# An EEG-Based Depression Detection Method Using Machine Learning Model

Ran Bai[1], Yu Guo[2], Xianwu Tan[1], Lei Feng[3,4,5], and Haiyong Xie[5,1]
[1]National Engineering Laboratory for Public Safety Risk Perception and Control by Big Data (NEL-PSRPC), Beijing, China
[2]Beijing University of Posts and Telecommunications, Beijing, China
[3]The National Clinical Research Center for Mental Disorders, Beijing, China
[4]Beijing Key Laboratory of Mental Disorders, Beijing Anding Hospital, Capital Medical University, Beijing China
[5]Advanced Innovation Center for Human Brain Protection, Capital Medical University, Beijing, China
Email: haiyong.xie@ustc.edu

*Abstract*—**Depression, different from usual mood fluctuations and short-lived emotional responses to challenges in everyday life, is a common illness worldwide, with more than 300 million people affected. Although there are known, effective treatments for depression, fewer than half of those affected in the world (in many countries, fewer than 10%) receive such treatments. The diagnose of depression is usually subject to doctors due to the lack of biomarkers of depression. Electroencephalogram (EEG) is an easy-to-use, cost-effective technique that records electrical activity in brain. In this study, 64-channel EEG data was collected from 213 subjects including 71 health controls and 142 depression patients. 13 different features were extracted from EEG signals from all 7 sub-bands of all channels. 3 different feature selection models were used to find the subset of features that best represents the characteristics of EEG signal and 6 machine learning models were applied on all subsets of features to find the model that gained the highest accuracy and recall on depression detection.**

*Index Terms*—**depression, MDD, EEG, machine learning**

## I. Introduction

Depression is a common mental illness characterized by persistent sadness and a loss of interest in activities that people normally enjoy, accompanied by an inability to carry out daily activities, for 14 days or longer [1]. The latest estimates from WHO show that more than 300 million people are now living with depression, and it has increased more than 18% between 2005 and 2015. The barriers stand between the patients and effective treatments mentioned by WHO including lack of resources, lack of trained health-care providers, social stigma associated with mental disorders and inaccurate assessment. In countries of all income levels, people who are depressed are often not correctly diagnosed [1]. To diagnose a patient, doctors usually conduct a psychiatric evaluation by asking questions about symptoms, thoughts, feelings and behavior patterns. Patients are sometimes asked to fill out questionnaires as well [2].

The lack of biomarkers makes diagnosing depression a subjective decision that depends highly on the experiences of doctors. Electroencephalogram (EEG)is an easy-to-use, cost-effective technique that records electrical activities in brain. Recent studies show that features extracted from EEG could be used as biomarkers for depression diagnosis. A number of studies have shown that linear features [3]-[6] and nonlinear features of EEG [7]-[10] could be used as significant differentiating factors between individuals with and without depression. Acharya, *et al.* used nonlinear features including fractal dimension, largest Lyapunov exponent. sample entropy and DFA has reached an accuracy of 98% [11]. Mumtaz, *et al.* extracted EEG alpha band inter-hemispheric asymmetry and spectral powers to construct a support vector machine classification model with an accuracy of 97% [12]. Hasanzadeh, Mohebbi, & Rostami used features including Lempel-Ziv complexity, Katz fractal dimension and power spectral density to reach an accuracy of 91.3% [13]. However, due to the difficulties of collecting data, especially patient data, most studies have used a relatively small data set. Acharya, *et al.* used 30 subjects including 15 health and 15 patients [11]. Mumtaz, *et al.* used 63 subjects including 30 health and 33 patients. [12] Hasanzadeh, Mohebbi, & Rostami used 46 patients including 23 treatment responders and 23 non-responders [13]. Smaller datasets are more likely to be biased and machine learning models are known to be more robust and convincible with larger training and testing dataset.

In this study, we have collected EEG data from 213 subjects, including 71 individuals without depression ("health controls") and 142 patients with depression ("depression patients"). We extracted a set of linear features, nonlinear features and power spectral features were extracted from the EEG signal. 3 feature selection models were then applied to find the subset of features that best represented the data. All 3 subsets of extracted features were used on 6 different machine learning methods including Support Vector Machine (SVM) and Logistic Regression (LR). The trained machine learning

Manuscript received August 26, 2020; revised November 19, 2020.

model could relate the EEG signal features to the subject groups, i.e., health controls and depression patients.

