Crucial Processing of Detect Depression with EEG

Xingyu Xiang
College of Life Sciences, Sichuan University, Sichuan, China
2020141240172@stu.scu.edu.cn

Abstract. Depression is a common mental disease that causes varying degrees of social impact. With the COVID-19 pandemic increasing the prevalence of depression, researchers are trying to develop better methods to diagnose and treat this disease. EEG, an emerging technology belonging to Brain-Computer Interface has been widely used in mental illnesses like Parkinson's and epilepsy monitoring. In recent years, many attempts have been made to apply EEG technology to the field of depression research. In this review, we focus on some major methods utilized in Depression detection with EEG technology as well as machine learning. We will talk about some mainstream methods in preprocessing, feature extraction, and classification, their features, advantages, drawbacks, and the parameters preferred in depression detection. Provide an overview for relevant researchers to promote their studies. Finally, the pipeline's drawbacks and defects, what can we do, how to make it better, and its potential in relative areas are discussed.

Keywords: EEG; Depression; Machine Learning.

1. Introduction

Depression is a common mental disorder around the world. By 2021, 5.0% of adults have been suffering from depression worldwide [1]. With depression comes a series of social hazards, high suicide rates, and reduced social vitality. The epidemic of COVID-19 exacerbates this trend, social restrictions, lockdowns, school, and business closures, loss of livelihood, decreases in economic activity, and shifting priorities of governments in their attempt to control COVID-19 outbreaks all have the potential to substantially affect the mental health of the population [2, 3].

Today's most common way of diagnosing depression is by questionnaire. Doctors decide on the result of the questionnaire and communicate with patients. It is a time-consuming and subjective process. Since 1980, treatment of depression has improved, while the incidence of depression has not reduced. Traditional inaccurate and untimely diagnoses are factors for this phenomenon[4].

Brain-Computer Interface (BCI) is a system to monitor brain activities and convert or decode the brain signal to an intuitive result. Electroencephalography (EEG) is one non-invasive way of BCI. It now has been deeply and maturely studied with EEG technology in mental illnesses such as Parkinson's Disease and epilepsy[5, 6]. For this kind of unintuitive data, artificial intelligence can help us make good use of them. The EEG-AI system, its accuracy, and its high efficiency could help us to solve the problem of depression diagnosis.

To use EEG signals we should first preprocess them, do the feature extraction, then import the processed data to the algorithm and get the result. The way of preprocessing is relatively rigid, including artifacts removal, frequency filtering, and channel selection. The preprocessing could help us to filter the fake signs and extract the most valuable part of the signals, which reduces computation and increases accuracy[7, 8].

The choice of feature extraction is varied, either linear or non-linear. The kind of feature extraction varies with the subsequent algorithms.

The selection of machine learning has gone through many attempts. There are some traditional classifiers in pattern recognition like support vector machine (SVM), k-nearest neighbor (KNN), and decision tree (DT). Or neuro networks in artificial intelligence like CNN and RNN[9]. These frameworks could give accurate depression diagnosis results with an accuracy of over 70%[10-13].

In this paper, I would talk about the basic pipeline in this area, its main structure for research, and how to choose parameters and tools preferred in depression detection. It could help readers get started quickly in this field. Details on how to perform certain algorithms in relative works can be found in cited papers.

2. Stages and Methods

2.1 Preprocessing

2.1.1 Artifacts Removal

In the natural state, the EEG voltage on the surface of the brain is only about 50μV. It causes the fact that EEG signals are vulnerable. It could be interrupted by EOG, ECG, sweating, or muscle movement. In this situation, correcting these deviations is important.

There are many ways to remove artifacts. First, with the given observed subject's unwanted behavior during experiments, we could remove the artifacts manually, and directly remove the data when the movement occurs. Then automatically remove the artifacts with algorithms, for example, analog methods, regression, adaptive filtering, BSS, frequency decomposition, or hybrid methods contain both of them[14]. For depression detection, BSS and ICA are commonly used.

2.1.2 Frequency Filtering

In EEG signals, not all frequency bands are meaningful. At the same time, the interference caused by EEG hardware is generally at 50Hz or 60Hz. Thus, we need to do filtering to reduce computation and increase accuracy. In depression detection, we tend to utilize frequency bands between 4 to 45 Hz, similar to that of emotions recognition. Between this range, signals are divided into alpha wave, beta wave, theta wave, and gamma wave.

