International Journal of Scientific Research in Computer Science, Engineering and Information Technology
© 2019 IJSRCSEIT | Volume 5 | Issue 2 | ISSN : 2456-3307
DOI : 10.32628/CSEIT1952196

Depression Detection Through Speech Analysis : A Survey

Aastik Malviya, Rahul Meharkure, Rohan Narsinghani, Viraj Sheth, Pratiksha Meshram
Department of Computer Science and Engineering, SVKM's NMIMS, Mukesh Patel School of Technology Management and Engineering, Shirpur, Maharashtra, India

ABSTRACT

Depression is a common and serious medical illness which affects the way how we think, feel and act. Although harmless in its initial stages, it can cause serious problems if detected at a later stage. Due to advancements in technology, it is now possible to detect signs of depression. Different implementation of machine learning algorithms has been worked upon to detect factors causing depression. It is found that speech of a person is dramatically affected and various vocal features are used to classify depression.

Keyword: Teager Energy Operator (TEO), Hamilton Depression Rating Scale (HAM-D), Major Depression Disorder (MDD), Support Vector Machine (SVM), Low-Level Descriptors (LLD), Gaussian Mixture Model (GMM), minimalredundancy-maximal relevance (mRMR), Sequential Forward Floating Selection (SFFS), Principal Component Analysis (PCA).

I. INTRODUCTION

Mental health issue is a common problem nowadays. Major Depression Disorder (MDD) and anxiety disorder are the most frequent [1]. It is seen quite a lot by observing the present state of an individual, inability to concentrate, lack of interest (low mood) which are the primary but severity of depression may lead to headaches, suicidal attempts which causes exponential rise the number of death toll.

Correspondingly it was found that depression is increasing in adolescents (i.e those aged 13-20 years). Therefore, prior detection of depression is of primary importance [5]. It is found that depression has an impact on the vocal characteristics which reflect the throughput of qualities like pitch, speech rate, hoarseness, energy [6], also flat speech and elongated pauses [10]. With recent development in machine learning and better sensing technology, many attempts have been made to automate depression detection using the vocal features [15]. There have also been approach to detect depression using the facial expression geometry along with speech analysis [18]. Hence an automated computer tool can be very useful for therapists and doctors for treating patients suffering from clinical depression [20].

II. LITERATURE REVIEW

A constant research is going on to remove the ongoing questionnaire method of detecting depression in presence of psychiatric doctor and automate by analyzing speech, detecting depression. It has been found that there are acoustic features that are glottal, prosodic, cepstral, spectral and Teager Energy Operator (TEO) based which are being divided among two broad category perceptual (include prosodic, spectral, glottal and cepstral) and physiological (include TEO).

All the related works have followed a common system architecture to classify the depressed patients. The architecture follows 5 steps for classification:

1. Database of Subject's speech- It involves the acquisition of speech samples of depressed and control patients and forming a database of it.
2. Pre-Processing- This stage includes the removal of noise from the audio files and separating the patient's audio from doctor's for analysis and feature extraction.
3. Feature Extraction- This stage includes the extraction of different features from the speech by using various feature extraction tools and identifying the features that are necessary for characterizing depression.
4. Classification Model- The features that are extracted are then used to train the classification model where different machine learning techniques are implemented to train the model.
5. Decision- The trained model is then used to classify the patients between depressed and healthy.

Figure 1. Process flow.

Machine language and sensor data are used for monitoring mental disorders like depression, anxiety and bipolar disorders. The World Mental Health Survey Consortium states that there are more mental health patients in developed countries as compared to less- developed ones but treatment need is largely unmet in the latter [1]. Data was collected in Nigeria to investigate the simultaneous identification of depression and co-occurring physical illness using a multi-dimensional Bayesian network classification (MBC) approach

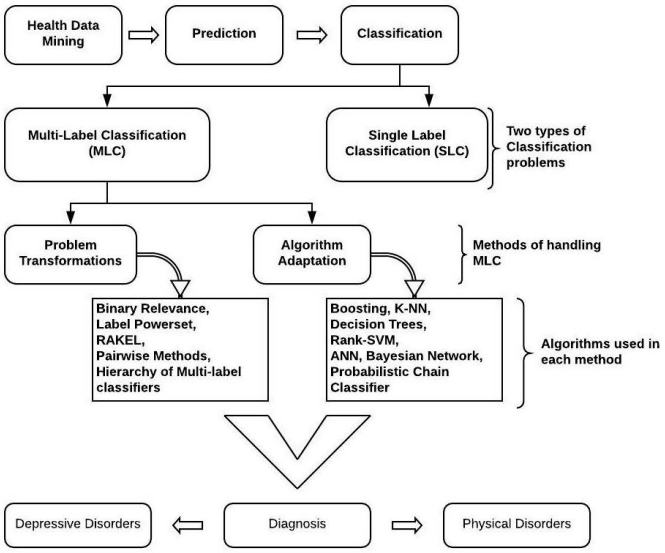

Figure 2. Framework for the study [3]

An experiment was conducted on high-risk (suicidal) patients, major depressed and non-depressed where two excitation-based speech parameters, namely vocal jitter and glottal flow spectrum. Vocal jitter is a method of frequency instability and glottal flow spectrum tells us about the air flow and how it changes the spectrum [4].

