Harnessing Speech Technology for Mental Health Assessment and Detection

Chitralekha Bhat, Sunil Kumar Kopparapu

TCS Research, Tata Consultancy Services Limited, India
bhat.chitralekha@tcs.com, sunilkumar.kopparapu@tcs.com

Abstract

One in eight people world wide live with some kind of mental disorder which results in significant disturbances in their thinking and emotional behaviour. While effective prevention and treatment options exist, early detection and intervention are essential for improving therapy outcomes and reducing the long-term impact of mental health conditions. Speech technology can be employed, effectively, to not only assess and monitor mental health conditions but also can contribute to the early detection of mental health. By analyzing speech patterns, including changes in speech rate, pitch, or other vocal features, algorithms can identify potential markers of conditions like depression, anxiety, or even neurodegenerative disorders such as Alzheimer's disease, dementia. Additionally, speech technology can provide insights into a person's emotional state, stress levels, facilitating proactive healthcare measures. In this paper, we present an overview of the speech processing technology-based research from the perspective of application to mental healthcare.

Index Terms: speech recognition, mental health, human-computer interaction, paralinguistic features

1. Introduction

According to world health organization (WHO), mental disorder is characterized by a clinically significant disturbance in an individual's cognition, emotional regulation, or behaviour leading to impairment in important areas of functioning. While effective prevention and treatment options exist, early detection and accurate assessment is crucial. In recent years, the cross-pollination of technology and mental healthcare has shown immense promise, paving way for early detection, intervention, and assessment. In addition to the linguistic and the biometric content, spoken speech of an individual carries a wealth of paralinguistic information which directly correlates to mental health. As a result speech technology has emerged as a transformative tool in the field of mental health assessment. Today, speech technology (speech analysis plus machine learning) can extract valuable insights into an individual's emotional well-being, stress levels, and even the presence of mental health conditions.

Traditional methods of mental health assessment often rely on self-reporting and clinical observation, which can be subjective, time-consuming, invasive and potentially prone to human biases. However, speech technology offers a more objective and efficient approach, enabling a deeper understanding of an individual's mental state through the analysis of their voice. Speech as a signal offers several advantages over other behavioral descriptors when it comes to identifying mental disorders. Firstly, it is difficult to hide symptoms in speech, making it a more reliable indicator. Secondly, speech directly expresses emotions and thoughts through both language content and paralinguistic flavour, providing valuable insights into a person's mental state. Variations in speech, both in terms of motor and acoustic aspects, indirectly reflect the mental state of a person. Another advantage of speech analysis is its potential for generalization across languages using paralinguistic analyses, making it useful across languages.

Speech technology allows for the early identification of potential mental health concerns, even before individuals may be fully aware of their own symptoms. By detecting subtle changes in vocal characteristics, we can raise red flags and trigger further assessment, potentially leading to timely interventions that can prevent the exacerbation of mental health issues and improve treatment outcomes. Acquisition of speech data is relatively effortless, as it can be easily captured using a microphone commonly available on smartphones, tablets, and computers. This makes it a cost-effective alternative to more expensive wearables (like EEG) or invasive neuro-imaging methods. Moreover, since many clinical interviews are already recorded, speech data is increasingly accessible for analysis. Therefore, speech technology has a very important role to play in mental health assessments and mental disorder identification. With the rise of teletherapy and remote healthcare, individuals can engage with virtual conversation bots, allowing speech algorithms to analyze their speech and provide valuable insights to healthcare professionals, regardless of geographical distances. This technological advancement brings mental health assessment to individuals who may face barriers to traditional in-person evaluations and social stigma, ensuring that support reaches a wider population.

