a specific sensory-motor or cognitive action to better understand brain function [31]. Many attempts have been made to quantify these connections through EEG analysis [32]. Functional
connectivity is an observable phenomenon that can be quantified by measuring statistical dependencies such as correlation or transfer entropy [33]. On the other hand, effective connectivity refers to the model parameters that attempt to explain the observed dependencies (functional connectivity). From this point of view, effective connectivity refers to the concept of coupling or direct causal influence by examining the direction of information dissemination [34, 35]. Bagherzadeh et al. achieved a high accuracy of 99% in the classification of five emotional states using effective EEG connectivity and convolutional neural networks [36]. In another study, Bagherzadeh et al. used EEG frequency effective connectivity maps based on the transfer learning technique and achieved an accuracy of 95% in the classification of five emotional states [37]. Although these studies have shown the high potential of EEG effective connectivity as a biomarker for human emotion recognition, few studies have used this important feature of brain signals to develop EEG-based emotion recognition systems.

Therefore, in this research, we aim to process EEG signals to estimate effective connectivity in brain resources and provide models to investigate how brain resources influence each other in order to detect three emotional states: neutral, negative, and positive. For this purpose, we investigated the potential of Granger causality and directed transfer function to estimate effective cortical connectivity in different emotions. Our research showed that these approaches can reveal noticeable effective connectivity patterns in each emotion, which can be classified in order to emotion recognition.

II. METHODS

Fig. 1 shows the work process of the current research. Considering that in this research, two concepts of Granger causality and directional transfer function are used in EEG data processing, it is necessary to briefly introduce these concepts first.

A. Granger Causality

Granger causality is a technique to derive specific kinds of causal dependence among stochastic samples by reducing the bias of predicting possible effects if past observations of the hypothesized causes are utilized to anticipate the effects plus previous observations of the possible effects. This concept was first proposed through Wiener and then reformulated through Granger based on linear autoregressive models. This algorithm considers two assumptions: (I) a cause should precede its effects, and (II) data on the past of a cause should enhance the anticipation of the effects above and beyond data on the collective past of the other observed samples. To estimate influences from the channel $x_i$ to $x_j$ for $n$ channel autoregressive processes, we considered $n$ and $n-1$ multivariate autoregressive models. The model is fitted to overall $n$-channel system, resulting in the residual variance $V_{i,n}(t) = \text{var}(E_{i,n}(t))$ for time series $x_i$. Then, a $n-1$ multivariate model is fitted for $n-1$ channels, except for channel $j$, which results in the residual variance $V_{i,n-1}(t) = \text{var}(E_{i,n-1}(t))$. So, Granger causality is given by:

$$Granger_{j\rightarrow i}(t) = \ln \left( \frac{V_{i,n}(t)}{V_{i,n-1}(t)} \right)$$  \hspace{1cm} (1)

B. Directed Transfer Function

This method is used to measure the direction and frequency content of brain activities. The directed transfer function is a multivariate approach that is formulated by multivariate autoregressive models to investigate the causal relationships between EEG channels and recognize the directed dissemination of EEG activities. This algorithm works in the frequency domain to characterize regional connectivity based on the factorization of the coherence between two EEG channels.

C. Dataset

In this research, we utilized an EEG dataset on neutral, positive, and negative emotional states from [38]. This dataset includes EEG signals recorded from 36 young adults in the neutral, positive, and negative emotional states induced by standard images from the International Affective Picture System (IAPS). 16 Ag/AgCl electrodes were utilized to capture brain signals at a sampling rate of 512 Hz based on a 10-20 international recording protocol.

![Flow chart of the proposed method for EEG-based emotion recognition based on effective connectivity.](image-url)
D. Data Preprocessing

First, we applied a Butterworth band-pass filter (order 4) with a frequency of 1–60 Hz to the signals to cancel unwanted noise. A 50 Hz notch filter was then applied to the signals to cancel power line interference. Next, an experienced neurologist checked all EEGs and rejected all parts of the signals contaminated with various artifacts, such as body motion and eye blinking. After obtaining a clean EEG signal for each subject, independent component analysis (ICA) was applied to the signals to decompose them into independent components. A fast ICA algorithm was utilized in this work to decompose 16 source components from EEG channels. However, we encountered a bad situation when using ICA because the order of the extracted components changes in every run of the algorithm. To solve this problem, Shannon's entropy of sources was used as a measure of information to sort brain sources. Sources are sorted according to Shannon entropy value and in ascending order. These brain sources obtained from ICA were used for further analysis.