The paper is organized as follow: Section II introduces the procedures of data collection, feature extraction, feature selection and building classification models; Section III lists and compares the results of feature selection and different machine learning classification models; Section IV concludes the paper and proposes future research plan.

## II. Materials and Methods

### A. Data Collection

EEG signals were recorded from 71 health controls and 142 depression patients from Beijing Anding Hospital, Capital Medical University, Beijing, China. A written consent form was obtained from each participant. Sixty-four channels of EEG data were recorded using 10-20 systems, and the recording frequency was set to 5k Hz. Subjects were asked to sit in a quiet room and the EEG signal was recorded for 3 minutes with subjects' eyes open and 3 minutes with subjects' eyes closed. To avoid eye-movement artifacts, the EEG data of closed-eyes were used. 2 channels were excluded from the data due to device defect. A band filter was used to remove all the frequencies above 40 Hz and the data was down sampled to 256 Hz for feature extraction. The data was then decomposed into six EEG sub-bands of interest: delta (0.5–4 Hz), theta (4–7 Hz), alpha1 (8–10 Hz), alpha2 (10–12 Hz), beta (13–30 Hz) and gamma (30–40 Hz) using the band-passed FIR filter.

### B. Feature Extraction

In this study, a total number of 13 features, specified in Table I, are extracted. The features include linear features, non-linear features and power spectral features. All features except for power spectral density, were calculated over a 3-minute data cut and each feature was extracted from delta band, theta band, alpha1 band, alpha2 band, beta band gamma band and full band of each channel. Power spectral density was extracted from each channel of all sub-bands above except for the full band, and for each sub-band, an average PSD was calculated over all channels. Therefore, a total of 6454 features (linear features: 7 features x 7 bands x 62 channels, non-linear features: 5 features x 7 bands x 62 channels, power spectral features: 1 feature x 6 bands x 62 channels, average PSD: 1 features x 6 bands) was extracted from the EEG data.

### C. Feature Selection

It is important to determine which subset of features, from the 6454 total features calculated, can best describe the difference between depression patients and health controls. In our study, we experimented 3 different feature selection models to find a better subset of features that would deliver the best accuracy and recall of classification model.

*1) L1-based feature selection*

The L1 based feature selection method takes advantages of the fact the linear models using L1 regularization have sparse solutions. L1 adds absolute values of coefficient as a penalty term. Due to the inherent linear dependence of the model parameters, regularization with L1 disables irrelevant features leading to sparse sets of features [14].

*2) Tree-based feature selection*

The tree-based feature selection method takes advantages of the interpretability of tree model. With each feature contributes to the final decision, the importance score of the feature is calculated. The method ranks all the importance scores, so the features with lower scores contributes less to the final decision and could be removed.

TABLE I. Features Extracted from EEG Signal

| Name | Property |
|---|---|
| Variance<br>Absolute centroid<br>Absolute power<br>Relative power<br>Activity<br>Kurtosis<br>Skewness | Linear Features |
| Spectral Entropy<br>Hjorth Complexity<br>Hjorth Mobility<br>HFD<br>DFA | Non-linear Features |
| Power Spectral density | Power Spectral Features |

### D. Classification

In this study, we deployed some of the most classic machine learning models to learn from the features extracted and selected above and make predictions. To gain a more accurate result, 5-fold cross validation was performed 10 times for each subset of features of each model. To better compare the results, we also train a machine learning model for each sub-bands of every calculated features and for all features combined together.

Accuracy and recall used to estimate the performance of the model are given by:

$$\text{Accuracy} = \frac{\text{True Postives+True Negatives}}{\text{Total Examples}} \quad (1)$$

$$\text{Recall} = \frac{\text{True Postives}}{\text{True Positives+False Negatives}} \quad (2)$$

The machine learning models used are support vector machine (SVM), KNN, decision tree, Naïve Bayes, random forest and Logistic Regression (LR).

*1) Support Vector Machine (SVM)*

Support Vector Machine (SVM) is a supervised machine learning model that could be used for classification. The algorithm of SVM creates a line or a hyperplane to separate the data into classes.