Audio–Visual Emotion Challenges (AVEC) [1] as well as Computational Paralinguistics Challenge[1] (ComParE) [2] - since 2009, are two major competitions that highlight the importance of the use of speech technology for mental health assessment and monitoring, indicative of the direction of automatic and objective assessment of mental health using speech as a medium. A recent survey [3] on machine learning and

[1] http://www.compare.openaudio.eu/

10.21437/SMM.2023-6

deep learning techniques for depression detection, suggests that speech is the best predictor to distinguish people who are healthy from individuals suffering from mental health conditions. By examining recent research findings, technological advancements, and ongoing initiatives, we aim to shed light on the transformative power of speech technology and its integration into mental healthcare. As the field continues to evolve, the effective utilization of speech technology stands to enhance early detection, improve treatment outcomes, and ultimately promote mental well-being for individuals. This paper explores the contribution of speech technology towards mental health assessment in terms of early detection. We look into the two major aspects of mental health; namely (1) mental disorders such as depression, anxiety and (2) neurological disorders such as Parkinson's disease, Alzheimer's disease.

2. Speech biomarkers for mental health assessment

Mental disorders and neurological disorders are two distinct categories of conditions that affect the brain and impact a person's cognitive, emotional, and behavioral functioning. Mental disorders primarily refer to conditions that affect an individual's thoughts, emotions, mood, and behavior. They are often characterized by disturbances in thinking, perception, mood regulation, and social functioning (eg. depression, anxiety, bipolar disorder, schizophrenia, and personality disorders). Whereas, neurological disorders primarily involve abnormalities or dysfunction in the structure, function, or chemistry of the brain and nervous system. They typically arise from specific neurological mechanisms, such as damage to brain cells, neurotransmitter imbalances, or genetic abnormalities. Neurological disorders encompass conditions like Alzheimer's disease, Parkinson's disease, epilepsy, multiple sclerosis, stroke, and traumatic brain injury. Figure 1 shows a complete list of mental disorders as mentioned in [4]. Mental health can manifest in various aspects of speech, as human thoughts, emotions, and cognitive processes are intricately connected to vocal expression. We explore some useful speech biomarkers below:

- **Speech Rate and Fluency:** Changes in speech rate and fluency can be indicative of certain mental health conditions. For example, individuals with anxiety may exhibit rapid and pressured speech, characterized by a fast pace and difficulty pausing. Conversely, individuals experiencing depression may have slow speech rate, with longer pauses and reduced verbal fluency or exhibiting speech rates alternating between slow and incoherently fast, in case of bipolar disorder and flat affect in case of schizophrenic patients [5, 6, 7, 8]. The importance of speech rate in mental health assessment is also evidenced by the use of it in the Young Mania Rating Scale [9]. An analysis of verbal fluency disruptions using the number of pauses and their duration, during various speech tasks, as a biomarker, was used for early detection of Parkinson's disease [10].
- **Pitch and Tone:** Alterations in pitch and tone of voice can convey emotional states and provide insights into mental health. For instance, individuals with depression may speak with a monotone or reduced vocal inflection, reflecting a lack of energy or enthusiasm. In contrast, heightened pitch and vocal intensity may be observed in individuals experiencing heightened emotional arousal or anxiety [11, 12].
- **Articulation and Pronunciation:** Changes in articulation and pronunciation may be present in individuals with certain mental health conditions. Speech may become slurred, unclear, or hesitant, particularly in conditions such as schizophrenia or certain neurological disorders. These alterations can affect the individual's ability to convey thoughts or communicate effectively [13, 14].
- **Vocal Quality and Resonance:** Variations in vocal quality, such as breathiness, tremors, or strained voice, can reflect emotional states or physiological changes associated with mental health conditions. These changes may occur in conditions such as post-traumatic stress disorder (PTSD) or during times of high stress and emotional distress [15, 16].
- **Content and Language Use:** The linguistic content of speech and the language usage can also provide clues about mental health. Individuals with depression may exhibit negative or self-critical thoughts, expressing feelings of hopelessness or worthlessness. In conditions like schizophrenia, individuals may exhibit disorganized or tangential speech, where their thoughts and ideas lack coherence or logical progression. Language is a potential source of predictors for suicidal thoughts and behaviors (STBs), as changes in speech characteristics, communication habits, and word choice may be indicative of increased suicide risk [17].
- **Emotional Expressiveness:** Emotional expressiveness in speech can vary depending on mental health. Some individuals may display reduced emotional expressiveness, where their speech lacks appropriate emotional modulation or fails to match the content of the conversation. Others may exhibit heightened emotional expressiveness, with exaggerated emotional reactions or difficulty controlling emotional responses [2, 18].

