

Review of Identification of Depression through Speech Analysis

**Prof. Nilesh Shelke¹, Ms. Ruchika Jadhav², Ms. Nikita Aldak², Ms. Neha Moon²,
Ms. Gayatri Gajbhiye², Ms. Namrata Patil²**

Assistant Professor, Department of Computer Science Engineering¹
Research Scholar, Department of Computer Science Engineering²
Priyadarshini College of Engineering, Nagpur, Maharashtra, India.
nileshrandd@gmail.com and ruchikajadhav106@gmail.com

Abstract: *Depression has developed into a crucial worldwide public health problem. This is a common psychological disturbance which affects the individuals physically and psychologically which leads to neurological illness. This effects on people in any age category. These makes researchers to work on this field so much. Traditionally, depression identification is performed by using semi-structured interviews of an individual and additional personality inventories that makes detection of depression is heavily depend on individual's response. Early treatment and identification of depression is needed to promote remission, prevention of relapse and decreasing an emotional tension of the disorder. It is difficult to detect depression at early stage of it using traditional processes. Studies in improvement of computational objective approaches indicates that speech signal of a speaker shows valuable relationship between depression and speech. Hence these acoustic features are used for diagnosis of depression. Enhancement in machine learning and deep learning techniques makes understanding of depression characteristics more rapid and convenient way which reduces the changes of clinical mistakes and labour costs. This paper shows study of various depression detection system or feature selection methods used by researchers in this field. This makes to detect depression at early stage and can be cured faster.*

Keywords: Speech, Depression Detection, Voice Quality Features, Machine Learning, Emotion, Deep Learning.

I. INTRODUCTION

Depression, a psychiatric illness which happens due to many factors like social, physical factors, etc. This expresses loss of interest, fatigue, irritability, psycho motor retardation, mood swings in daily every joyful activity and affects individual's work productivity and thinking power. By World Health Organization (WHO), over 350 million people worldwide are recognized with depression. Nearly one of the five women and one of the twelve men are affected by these major depressive disorders [1]. It has developed into a crucial worldwide public health problem. During COVID-19 pandemic, depression and anxiety show generality in population worldwide [2,3]. By 2030, depression will become second most affecting disorder in general population [4]. Suicide is big result of depression which makes over 8,00,000 of people passed every year [5]. Symptoms for depression may be seen as

1. Deficit of commitment in day-to-day actions.
2. Grief, emptiness or thinking downward.
3. Discouragement.
4. Annoyed and lacking spirit.
5. Little confidence, ego assessment or sensing impotent.
6. Disturbance concentrating and struggling to make judgments.
7. Irritation or enormous exasperation.

Traditionally, Diagnostic Statistical Manual (DSM) was the depression detection standard protocol which includes Patient Health Questionnaire Depression (PHQ) [6], Montgomery-Asberg Depression Rating Scale (MADRS) [7] and Beck's Depression Inventory (BDI) [8] plus by doctor's judgment. According to their answers, scores are assigned automatically. But all these are subjective measures are bias and requires physician's experience. These all are patients self-

report symptoms and which makes to identify only half of the depression [9]. Most of the people don't admit their emotional instability and conditions of mental illness and public health. But depression can be cured. Now mental health disorder identification and recognition have put their step in artificial intelligence and computer vision community. WHO states that, even though having treatments for depression, barriers in effective treatments and cure for depression is due to inaccurate assessment. Hence the development in easily accessible tools which enable reliable treatment of depressive symptomatology is needed. There are several systems have been involved to learn human-machine interactions which automatically access mental state and emotions of person objectively [10].

It is shown that voice signals carry significant information about the mental health disorder in speaker's speech [11-13]. The continuous changes in emotion state also affects acoustic properties of voice. Depression diagnosis using speech characterises is more effective and convenient way. It is different from normal people's speech characteristics. Distortions in path of speech, increase in number of pauses monotonous pitch and lower intensity of loudness are shown to be dependable indication of the depression [14-15]. These all are relatively objective features. The spectral feature and features related to energy of voice can be used for detection in depression [16]. The fundamental frequency (F0), intensity, speed, energy distribution and cepstral features are also considered to be good features for detection of depression. The variance in audio amplitude, bandwidth, and energy is also reduced in person. Along with this, speech signal of patient having depression do not show strong emotions like happiness and anger but they commonly show sadness, calm emotions. The scale of pitch, volume get decreased in depression. These are possibly also measured in cheaply, remotely and non-intrusively way [17]. This paper represents the study of related works done on depression identification and recognition of depression through speech. The various diagnosis or recognition systems and selection of relevant features methods which predicts the appropriate results estimated by researchers have discussed. This also shows works on different kind of datasets.

