Detection of depression based on speech signal samples

Marina Galanina, Anna Rekiel, Samuel Szurman, Anna Bączyk
Faculty of Electronics Telecommunications and Informatics, Gdansk University of Technology
Gdansk Siedlicka street 5a, Poland
s183053@student.pg.edu.pl, s175635@student.pg.edu.pl, s175702@student.pg.edu.pl, s180489@student.pg.edu.pl

Abstract— Depression is a significant mental health issue and presents a challenge for the machine learning field in the detection of this illness. This study investigates automatic depression classification using methods like SVM (Support Vector Machine), Random Forest, MLP (Multilayer Perceptron), DepAudioNet, autoencoders, and spectrogram analysis. It analyzes four datasets - DAIC-WOZ, EATD Corpus, D-Vlog and EMU- which exhibit variation in language (English and Chinese), depression classification scales and gender content proportion. Moreover, two of the datasets (EMU, D-Vlog) contain features extracted from recordings, whereas others are raw speech recordings. First, feature extraction was performed consisting of parameters, such as formant-related, MFCCs (Mel Frequency Cepstral Coefficients), and jitter. An in-depth analysis showed that the combination of a feature vector and Random Forest returns the best classification results in terms of accuracy, sensitivity, and F1 score metrics.

Keywords—mental disorders, depression, speech, natural language processing, artificial neural networks, machine learning

I. Introduction

Depression is one of the most common mental disorders in the world. According to the World Health Organization (WHO), more than 264 million people worldwide and 5% of adults suffer from depression [1]. Moreover, depression affects people of all ages: about 2 million children between the ages of 3 and 17 suffer from the disease [20]. To put this into perspective, this number is equivalent to the population of a major city like Warsaw. In 2021, data from the National Health Service disclosed that services with a principal or co-existing diagnosis of depression were provided to a staggering 682,000 patients. In psychiatry, depressive states are referred to as "endogenous depression," "major depression," or "depressive episode." They are characterized by lowered mood, loss of interest or pleasure, changes in weight or appetite, sleep disturbances, fatigue, feelings of worthlessness, difficulty concentrating, and thoughts of death or suicide [2], [3]. Alarmingly, less than 25% of people with depression receive adequate treatment, and the lifetime risk of suicide is nearly 20% for people with untreated major depression [4]. According to the 2021 U.S. National Survey on Drug Use and Mental Health, an estimated 0.7% of adults aged 18 and older have made at least one suicide attempt [5]. Adding to the gravity of the situation, a survey of college students in Poland, showed alarming and concerning findings: 23.5% of those surveyed had scores indicative of mild depression, and 6.5% of those surveyed indicated symptoms of moderate to severe depression [19]. Looking ahead, it is estimated (WHO, 2017) that by 2030, depression may be at the top of the disease incidence.These numbers emphasize the urgent need for a machine learning solution to address the challenge of depression detection on a global scale.

On the basis of epidemiological data and screening, it is estimated that depression is, on average, two to three times more common in women than in men in developed Western societies [6]. Furthermore, the WHO highlights that women who have given birth experience depression in approximately 10% of cases, underscoring the importance of addressing this specific demographic within the realm of mental health care and research. Disparities in the number of diagnoses of depression in women and men are explained by reference to biological and psychosocial factors. In contrast, the different symptoms of depression in women and men are seen as a result of socialization into gender roles. In the course of "female depression," socially or personally unacceptable anger is often masked by sadness and anxiety. In the case of "male depression," sadness and anxiety are masked by anger, and tension is relieved through impulsive behavior [7].

The traditional method of diagnosing depression is performed by a psychiatrist or psychotherapist and involves interviewing the patient. There are also various tests, such as the PHQ-8 and PHQ-9, which involve filling out a form, allowing diagnosis of depression with a high degree of certainty [8], [9]. In the study performed, we focus on diagnosing depression based on the extracted parameters of audio recordings, building on research showing distinct speech patterns in depression across languages [10], [11], [12]. Characterized by a lower, monotonous tone, slower pace, and increased pauses, these speech features are integral to our analysis[13], [14]. From existing studies on depression detection based on speech signal analysis, it has been noted that gender- and age-dependent models have better results [15]. Our study aims to leverage these insights, focusing on gender factors in speech signal analysis to enhance the accuracy of depression detection.