3. Early Detection

Early detection of mental disorders allows for timely intervention and better treatment outcomes. Speech technology can contribute to the early detection of mental disorder by analyzing various speech biomarkers mentioned in the previous section. By analyzing and machine learning these speech biomarkers, one can identify potential markers of conditions like depression, anxiety, or even neurodegenerative disorders.

3.1. Mental Health Disorders

Advances in speech technology for eliciting and analyzing speech samples along with the machine learning capabilities have paved way to several research works for automatic detection and assessment of mental health disorders using speech biomarkers [19, 20, 21]. An early study using voice acoustic measures to assess depression severity and response to treatment using telephonic speech showed the validity of using this method [22]. Suitability of speech biomarkers for depression and suicide risk has been highlighted in [23] and presents an overview of the investigations that have been carried out into

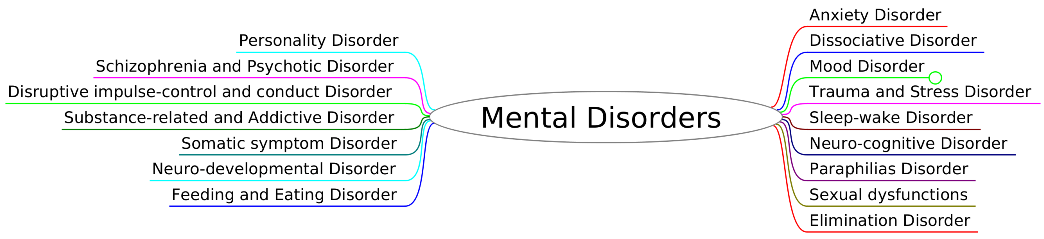

Figure 1: *List of Mental Disorders [4].*

the automatic analysis of speech as a predictor of suicidality and depression. Feature selection strategies for the automatic detection of depression from speech using a combination of hand crafted and deep-learned features from Deep Convolutional Neural Networks (DCNN) are used for joint fine-tuning of layers to combine the raw and spectrogram features based DCNN, to boost the depression recognition performance in [24].

A system based on ensemble learning whose individual classifiers are Convolutional Neural Networks (CNNs) and the inputs to them are raw log-spectrograms derived from speech are used to classify whether a speaker is depressed or not in [25]. Three timing-related biomarkers, namely: speech rate, pause time, and response time are used as metric to automatically assess depression severity levels in patients with major depressive disorder (MDD) and bipolar disorder in [8]. The study indicated that mentally depressed individuals showed longer response time, longer pause time, and slower speech rate than healthy individuals, all of which were suggestive of psychomotor retardation. Additionally, non-verbal speech biomarkers such as prosody, formant, source-filter and spectral analysis features were extracted from speech and machine learning models were applied to predict the existence or the severity of various mental disorders such as depression, PTSD, autism and schizophrenia [7, 26]. A 3-minute speech task was used to assess the speech of young children with internalizing disorders (such as anxiety and depression) by using speech biomarkers such as low-pitch voices, repeatable speech inflections and content, and response to surprising stimuli [27].

3.2. Neurological Disorders

Speech technology for early detection Parkinson's disease (PD), Alzheimer's disease (AD) and dementia, affecting a very large aging population world wide, has attracted a lot of research attention The argument in favor of using speech technology is non-invasive, less time consuming, cost efficiency [28] and more importantly social stigma [29]. It was shown that 78% of early untreated PD subjects indicate some form of vocal impairment in the study [30], also the variations in fundamental frequency of speech performed best in classifying PD and normal speech. Speech biomarkers typically phonation, articulation, and prosody plays a key role in the characterization of speech and along with signal processing algorithms are useful for early detection of PD [30, 31, 32]. The characterization of speech and it's classification using various available machine learning tools, from a PD detection standpoint has been an active area of investigation and research [33, 34].