*2) Decision tree*

Decision Tree is a tree-like predictive model. In a decision tree, each interior node represents an input feature, the leaf node represents the class label, and the branches represents the decision-making progress from nodes to leaves.

*3) Random forest*

Random forests, shown in Fig. 1, is a combination of tree predictors such that each tree depends on the values of a random vector sampled independently and with the same distribution for all trees in the forest [15]. It is an ensemble learning method for classification. Random forest grows many decision trees. When classifying, the input was put to each decision dree and each tree returns a classification result, and the trees "vote" for the final result. The forest then returns the classification with the most votes [16].

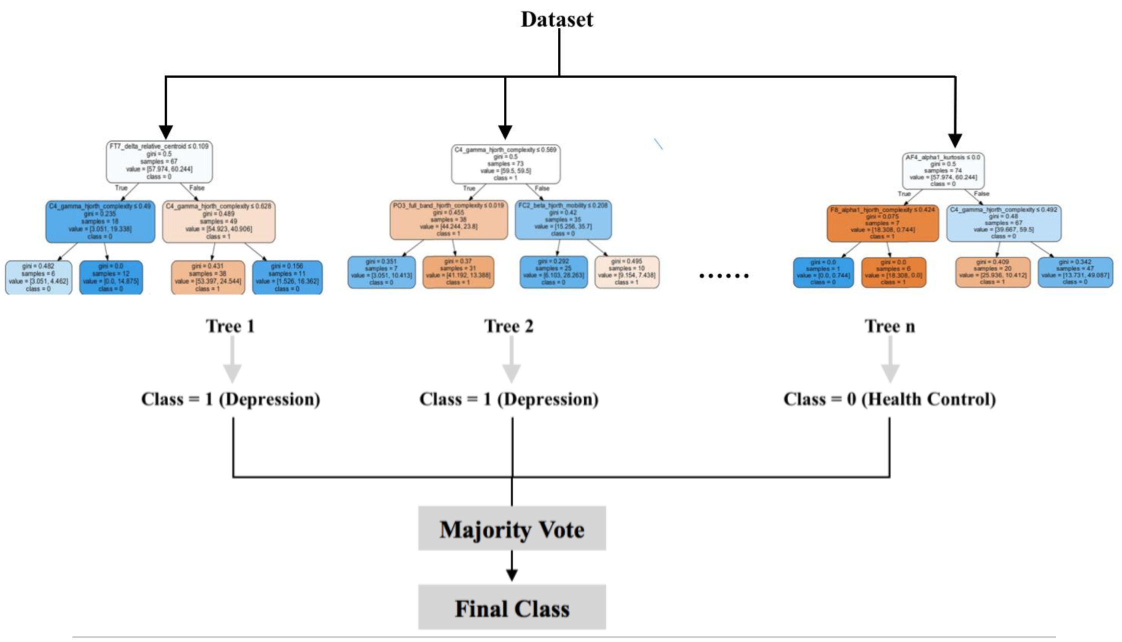


Figure 1. Random forest model.

## III. Results and Discussion

In this paper, 13 linear, non-linear and power spectral features were calculated on each channels of EEG data on each sub-band, and a total of 6454 features were extracted. 3 different feature selection models were used to select the feature subset that best describes the data (see Table II). 6 machine learning methods were deployed for each calculated feature, selected feature subsets and for all features combined. A 5-fold cross-validation was performed 10 times for each model and the average accuracy and recall for each features and feature subsets and all feature combined are calculated, shown in Table III and Table IV.

Table III shows the models with best performances for each feature calculated. Among all the linear features, Skewness on gamma band achieved the highest accuracy of 68.28% with a recall of 88.48% using the SVM model. Activity on theta band achieved the highest recall of 92.17% with accuracy of 67.01% using the random forest model.

For non-linear features, Complexity on gamma band achieved the highest accuracy of 79.63% using the KNN model, with a recall of 88.42%. HFD on beta band achieved the highest recall of 89.63% with an accuracy of 65.94% using random forest model.

