International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395-0056
Volume: 07 Issue: 02 | Feb 2020 www.irjet.net p-ISSN: 2395-0072

SURVEY ON DIFFERENT APPROACHES OF DEPRESSION ANALYSIS

Swathy Krishna[1], Anju J[2]

[1] *MTech Student, Dept. Of Computer Science & Engineering, LBS Institute of Technology for Women, Kerala, India*
[2]*Professor, Dept. Of Computer Science & Engineering, LBS Institute of Technology for Women, Kerala, India*

---***---

Abstract - *Clinical depression has been a common but a serious mood disorder nowadays affecting people of different age group. Since depression affects the mental state, the patient will find it difficult to communicate his/her condition to the doctor. Commonly used diagnostic measures are interview style assessment or questionnaires about the symptoms, laboratory tests to check whether the depression symptoms are related with other serious illness. With the emergence of machine learning and convolutional neural networks, many techniques have been developed for supporting the diagnosis of depression in the past few years. Since depression is a multifactor disorder, the diagnosis of depression should follow a multimodal approach for its effective assessment. This paper presents a review of various unimodal and multimodal approaches that have been developed with the aim of analyzing the depression using emotion recognition. The unimodal approach considers either of the attributes among facial expressions, speech, etc. for depression detection while multimodal approaches are based on the combination of one or more attributes. This paper also reviews several depression detection using facial feature extraction methods that use eigenvalue algorithm, fisher vector algorithm, etc. and speech features such as spectral, acoustic feature, etc. The survey covers the existing emotion detection research efforts that use audio and visual data for depression detection. The survey shows that the depression detection using multimodal approach and deep learning techniques achieve greater performance over unimodal approaches in the depression analysis.*

Key Words: *CNN, Depression, Facial Action Coding System(FACS), Active Appearance Model(AAM), Support Vector Machine(SVM), Teager Energy Operator (TEO)*

1. INTRODUCTION

Nowadays, many people are affected by depression and making their life miserable. Since depressed people are less sociable i.e they always keep them away from others, and may be introverted in nature, thus making the detection of the disease becomes difficult. Current diagnosis of depression is based on an evaluation by a psychiatrist supported by questionnaires to screen candidates for depression. But, these questionnaires need to be administered and interpreted by the concerned therapist. There is a need to develop automatic techniques for the detection of the presence and severity of depression. Depression has no dedicated laboratory tests and hence, there is difficulty in diagnosing depression. Recently with the emergence of machine learning and artificial neural networks, there has been an improvement in the diagnosis of depression. Thus several methods have been developed in recent years for supporting clinicians during the diagnosis and monitoring of clinical depression. In the long term, such a system may also become a very useful tool for remote depression monitoring to be used for doctor-patient communication in the context of e-health infrastructure. Clinical assessment of patients with depression relies heavily on two domains—the clinical history (i.e., history of presenting symptoms, prior episodes, family history etc.) and the mental state examination (appearance, speech, movement). The latter should be focused more for the effective diagnosis. In particular, the analysis of audio-visual data is found to be much more effective for assessing the mental state .Behaviors, poses, actions, speech and facial expressions; these are considered as channels that convey human emotions. Emotions are an important part of human life and much research has been carried out to explore the relationships between these channels and emotions.

The expressions on a face are a way of non verbal communication. The face is not only responsible for communicating ideas but also emotions. Generally, facial expressions can be classified into neutral, anger, fear, disgust, happiness, sadness and surprise. Emotions are extracted from the facial landmarks such as eyes, eyebrows, mouth, nose etc which are used to localize and represent salient regions of the face. Many studies have been conducted to identify the precise facial expressions that are related to depression. Some studies have been developed for recognizing facial expressions based on Action Units [1], eye movements [2], and so on.

Speech is a source of information for emotion recognition. Speech is an emotion which truly identifies one's mental state. A depressed individual will more clearly express emotion through speech. Speech features can be grouped into four types: continuous features, qualitative features, spectral features, and TEO (Teager energy operator)-based features. An important issue in the process of a speech emotion recognition system is the extraction of suitable features that appropriately characterize the variation in the emotions. Since pattern recognition techniques are rarely independent of the problem domain, it is believed that a proper selection of features significantly affects the classification performance.