Although Alzheimer's disease (AD), a neurological disorder, is typically characterized by memory impairment, language impairment can be an important marker. In [35], Word2Vec and Bidirectional Encoder Representations from Transformers are used to analyze acoustic and linguistic biomarkers from audio recordings and their transcriptions, to assess an early onset of genetic AD. A systematic comparison of different methods for detecting cognitive impairment using narrative speech from a picture description task has been proposed in [36]. They analyze speech biomarkers and train ML classifiers, generate transcripts using automatic speech recognition (ASR), and extract linguistic and psychological features for dementia screening. The findings contribute to early detection and intervention in cognitive impairment cases. Using content-free biomarkers like speech rate, turn-taking patterns extracted from speech interactions, [37] proposed the use of additive logistic regression to build a reliable predictive model for AD. Combining acoustic feature processing with NLP techniques [38] employs conversational analysis for early dementia detection. Their system utilizes long short term memory (LSTM) and grated recurrent unit (GRU) models to capture temporal features and long-term dependencies. Non-linguisitc audio descriptors from the openSMILE toolkit [39] have been used to identify and classify AD [40, 41]

These studies provide insights into the recent advancements in using speech technology for detecting mental disorders as well as neurological disorders such as AD and PD. They highlight the utilization of diverse acoustic and linguistic biomarkers, along with advanced ML algorithms, to improve the accuracy and effectiveness of early detection of disorders. They cover areas such as open-source feature extraction, paralinguistic speech analysis, and automated detection/assessment of disorders using speech signals as the medium. Extensive study on the use of speech technologies for AD [42] and psychiatric disorders [29] detection has also been reviewed. Classification of voice features and ML-based detection algorithms for various mental health and neurological disorders impacting speech, is provided in Table 3 of article [43]. The viability of leveraging speech technologies to facilitate the detection, assessment, clinical monitoring purposes of mental health disorders have been explored in [44, 45]. However, preserving patient privacy is a major concern. Speech techniques, that retain relevant mental health information conveyed by the speaker's speech and accurately represent the phonetic aspects of the speech, while minimizing any speaker-specific traits [46] is also being investigated.

4. Conclusions

According to world health organization one out of every eight people[2] world wide suffer from some form of mental disorder. While effective prevention and treatment options exist, traditional methods of mental health assessment fail because of the sheer scale of people suffering from mental disorders. The combination of speech technology and the knowledge and skills of mental health professionals can lead to more accurate and comprehensive assessments, ultimately guiding the development of tailored treatment plans for people suffering from mental disorder. However, it is important to acknowledge the ethical considerations associated with speech technology in mental health assessment. Issues such as data privacy, consent, and the potential for misinterpretation of speech data pose serious challenges to the application of these technologies for the assessment of mental health. Future direction could be exploring the limitations of the above studies in terms of what models/choices of AI/ML techniques can be made with particular datasets in particular contexts, especially to understand when certain approaches or models overfit in order to make intelligent recommendations.

5. References

[1] F. Ringeval, B. Schuller, M. Valstar, N. Cummins, R. Cowie, and M. Pantic, "Avec'19: Audio/visual emotion challenge and workshop," in *Proceedings of the 27th ACM International Conference on Multimedia*, ser. MM '19. New York, NY, USA: Association for Computing Machinery, 2019, p. 2718–2719. [Online]. Available: https://doi.org/10.1145/3343031.3350550

[2] B. Schuller, A. Batliner, S. Steidl, and D. Seppi, "Recognising realistic emotions and affect in speech: State of the art and lessons learnt from the first challenge," *Speech Commun.*, vol. 53, no. 9–10, p. 1062–1087, nov 2011. [Online]. Available: https://doi.org/10.1016/j.specom.2011.01.011

[3] M. Squires, X. Tao, S. Elangovan, R. Gururajan, X. Zhou, U. R. Acharya, and Y. Li, "Deep learning and machine learning in psychiatry: a survey of current progress in depression detection, diagnosis and treatment," *Brain Informatics*, vol. 10, 2023.

[4] F. Ma, "Diagnostic and statistical manual of mental disorders-5 (dsm-5)," in *Encyclopedia of Gerontology and Population Aging*. Springer, 2022, pp. 1414–1425.

[5] E. Kraepelin, "Manic depressive insanity and paranoia." *The Journal of Nervous and Mental Disease*, vol. 53(4), p. 350, 1921.