TABLE II. Feature Selection Models and Selected Features

| Feature Selection Models | Number of Feature Selected | Selected Features |
|---|---|---|
| L1-based feature selection | 97 | Features including Kurtosis,<br>Hjorth Complexity of beta band, gamma band, delta band and full band from 57 channels |
| Tree-based feature selection | 13 | Features from 11 channel including<br>AF4(Kurtosis on alpha1),<br>F8(complexity on alpha1),<br>CPz (peak on gamma, DFA on alpha2),<br>FC2(mobility on beta),<br>FC3(absolute centroid on theta),<br>FC3(mobility on delta),<br>FT7(relative centroid on delta),<br>FT8(complexity on alpha2),<br>C4(complexity on alpha1),<br>PO3(complexity on full band),<br>TP9(complexity on gamma)<br>and P7(spectral entropy on full band). |
| FDR - based feature selection | 453 | Features mostly including<br>Hjorth Complexity<br>Hjorth Mobility,<br>Activity from delta, theta, alpha1, alpha2, beta, gamma and full band |

TABLE III. SINGLE FEATURE CLASSIFICATION RESULTS

| Name | Accuracy | Recall | Machine Learning Model | Sub-band Used |
|---|---|---|---|---|
| Variance | 67.90% | 91.60% | SVM | Full-Band |
| Absolute centroid | 67.51% | 90.37% | KNN | Gamma |
| Absolute power | 66.26% | 91.47% | Random Forest | Gamma |
| Relative power | 65.10% | 88.45% | Random Forest | Delta |
| Activity | 67.01% | 92.17% | Random Forest | Theta |
| Kurtosis | 69.29% | 90.20% | Random Forest | Gamma |
| Skewness | 68.28% | 88.48% | SVM | Gamma |
| Spectral Entropy | 70.24% | 84.58% | SVM | Gamma |
| Hjorth Complexity | 79.63% | 88.42% | KNN | Gamma |
| Hjorth Mobility | 76.29% | 86.00% | Random Forest | Alpha1 |
| HFD | 65.94% | 89.63% | Random Forest | Beta |
| DFA | 69.43% | 88.02% | Random Forest | Alpha2 |
| Power Spectral density | 66.95% | 90.13% | Random Forest | Gamma |

TABLE IV. CLASSIFICATION RESULTS OF SELECTED FEATURE

| Feature Set | Number of Feature Selected | Accuracy/ Recall | Model Names | | | | | |
|---|---|---|---|---|---|---|---|---|
| | | | SVM | KNN | Decision Tree | Naïve Bayes | Random Forest | Logistic Regression |
| All Features | 6454 | Accuracy | 69.20% | 64.54% | 67.34% | 57.10% | 76.97% | 68.50% |
| | | Recall | 79.30% | 84.02% | 75.80% | 53.06% | 89.54% | 77.24% |
| L1-based feature selection | 97 | Accuracy | 68.31% | 64.69% | 62.39% | 63.33% | 66.92% | 67.62% |
| | | Recall | 72.29% | 85.25% | 71.69% | 77.58% | 94.82% | 71.16% |
| Tree-based feature selection | 13 | Accuracy | 63.17% | 67.73% | 73.17% | 63.56% | 81.16% | 64.44% |
| | | Recall | 57.90% | 79.64% | 78.78% | 68.50% | 91.96% | 64.65% |
| FDR - based feature selection | 453 | Accuracy | 68.68% | 74.16% | 69.70% | 64.49% | 78.35% | 68.78% |
| | | Recall | 63.64% | 78.91% | 76.84% | 55.91% | 88.81% | 66.75% |

Power Spectral Density reached the highest accuracy of 66.95% with a recall of 90.13% on gamma band using random forest model.

Among models that uses all 6454 features, the random forest model has the highest accuracy of 76.97% and the highest recall of 89.54% using as shown in Table IV. From the 3 subsets of features selected using different feature selection models, a subset of 13 features selected by extra tree classifier performed the best, with an accuracy of 81.16% and a recall of 91.96% using random forest model. These 13 features are from 11 channels: AF4(Kurtosis on alpha1), F8(complexity on alpha1), CPz (peak on gamma, DFA on alpha2), FC2(mobility on beta), FC3(absolute centroid on theta, mobility on delta), FT7(relative centroid on delta), FT8(complexity on alpha2), C4(complexity on alpha1), PO3(complexity on full band), TP9(complexity on gamma) and P7(spectral entropy on full band). The performance of this model is shown in Fig. 2, using receiver operation characteristics (ROC). The electrodes shaded on the electrode map shown in Fig. 3 are the 11 channels used to calculate features for the model. This model is considered having the best performance with overall highest accuracy and recall. It also has the least amount of features and channels being used.