E. Effective Connectivity Estimation

EEGLAB and SIFT toolbox were utilized to estimate effective connectivity [39]. The first step to estimate the connectivity value is to fit a model to the data using Adaptive Multivariate Autoregressive (AMVAR) modeling, which is acceptable for this purpose. To adapt this model, the length of the window, the step of the window, and the order of the model should be selected. After selecting the values of these parameters, the accuracy of the fitted model is evaluated. Table I shows the selected values for these parameters.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Window length (second)</td>
<td>5</td>
</tr>
<tr>
<td>Window step (second)</td>
<td>1</td>
</tr>
<tr>
<td>Model order</td>
<td>10</td>
</tr>
</tbody>
</table>

It should be noted that the higher order of the model causes complexity, and the short length of the window leads to an increase in calculations. As a result, a trade-off should be made in choosing these parameters. If the model is sufficiently adapted to the data, the residual coefficients of the model should be small and uncorrelated relative to the real data. The presence of correlation in the residuals indicates the presence of correlated structures in the data that the model is unable to provide. To solve this issue, a null hypothesis with a significance level is considered, and the model is evaluated through it. To evaluate the whiteness of the residuals in this research, the autocorrelation function method with a significance level of 90% was utilized, as well as the consistency assessment through the percent consistency [40]. The stability of the model was evaluated with the stability index. A VAR model is stable if the augmented coefficient matrix of all stability indices is less than one [41].

Granger causality and direct transfer function methods were used to estimate the effective connectivity. In this study, effective connectivity was estimated in four EEG frequency bands: delta (1–4 Hz), theta (4–8 Hz), alpha (8–13 Hz), and beta (13–30 Hz). In fact, for six sources sorted by Shannon entropy, effective connectivity was estimated in these frequency bands.

F. Performance Evaluation

To evaluate our proposed approach for emotion recognition, we utilized various classification metrics, including accuracy, precision, recall and F1-score. These metrics are given by:

\[
\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \tag{2}
\]

\[
\text{Precision} = \frac{TP}{TP + FP} \tag{3}
\]

\[
\text{Recall} = \frac{TP}{TP + FN} \tag{4}
\]

\[
F1 = \frac{2 \times \text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \tag{5}
\]

where TP, TN, FP and FN denote true positive, true negative, false positive and false negative elements for a specified emotional class obtained from the confusion matrix.

III. RESULTS

First, the accuracy, stability, and consistency of the model were checked using the mentioned criteria. The results showed that the stability of all models was fully established, and the average model stability in all samples was about 78%. The average rejection of the null hypothesis was 93%. For each emotional state, the maximum number of connections in each frequency band was calculated. In each sample, a source can receive information from other sources. Fig. 2 shows a schematic of the calculation of the connectivity between sources as a sender or receiver of information.

Based on the schematic shown in Fig. 2, the sources with the most receiving and sending information in three emotional states were determined based on averaging in different frequency bands as shown in Table II.

![Fig. 2](image-url)
models, the information related to the location of brain sources is not included, so they cannot be used to identify the location of connections. To solve this problem, topographic images of sources were used.

<table>
<thead>
<tr>
<th>Emotional state</th>
<th>Sending sources</th>
<th>Receiving sources</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Source</td>
<td>Percentage</td>
</tr>
<tr>
<td>Neutral</td>
<td>5</td>
<td>54.2</td>
</tr>
<tr>
<td></td>
<td>7</td>
<td>48.4</td>
</tr>
<tr>
<td>Negative</td>
<td>6</td>
<td>47.1</td>
</tr>
<tr>
<td></td>
<td>10</td>
<td>44.3</td>
</tr>
<tr>
<td>Positive</td>
<td>10</td>
<td>66.2</td>
</tr>
<tr>
<td></td>
<td>9</td>
<td>43.1</td>
</tr>
</tbody>
</table>

In this model, each circle represents a brain source, which is assigned a number based on Shannon entropy ranking. Arrows go from the source of the information transmitter to the source of the information receiver. The larger the diameter of an arrow, the greater the amount of information spread.

In addition to separating different sources from each other, ICA provides us with information about the location of these sources on the scalp by producing images called the topography of a source. These images are important in that they can show how a source spatially affects the entire scalp. As a result, by considering the assumptions about the location that is activated during specific stimuli, the location of their sources can be identified. Accordingly, the topography of the sources whose models were created in the previous section was determined, and their average topography was examined in all samples. It is worth mentioning that the selected sources are the same in terms of entropy. The topography of source 10 in the positive emotion belongs to source 10 in terms of entropy in all samples. Table II and Fig. 3 specify the sources with the most output and input connections in each emotion. Based on this, Fig. 4 shows the average topography of these sources for positive, negative, and neutral emotional states.