[6] A. Maxhuni, A. Muñoz-Meléndez, V. Osmani, H. Perez, O. Mayora, and E. F. Morales, "Classification of bipolar disorder episodes based on analysis of voice and motor activity of patients," *Pervasive and Mobile Computing*, vol. 31, pp. 50–66, 2016.

[7] Y. Tahir, D. Chakraborty, J. Dauwels, N. Thalmann, D. Thalmann, and J. Lee, "Non-verbal speech analysis of interviews with schizophrenic patients," in *2016 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2016, pp. 5810–5814.

[8] M. Yamamoto, A. Takamiya, K. Sawada, M. Yoshimura, M. Kitazawa, K.-c. Liang, T. Fujita, M. Mimura, and T. Kishimoto, "Using speech recognition technology to investigate the association between timing-related speech features and depression severity," *PLOS ONE*, vol. 15, no. 9, pp. 1–10, 09 2020.

[9] R. C. Young, J. T. Biggs, V. E. Ziegler, and D. A. Meyer, "A rating scale for mania: Reliability, validity and sensitivity," *The British Journal of Psychiatry*, vol. 133, no. 5, p. 429–435, 1978.

[10] L. Jeancolas, G. Mangone, D. Petrovska-Delacrétaz, H. Benali, B.-E. Benkelfat, I. Arnulf, J.-C. Corvol, M. Vidailhet, and S. Lehéricy, "Voice characteristics from isolated rapid eye movement sleep behavior disorder to early parkinson's disease," *Parkinsonism & Related Disorders*, vol. 95, pp. 86–91, 2022.

[11] S. Scherer, G. Stratou, G. Lucas, M. Mahmoud, J. Boberg, J. Gratch, A. S. Rizzo, and L.-P. Morency, "Automatic audiovisual behavior descriptors for psychological disorder analysis," *Image and Vision Computing*, vol. 32, no. 10, pp. 648–658, 2014, best of Automatic Face and Gesture Recognition 2013. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0262885614001000

[12] A. K. Alimuradov, A. Y. Tychkov, A. V. Kuzmin, P. P. Churakov, A. V. Ageykin, and G. V. Vishnevskaya, "Improved ceemdan based speech signal analysis algorithm for mental disorders diagnostic system: Pitch frequency detection and measurement," *International Journal of Embedded and Real-Time Communication Systems (IJERTCS)*, vol. 10, no. 1, pp. 22–47, 2019.

[13] J. Emerson and P. Enderby, "Prevalence of speech and language disorders in a mental illness unit," *European Journal of Disorders of Communication*, vol. 31, no. 3, pp. 221–236, 1996.

[14] A. Koskela, A. Neittaanmäki, K. Rönnberg, A. Palotie, S. Ripatti, and T. Palotie, "The relation of severe malocclusion to patients' mental and behavioral disorders, growth, and speech problems," *European Journal of Orthodontics*, vol. 43, no. 2, pp. 159–164, 2021.

[15] S. Scherer, G. Stratou, J. Gratch, and L.-P. Morency, "Investigating voice quality as a speaker-independent indicator of depression and ptsd." in *Interspeech*, 2013, pp. 847–851.

[16] C. W. Espinola, J. C. Gomes, J. M. S. Pereira, and W. P. dos Santos, "Detection of major depressive disorder using vocal acoustic analysis and machine learning—an exploratory study," *Research on Biomedical Engineering*, vol. 37, pp. 53–64, 2021.

[17] S. Homan, M. Gabi, N. Klee, S. Bachmann, A.-M. Moser, M. Duri', S. Michel, A.-M. Bertram, A. Maatz, G. Seiler, E. Stark, and B. Kleim, "Linguistic features of suicidal thoughts and behaviors: A systematic review," *Clinical Psychology Review*, vol. 95, p. 102161, 2022.

[18] Y. Ziwei, P. Zheng, and D. Pin, "Multiparameter voice assessment for voice disorder patients: a correlation analysis between objective and subjective parameters," *Journal of Voice*, vol. 28, no. 6, pp. 770–774, 2014.

