EAI Endorsed Transactions
on Pervasive Health and Technology
Research Article EAI.EU

A Comprehensive Study on Mental Illness Through Speech and EEG Using Artificial Intelligence

Sanjana Bhat[1] and Reeja S R[2*]

[1,2]School of Computer Science Engineering, VIT-AP University, Amaravati, Andhra Pradesh, India

Abstract

A typical mental ailment is depression that considerably harms an individual's everyday activities as well as their mental health. In light of the fact that mental health is one of the biggest problems facing society, researchers have been looking into several strategies for efficiently identifying depression. Mental illness can now be identified through speech analysis thanks to modern artificial intelligence. The speech aids in classifying a patient's mental health status, which could benefit their new study. For the purpose of identifying depression or any other emotion or mood in an individual, a number of past studies based on machine learning and artificial intelligence are being studied. The study also examines the effectiveness of facial expression, photos, emotional chatbots, and texts in identifying a person's emotions. Naive-Bayes, Support Vector Machines (SVM), Linear Support Vectors, Logistic Regression, etc. are ML approaches from text processing. Artificial Neural Network (ANN) is a sort of artificial intelligence method used to extract information from photos and classify them in order to recognise emotions from facial expressions.

Keywords: Artificial Neural Network, Mental Illness, Facial Expressions

Received on 10 December 2023, accepted on 01 March 2024, published on 07 March 2024

doi: 10.4108/eetpht.10.5328

*Corresponding author. Email: reeja.sr@vitap.ac.in

1.Introduction

Finding a suitable treatment for mental illness and detecting it are crucial. SOCIAL NETWORK MENTAL DISORDER DETECTION is a technique that can be used to de-stigmatize mental health and despair. On various AI and ML, tests can be run in a variety of emotional imbalance scenarios.
Using AI-based approaches, text-based emotion identification can be done (for example, determining the user's mood by keeping an eye on tweets and postings on various social media platforms and anticipate the risk of suicidal thoughts in the user).

Through a confusion matrix (a summary of the results of the predictions on a classification task), ML- based methods like Support Vector Machine and Naive Bayes can be assessed. A highly accurate algorithm aids in forecasting sentiment, which can be either positive or negative.

2. Related work

Numerous studies have been conducted on Electroencephalography (EEG), and the signals are frequently used to identify illnesses like seizure, Alzheimer's, and schizophrenia. This problem can be resolved by the EEG signal classification method, which also offers benefits including high temporal resolution, high sensitivity, relative affordability, ease of use, and convenience of recording [1]. A lot of study in the area of computer-aided depression diagnosis utilizing EEG data has been made possible because to developments in computer technology. Researchers use EEG inputs with a variety of other strategies, including machine learning algorithms, pattern recognition methods, and fusion with other neuroimaging methods. These investigations enhance the precision and effectiveness of depression diagnosis and provide new information about the mechanisms behind depression, but more study and validation of these techniques are required before they can be used in therapeutic settings [2]. Researchers are becoming more interested in using deep learning techniques to analyses EEG signals, which capture brain activity, as they have demonstrated great success in a number of domains, including computer vision, speech recognition, and natural language processing. Even though deep learning has been effective in research areas, there is comparatively limited research on utilizing deep learning methods for recognizing depression based on EEG signals. Furthermore, both traditional machines learning and deep learning approaches involve manual feature extraction from EEG signals, without optimizing the representation of these features. As a result, the challenge of identifying more effective feature representations for improved depression recognition using EEG signals remains unresolved [3]. This study uses EEG transformation to improve depression recognition as its main goal. Activity and power spectral density were taken out as the two main features. An ensemble learning method and a deep learning method were also introduced as two separate strategies. [4].

3. Datasets, features and methods used

3.1. Training and testing data set

Subject-dependent and subject-independent categorization techniques can be used to apply classification algorithms for activities including emotion, mental illnesses, and motor imagery. In the subject-dependent scheme, a classifier is trained specifically for each distinct subject, whereas in the subject-independent method, a classifier is learnt utilizing data from multiple individuals. Prior research has shown that subject-dependent classifiers are frequently substantially more accurate than subject-independent classifiers at classifying human mental states. The subject-independent strategy, which acknowledges depression, calls for a strictly defined separation of the training set and testing set. [10]. To avoid inflating classification accuracy artificially, it is essential to make sure samples from the same individual are not present in both the training and testing sets. To solve this issue and get reliable results, the training set is strictly separated from the testing set based on the participants. To ensure that each participant's sample appears twice in the testing set, a 7-fold cross-validation approach was applied. The training set provided 70% of the data. To fully examine both strategies, this procedure was carried out ten times. In order to assess these outcomes, the standard deviation and average accuracy was computed.