There are also many multimodal approaches that came up in the past few years in which different uni-modalities are fused, such as body movement, facial expression and speech prosody, etc. It is found that the multimodal performs much better in the detection of depression as it considers several factors for the classification of emotions. Single modality may give only one sided information or may miss out an inherent parameter. This paper discusses the various approaches that have been introduced for the effective determination of depression which may be unimodal (considering any of the parameters such as face, speech, behaviour etc) or multimodal (i.e any possible combinations of audio, video, body posture etc). Section 2 describes some of the existing unimodal approaches and Section 3 reviews some of the multimodal approaches that have been developed for the depression analysis.

2. Depression Analysis Based on Unimodal Feature

Depression can be analyzed in two ways : Unimodal approach which considers a single attribute or feature that greatly contributes in assessing the emotional state and Multimodal approach which considers two or more attributes for assessing the depression. The unimodal approach considers either of the one feature among facial expressions, speech features, transcriptions, body postures etc. Here we are discussing some of the depression detection systems that use either facial expressions or speech features as the attribute for emotion recognition of a depressed individual.

2.1 Based on Facial Expressions

As the facial expressions of a depressed individual are quite different from that of a normal person, the identification of the micro expressions is of great importance in the diagnosis of depression. Many studies have been conducted in recent years to find out the Action Units (AU) for recognizing the emotions exhibited by depressed patients using Facial Action Coding System (FACS) [3]. According to the studies, the negative emotions such as sadness, contempt, disgust are found to be more prominent in depressed people.

For implementation of a depression detection method, Kulkarni et al.[4] use two algorithms named as Fisher vector algorithm and Local tetra pattern (LTrP). The feature extraction method using LTrP is used to extract the facial features. Local tetra pattern shows the connection between the pixel present in center and that pixel's neighborhood according to the directions and then gives a pattern. Fisher vector is an effective image classifier algorithm. It encodes the face values. After feature extraction, the Fisher vector method is used. Fisher vector gives the classification result. For finding and comparing the testing data with training data, KNN classifier is used. It compares the output image with the input image and according to that it classifies the result in between 'Depressed' and 'Non depressed'.

In [5] Khan et al. introduces Sobel operator and the Hough transform followed by Shi Tomasi corner point detection for facial landmark detection and feature extraction. First, the face is detected using face detection algorithm and then the facial fiducial landmarks are extracted using the Sobel horizontal edge detection method and the Shi Tomasi corner point detection method. Then the input feature vectors were calculated out of the facial feature extracted points and given as input to the Multi Layer Perceptron (MLP) neural network which classifies the human emotions. After training the network, testing sets of images were taken from the Karolinska Directed Emotional Faces (KDEF) database to check the performance. This system was found to have an accuracy of 95%. But it doesn't consider the cases of incidence of facial hair, occlusions etc while determining the facial expression.

Sarmad et al. [6] developed a system to detect depression on the basis of eye blink features. Eye blink features were extracted from video interviews using facial landmarks. It uses an adaboost classifier. From these features, eye state is obtained by finding the distance between eyelids across successive video frames and the rapid distance change between the eyelids is detected as blinks. Facial tracking is done by estimating the vertical distance between eyelids, followed by signal processing to filter out noise and finally eye blink features extraction. .As the eye tracking signal is represented by a set of consecutive value , averaging filter then, SG filter is applied for the processing of the signal. The system is found to have 88% accuracy. The limitation is that the AVEC 2014 and 2013 dataset which was used for the experiment does not contain ground truth annotation of eye blink rate.