[19] N. Cummins, J. Epps, M. Breakspear, and R. Goecke, "An investigation of depressed speech detection: Features and normalization," in *Twelfth Annual Conference of the International Speech Communication Association*, 2011.

[2] https://www.who.int/news-room/fact-sheets/detail/mental-disorders

[20] M. Valstar, B. Schuller, K. Smith, T. Almaev, F. Eyben, J. Krajewski, R. Cowie, and M. Pantic, "Avec 2014: 3d dimensional affect and depression recognition challenge," in *Proceedings of the 4th international workshop on audio/visual emotion challenge*, 2014, pp. 3–10.

[21] Z. Liu, B. Hu, L. Yan, T. Wang, F. Liu, X. Li, and H. Kang, "Detection of depression in speech," in *2015 International Conference on Affective Computing and Intelligent Interaction (ACII)*, 2015, pp. 743–747.

[22] 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, pp. 50–64, 2007.

[23] N. Cummins, S. Scherer, J. Krajewski, S. Schnieder, J. Epps, and T. F. Quatieri, "A review of depression and suicide risk assessment using speech analysis," *Speech Communication*, vol. 71, pp. 10–49, 2015. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0167639315000369

[24] L. He and C. Cao, "Automated depression analysis using convolutional neural networks from speech," *Journal of Biomedical Informatics*, vol. 83, pp. 103–111, 2018. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S153204641830090X

[25] A. Vázquez-Romero and A. Gallardo-Antolín, "Automatic detection of depression in speech using ensemble convolutional neural networks," *Entropy*, vol. 22, p. 688, 06 2020.

[26] N. Cummins, J. Epps, V. Sethu, M. Breakspear, and R. Goecke, "Modeling spectral variability for the classification of depressed speech," in *Proc. Interspeech 2013*, 2013, pp. 857–861.

[27] E. W. McGinnis, S. P. Anderau, J. Hruschak, R. D. Gurchiek, N. L. Lopez-Duran, K. Fitzgerald, K. L. Rosenblum, M. Muzik, and R. S. McGinnis, "Giving voice to vulnerable children: Machine learning analysis of speech detects anxiety and depression in early childhood," *IEEE Journal of Biomedical and Health Informatics*, vol. 23, no. 6, pp. 2294–2301, 2019.

[28] A. Tsanas, M. A. Little, P. E. McSharry, and L. O. Ramig, "Accurate telemonitoring of parkinson's disease progression by noninvasive speech tests," *IEEE Transactions on Biomedical Engineering*, vol. 57, no. 4, pp. 884–893, 2010.

[29] D. M. Low, K. H. Bentley, and S. S. Ghosh, "Automated assessment of psychiatric disorders using speech: A systematic review," *Laryngoscope Investigative Otolaryngology*, vol. 5, no. 1, pp. 96–116, 2020.

[30] J. Rusz, R. Cmejla, H. Ruzickova, and E. Růžička, "Quantitative acoustic measurements for characterization of speech and voice disorders in early untreated parkinson's disease." *Journal of the Acoustical Society of America*, 2011.

[31] M. A. Little, P. E. McSharry, E. J. Hunter, J. Spielman, and L. O. Ramig, "Suitability of dysphonia measurements for telemonitoring of parkinson's disease," *IEEE Transactions on Biomedical Engineering*, vol. 56, no. 4, pp. 1015–1022, 2009.

[32] J. R. Orozco-Arroyave, S. M. Rendón, J. F. Vargas-Bonilla, E. Delgado-Trejos, J. D. Arias-Londoño, and G. Castellanos-Domínguez, "Nonlinear dynamics for hypernasality detection," in *Non-Linear Speech Processing*, 2011.

[33] J. R. Orozco-Arroyave, F. Hönig, J. Arias-Londoño, J. Vargas-Bonilla, K. Daqrouq, S. Skodda, J. Rusz, and E. Nöth, "Automatic detection of parkinson's disease in running speech spoken in three different languages," *The Journal of the Acoustical Society of America*, vol. 139, no. 1, pp. 481–500, 2016.