II. LITERATURE REVIEW

Physically how you appear is not just sufficient to say that you are physically well but also it is depended on how is your mental state. Mental state is also a very big problem to deal with such as Depression. Health care treatments by professionals are available but they are time consuming. There are several scientific literature research papers for depression detections which uses different ways according to nature of data like speech, handwriting and drawing which focuses on shape of the lines drawn, video analysis, contents of words written and spoken, electroencephalogram (EEG) and multimodality. Here focusing on relevant features extracted from speech only. If anyone suffering from Depression, there is changed can be found in speaking, pitch of voice, expressions of an individual. Studies have shown that speech show different voice characteristics of depressed person for depression detection diagnosis. Emotions indicated through speech may be analyzed in three distinct levels as follows:

- Physiological level: shows impulse of nerves or muscles consciousness in speech production process.
- Phonatory-articulatory level: shows movements and position of vocal folds or cords in voice of a person.
- Acoustic level: describes characteristics in the audio signals produced by speech.

There are many literatures which used acoustic and prosodic characteristic extracted from speech for depression detection [18, 19, 20, 21]. There are many works done on automatic depression recognition and analysis [22]. The sizeable cross-label differences but the compact inner-label differences must be manifested by feature. Most of the methods in literatures are regarding automatic depression detection (ADD) methods. They examine two important sources of knowledge as either combined or independent such as visual and audio methods. Even though here visual features are not considered for this research but audio features for depression diagnosis is considered. A heterogeneous symbol-used method for detecting the depression in speech was used by Zhaocheng Huang [23]. Sudden changes and acoustic regions are uniquely and inclusively figured out to meet among various embedding ways.

The Gaussian Mixture Model (GMM) and Support Vector Machine (SVM) algorithm are most used Machine Learning methods applied for depression detection model because of their potential to manage sparse dataset dynamically with comparably small cost and presence in more free libraries (Cummins et al., 2015 [24]). GMM gain 77% accuracy by using the Mel Frequency Cepstral Coefficients (MFCCs) only as feature, while SVM gain more accuracy than GMM but might gain worst accuracy if unsuitable kernel is taken. Most of the research workers have depended on support vector machine (SVM) classification to identify person with mental illness or depression except to find severity of depression with regression classifiers (Schwartz et al., 2014). They researched essential characteristics of paralinguistic (study of vocal



signals beyond the normal verbal speech) speech. By this the patterns for regression and classification issues were analysed. They have also discussed on current challenges and limitation in speech diagnosis.

The depression diagnosing through only speech has observed in Interspeech Conference held on 2018 in Hyderabad, India. There are many machine learning methods which are multimodal in nature like text, video, audio, etc. But depression detection using speech has also seen. Afshan et al. [25] concentrated on the efficacy of features of voice quality for depression detection. The features like F0, F1, F2, MFCCs, F3, H2-H4, H1-H2, A1, A2, A3, H4-H2k and CPP are used for voice quality in their research. The samples used in was of 10s and its accuracy of 77%, the accuracy also depend on the length of sample audios. In the [26] researchers explored correlation between voiced metrics and changes in seriousness of depression with period of time. The hypothesis presented by them shows that quantitative features in voiced prosody can be used for depression detection classification. The results demonstrated that voiced prosody analysis is worthwhile for analysis of depression. Sharifa, Goecke, Roland, Wagner, Michael, Parker, Gordon, Breakspear & Michael (2013) Tzirakis et al. used deep learning methodology to evaluate depression severity of a person using CNN on audios and Deep Residual System (ResNet) having 50 layers on visual information.

As discussed in above related studies, researchers have used various Machine Learning and deep learning models, different features of audio, selection of relative features and use of proper recognition pattern to be used in depression detection. It is found to be difficult having literatures which compares all these recognition patterns.

Actually, as per [27,28] the relevant features for audio analysis task are intensity, shimmer, harmonic, pitch, rates of speech, formants still all these features are not required for analysis. The required ones are used for depression detection using speech.