Pediaditis et al. [7] suggests multiple methods that analyse each facial region individually for elaborating the number of facial features for the detection of stress or anxiety. It introduces a different approach of stress and anxiety assessment consisting of the mouth activity, head motion, heart rate, blink rate and eye movements. Here the analysis was carried out on a video sequence taken during a session where participants were allowed to watch videos which elicit feelings of stress and anxiety. The face detection is done using Viola Jones detection algorithm [8] which uses a rejection cascade of boosted classifiers. Head motion is measured in terms of horizontal and vertical deviation from a specified reference point. The heart rate is estimated by detection of subtle color changes in facial blood vessels during the cardiac cycle using Joint Approximate Diagonalization of Eigen matrices (JADE).Features from eyes movements' dynamics used are eye blinks and eye opening. The mouth related features are detected using optical flow method for mouth motion estimation and tracking Eigen features for mouth opening detection. It uses a multilayer Perceptron artificial neural network (ANN) as a classifier. The results indicate that ANN outperforms other classifiers (Na� e Bayes, Bayes network, SVM, Decision tree) with an accuracy of 73%. The method is not feasible on large datasets.

Wang et.al [9] demonstrates the use of Active Appearance Model (AAM) [10] for interpreting face images and image sequences in the depressed patients. The AAM contains a statistical, photo-realistic model of the shape and grey-level appearance of faces. The video of depressed patients were analyzed and compared with the person specific AAM model.

The facial feature points were marked according to FACS encoding system. Along with the facial expression analysis, specific eye feature points like eye pupil movement, blinking frequency, and movement changes of bilateral eyebrows and corners of mouth were also considered. SVM, which is known to be one of the best separation hyperplane in the feature space, was used for classification of features. Radial Basis Function is used in the SVM classifier. In addition to eyebrow, corners of mouth, they also included features such as pupil movement, blink frequency which is found to be very effective in identifying the depression. So considering the specific features in the face such as eyes, mouth, etc. has improved the accuracy in the depression analysis.

Table -1: Analysis of depression using facial features

REF NO.	DATASET	FEATURE EXTRACTION METHOD	CLASSIFIER
[4]	AVEC 2013	Local Tetra Pattern	KNN
[5]	KDEF	Hough transform and Sobel edge detection	MLP neural network
[6]	AVEC 2013, AVEC 2014	Vertical calculation by SG filter	Adaboost
[7]	Video recorded while watching anxiety/stress elicited videos	AAM, Optical flow and eigenvector method	MLP
[9]	Clinical video samples at Shandong Mental Health Centre	AAM	SVM

2.2 Based on Speech Features

The depression analysis based on speech includes processing of the audio signal and extracting the various speech features. Mainly there are four types of speech features and they are continuous, qualitative, spectral, and TEO based features. The depressed patients normally speaks in a low voice, slowly, sometimes stuttering, whispering, trying several times before they speak up or become mute in the middle of a sentence. The speech features such as prosodic features, acoustic features etc are extracted from the speech sample and classification or score prediction is done through statistical models such as Support Vector Machines (SVM), Gaussian Mixture Models(GMM), Linear/Logistic Regression etc.

Kevin et al. [11] analyses the speech to detect candidates' stress during HR (human resources) screening interviews. Automating some of the HR recruitment processes alleviates the tedious HR tasks of screening candidates for hiring new employees. This paper summarizes the results of applying several speech analysis approaches to determine if the interviewed candidates are stressed. Using the mean energy, the mean intensity and Mel- Frequency Cepstral Coefficients (MFCCs) as classification features, the stress in speech is detected. The Berlin Emotional Database (EmoDB), the Keio University Japanese Emotional Speech Database (KeioESD) and the Ryerson Audio-Visual Database of Emotional Speech and Song (RAVDESS are used as datasets. In order to identify different emotions in a human speech, features like pitch, articulation rate, energy or Mel- Frequency Cepstral Coefficients (MFCCs) are used. They use Support Vector Machine and Artificial Neural Network as the machine learning techniques for classification of emotions. The best results were obtained with neural networks with accuracy scores for stress detection of 97.98% (EmoDB).

In [12] Lang et al. uses a combination of handcrafted and deep learned features which can effectively measure the severity of depression from speech was used. Deep Convolutional Neural Networks (DCNN) are firstly built to learn the deep learned feature from spectrograms and raw speech waveforms. Then the median robust extended local binary patterns are extracted from spectrograms. Here AVEC 2014 dataset was used. Finally, they also proposed to adapt joint tuning layers, to combine the raw and spectrogram DCNN which can improve the performance of depression recognition. When both hand crafted and deep learning models were used, the performance of joint tuning was improved. They also combine AVEC 2013 and AVEC 2014 dataset to get a new enlarged database for training for improving the prediction of depression scores.