Our paper is organized as follows. Section II explores the depression in number. In Section III we delve into Acoustic Differences in Speech, highlighting the various vocal characteristics associated with depressive states. Section IV provides insights into the databases, section V shows the Differences Between Women and Men in terms of how depression may affect their speech. In Section VI, we conduct an in-depth Analysis of Methods, whereas Section VII focuses on the Analysis of spectrograms. Finally, in Section VIII, we present the Analysis of the Best Methods and discuss the most effective approaches for depression detection.

II. Acoustic Differences in Speech

The human voice is not only a vehicle for speech but also contains information about gender and emotions,

XXX-X-XXXX-XXXX-X/XX/$XX.00 ©20XX IEEE

among other features, and is unique to each person. Differences in acoustic characteristics between male and female voices are internationally noticeable but may undergo some variation depending on language and culture. In Anglo-Saxon languages, such as English, there are distinctive acoustic characteristics that affect the perception of gender in speech. The female voice exhibits higher levels of respiratory noise, causing it to be perceived as more "rustling" than the male voice [23], [24]. The lower resonant frequencies of the larynx in men result in a more bassy voice, while the higher resonant frequencies of the larynx in women give their voice a brighter, soprano tone. These differences are due to biological differences between the genders. The larynx, which contains the vocal folds, differs in structure between men and women. In men, the vocal folds tend to be longer and thicker, generating lower sounds. In women, the folds are shorter and thinner, which promotes higher sounds. Also, men tend to have more muscle mass in the larynx, which affects sound resonances [25]. Different languages have specific phonetic features that affect gender perception in speech. For example, tonal languages, such as Mandarin or Japanese, are characterized by tonal accent, meaning that a change in tone can alter the meaning of a word. In such languages, even though there are similarities in biological gender differences, tonality makes the difference in perception compared to Anglo-Saxon languages.

Examples of spectrograms are shown in Figure 1 a) (female speech sample), and Figure 1 b) (male speech sample). The recordings are from the RAVDESS database [33]. This is based on the same utterance "Kids are talking by the door".

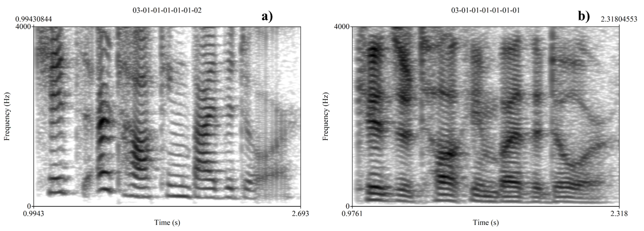

Figure 1: *Image of narrowband Spectrogram based on the same utterance "Kids are talking by the door" a) for Female Voice, b) for Male Voice.*

The spectrograms shown in Figure 1 a) and b) differ from each other. There are more prominent harmonic components and formant structures in the spectrogram of the male voice. This is due to the lower tone and resonance.

III. Databases

In the study several datasets were employed, which contain recordings of individuals diagnosed with depression and the control group without depression. They are as follows: DAIC-WOZ [34], EATD Corpus [37], D-Vlog [35], and EMU [36].

DAIC-WOZ (Distress Analysis Interview Corpus) contains recordings in English of interviews conducted as part of the diagnosis of mental disorders such as anxiety, depression, and post-traumatic stress syndrome [38]. The interviews were recorded in both audio and video form, and the statements were also transcribed in text form. Each interview is annotated with both verbal and non-verbal features. The entire database consists of 189 folders, and each folder contains data on one interview. The criterion that assessed whether or not a person was suffering from depression was a score on the PHQ-8 (Personal Health Questionnaire Depression Scale) test [27]. If a patient scored 10 or more on this test, he or she was considered to be suffering from depression.

EATD corpus is a Chinese database containing audio files and transcriptions from interviews with 162 willing students from Tongji University [28]. Each volunteer randomly selected 3 questions and completed the SDS (Self-Rating Depression Scale) questionnaire [29], [30], [31]. It contains 20 questions on various aspects of depression. A score from this questionnaire multiplied by 1.25 equal to or greater than 53 indicates depression. Relative to this criterion, 30 people surveyed suffer from depression, and the remaining 132 people have insufficient symptoms. The entire database contains about 2.26 hours.