The images in Fig. 4 contain spatial information about brain sources. By combining these images with the information of the models from Fig. 3, other new models are presented that also contain spatial information about brain sources. Fig. 5 shows the presented model for effective connectivity between brain sources in positive, negative, and neutral emotional states with spatial information.

---

**Fig. 3.** Proposed model based on Granger causality and directed transfer function for effective connectivity for (A) positive, (B) neutral, and (C) negative emotional states.

**Fig. 4.** Average topography images of 36 samples for sources with the most input and output connections, as specified in Fig. 3, extracted with independent component analysis, in (A) positive, (B) negative, and (C) neutral emotional states.
In this model, each circle represents a brain source, which is assigned a number based on Shannon entropy ranking. Arrows go from the source of the information transmitter to the source of the information receiver. The larger the diameter of an arrow, the greater the amount of information spread.

The models presented in Fig. 5 are the average of all samples. In the next step, similarity measurement and labeling operations were performed for each sample. In other words, each individual’s connectivity model was compared with the average models presented in Fig. 5 and labeled based on its similarity to each of the average models. The total number of samples is 36 subjects with three emotional states: positive, negative, and neutral. Therefore, there are 108 models with similar distribution in different emotional states. To measure the similarity of each sample in all emotional states and frequency bands, first, the average input and output information of all sources of that sample was calculated. The rate of receiving and sending information from all sources was determined for each sample. Then, in each sample, two sources that had the highest amount of receiving and sending information were selected. In the next step, the emotional state label of 108 samples was removed, and the samples were shuffled. Two selected sources from each sample were compared with the proposed models and given a label according to their similarity to each of the models as shown in Table III.

### Table III. The Result of Labeling Samples for Positive, Negative, and Neutral Emotional States Based on the Proposed Effective Connectivity Models

<table>
<thead>
<tr>
<th>Emotional states</th>
<th>Positive</th>
<th>Neutral</th>
<th>Negative</th>
</tr>
</thead>
<tbody>
<tr>
<td>Real label</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Positive</td>
<td>28</td>
<td>3</td>
<td>5</td>
</tr>
<tr>
<td>Neutral</td>
<td>2</td>
<td>26</td>
<td>8</td>
</tr>
<tr>
<td>Negative</td>
<td>3</td>
<td>6</td>
<td>27</td>
</tr>
</tbody>
</table>

Table III shows the confusion matrix obtained by the proposed approach based on effective connectivity to identify positive, negative, and neutral emotions. According to the obtained confusion matrix, our method yielded an overall accuracy of 75% in recognizing three emotional states. Table IV shows the obtained classification criteria for each class. As shown, positive emotion was recognized with the highest accuracy of 87.96% (precision = 0.78, recall = 0.78, and F1-score = 0.81), followed by neutral (accuracy = 82.41%) and negative (accuracy = 79.63%) emotions.

### Table IV. Calculated Classification Criteria for Each Positive, Negative, and Neutral Emotional State

<table>
<thead>
<tr>
<th>Class</th>
<th>n (truth)</th>
<th>Accuracy (%)</th>
<th>Precision</th>
<th>Recall</th>
<th>F1-score</th>
</tr>
</thead>
<tbody>
<tr>
<td>Positive</td>
<td>33</td>
<td>87.96</td>
<td>0.78</td>
<td>0.85</td>
<td>0.81</td>
</tr>
<tr>
<td>Neutral</td>
<td>35</td>
<td>82.41</td>
<td>0.72</td>
<td>0.74</td>
<td>0.73</td>
</tr>
<tr>
<td>Negative</td>
<td>40</td>
<td>79.63</td>
<td>0.75</td>
<td>0.68</td>
<td>0.71</td>
</tr>
</tbody>
</table>

### IV. Discussion

The aim of this study was to estimate cortical effective connectivity from the EEG signal of emotions based on Granger causality and directed transfer function for the recognition of different human emotions. Our proposed method achieved the highest recognition accuracy for positive emotion. In the proposed model for positive emotion, source 10 had the highest amount of receiving information from other sources. This finding is consistent [38], where source 10 achieved the highest recognition accuracy for positive emotion. In the proposed model for negative emotion, source 8 had the highest amount of sending information to other sources. Abdolssalehi et al. also achieved the highest recognition accuracy for negative emotion through source 8 [38]. However, in [38], the authors relied on recurrence quantification analysis for pattern recognition from brain sources.