Betty et al. [13] *u*se deep learning and image classification methods to recognize emotion and classify the emotion according to the speech signals instead of using machine learning methods. All images labeled with respective emotions are prepared for training the model. The IEMOCAP database is used as a dataset. The spectrogram images generated from the IEMOCAP are resized to 500 x 300. This model uses the inception net for solving emotion recognition problems. Accuracy rate of about 38% is achieved.

Instead of focusing on acoustic features, Jingying et al.[14] focuses on identifying comorbidities from depressed people. The comorbidities focused here are Generalized Anxiety Disorder (GAD) and dysthymia. Classification algorithm was applied to predict the group labels of voice clips. All patients were divided into two groups on the basis of whether they have been diagnosed with comorbidity or not. The group labels were considered as golden standard in classification: patients with GAD comorbidity labeled 1, with dysthymia comorbidity labeled 2, without comorbidity labeled 0. To identify whether a voice clip is accompanying with a comorbidity or not, they classified 1 v/s 0 and 2 v/s 0, respectively. SVM was used for classification. Partial correlation was used to figure out whether there are salient relationships between the independent variables "number of features" and "sample size" and the dependent variable "predictive effects".

The recent development in the diagnosis of clinical depression focuses mainly on the objective assessment. Alghowinem et al.[15] explores the general characteristic of clinical depression which can greatly contribute to the effective assessment. It evaluates discriminative power of read versus spontaneous speech (in an interview / conversation) for the task of detecting depression. The read and spontaneous speech were classified using Support Vector Machines (SVM). The results show that the spontaneous rate has more variability which increases the recognition rate of depression and the MFCC feature group gave good results for classifying depression in spontaneous speech.

Karol et al.[16] proposes a novel method to detect depression using deep convolutional neural networks. The experiments were done on Distress Analysis Interview Corpus (DAIC) and uses ResNet-34 for the classification. The results suggest a promising new direction in using audio spectrograms for preliminary screening of depressive subjects by using short voice samples. The increase in the resolution of spectrogram would not significantly improve the assessment. The Test Time Augmentation (TTA) which creates predictions based on original spectrograms from the dataset and four augmentations of it. The TTA improves the performance of the classifier by 7%.And also the use of small samples diminished the effect of noise.

Table -2: Analysis of depression using speech features

REF NO.	DATASET	FEATURES EXTRACTED	CLASSIFIER
[11]	EmoDB, KeioESD, RAVDESS	Mean Energy, MFCC	SVM, ANN
[12]	AVEC 2013 and AVEC 2014	Low level descriptors, MRLEB	DCNN
[13]	IEMOCAP	Acoustic features	CNN
[15]	Real World clinical dataset	Low level descriptors and statistical features	SVM
[16]	DAIC –WOZ	Spectrogram features	ResNet

3. DEPRESSION ANALYSIS BASED ON MULTIMODAL FEATURES

Multimodal approach for depression detection is significant because depression is a complex and multi-factor disorder and to consider information from multiple modalities will be essential in addressing automatic depression assessment. Some of the multimodal approaches that have been developed for the effective assessment of depression are discussed as follows:

In [17] Abhay et al. proposes a speech emotion recognition method based on speech features and speech transcriptions(text). Emotion related low-level characteristics in speech are detected using speech features such as Spectrogram and Mel-frequency Cepstral Coefficients (MFCC) and text helps to extract semantic meaning. They experimented with several Deep Neural Network (DNN) architectures, in which transcriptions along with its corresponding speech features, Spectrogram and MFCC, which together provide a deep neural network. Experiments have been performed on speech transcriptions and speech features independently as well as together in an attempt to achieve greater accuracy than existing methods. The combined MFCC-Text Convolution Neural Network (CNN) model proved to be the most accurate in recognizing emotions in IEMOCAP data. Better results are observed when speech features are combined with speech transcriptions. The combined Spectrogram-Text model gives a class accuracy of 69.5% and an overall accuracy of 75.1% whereas the combined MFCC-Text model also gives a class accuracy of 69.5% .