Rachelle Horwitz et al.[6]suggested an "on body" system to monitor MDD and evaluate its treatment. The database they used includes telephonic conversations collected from patients during office visits. The study uses 17-item Hamilton Depression Rating Scale (HAM-D) to rate the severity of the depression of the patient. [6]

Sharifa Alghowinem et al. [7] used a hypothesis that classification of clinical depression is obtained better by spontaneous speech rather than read speech. They used Support Vector Machine (SVM) and Leave-one-out cross-validation scheme for the same. An open source software named "openSMILE" to extract several Low-Level Descriptors (LLD) which can be calculated frame -by frame. It was concluded that SVM produced a better classification and hence the overall recognition rate of spontaneous speech was higher than that of read speech.

Lu-Shih Alex Low et al. [5] proposed a system where they used two different machine learning techniques i.e. Gaussian Mixture Model (GMM) and Support Vector Machine (SVM) to extract acoustic features of adolescent patients which was then modelled in to two different classes of depressed and control class. The feature extraction was performed on the voice database provided by Oregon Research Institute which consisted of audio recordings of 139 adolescents. A total of 14 acoustic features were identified to clearly characterize the differences in speech of depressed and control patients. A number of experiments were conducted while combining different features and machine learning techniques on male and female patients where average classification accuracy of male is 83% while female is 75%.

Zhenyu Liu et al.[19] proposed a system with main objective of finding out a set of speech features which will be helpful to differentiate people those who are depressed through comparative analysis by feature selection method and get the top k features, for feature selection they used classify them into filters, wrappers and embedded solutions, the method of selecting the features is of two stage combining the minimal-redundancy-maximal relevance (mRMR) criterion as the filter approach and the Sequential Forward Floating Selection (SFFS).Secondly to figure out which feature set will help to severity of depressive disorder and even to further predict by follow up study and thirdly to discover which interview manner is better on the basis of finding they implemented effective sensing support system to support clinicians. Lately classification was executed by using K-mean and Support Vector Machine (SVM).

A. Pampouchidou et al. [18], which created system by using facial as well as speech to detect depression. The following shows the system works.

For feature selection they used Principal Component Analysis (PCA) for video as well as audio and coming to conclusion by using cutoff points on BDI-II.

Shweta Oak, [17] that talked about finding out the reason of depression by creating a chatbot, at same time using user voice to analyze as well as text whether he is depressed. Using Radial Bias Function Network to find the root cause. Other work followed the way of acquiring the data by physical means or by means of an organization (like DVAC), followed by preprocessing and features selection (i.e prosodic, glottal, spectral, TEO, cepstral), then by using classification algorithm SVM, K-mean, 5-fold cross validation, regression and by using mathematical means. [10-16]

Various methods were discussed upon and implemented and each has produced their own result. First method was the analyzing of voice sample data obtained through lifeline numbers [21]. Another method sees the implementation of the k mean algorithm which although produced results in the right direction but it proved to be, “very difficult to quantify and monitor” [22]. A neural network model was trained that could depict the severity of depression by recognizing “red alert signals” [23]. A logistic regression model proved successful but it was gender biased [24]. The final method saw the use of a BDI questionnaire and speech recognition using a combination of k mean and Google API with the drawbacks of low accuracy and dependency on the questions [25].

Sr. No.	Speech Features	Classifier	Accuracy
1	MFCC + formant [1]	GMM	80%
2	Prosodic + Vocal tract + glottal [5]	PCA	86.7%(Male) 95.6%(Female)
3	Log energy and simmer [6] Loudness MFCC + energy + Intensity	HMM GMM	68%(Female) 57%(Male) 64%(Mixed)
4	TEO+ Glottal + prosodic + spectral[2]	GMM SVM	81%-87%(Male) 72%-79%(Female) 67%-69%(Male) 70%-75%(Female)
5	Glottal + prosodic + spectral + TEO[7]	GMM	73%

Figure 3. Various classifier with accuracy [20]

III. LIMITATION

- Data collection process that was incorporated were subject to read paragraph, picture description and interview [10-16].
- It was found that gender dependent was the key aspect for the models to perform better.
- It was rare to find model based only on voice and they took both video and voice data for the analysis.
- For feature selection process very, efficient search algorithm where required which resulted in high computational cost and data collection process their incorporated subject to read paragraph, picture description and interview. No mention of accuracy.