[34] H. P. Rowe, S. Shellikeri, Y. Yunusova, K. V. Chenausky, and J. R. Green, "Quantifying articulatory impairments in neurodegenerative motor diseases: A scoping review and meta-analysis of interpretable acoustic features," *International Journal of Speech-Language Pathology*, vol. 0, no. 0, pp. 1–14, 2022.

[35] P. A. Pérez-Toro, J. C. Vásquez-Correa, T. Arias-Vergara, P. Klumpp, M. Sierra-Castrillón, M. E. Roldán-López, D. Aguillón, L. Hincapié-Henao, C. A. Tóbon-Quintero, T. Bocklet, M. Schuster, J. R. Orozco-Arroyave, and E. Nöth, "Acoustic and linguistic analyses to assess early-onset and genetic alzheimer's disease," in *ICASSP 2021 - 2021 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2021, pp. 8338–8342.

[36] J. Chen, J. Ye, F. Tang, and J. Zhou, "Automatic detection of alzheimer's disease using spontaneous speech only," in *Interspeech*, vol. 2021. NIH Public Access, 2021, p. 3830.

[37] S. Luz, S. De La Fuente Garcia, and P. Albert, "A method for analysis of patient speech in dialogue for dementia detection," in *Resources and ProcessIng of linguistic, para-linguistic and extra-linguistic Data from people with various forms of cognitive impairment*. European Language Resources Association (ELRA), Jun 2018, pp. 35–42.

[38] N. Shivhare, S. Rathod, and M. R. Khan, "Automatic speech analysis of conversations for dementia detection using lstm and gru," in *2021 International Conference on Computational Intelligence and Computing Applications (ICCICA)*, 2021, pp. 1–7.

[39] F. Eyben, M. Wöllmer, and B. Schuller, "Opensmile: the munich versatile and fast open-source audio feature extractor," in *Proceedings of the 18th ACM international conference on Multimedia*, 2010, pp. 1459–1462.

[40] C. Bhat and S. K. Kopparapu, "Identification of alzheimer's disease using non-linguistic audio descriptors," in *2019 27th European Signal Processing Conference (EUSIPCO)*. IEEE, 2019, pp. 1–5.

[41] A. Triapthi, R. Chakraborty, and S. K. Kopparapu, "Dementia classification using acoustic descriptors derived from subsampled signals," in *2020 28th European Signal Processing Conference (EUSIPCO)*. IEEE, 2021, pp. 91–95.

[42] I. Martínez-Nicolás, T. E. Llorente, F. Martínez-Sánchez, and J. J. G. Meilán, "Ten years of research on automatic voice and speech analysis of people with alzheimer's disease and mild cognitive impairment: A systematic review article," *Frontiers in Psychology*, vol. 12, 2021.

[43] M. Sayadi, V. Varadarajan, M. Langarizadeh, G. Bayazian, and F. Torabinezhad, "A systematic review on machine learning techniques for early detection of mental, neurological and laryngeal disorders using patient's speech," *Electronics*, vol. 11, no. 24, 2022. [Online]. Available: https://www.mdpi.com/2079-9292/11/24/4235

[44] T. B. Holmlund, C. Chandler, P. W. Foltz, A. S. Cohen, J. Cheng, J. C. Bernstein, E. P. Rosenfeld, and B. Elvevåg, "Applying speech technologies to assess verbal memory in patients with serious mental illness," *npj Digital Medicine*, 2020.

[45] M. Song, A. Triantafyllopoulos, Z. Yang, H. Takeuchi, T. Nakamura, A. Kishi, T. Ishizawa, K. Yoshiuchi, X. Jing, V. Karas, Z. Zhao, K. Qian, B. Hu, B. W. Schuller, and Y. Yamamoto, "Daily mental health monitoring from speech: A real-world japanese dataset and multitask learning analysis," in *ICASSP 2023 - 2023 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2023, pp. 1–5.

[46] V. Ravuri, R. Gutierrez-Osuna, and T. Chaspari, "Preserving mental health information in speech anonymization," in *2022 10th International Conference on Affective Computing and Intelligent Interaction Workshops and Demos (ACIIW)*, 2022, pp. 1–8.