Alpha waves occur in the frequency range of 8 to12 Hz. It happens when a person is relaxed or in a resting state. Beta waves occur in the frequency range of 12 to 28 Hz. This frequency band increases when a person is in a state of stress. Theta waves occur in the frequency range of 4 to7 Hz. It can reflect that an adult is in a state of exhaustion. Gamma waves occur in the frequency band over 30Hz. When a person is concentrating on some complex tasks, it increases. It is worth mentioning that Delta waves, occurring between 1 to 3 Hz, mainly happen during sleeping time, which is the reason we filter off this frequency range.

2.1.3 Channel Choosing

In EEG signal studies, channel choosing in some areas like Motor Imagery (MI) is important. Because the brain is a spatial system, analyzing data with location information is required. However, in depression, few studies provide their electrode selections in paper[10, 15], suggesting this process is not a necessity. In experiments, we usually choose to delete bad electrodes, those electrodes with loud noise or whose impedance exceeds the threshold, preventing them from giving biased information.

Overall, the significance of preprocessing is to initially mine the value of the data, reduce the noise inside, and facilitate subsequent data analysis. In the case of different algorithms, some studies will selectively give up some preprocessing, such as not performing artifacts removal[16].

2.2 Feature Extraction

Except deep learning can extract features automatically, before entering data into the artificial intelligence algorithm model, we need to do some feature extraction first, which will provide some convenience for their subsequent classification.

Feature can be divided into linear features and non-linear features. Linear feature is easy to calculate, for instance, for a wave signal under specific time series, its mean as well as standard deviation, or its difference. Difference depicts a change between discrete quantities. Difference can be expressed by the following formula. x_t represents the time series. i could be 1 or 2, respectively calculate 1st difference and 2nd difference.

$$DF = \frac{1}{n-i}\sum_{t=1}^{n-i}(|x_{t+i} - x_t|) \tag{1}$$

It is difficult for linear features to contain all the information in EEG signals. Non-linear features include the time domain feature, frequency domain feature, and time-frequency domain feature. More details could be seen in [17], here we would like to talk about some commonly used features respectively in three of these.

2.2.1 Time Domain Feature

Fractal dimension is a time domain feature, it can measure the complexity of a signal. Besides Higuchi algorithm[18], there are Fractal Brownian Motion[19], Box-counting[20], and so on. Here is the calculation of Higuchi Fractal Dimension (HFD)

For a discrete time series $\{x_1, x_2, \ldots, x_n\}$, we can reform it as the following expression.

$$x_m^k = \left\{x_m, x_{m+k}, \ldots, x_{m+int(\frac{n-m}{k})\times k}\right\} \tag{2}$$

Thus, the length of this new sequence $L_m(k)$

$$L_m(k) = \frac{1}{k}\left\{\sum_{i=1}^{int(\frac{n-m}{k})}\left|x_{m+ik} - x_{m+(i-1)k}\right|\right\} \times \frac{N-1}{int(\frac{n-m}{k})} \tag{3}$$

Its mean length $L(k)$ should be as follow

$$L(k) = \frac{1}{k}\sum_{m=1}^{k} L_m(k) \tag{4}$$

Then we can get HFD

$$HFD = \frac{\ln(L(k))}{-\ln(k)} \tag{5}$$

2.2.2 Frequency Domain Feature

Fourier Transform is commonly used in signal processing as a frequency analyzing method. Discrete Fourier transform (DFT) could be calculated with Fast Fourier transform (FFT) with a faster speed.

For a limited time series x_t long M, we can transform it into the following form

$$X(k) = \sum_{n=0}^{N-1} x_n e^{-j\frac{2\pi}{N}}, 0 \leq k \leq N-1 \tag{6}$$

N is the length of the DFT transform interval, $N \geq M$. However, on most occasions, DFT is more suitable for statistically stationary signals, so we import the time window and we can have STFT as follow, $\omega(m\text{-}n)$ is the short-time analysis window.

$$X(n, \omega) = \sum_{n=-\infty}^{n=\infty} x_n(\omega(m-n))e^{-j\omega n} \tag{7}$$

2.2.3 Time-Frequency Domain Feature

Time information is lost in Fourier transformation. Wavelet transform can solve this defect. Discrete Wavelet Transform (DWT) decomposes the signal in different approximation and detail levels corresponding to different frequency ranges while conserving the time information of the signal. It works by repeating filtering and down-sampling signals. Details should be seen in [21]. The choice of wavelet will also bring some changes to the results.

To sum up, the purpose of feature extraction is to prevent the signal from being directly input into the classifier, which may cause the classifier to only pay attention to the superficial time domain characteristics of the signal and ignore its other characteristics. In addition to these methods above, there are many methods analyzed in digital signal processing and informatics, which can be used in EEG signal feature extraction. In practice, feature extraction can have many options. In depression detection it is hard to define which feature can best depict the signal, the features in different papers vary.