IV. CONCLUSION

Professionals evaluation is varying door to door depending on their diagnostic method and expertise. The above process can be automated through speech and machine learning. The psychological state of a person can be easily known by voice, as remarkable characteristics like low voice, monotonously, shuttering, jitter etc. At the same moment voice samples are easy to collect. Hence, we conclude speech analysis.

V. REFERENCES

[1]. Enrique Garcia-Cejaa et al. "Mental health monitoring with multi modal sensing and machine learning: A survey" in Elsevier, 2018.

[2]. Amanda Cibelly Brito Gois et al. "Factors associated with voice disorders among the elderly: a schematic review" in Brazilian Journal of Otorhinolaryngology, 2017.

[3]. Blessing Ojeme et al. "Selecting Learning Algorithms for Simultaneous Identification of Depression and Comorbid Disorders" at 20th International Conference on Knowledge Based and Intelligent Information and Engineering Systems, 2016.

[4]. Asli Ozdas et al. "Investigation of Vocal Jitter and Glottal Flow Spectrum as Possible Cues for Depression and Near-Term Suicidal Risk" at IEEE Transactions on Biomedical Engineering, Vol. 51, No. 9,2004.

[5]. Lu-Shih Alex Low "Detection of Clinical Depression in Adolescents' Speech During Family Interactions" at IEEE Transactions on Biomedical Engineering, Vol. 58, No. 3,2011.

[6]. Rachelle Horwitz et al. "On the Relative Importance of Vocal Source, System and Prosody in Human Depression" at 2013 IEEE International Conference on Body Sensor Networks, 2013.

[7]. Sharifa Alghowinem "Detecting Depression: A Comparison Between Spontaneous and Read speech" at IEEE ICASSP, 2013.

[8]. Andrea Guidi "Voice Quality in patients suffering from Bipolar Disease" at 37th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), 2015.

[9]. Andrea Guidi et al. "Analysis of running speech for the characterization of mood state in bipolar patient" at AEIT International Annual Conference, 2015.

[10]. A. Pampouchidou et al. "Facial Geometry and Speech Analysis for Depression Detection" at 39th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), 2017.

[11]. Jingying Wang "Identifying Comorbidities from Depressed People via Voice Analysis" at IEEE Conference on BIBM, 2017.

[12]. Yasuhiro Omiya "Difference in Speech Analysis Results by Compression" at IEEE ICIIBMS, 2017.

[13]. Miss. Pallavi R.Parekh et al. "Clinical Depression Detection for Adolescent by Speech Features" at ICECDS, 2017.

[14]. Yalin Li et al. "EEG-based Mild Depressive Detection Using Differential Evolution" at IEEEAccess Vol. 7, 2018.

[15]. Hailiang Long et al. "Detecting Depression in Speech: A Multi-classifier System with Ensemble Pruning on Kappa-Error Diagram" at Journal of Health and Medical Informatics, 2017.

[16]. Nicholas Cummins et al. "An Investigation of Depressed Speech Detection: Features and Normalization" at 12th Annual Conference of the International Speech Communication Association, Florence, Italy, August 27-31, 2011.

[17]. Shweta Oak "Depression Detection and Analysis" at The AAAI Spring Symposium on Wellbeing AI: From Machine Learning to Subjectivity Oriented Computing Technical Report, 2017.

[18]. A. Pampouchidou et al. "Facial Geometry and Speech Analysis for Depression Detection" at 39th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC'17) , Jul 2017, Jeju Island, South Korea. IEEE, 2017.

[19]. Zhenyu Liu, Bin Hu, Lihua Yan, Tianyang Wang, Fei Liu, Xiaoyu Li, Huanyu Kang "Detection of Depression in Speech" at 2015 International Conference on Affective Computing and Intelligent Interaction (ACII), 2015.

[20]. Shamla Mantri, Dr. Pankaj Agrawal, Dr. S.S.Dorle, Dipti Patil, Dr. V.M.Wadhai "Clinical Depression analysis Using Speech Features" at 6th International Conference on Emerging Trends in Engineering and Technology, 2013.

[21]. "Voice Analysis Aids Fight Against Suicide Depression" at https://www.cio.com.au/article/648076/voice-analysis-aid-fight-against-suicide-depression/

[22]. David Nield "Computer Program Tells When Someone Is Depressed by Speech Patterns" at Sciencealert, 2016.

[23]. Rob Matheson "Neural Network Model to Detect Depression in Conversations" at MIT News Office.

[24]. Prathamesh Raut et al. "Depression Detection Using BDI And Speech Recognition" at IJRASET, 2018.

[25]. Haihua Jiang et al. "Detecting Depression Using an Ensemble Logistic Regression Model Based on Multiple Speech Features" at Computational and Mathematical Methods in Medicine Vol. 2018, 2018.