3.2. Task for emotional face stimuli

Past research has proven that people with depression frequently have specific traits, such as a decreased response to pleasant stimuli and a heightened sensitivity to negative stimuli. This indicates that people who are depressed examine images or photos that arouse emotions; they may pay more attention to or concentrate more on negative images than on happy ones [5].

Additionally, research has revealed that people are generally better at identifying facial expressions in others of their own race. The "own-race bias" or "cross-race effect" are terms used to describe this phenomenon. It implies that people may be better able to recognize the facial expressions of members of their own racial or ethnic group than members of other racial or ethnic groupings. This bias may be present in individuals with or without depression.

The Chinese Facial Affective Picture System (CFAPS) was utilized in the experiment's stimuli. [6]. The experiment involved 19 trials, each with 5 types of stimuli combinations that included four facial affective pictures namely - happy, sad, angry, astounded faces along with a neutral face. The trial began with a 1second display of a cross symbol, indicating rest, followed by a 10 second display of the combination of facial affective pictures. Participants were instructed to observe the pictures during this period.

3.3. EEG signal collection

EEG data was obtained by using the Net Station Software to record a 128 channel HydroeCel Geodesic Sensor Net (HCGSN) at a sampling rate of 250Hz. All of the electrode impedances were set to less than 70 k and all electrodes were positioned in accordance with the worldwide 10-20 placement standard, with Cz acting as reference electrodes [7]. The recording took place in a calm, enclosed, air-conditioned setting with the participant seated 60 cm from the screen to minimize electromagnetic interference. The middle 8 seconds of every observation throughout each trial were preserved to lessen the effect of the signal-denoising technique regarding the border value. When processing EEG data, this method of cropping the data is frequently employed to expand the sample size and increase the decoding precision. Due to the sampling rate of 250Hz, the cropping approach required 200 samples points for each 8 second trial. In order to be consistent with earlier studies, 4 seconds (1000 sample points) window was moved for each 125 sample points. In order to analyze the EEG data in the study, each 8-second trial was divided into 9 segments that lasted 4 seconds each using a shifting window method. This produced a total of 28 participants, whose EEG data were examined. There were 19 trials used in the experiment, as previously described in the section of Emotional face stimuli task, meaning that each individual had 19 trials, yielding a trial of 171(9*19) samples. Thus, an overall of 4788 (171*28) was used in the investigation [8]. Many researchers choose to employ fewer electrodes for their investigations in order to address issues with computing performance and efficiency in real-world applications. A common set of 16 electrodes (Fp1, Fp2, F3, F4, F7, F8, C3, C4, T3, T4, P3, P4, T5, T6, O1, O2) have been recognized in study using EEG data to examine depression. So, for the study's EEG analysis, 16 electrode sets were made use for the same.

3.4. Signal denoising

However, during the recording procedure, the existence of artefacts that do not reflect brain bioelectricity could have a detrimental effect on the data's quality. EEG signals contain both pathological and physiological information. As a result, before analysis, to remove these artefacts, the EEG data needs to be filtered. [9]. To filter, methods were adhered as listed below:
Step 1: The LMS technique was used to effectively remove noise from the raw EEG data by applying an adaptive noise canceller.

Step 2: Electromyography (EMG) signals were filtered out using a 0.5-40Hz band-passing filter.

Step 3: To reduce the impact of the Electrooculogram (EOG), the FastlCA algorithm was employed for artefact reduction.

3.5. Deep learning method

Time domain and frequency domain parameters are frequently considered in conventional machine learning methods for interpreting EEG signals. The suggested ensemble model also makes use of these features, for the detection of depression, in particular the time-domain activity and the frequency-domain power spectral density. But in this process, the spatial data that EEG caps provide is frequently ignored. The goal of this project is to look into ways to include spatial information in the process of identifying depressions.

3.6. T-SNE method

The understanding of the display of nonrepresentational qualities is improved by the T-distributed stochastic neighbor embedding (t-SNE) method. T-SNE enables a deeper understanding of the data by presenting high-dimensional data points on a two- or three-dimensional map. In order to conduct the study, two-dimensional maps of these feature vector occurrences were produced using t-SNE. The resulting representation, displayed in a graphic, makes use of various hues to differentiate between depressed subjects and healthy participants.