Research in the field of Electroencephalogram (EEG) signal analysis has shown that, EEG signals can be used to distinguish between depressive patients and normal healthy persons. Yashika et al.[18] explore the EEG signal properties along with video signal for improving the depression assessment. The waveform generated while capturing these EEG signals can be used for identifying the medical abnormalities of a patient. Independent component analysis and wavelet packet decomposition is used for preprocessing and extracting features of EEG signal respectively. The facial expression analysis is done by face detection using Viola Jones Algorithm followed by the facial extraction using eigenvector method. The classification is done using neural networks.

Pampouchidou et al.[19] proposed a system with the potential of serving as a decision support system, based on novel features extracted from facial expression geometry and speech, by interpreting non-verbal manifestations of depression. The system has been tested both in gender independence and with different fusion methods. The framework achieved a reasonable accuracy for detecting persons achieving high scores on self-report scale of depressive symptomatology. The algorithm employed here follows the standard pipeline which consists of pre processing, feature extraction, selection, fusion and classification. The first step of the algorithm is preprocessing, followed by feature extraction. Two types of features were extracted, video based and audio-based. Subsequently, feature selection was employed to reduce the dimensions of the feature vectors, through Principal Components Analysis (PCA), for both video and audio. Video and audio features are then fused, and then classification follows, to reach a decision regarding presence or absence of depressive symptomatology. Optimal system performance was obtained using a nearest neighbour classifier on the decision fusion of geometrical features and audio based features in the gender based mode.

Shubham et al.[20] studied DAIC- WOZ dataset by utilizing text, audio and visual data. Studies have shown that depressed people behave differently compared to normal people and these differences can be detected by analyzing their audio and visual data. The audio sample was processed to obtain the voice of the patient only. Gaussian Mixture Model (GMM) clustering and Fisher vector approach were applied on the visual data, low level audio features and head pose and text features were also extracted. With the results obtained on the features from both the approaches, SVM and Neural Networks, Decision Level Fusion was applied on the results of different modalities.SVM was classifier was applied separately on the extracted features. In all the networks, Adam optimiser over Stochastic Gradient Descent was used as it was faster as well as efficient. Finally, the output of Audio and Text were fused together. Since the fused outputs improved accuracy, future work would be devoted towards fusing the outputs in a more generic manner.

Table 3: Analysis of depression using multiple attributes

REFNO.	ATTRIBUTES CONSIDERED	DATASET	CLASSIFIER
[17]	Speech, Text	IEMOCAP	CNN
[18]	EEG, Face	EEG database, From recorded video	ANN
[19]	Face, Speech	AVEC 2013,AVEC 2014	Nearest neighbour classifier
[20]	Text, audio, visual data	DAIC- WOZ	SVM and NN

4. CONCLUSION

This paper reviewed various kinds of unimodal and multimodal approaches that have been developed in the field of depression detection. According to the research so far, it has been found that multimodal approaches give best results and also deep learning approaches are more preferable than conventional machine learning approaches. Thus there have been a number of methods that came up with depression analysis from unimodal and multimodal approaches and still various approaches are developing in order to improve the accuracy in assessing the depression severity.

REFERENCES

[1] J. J. Lien, T. Kanade, J. F. Cohn and Ching-Chung Li, "Automated facial expression recognition based on FACS action units," Proceedings Third IEEE International Conference on Automatic Face and Gesture Recognition, Nara, 1998,pp.390-395.doi10.1109/AFGR.1998.670980

[2] Z. Pan, H. Ma, L. Zhang and Y. Wang, "Depression Detection Based on Reaction Time and Eye Movement," 2019 IEEE International Conference on Image Processing (ICIP), Taipei, Taiwan, 2019, pp. 2184-2188

[3] Ekman, Paul and Wallace V. Friesen. “Facial Action Coding System: Manual.” (1978).