### Table V. Comparing the Findings of the Present Study with Previous Studies

<table>
<thead>
<tr>
<th>Reference</th>
<th>Algorithm</th>
<th>Results</th>
</tr>
</thead>
<tbody>
<tr>
<td>[42]</td>
<td>Spectral features and SVM classifier</td>
<td>49.4% like, 55.7% arousal, 58.5% valence</td>
</tr>
<tr>
<td>[43]</td>
<td>Spectral features and SVM classifier</td>
<td>79.59% sadness, 74.11% anger, 86.15% joy, 83.59% pleasure</td>
</tr>
<tr>
<td>[44]</td>
<td>Spectral and statistical features and LDA classifier</td>
<td>62% anxiety, 50% engagement, 57% boredom</td>
</tr>
<tr>
<td>[45]</td>
<td>Spectral asymmetry index, wavelet entropy and SVM classifier</td>
<td>82.5% negative excitement, 64% neutral</td>
</tr>
<tr>
<td>[46]</td>
<td>Spectral features, Gabor transform and probabilistic neural network</td>
<td>62.97% sadness, 69.74% disgust, 73.64% anger, 56.79% fear</td>
</tr>
<tr>
<td>[47]</td>
<td>Spectral features and SVM classifier</td>
<td>50.5% valence, 62.1% arousal</td>
</tr>
<tr>
<td>[48]</td>
<td>Spectral features and SVM classifier</td>
<td>61% sadness, 58% fear, 53% anger, 51% joy</td>
</tr>
<tr>
<td>Our proposed approach</td>
<td>Effective connectivity and confusion matrix</td>
<td>87.96% positive, 82.41% neutral, 79.63% negative</td>
</tr>
</tbody>
</table>

In contrast, we utilized the effective connectivity between these sources and improved the results obtained in
Abdolssalehi et al.’s work. In addition, as mentioned, sources 10 and 8 had the highest amount of information exchange in positive and negative emotional states, respectively. The average topography of sources 10 and 8 of all samples was located in the left posterior hemisphere and the right posterior hemisphere, respectively. This finding is consistent with the valence hypothesis, which assumes the opposite predominance of the right hemisphere for negative emotions and the left hemisphere for positive emotions [49, 50]. Table V compares the findings of the present study with previous studies. As can be seen, our proposed system performs better than most of the previous methods.

Most of the studies that have tried to recognize human emotions using EEG signals have used a variety of machine learning methods and artificial neural networks that provide us with a black box-like function [51-53]. Although some of these studies have reported very high detection accuracies, the process of pattern recognition in them is unclear and ambiguous [54]. However, in this research, we tried to choose and propose a transparent work process to overcome this important limitation of previous studies. Although we did not use learning machines and artificial neural networks for classification and pattern recognition, the obtained results are quite promising. Therefore, our proposed model can be very useful in various research fields, such as psychiatry, psychology, neuroscience, and cognitive science. The important thing to consider about our proposed model is that this model can work completely depending on a person, and it is a personalized model. Individual and cultural differences are very important issues in the development of emotion recognition systems. At the same time, most previous techniques ignore this important issue. However, the proposed model in the present study has the ability to identify emotions in a completely personalized way based on neurobiological data. In the future, the proposed approach in the present study can be integrated with machine learning and neural network methods and improve the proposed model.

V. CONCLUSION

In this study, a new method based on effective connectivity using Granger causality and directed transfer function was able to successfully extract connectivity patterns related to positive, negative and neutral emotions and led to the detection of these emotions based on EEG signal analysis. The obvious advantage of our method is its transparency in all stages of analysis and not using black boxes related to machine learning and neural networks, which increases its clinical applicability compared to previous works. However, this method needs further validation using different databases. In addition, our proposed method should be integrated with an artificial intelligence system to automate emotion recognition. In this study, only three emotions were investigated, and future studies should evaluate this method for other emotions as well.

ACKNOWLEDGMENT

This work was supported by 2022 Guangxi University Young and Middle aged Teachers’ Basic Research Ability Improvement Project: Research and Practice of Applied Undergraduate Colleges Serving Local Economic Development - Taking Guangxi Normal University of Science and Technology as an Example (No.: 2022KY0847) and Guangxi Education Science “14th Five Year Plan” 2022, Guangxi Financial Literacy Education Research Special Project: Research and Practice on Financial Literacy Education for College Students in Ethnic Regions of Guangxi (No.: 2022ZJY2595).

REFERENCES