4. Methodology

4.1. Baseline

First, the effectiveness of the traditional classification methods is assessed for each of the feature vectors employed in the study. Three classifiers are used in the strategy: K-Nearest Neighbor (KNN), the Random Forest (RF), and the Support Vector Machine (SVM). The results show that using the activity feature in conjunction with the SVM classifier for the entire frequency range resulted in the highest average accuracy, which reached 82.06%. When analyzing individual frequency bands with the random forest classifier, the alpha frequency band achieved the maximum precision of 84.86%. [11]. However, it should be noted that the majority of feature vectors in both the separated and whole frequency bands were unable to distinguish between healthy people and patients with depression.

4.2. Ensemble learning

Second, the ensemble model appraised the repaired performance of the feature and found the results in accordance with techniques for reusing the fixed feature to create a new feature. Four different feature vectors were employed to represent the power spectral density: one for in terms of power spectral density, a 0.2 s time frame, one for a 0.5 s time window, one for a 1 s time window, and one for a 2s time window shown in table.1.

Table 1. To create new features, three training approaches were used.

Method 1	With 0.2 s, 0.5 s, 1 s, and 2 s time windows, feature vectors of the power spectral density of the full frequency band were used for datasets 1-4.
Method 2	Feature vectors of distinct frequency bands (such as alpha, beta, or theta) are selected for each training procedure in this method rather than the entire frequency band that was utilised in Method 1.
Method 3	Alpha, beta, and theta frequency band power spectral densities were used as feature vectors for datasets 1-3, and for each training procedure, distinct time window features were used. Instead of using SVM, other classifiers such as RF and KNN for comparison were tested.

The overall feature's average accuracy acquired from the separated and combined frequency bands were displayed and the same training techniques were created for the power spectral density to further examine the ensemble model's performance. The majority of the new features distinguished between depression patients and healthy people when compared to the original feature vectors were displayed as well. In both Methods 1 and 2, the power spectral density and average SVM accuracies perform better than typical RF and KNN accuracies. Three classifiers achieve accuracy levels of greater than 86% throughout the whole frequency band, with SVM achieving the highest accuracy of 89.02%. Alpha frequency band accuracy is superior to beta and theta frequency band for an independent frequency band.

Each of the frequency bands, alpha, beta, and theta, has typical accuracy values of over 84%, 69%, and 68%. On the alpha, beta, and theta frequency bands, SVM achieves the greatest accuracy of 87.42%, 73.04%, and 69.05%. Additionally, when using technique 3, the results of various time periods are approximations. Three classifiers have average accuracy ranges between 78.96% and 85.96%. Using feature vectors with a 0.5 s time window, the greatest accuracy of 85.66% is attained. SVM achieves the best accuracy for the activity, 88.76% on the whole frequency range. The range of the alpha frequency is where strategy 2's accuracy is at its highest, 87.06%. With feature vectors with a 0.5 s time window, method 3's greatest accuracy is 86.05%. Aside from that, it's important to note that the ensemble model's accuracy is typically higher than the baseline technique.

Additionally, the ensemble model was used to evaluate how well the specified timeframe performed and achieved results using fixed time frame with new creation techniques. It was found that the feature vectors with a 0.5-second time frame displayed superior performance employing unique feature generating approaches the fixed feature and baseline approach. The power spectral density and activity from a 0.5-second time window made up these feature vectors. Three different training methodologies were used in the analysis.

Table 2. Different training methodologies used in the analysis.

Method 1	For the datasets 1-2, both features of the whole frequency band were employed.
Method 2	For datasets 1-2, both features of distinct frequency bands were employed. The frequency band (alpha, beta, or theta) for each training process was altered.
Method 3	For dataset 1-3, features from the alpha, beta, and theta frequency ranges were used. Merely altered the feature type for each training process and different classifiers (RF, KNN) were utilized for comparison.

For the whole frequency range, the classifiers' accuracy rates above 88%, with SVM doing the best at 88.73%. The alpha band regularly showed greater accuracy when comparing the beta and theta bands to the separated spectral ranges. Particularly, the accuracy values for the alpha, beta, and theta bands were over 87%, 73%, and 67%, respectively. Approach 3 produced the highest precision for the power spectrum and activity characteristics, with 85.96% and 86.05%, respectively. Similarly, according to Method 2 shown in table.2, the ensemble model's accuracy, according to which a fixed time window, is generally greater unlike the baseline technique; nevertheless, the variation is not immediately apparent on the frequencies in the beta and theta ranges with fresh features [12] in table.3

Table 3. The average accuracy of new feature creation methods with a defined time window