D-Vlog is a multimodal collection designed to diagnose depression based on non-verbal signals [32]. It contains acoustic and visual features extracted from 961 vlogs in English downloaded from YouTube. They total about 160 hours and come from 816 different speakers. Classification of whether or not a speaker was depressed was based on the phrases the speakers used and whether or not they had symptoms of depression during the vlog. Those classified as suffering from depression used phrases such as "I have suicidal thoughts" or "I finally changed my medication." The final database consists of 555 vlogs from people diagnosed with depression (182 men and 373 women) and 406 vlogs from people without depression (140 men and 266 women).

EMU is a database that is designed for both depression and generalized anxiety syndrome diagnoses [39]. The data include the results of the PHQ-9 questionnaire for depression and the results of the GAD-7 [40] questionnaire for generalized anxiety syndrome. For each subject, two types of audio recordings were registered. In the first case, the task was to read a Shakespeare quote, and in the second case, the task was to describe a good friend in their own words. Finally, the database contains both the acoustic features of the recordings and personal data, such as age, gender, and education.

IV. Differences Between Women and Men

The purpose of this Section is to explore in depth how gender affects the speech of people suffering from depression. Several voice features such as, e.g., tone of voice, rate of speech and overall communication style change in the context of the depressive experience. Symptoms of depression in men and women will also be examined in order to understand the social differences that can directly affect speech. As already said, depression, as a disorder affecting both sexes, manifests itself in different ways, which is also reflected in the way they communicate. Understanding these differences is key to better diagnosing and treating depression and tailoring therapy to patients' needs.

The main effect of psychological gender is often noted in many studies. Psychological gender is a person's self-identification in independence of biological sex. The statistics are shown below in Table 1.

The numbers shown in parentheses ($F[3,53] = 7.77$) indicate the value of the F statistic in the ANOVA test. $F[3.53]$ means that 3 groups were used in the test (this number may refer to different levels of psychological gender or other variables in the model) and that there were 53 degrees of freedom in the study (total number of observations minus the number of groups). The value of 7.77 is the F-statistic, which is relatively high, suggesting statistically significant differences between groups. A "$p < 0.001$" value indicates that the result is statistically significant. In practice, p-values less than 0.05 (or 5%) are usually considered statistically significant, meaning that the observed differences are probably not random. In contrast, $\eta^2 = 0.31$ refers to a measure of effect size, in this case, the eta squared coefficient. A value of 0.31 indicates a moderately strong effect, meaning that psychological gender has a significant effect on levels of depression [21].

In general, many symptoms of depression manifest differently according to gender, as can be seen in Table 1 [22]. Women have stronger clinical symptoms compared to men. Based on the IDS (Depression Inventory) scale, it could be deduced that the symptoms of depression for women – 36.3±11.4, while for men, it is 23.7±11.3 on the above scale. The difference is about 10% [22].

Table 1 : *Psychological gender impact on depression symptoms.*

Symptom	Men [%]	Women [%]
Increased Appetite	16.3	24.1
Weight Gain	19.4	25.7
Energy/Exhaustion	84.9	91.4
Somatic Complaints (pain)	73.2	79.8
Gastrointestinal Perception	32.2	45.3
Interpersonal Sensitivity	55.9	64.7

In the context of gender, it is also noteworthy that rarely are sets balanced in the context of gender. For example, the unequal sex ratio in the DAIC-WOZ dataset is an important factor that can affect analyses and conclusions based on these data. Moreover, in the case of gender bias, such inappropriateness can influence the quality of the research and the reliability of the results. In DAIC-WOZ, the distribution of classes reveals notable patterns. It can be seen that the ratio of female D:ND (Depression:Not Depressed) is roughly 5:8, and the ratio of male D:ND in the dataset is 3:8.

The incidence of people with depression in the dataset varies by gender, such that $p(D \mid g = f) > p(D \mid g = m)$, where g is gender, f stands for female, and m stands for male.

Some publications suggest that it is possible to divide a given subset into two groups: men and women, separately, and then perform the data processing [26]. Unfortunately, this work's goal is to develop a set of gender-dependent vowel-level features instead of focusing on a broad range of speech processing techniques.