Paper	Concept
Natural Language Processing Methods for Acoustic and Landmark Event based Features in Speech-based Depression Detection [23]	Speech features using proven natural language processing methods
A Review of Depression and Suicide Risk Assessment Using Speech Analysis [24]	Effect of depression on paralinguistic speech characteristics
Effectiveness of voice quality features in detecting depression [25]	Automatic assessment of depression using conventional cepstral features
Detecting depression from facial actions and vocal prosody [26]	The feasibility of automatic detection of depression
Inferring clinical depression from speech and spoken utterances [27]	Features from harmonic model improve the performance of detecting depression from spoken utterances
Vocal-source biomarkers for depression: A link to psychomotor activity [28]	Vocal biomarkers are introduced for detecting different cognitive load conditions

Table 1: Previous works in depression detection using speech

III. CONCLUSION

The speech of a person shows useful characteristics for detecting depression or mental stress in an individual. This paper shows the study of literature works on this field of depression detection using speech or voice samples of people. Content of this research paper will provide guidelines to the researchers to work upon more research in area of depression detection using speech.

REFERENCES

- [1]. American Medical Association, "for Treatment of Mental Disorders in the World Health Organization," *Jama*, vol. 291, no. 21, pp. 2581–2590, 2004.
- [2]. N. Salari, A. Hosseinian-Far, R. Jalali, A. Vaisi-Raygani, S. Ra-soulpoor, M. Mohammadi, S. Rasoulpoor, and B. Khaledi-Paveh, "Prevalence of stress, anxiety, depression among the general population during the covid-19 pandemic: a systematic review and meta-analysis," *Globalization and health*, vol. 16, no. 1, pp. 1–11, 2020.
- [3]. R. Barzilay, T. M. Moore, D. M. Greenberg, G. E. DiDomenico, L. A. Brown, L. K. White, R. C. Gur, and R. E. Gur, "Resilience, covid-19-related stress, anxiety and depression during the pandemic in a large population enriched for healthcare providers," *Translational psychiatry*, vol. 10, no. 1, pp. 1–8, 2020.