Methods	Frequency Bands	SVM	RF	KNN
Method 1	Total(q)	88.73% ± 06.74% = 95.47% / 81.99%	88.69% ± 06.73% = 95.42% / 81.96%	88.61% ± 06.74% = 95.35% / 81.87%
	Alpha(r)	87.87% ± 07.93% = 95.80% / 79.94%	87.77% ± 07.82% = 95.59% / 79.95%	87.68% ± 07.72% = 95.40% / 79.96%
Method 2	Beta(s)	74.02% ± 13.83% = 87.85% / 60.19%	73.43% ± 13.82% = 87.25% / 59.61%	73.42% ± 13.94% = 87.36% / 59.48%
	Theta(t)	67.81% ± 12.37% = 80.18% / 55.44%	67.41% ± 11.07% = 78.48% / 56.34%	67.33% ± 10.99% = 78.32% / 56.34%
Method 3	Psd(u)	85.96% ± 07.34% = 93.30% / 78.62%	83.04% ± 11.15% = 94.19% / 71.89%	83.94% ± 09.81% = 93.75% / 74.13%
	Activity(v)	86.05% ± 07.24% = 93.29% / 78.81%	82.79% ± 11.56% = 94.35% / 71.23%	84.19% ± 09.57% = 93.76% / 74.62%

4.3. Deep Learning Method

Deep learning methods were well assessed and performed on the endeavour characteristics including power spectral density. Because they performed well, feature vectors from a 0.5-second time window specifically are chosen for this phase. These feature vectors, which indicate the power spectrum density and activity, were utilised to create images. The CNN parameters were then initialised using the Xavier initialization technique after a batch size of 32 was set. [13]. Dropout was utilised to prevent overfitting, and the Adam algorithm was employed to optimise. For purposes of evaluation, the average accuracy was determined of the distinct and comprehensive frequency bands. [14].
Maximum accuracy of 82.36% was achieved using an 8-layer CNN in combination with the activity feature, exceeding the frequency spectrum accuracy of 76% overall. With respect to the separated frequency bands, the alpha band consistently displayed greater accuracy when compared to the beta and theta bands, with the alpha band having the greatest precision of 84.75%. Either a 7-layer or an 8-layer CNN may be used to apply the deep learning methodology, and both showed performance that was comparable to the baseline approach.

5. Analysis

5.1. Analysis of depression using artificial intelligence and other machine learning approaches

The complete analysis with comparison study of various approaches is shown in Table.4. RBFN (Radial basis function networks) gives 71.4% accuracy.

Table 4. Analysis of depression using artificial intelligence and other machine learning approaches

Objective	Criteria	Data set used	Detection techniques for depression	Results	Parameters of performance
Determine whether a person is depressed using ML classifiers and the Twitter dataset. [15]	Sentiment analysis using Twitter and text processing	Twitter dataset	The following algorithms are used to classify data: LSTM, CNN, linear support	The output of various classifiers, such as the linear support vector	Precision, Recall, F1-measure

			vector classifier (SVC), multinomial naive-bayes, logistic regression, random forest classifier, and gradient boosting classifier.	classifier (SVC), logistic regression, multinomial naive-bayes, and ensembles like the random forest classifier and gradient boosting classifier, are compared with the findings of the LSTM and CNN classifiers.	
A comparison of various methods for interpreting emotions from facial expressions. [16]	The processing of images and videos of facial expressions		SVM, K-means clustering, PCA, Gabor Filters, etc.	By using Feature extraction for gabor, facial traits can be retrieved. To recognise faces, a neural network is trained. Emotion classification is done using SVM.	

A mental evaluation system powered by artificial intelligence can determine a patient's level of depression using deep learning. [17]	Application of chatbots, emotional artificial intelligence, and inputs that integrate text, audio, image, and video (Image, Video)	Webcam and microphone recordings of 671 participants' responses to interview inquiries	Multimodal Deep Learning Model	The responses to a patient health questionnaire are utilised to help diagnose depression in the participants. To identify depression, an integrated deep learning model is created. It uses a confusion matrix to evaluate the findings.	Precision: 68.61, NPV: 67.95, Sensitivity: 68.59, Specificity: 67.46 and FScore: 67.66
Several methods of depression detection, including deep learning models and the integration of AI and OCR. [18]	Combining text, voice, image, and video as inputs as well as chatbots and emotional AI		Image processing, AI, deep learning models, and voice or speech recognition applied together	Combining optical character recognition (OCR) and artificial intelligence (AI) makes it possible to identify facial expressions and tell if someone is smiling or dozing off (with their eyes closed) in a picture or a	

				video, which can aid in the early diagnosis of depression and the prevention of suicide.	
The RBFN's ability to identify depression in a person. Text and spoken input are considered. To identify depression from vocal expressions, partial least squares technique is utilised. [19]	Combining text, voice, image, and video as inputs, emotional AI, and chatbots	53 Volunteers	RBFN (Radial basis function networks)	Resemblance between the user's questionnaire and the RBFN model's results while assessing for depression. Users can talk about problems and get responses that seem human when they utilise a chatbot application. It was also shown that text communication is more user-friendly than speech.	Accuracy: 71.4%