[4] P. B. Kulkarni and M. M. Patil, "Clinical Depression Detection in Adolescent by Face," 2018 International Conference on Smart City and Emerging Technology (ICSCET), Mumbai, 2018, pp. 1-4. doi: 10.1109/ICSCET.2018.8537268

[5] Khan, F. (2018). Facial Expression Recognition using Facial Landmark Detection and Feature Extraction via Neural Networks. ArXiv, abs/1812.04510.

[6] Sarmad. Al-gawwam and M. Benaissa, "Depression Detection From Eye Blink Features," 2018 IEEE International Symposium on Signal Processing and Information Technology (ISSPIT), Louisville, KY, USA, 2018, pp. 388-392.

[7] Pediaditis, Matthew & Giannakakis, Giorgos & Chiarugi, Franco & Manousos, & Tsiknakis, Manolis. (2015). Extraction of Facial Features as Indicators of Stress and Anxiety. Conference proceedings: Annual International Conference of the IEEE Engineering in Medicine and Biology Society. IEEE Engineering in Medicine and Biology Society. Conference. 2015. 10.1109/EMBC.2015.7319199.

[8] P. Viola and M. Jones, "Rapid object detection using a boosted cascade of simple features," Proceedings of the 2001 IEEE Computer Society Conference on Computer Vision and Pattern Recognition. CVPR 2001, Kauai, HI, USA, 2001, pp. I-I.

[9] Wang, Qingxiang, Huanxin Yang, and Yanhong Yu. "Facial expression video analysis for depression detection in Chinese patients." Journal of Visual Communication and Image Representation 57 (2018): 228-233.

[10] Edwards, Gareth J., Timothy F. Cootes, and Christopher J. Taylor. "Face recognition using active appearance models." European conference on computer vision. Springer, Berlin, Heidelberg, 1998.

[11] Kevin Tomba , Joel Dumoulin .(2018) ”Stress Detection Through Speech Analysis”,:International Conference on Signal Processing and Multimedia Applications.

[12] Lang He, Cui Cao.(2018) “Automated depression analysis using convolutional neural networks from speech”,ELSEIVER Journal of Bio Informatics.

[13] Betty.P ,Nithya Roopa S., Prabhakaran M (2018) “Speech Emotion Recognition using Deep Learning” International Journal of Recent Technology and Engineering (IJRTE) ISSN: 2277-3878, Volume-7 Issue-4S, November 2018.

[14] J. Wang, X. Sui, T. Zhu and J. Flint, "Identifying comorbidities from depressed people via voice analysis," 2017 IEEE International Conference on Bioinformatics and Biomedicine (BIBM), Kansas City, MO, 2017, pp. 986-991.

[15] S. Alghowinem, R. Goecke, M. Wagner, J. Epps, M. Breakspear and G. Parker, "Detecting depression: A comparison between spontaneous and read speech," 2013 IEEE International Conference on Acoustics, Speech and Signal Processing, Vancouver, BC, 2013, pp. 7547-7551.doi: 10.1109/ICASSP.2013.6639130

[16] Chlasta, Karol, Krzysztof Wołk, and Izabela Krejtz. "Automated speech-based screening of depression using deep convolutional neural networks." arXiv preprint arXiv:1912.01115 (2019).

[17] Tripathi, Suraj, Abhay Kumar, Abhiram Ramesh, Chirag Singh and Promod Yenigalla.(2019) “Deep Learning based Emotion Recognition System Using Speech Features and Transcriptions.” ArXiv abs/1906.05681

[18] Y. Katyal, S. V. Alur, S. Dwivedi and Menaka R, "EEG signal and video analysis based depression indication," 2014 IEEE International Conference on Advanced Communications, Control and Computing Technologies, Ramanathapuram, 2014, pp. 1353-1360. doi: 10.1109/ICACCCT.2014.7019320

[19] A. Pampouchidou, O. Simantiraki .(2017) ” Facial geometry and speech analysis for depression detection”, Annual International Conference of the IEEE Engineering in Medicine and Biology Society.

[20] Dham, S. Sharma, A., & Dhall, A. (2017). Depression Scale Recognition from Audio, Visual and Text Analysis. ArXiv, abs/1709.05865.