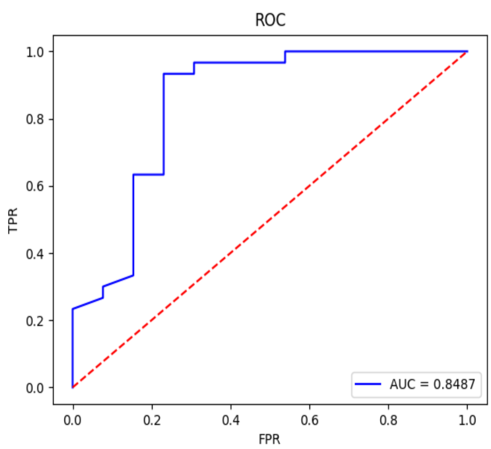


Figure 2. ROC plot for random forest model using 13 features.

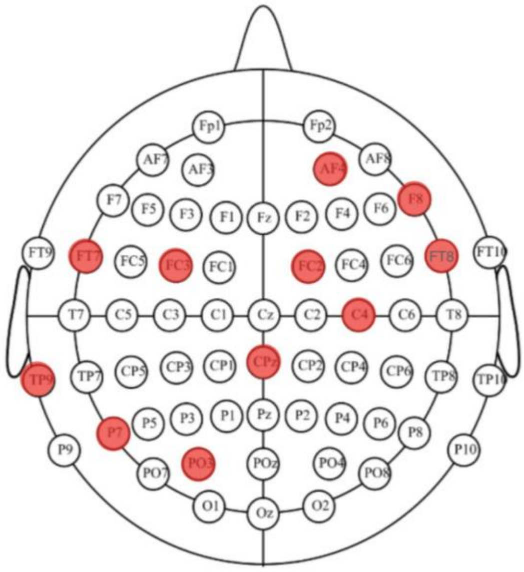


Figure 3. 11 channels used by random forest model on an electrode map.

It is shown in Table IV that the recalls from most of the models are higher than the accuracies. We consider that it is because the imbalance of the dataset. With 71 health controls and 142 patients, the size of the patient data is twice of the health controls. Since recall reflects how many of the actual patients our models captured by labeling it as patients. With more patient data, the models learn more about characteristics of patients. We believe that with a more balanced dataset, it is possible that the accuracies of the models could be improved as well as the recalls.

## IV. Conclusion

In this study, we have looked at different machine learning models using features extracted from EEG signal on a larger dataset compared to previous studies. The best model used 11 channels out of 64 channels to calculate features that could be used to detect depression. Although the accuracy and recall gained from the selected model could still be improved, it shows that the use of machine learning model to classify depression patients and healthy people is a promising approach. The results show a higher rate of recall than accuracy, which we believe may be caused by the imbalance of the dataset. With a larger heath control dataset, the model can learn more about the features and characteristics of healthy people, thus generating a higher accuracy rate. In the future, we hope to collect more healthy people data to gain a larger and more balanced dataset to train a more accurate model.

## Conflict of Interest

The authors declare no conflict of interest.

## Author Contributions

In this study, Ran Bai conducted the research, analyzed the data and wrote the paper; Yu Guo extracted features and built machine learning models; Xianwu Tan processed the data; Lei Feng labeled the patient data; Haiyong Xie designed the study; all authors had approved the final version.

## Acknowledgment

This research is sponsored by National Science and Technology Major Project for IND (Investigational New Drug) 2018ZX09201-014, supported in part by National Key Research and Development Project (Grant No. 2017YFC0820504) and supported in part by Capital's Funds for Health Improvement and Research 2018-1-2121.