2.3 Machine Learning

Here comes the main part of depression detection. With the processed data above, we should use the power of computers to go further. EEG signals are always big data. Although feature extraction could reduce the amount of data, with multiple channels, it is still impossible to calculate manually. Machine learning can put these data to good use.

2.3.1 Pattern Recognition

For a set of data X, which includes features of an object, we could use pattern recognition to classify X. We can define the classifier as F, then we have the formula $Y = F(X)$. Here Y is the result after classification.

The classifier is divided according to its corresponding algorithm. Usually, we choose SVM, KNN, and DT.

SVM[22] is to upscale the data and then compute a hyperplane that classifies them appropriately. It is suitable for binary classification, here which means detecting whether depressed or not. In the classification problem, when linear inseparability is encountered, the kernel function is generally used to map it to a higher-dimensional space for classification. The selection of the kernel function directly affects the quality of the SVM classification performance. However, selecting the kernel function is based on personal experience. Thus, the SVM classifier still has great limitations.

KNN[23] is to calculate the nearest K nodes close to new sample x, the distances here could be Euclidean distance, Manhattan distance, and Minkowski distance. K is important to the result. If K is a small number, it means that only a few nodes around x would be learned. When the nodes are with loud noises, the result would not be charming. The advantage of KNN is its lower time complexity.

DT makes the decision by representing the relationship between features and categories. The final formed decision tree could be intuitive. But the drawback is that its algorithm structure is not stable, sometimes it can lead to overfitting. It doesn't perform well when data relationships get complicated.

Although the detail theory in the algorithms above could be complex, running them on a computer is not difficult. For example, by importing the *sklearn* module in Python, module functions will help us quickly implement these methods, we only need to input the data. None of the algorithms above is the best, still, we can hybrid them and try to make the best result.

2.3.2 Deep Learning

Compared with the traditional methods above, deep learning is special because it can automatically extract data features. Among a large number of deep learning frameworks, CNN is the most widely used one in depression detection[24].

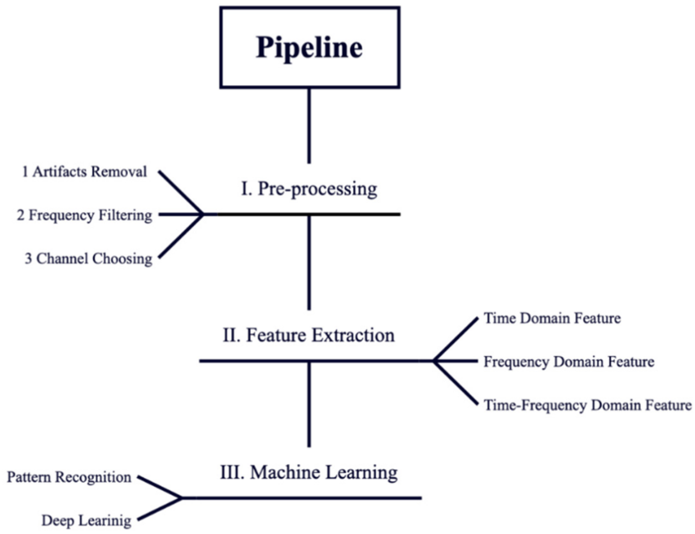

Fig 1. Pipeline of depression detection with EEG technology

Besides the input layer and output layer, CNN is divided into 3 parts, including the convolution layer, pooling layer, and fully connected layer. The convolution layer is for feature extraction. After the convolution operation, it needs to go through the activation function for nonlinear transformation. After this, the pooling layer compresses the convolved data to prevent overfitting due to too many parameters. The process of convolution-pooling can be repeated for times, which is the number of layers of the neural network. Finally, the fully connected layer could make the decision.

Similar to the methods in pattern recognition, running a CNN network is easy. For example, by importing *Pytorch* or *Tensorflow* module in Python, there are module functions to build these layers.

All of the above is the mainstream pipeline of depression detection with EEG technology, Fig. 1 could give a more intuitive illustration.

3. Discussion

Depression is a mental illness, which is the main reason why detecting it with the EEG signal makes sense. And the fact is that nowadays technology can give a fairly accurate result with EEG and computer power. It is not to hard build a detecting system with an accuracy of over 90%. However, there are still many aspects we can make better.