In the first step, metrics were analyzed for two identical models dedicated to women and men. The same model was used to accurately reflect the differences in results. The RandomForest algorithm was tested for the parameters *n*_estimators=140 and random_state=42. The results are presented in the form of a confusion matrix for women – Figure 3 and for men – Figure 4.

The link to the social context of depression is crucial. Men often try to hide their condition, which can result in unpredictable behavior and even suicide. In contrast, depression in women is usually not as strongly stigmatized. When faced with depression, men often turn to alcohol or drugs [21], which can affect slowed speech, monotony and lack of emotional expression. In order to more accurately predict depression in gender groups, it is recommended to diversify the datasets with recordings that include both men and women to better reflect social diversity and improve the overall efficiency of the analysis.

In the case of the EMU database, there was also a problem of gender inequality among individuals, but in this case there were acutely fewer women than men. Analogous training was performed using the RandomForest algorithm for both the entire base and separately for women and separately for men. Separate classification yielded better results.

V. Analysis of Methods

To address the issue, various methods were tested, initially without gender division and subsequently by dividing the dataset based on gender. Not all methods were applied to every dataset. The tested methods encompassed the Multilayer Perceptron (MLP), Random Forest (RF), and Support Vector Machine (SVM). Furthermore, the MLP underwent an additional evaluation on spectrograms of audio recordings.

For the autoencoder, its role was to perform feature extraction by transforming data from recordings into a latent space. Subsequently, this representation was utilized for classification purposes employing the Support Vector Machine (SVM) algorithm.

Table 2 : *Comparisons of the most successful machine learning classifiers.*

Method	DAIC-WOZ	EATD Corpus	D-Vlog	EMU
SVM	66.15%	71.05%	63.16%	80%
Random Forest with gender split	86.84%	56.8%	86.84%	77.78%
DepAudioNet-like	72.34%	60.57%	61.82%	63.78%
Perceptron with gender split	53.478%	70%	56.79%	62.67%

Our first proposed model will be based on “DepAudioNet”, which was one of the best models in AVEC 2016 to utilize deep learning with audio-only processing for depression detection. In the DepAudioNet model, convolutional layers were used, but the size, number of filters, and activation functions were not specified. In our architecture a Conv1D layer with 64 filters, a kernel size of 3, and a ReLU activation function was employed, followed by a second Conv1D layer with 128 filters. necessary to mention, that DepAudioNet includes a single MaxPooling layer, whereas the provided architecture consists of two MaxPooling layers. Changes resulted in performance of the model.

As an alternative approach, Support Vector Machines (SVMs) were employed. SVMs are particularly effective for binary classification tasks. Moreover, they can handle high-dimensional feature spaces effectively, making them suitable for speech processing where many features are involved.

In the case of Random Forest, we investigated two versions: one with gender-based data separation and one without it. The reason for this was that initially, metrics were analyzed for two identical models, one for females and one for males and it was observed that the male dataset, under the same conditions, resulted in lower accuracy compared to the female dataset, specifically when using the RandomForest classifier. The testing was conducted with the following parameters:n_estimators=140, random_state=42. For the female dataset, the accuracy was approximately 76.47%, while for the male dataset, it was only 57.14%. This discrepancy in accuracy between genders prompted further investigation into potential factors contributing to this difference.

The next applied algorithm was the MLP, both for data with gender division and without division. To achieve better results, we employed the Synthetic Minority Over-sampling Technique (SMOTE) on the training set and restricted recordings of individuals without depression. This MLP network is structured with five layers. Crucially, dividing the dataset based on gender significantly augmented the model's efficacy, elevating accuracy from 59.574% to 70%. The gender-based dataset division positively influenced the model's effectiveness in distinguishing individuals with and without depression.