- [4]. C. D. Mathers and D. Loncar. "Projections of global mortality and burden of disease from 2002 to 2030," *PLoS medicine*, vol. 3, no. 11, pp. e442, 2006.
- [5]. W. H. Organization et al., "Preventing suicide: A global impera-tive," 2014.
- [6]. K. Kroenke, R. L. Spitzer, and J. B. Williams, "The phq-9: validity of a brief depression severity measure," *Journal of general internal medicine*, vol. 16, no. 9, pp. 606–613, 2001.
- [7]. A. T. Beck, C. H. Ward, M. Mendelson, J. Mock, and J. Erbaugh, "An inventory for measuring depression," *Archives of general psy-chiatry*, vol. 4, no. 6, pp. 561–571, 1961.
- [8]. S. Montgomery and M. Asberg, "A new depression scale designed to be sensitive to change," *Acad. Department of Psychiatry, Guy's Hospital*. 1977.
- [9]. A. J. Mitchell, A. Vaze, S. Rao, "Clinical diagnosis of depression in primary care: a meta-analysis" *The Lancet*, vol. 374, no. 9690, pp. 609–619, 2009.
- [10]. F. Dornaika, B. Raducanu, "Inferring facial expressions from videos: Tool and application," *Signal Processing: Image Com-munication*, vol. 22, no. 9, pp. 769–784, 2007.
- [11]. France DJ, Shiavi RG, Silverman S, Silverman M, Wilkes DM, "Acoustical properties of speech as indicators of depression and suicidal risk," *IEEE Trans Biomed Eng*, vol. 47, no. 7, pp. 829–37, Jul 2000.
- [12]. Mundt JC, Snyder PJ, Cannizzaro MS, Chappie K, Geralts DS, "Voice acoustic measures of depression severity and treatment response collected via interactive voice response (IVR) technology," *J Neurolinguist*, vol. 20, no. 1, pp. 50–64, Jan 2007.
- [13]. Stasak B, Epps J, Goecke R., "Elicitation design for acoustic depression classification: An investigation of articulation effort, linguistic complexity, and word affect," *Interspeech; 2017 Aug 20–24, Stockholm*, pp. 834–8.
- [14]. J. C. Mundt, P. J. Snyder, M. S. Cannizzaro, K. Chappie, and D. S. Geralts, "Voice acoustic measures of depression severity and treatment response collected via interactive voice response (IVR) technology," *Journal of Neurolinguistics*, vol. 20, no. 1, 2007, pp. 50–64.
- [15]. J. C. Mundt, A. P. Vogel, D. E. Feltner, and W. R. Lenderking, "Vocal acoustic biomarkers of depression severity and treatment response," *Biological Psychiatry*, vol. 72, no. 7, 2012, pp. 580– 587.
- [16]. Tolkmitt, F., Helfrich, H., Standke, R., and Scherer, K., "Vocal indicators of psychiatric treatment effects in depressives and schizophrenics," *J. Commun. Disord*, vol. 15, pp. 209–222, 1982, doi: 10.1016/0021-9924(82)90034-X.
- [17]. N. Cummins, J. Epps, M. Breakspear, R. Goecke, "An investigation of depressed speech detection: Features and normalization," In: *Twelfth Annual Conference of the International Speech Communication Association*, 2011.
- [18]. N. Cummins, "Automatic assessment of depression from speech: Paralinguistic analysis, modelling and machine learning," Ph.D. dissertation, The University of New South Wales, 2016.
- [19]. J. Williamson, T. Quatieri, B. Helfer, G. Ciccarelli, and D. Mehta, "Vocal and facial biomarkers of depression based on motor incoordination and timing," in *Proceedings of AVEC'14*, 2014.
- [20]. P. Lopez-Otero, L. Docio-Fernandez, and C. Garcia-Mateo., "A study of acoustic features for depression detection," in *Proceedings of IWBF*, pp. 1–6, 2014.
- [21]. P. Lopez-Otero, L. Fernández, C. Garcia-Mateo, "A study of acoustic features for the classification of depressed speech," in *Proceedings of MIPRO*, pp. 1331–1335, 2014.
- [22]. A. Pampouchidou, P. Simos, K. Marias, F. Meriaudeau, F. Yang, M. Pediaditis, M. Tsiknakis, "Automatic assessment of depres-sion based on visual cues: A systematic review," *IEEE Transac-tions on Affective Computing*, 2017.
- [23]. Z. Huang, J. Epps, D. Joachim, and V. Sethu, "Natural Language Processing Methods for Acoustic and Landmark Event-based Features in Speech-based Depression Detection," *IEEE Journal on Selected Topics in Signal Processing*, vol. 14, no. 2, pp. 435-448, 2019.
- [24]. Cummins, N., Scherer, S., Krajewski, J., Schnieder, S., Epps, J., & Quatieri, T. F., "A Review of Depression and Suicide Risk Assessment Using Speech Analysis," *Speech Communication*, vol. 71, pp. 10–49, 2015. <http://doi.org/10.1016/j.specom.2015.03.004>

- [25]. Afshan, Amber, Jinxi Guo, Soo Jin Park, Vijay Ravi, Jonathan Flint, and Aber Alwan, "Effectiveness of voice quality features in detecting depression," In Proc. Interspeech, pp. 1676-1680, 2018.
- [26]. J.F. Cohn, T.S. Kruez, I. Matthews, Y. Yang, M.H. Nguyen, M.T. Padilla, F. Zhou, F. De, la Torre, "Detecting depression from facial actions and vocal prosody," International Conference on Affective Computing and Intelligent Interaction and Workshops, pp. 1-7, 2009.
- [27]. Asgari, M.; Shafran, I.; Sheeber, L.B., "Inferring clinical depression from speech and spoken utterances," In Proc. IEEE International Workshop on Machine Learning for Signal Processing (MLSP), Reims, France, 21-24 September 2014, pp. 1-5.
- [28]. Quatieri, T.F., Malyska, N., "Vocal-source biomarkers for depression: A link to psychomotor activity," In Proc. Thirteenth Annual Conference of the International Speech Communication Association, Portland, OR, USA, September 2012, pp. 9-13.
- [29]. karol j. piczak, "Environmental sound classification with convolutional neural networks," 2015 IEEE International Workshop on Machine Learning for Signal Processing, pp. 17-20.
- [30]. Nguyen T., Pernkopf, "Acoustic scene classification using a convolutional neural network ensemble and nearest neighbour filters," Detection and Classification of Acoustic Scenes and Events 2018 Workshop (DCASE2018), Surrey.