When patients are suffering from depression, these changes can not only be manifested at the electrophysiological level. Changes also manifest at the biochemical level[25], such as concentration changes of N-acetyl-aspartate and Choline, those important brain biochemical metabolites. By monitoring the concentration changes of these substances, we can paint a more complete picture of what happens in the brain. Further, other features related to the brain besides chemical variation or data from electrodes are also useful for this progress. For example, by analyzing brain connectivity and function network imaged by DTI and fMRI[26], we can calculate their correlation with depression.

EEG signals originate from electrical signals previously produced by neurons. However, through the process of its origin from the brain and then to the scalp, to what extent its information is distorted remains uncertain. The brain is a black box to us, it is too complex to know what is the precise structure inside, let alone for us to quantify the distortion of the EEG signal. Thus, more research on brain science is needed.

There are differences between people in different environments, and that is also the reason why the present models can't detect depression with 100% accuracy. To go further, we can develop an auto-adapting system, which can vary its intrinsic parameters with different groups of samples. Similarly, mental diseases share similarities. Schizophrenia, autism, phobia, and so on, there is a lack of research related to these relatively rare diseases. We can transfer models based on depression detection to these diseases to help these studies go deeper. It is approachable like we modify parameters in MI to make it better in certain individuals. It is always good to fit the brain-related model to the samples.

4. Conclusion

From the pipeline above, it is clear to see that, for short, depression detection is the process of processing and classifying EEG data. The preprocessing and feature extraction lays the foundation and provides convenience for the final classification.

In each part, multiple methods are offered, each of them suitable for different situations. More methods can be found in relative papers and novel methods can apply with advanced technology. In practice, experiments can't be smooth sailing. It is necessary to compare the results of different methods to determine the final process. The choice of parameters is actually the same. It is often essential to iteratively determine the parameters that can realize the highest accuracy rate.

Still, there is a long way to go, either in improving better algorithms or deepening the understanding of the brain.

References

[1] WHO, "Depression," https://www.who.int/news-room/fact-sheets/detail/depression. [Online]. Available: https://www.who.int/news-room/fact-sheets/detail/depression.

[2] M. T. Hawes, A. K. Szenczy, D. N. Klein, G. Hajcak, and B. D. Nelson, "Increases in depression and anxiety symptoms in adolescents and young adults during the COVID-19 pandemic," Psychol Med, pp. 1-9, Jan 13 2021, doi: 10.1017/S0033291720005358.

[3] WHO, "The impact of COVID-19 on mental health cannot be made light of," https://www.who.int/news-room/ feature-stories/detail/the-impact-of-covid-19-on-mental-health-cannot-be-made-light-of. [Online]. Available: https://www.who.int/news-room/feature-stories/detail/the-impact-of-covid-19-on-mental-health -cannot-be-made-light-of.

[4] J. Ormel, S. D. Hollon, R. C. Kessler, P. Cuijpers, and S. M. Monroe, "More treatment but no less depression: The treatment-prevalence paradox," Clin Psychol Rev, vol. 91, p. 102111, Feb 2022, doi: 10.1016/j.cpr.2021.102111.

[5] S. L. Oh et al., "A deep learning approach for Parkinson's disease diagnosis from EEG signals," Neural Computing & Applications, vol. 32, no. 15, pp. 10927-10933, Aug 2020, doi: 10.1007/s00521-018-3689-5.

[6] U. R. Acharya, F. Molinari, S. V. Sree, S. Chattopadhyay, K. H. Ng, and J. S. Suri, "Automated diagnosis of epileptic EEG using entropies," Biomedical Signal Processing and Control, vol. 7, no. 4, pp. 401-408, Jul 2012, doi: 10.1016/j.bspc.2011.07.007.

[7] N. Bigdely-Shamlo, T. Mullen, C. Kothe, K. M. Su, and K. A. Robbins, "The PREP pipeline: standardized preprocessing for large-scale EEG analysis," Front Neuroinform, vol. 9, p. 16, 2015, doi: 10.3389/ fninf. 2015. 00016.

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

[9] S. M. Alarcao and M. J. Fonseca, "Emotions Recognition Using EEG Signals: A Survey," Ieee Transactions on Affective Computing, vol. 10, no. 3, pp. 374-393, Jul-Sep 2019, doi: 10.1109/ taffc. 2017. 2714671.

[10] B. Hosseinifard, M. H. Moradi, and R. Rostami, "Classifying depression patients and normal subjects using machine learning techniques and nonlinear features from EEG signal," Computer Methods and Programs in Biomedicine, vol. 109, no. 3, pp. 339-345, Mar 2013, doi: 10.1016/j.cmpb.2012.10.008.