6. Conclusion

This paper focuses on analyzing EEG signals, with an emphasis on their temporal, spectral, and spatial information. Feature extraction involves capturing activity spectrum and power density. Deep learning and ensemble modelling are used to detect depression via an EEG. While the ensemble learning method is initially applied, the spatial information from the electrode cover is disregarded. To incorporate spatial details, the feature vectors are transformed into deep learning-based images techniques. Deep learning utilized to recognize depressive disorders by including spatial information from EEG data, setting it apart from conventional techniques that focus solely on the time and frequency domains. Spatial information is incorporated through the Clough-Tocher interpolation strategy and the AEP method for coordinate modification. This enables the generation of EEG images with preserved topology and enhanced relevant information. By combining EEG images with a neural network model, the approach provides a novel perspective for analyzing EEG data.

References

[1] Alexander, Lindsay M., et al. "An open resource for transdiagnostic research in pediatric mental health and learning disorders." *Scientific data* 4.1 (2017): 1-26.

[2] Lee, Ellen E., et al. "Artificial intelligence for mental health care: clinical applications, barriers, facilitators, and artificial wisdom." *Biological Psychiatry: Cognitive Neuroscience and Neuroimaging* 6.9 (2021): 856-864.

[3] Amrani G, Adadi A, Berrada M, Souirti Z, Boujraf S. EEG signal analysis using deep learning: A systematic literature review. In2021 Fifth International Conference on Intelligent Computing in Data Sciences (ICDS) 2021 Oct 20 (pp. 1-8).

[4] Su, Chang, et al. "Deep learning in mental health outcome research: a scoping review." *Translational Psychiatry* 10.1 (2020): 116.

[5] Wang, X., Yang, X., Zhou, X. (2023) International Journal of Clinical and Health Psychology

[6] Gong, Xu, et al. "Revision of the Chinese facial affective picture system." *Chinese mental health journal* (2011).

[7] Gong, Xu, et al. "Revision of the Chinese facial affective picture system." Chinese mental health journal (2011).

[8] Li, Junhua, et al. "Feature learning from incomplete EEG with denoising autoencoder." Neurocomputing 165 (2015): 23-31.

[9] J. Shen et al., "Depression Recognition from EEG Signals Using an Adaptive Channel Fusion Method Via Improved Focal Loss," in IEEE Journal of Biomedical and Health Informatics, doi: 10.1109/JBHI.2023.3265805.

[10] Pedregosa, Fabian, et al. "Scikit-learn: Machine learning in Python." the Journal of machine Learning research 12 (2011): 2825-2830.
[11] Glorot, Xavier, and Yoshua Bengio. "Understanding the difficulty of training deep feedforward neural networks." Proceedings of the thirteenth international conference on artificial intelligence and statistics. JMLR Workshop and Conference Proceedings, 2010
[12] Kingma, Diederik P., and Jimmy Ba. "Adam: A method for stochastic optimization." arXiv preprint arXiv:1412.6980 (2014).
[13] Shetty, Nisha P., et al. "Predicting depression using deep learning and ensemble algorithms on raw twitter data." International Journal of Electrical and Computer Engineering 10.4 (2020): 3751.
[14] Joshi, Manju Lata, and Nehal Kanoongo. "Depression detection using emotional artificial intelligence and machine learning: A closer review." Materials Today: Proceedings 58 (2022): 217-226.
[15] Victor, Ezekiel, et al. "Detecting depression using a framework combining deep multimodal neural networks with a purpose-built automated evaluation." Psychological assessment 31.8 (2019): 1019.
[16] Zohuri, Bahman, and Siamak Zadeh. "The utility of artificial intelligence for mood analysis, depression detection, and suicide risk management." Journal of Health Science 8.2 (2020): 67-73.
[17] Oak, Shweta. "Depression detection and analysis." 2017 AAAI spring symposium series. 2017.