## References

[1] "Depression: Let'S talk" says WHO, as depression tops list of causes of ill health. (2017). [Online]. Available: https://www.who.int/news/item/30-03-2017--depression-let-s-talk-says-who-as-depression-tops-list-of-causes-of-ill-health

[2] Depression (major depressive disorder). (2018). [Online]. Available: https://www.mayoclinic.org/diseases-conditions/depression/diagnosis-treatment/drc-20356013

[3] A. C. Deslandes, H. D. Moraes, F. A. Pompeu, P. Ribeiro, M. Cagy, C. Capitão, *et al.*, "Electroencephalographic frontal asymmetry and depressive symptoms in the elderly," *Biol. Psychol.*, vol. 79, pp. 317-322, 2008.

[4] V. A. G. Yatsenko, I. Baas, V. A. Ponomarev, and J. D. Kropotov, "Independent component approach to the analysis of EEG recordings at early stages of depressive disorders," *Clin. Neurophysiol.*, vol. 121, pp. 281-289, 2010.

[5] D. V. Iosifescu, S. Greenwald, P. Devlin, D. Mischoulon, J. W. Denninger, J. E. Alpert, *et al.*, "Frontal EEG predictors of treatment outcome in major depressive disorder," *Eur. Neuropsychopharmacol.*, vol. 19, pp. 772-777, 2009.

[6] C. Salustri, F. Tecchio, F. Zappasodi, G. Bevacqua, M. Fontana, M. Ercolani, *et al.*, "Cortical excitability and rest activity properties in patients with depression," *J. Psychiatry Neurosci.*, vol. 32, pp. 259-266, 2007.

[7] M. Ahmadlou, H. Adeli, and A. Adeli, "Fractality analysis of frontal brain in major depressive disorder," *Int. J. Psychophysiol.*, vol. 85, pp. 206-211, 2012.

[8] M. Ahmadlou, H. Adeli, and A. Adeli, "Spatiotemporal analysis of relative convergence of EEGs reveals differences between brain dynamics of depressive women and men," *Clin. EEG Neurosci.*, vol. 44, pp. 175-181, 2013.

[9] H. Hinrikus, A. Suhhova, M. Bachmann, K. Aadamsoo, U. Võhma, H. Pehlak, *et al.*, "Spectral features of EEG in depression," *Biomed. Tech.*, vol. 55, pp. 155-161, 2010.

[10] O. Faust, A. P. C. Alvin, D. P. Subha, and P. Joseph, "Depression diagnosis support system based on EEG signal entropies," *J. Mech. Med. Biol.*, vol. 14, pp. 1450035-1450041, 2014.

[11] U. R. Acharya, V. K. Sudarshan, H. Adeli, J. Santhosh, J. E. Koh, S. D. Puthankatti, *et al.*, "A novel depression diagnosis index using nonlinear features in EEG signals," *European Neurology*, vol. 74, pp. 79-83, 2015.

[12] W. Mumtaz, L. Xia, S. S. A. Ali, M. A. M. Yasin, M. Hussain, and A. S. Malik, "Electroencephalogram (EEG)-based computer-aided technique to diagnose major depressive disorder (DD)," *Biomedical Signal Processing and Control*, vol. 31, pp. 108-115, 2017.

[13] F. Hasanzadeh, M. Mohebbi, and R. Rostami, "Prediction of rTMS treatment response in major depressive disorder using machine learning techniques and nonlinear features of EEG signal," *Journal of Affective Disorders*, vol. 256, pp. 132-142, 2019.

[14] D. K. Ozgur, K. Mayumi, A. Tatsuya, and W. K. Ernst, "Prediction using step-wise L1, L2 regularization and feature selection for small data sets with large number of features," *BMC Bioinformatics*, vol. 12, article 412, 2011.

[15] L. Breiman, "Random Forests," *Mach. Learn.*, vol. 45, pp. 5-32, 2001.

[16] Random forests Leo Breiman and Adele cutler. [Online]. Available: https://www.stat.berkeley.edu/~breiman/RandomForests/cc_home.htm#giniimp

**Ran Bai** was born in Beijing, China. She received her Master's degree in computer science from Yale University, New Haven, Connecticut in 2017, and her Bachelor's degree in computer science from Purdue University, West Lafayette, Indiana in 2016.
She is currently an Assistant Engineer at National Engineering Laboratory for Public Safety Risk Perception and Control by Big Data (NEL-PSRPC), Beijing, China.