[11] S. C. Liao, C. T. Wu, H. C. Huang, W. T. Cheng, and Y. H. Liu, "Major Depression Detection from EEG Signals Using Kernel Eigen-Filter-Bank Common Spatial Patterns," Sensors, vol. 17, no. 6, Jun 2017, Art no. 1385, doi: 10.3390/s17061385.

[12] U. R. Acharya et al., "A Novel Depression Diagnosis Index Using Nonlinear Features in EEG Signals," European Neurology, vol. 74, no. 1-2, pp. 79-83, 2015, doi: 10.1159/000438457.

[13] M. Bachmann, J. Lass, and H. Hinrikus, "Single channel EEG analysis for detection of depression," Biomedical Signal Processing and Control, vol. 31, pp. 391-397, Jan 2017, doi: 10.1016/j. bspc. 2016. 09. 010.

[14] W. Mumtaz, S. Rasheed, and A. Irfan, "Review of challenges associated with the EEG artifact removal methods," Biomedical Signal Processing and Control, vol. 68, Jul 2021, Art no. 102741, doi: 10. 1016/ j. bspc. 2021.102741.

[15] W. Wu et al., "An electroencephalographic signature predicts antidepressant response in major depression," Nature Biotechnology, vol. 38, no. 4, pp. 439-447, 2020/04/01 2020, doi: 10.1038/s41587-019-0397-3.

[16] H. L. Chang, Y. Zong, W. M. Zheng, C. G. Tang, J. Zhu, and X. J. Li, "Depression Assessment Method: An EEG Emotion Recognition Framework Based on Spatiotemporal Neural Network," Frontiers in Psychiatry, vol. 12, Mar 2022, Art no. 837149, doi: 10.3389/fpsyt.2021.837149.

[17] R. Jenke, A. Peer, and M. Buss, "Feature Extraction and Selection for Emotion Recognition from EEG," Ieee Transactions on Affective Computing, vol. 5, no. 3, pp. 327-339, Jul-Sep 2014, doi: 10.1109/ taffc. 2014. 2339834.

[18] Y. Liu and O. Sourina, "Real-time fractal-based valence level recognition from EEG," in Transactions on Computational Science XVIII: Special Issue on Cyberworlds, 2013, vol. 7848: Springer Verlag, in Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), pp. 101-120, doi: 10.1007/978-3-642-38803-3-6. [Online]. Available: http://dx.doi.org/10.1007/978-3-642-38803-3-6.

[19] R. Khosrowabadi and A. Wahab bin Abdul Rahman, Classification of EEG correlates on emotion using features from Gaussian mixtures of EEG spectrogram (Proceedings of the 3rd International Conference on Information and Communication Technology for the Muslim World 2010). 2010, pp. 6 pp.-6 pp.

[20] O. Sourina and Y. Liu, "A FRACTAL-BASED ALGORITHM OF EMOTION RECOGNITION FROM EEG USING AROUSAL-VALENCE MODEL," in International Conference on Bio-inspired Systems and Signal Processing, Rome, ITALY, 2011, pp. 209-214.

[21] M. Akin, "Comparison of wavelet transform and FFT methods in the analysis of EEG signals," J Med Syst, vol. 26, no. 3, pp. 241-7, Jun 2002, doi: 10.1023/a:1015075101937.

[22] C. Cortes and V. Vapnik, "Support-vector networks," Machine Learning, vol. 20, no. 3, pp. 273-273, 1995, doi: 10.1023/A:1022627411411.

[23] T. M. Mitchell, "Machine Learning and Data Mining," Communications of the ACM, vol. 42, no. 11, pp. 31-36, 1999.

[24] U. R. Acharya, V. K. Sudarshan, H. Adeli, J. Santhosh, J. E. W. Koh, and A. Adeli, "Computer-Aided Diagnosis of Depression Using EEG Signals," European Neurology, vol. 73, no. 5-6, pp. 329-336, 2015, doi: 10.1159/000381950.

[25] B. Penninx, Y. Milaneschi, F. Lamers, and N. Vogelzangs, "Understanding the somatic consequences of depression: biological mechanisms and the role of depression symptom profile," Bmc Medicine, vol. 11, May 2013, Art no. 129, doi: 10.1186/1741-7015-11-129.

[26] S. Gu et al., "Network controllability mediates the relationship between rigid structure and flexible dynamics," (in English), Netw Neurosci, vol. 6, no. 1, pp. 275-297, Mar 16 2022. [Online]. Available: <Go to ISI>://WOS:000829501300003.