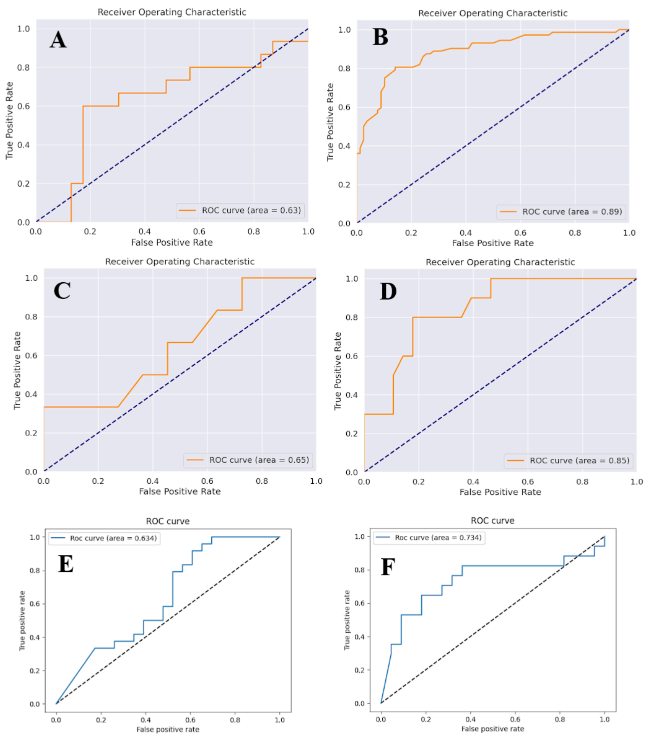

Figure 1: *Comparisons of the most successful machine learning classifiers with and without gender subgroups A) DepAudioNet model on DAIC-WOZ corpus, without subgroups, B) DepAudioNet model on DAIC-WOZ corpus, with subgroups C) RF classifier, without subgroups D) RF classifier, with subgroups E) MLP classifier, without subgroups D)MLP classifier, with subgroups.*

VI. Analysis of spectrograms

Each recording from DAIC-WOZ dataset and EATD Corpus was converted to a spectrogram and then given as input to the convolutional neural network. The manner in which the spectrogram was generated had little effect on the effectiveness of the network. Better results were obtained for the EATD Corpus compared to the DAIC-WOZ dataset. This may be the result of more advanced processing of the recordings, where recordings from the DAIC-WOZ were cut to remove sections with questions from the interviewee (i.e., doctor), as well as processed to eliminate noise. Processing these recordings could potentially have negatively affected the quality of the generated spectrograms. Another major problem is the limited amount of data available to train the models. The spectrograms have a much larger amount of data than the extracted feature vectors, making it more difficult to select relevant information. In contrast, feature vectors may contain more crucial and precise information necessary for effective depression diagnosis.

However, the classification of depression based on spectrograms seems to be an interesting area for further research. Increasing the amount of data available for learning and developing a dedicated neural network architecture for analyzing spectrogram images could yield very good results.

VII. Conclusions and Summary

The research study to develop a method for diagnosing depression based on a recording has produced noteworthy results and conclusions. The results suggest that including gender in the analysis of depression is important, and that models focused on gender subgroups may be more effective. However, data volume limitations and differences in spectrogram analysis are areas for further research and study development. Introducing a greater variety of data and adapting a dedicated neural network architecture to analyze spectrogram images could further improve the system's performance.

Splitting the dataset by gender has significantly improved the effectiveness of models in diagnosing depression. There is a noticeable improvement in accuracy, sensitivity, and precision. The inclusion of gender in the analysis of depression, both in terms of acoustic features and symptoms, may be a key component of effective diagnostic models. As already mentioned, when working with DAIC-WOZ, there is a need to pay special attention to the unequal gender distribution in this database. To optimize the analysis, one can consider extracting different features of the speech signal for men and women in the OpenSmile tool, for example, for women: shimmerLocal − speed and variety of modulation, pauseDuration − length of pauses, F0final_sma − treble tones. For men: F0env − monotonicity, F0final_sma − low tones.

In conclusion, the presented research project shows significant potential for using analysis of acoustic speech features and gender considerations in automatic depression diagnosis. The introduction of a greater variety of data and more sophisticated audio processing techniques could further develop effective tools to support the diagnosis and treatment of depression.

VIII. Future works

The next step in our project involves creating a tool for psychiatrists to use in their daily work—a monitoring application that enables the assessment of patients' mental health not only during visits but also online, through responses to a set of recorded questions. Such an application could facilitate the inclusion of a greater variety of data and additional languages, such as Polish. Additionally, the proposed introduction of depression detection methods that consider psychological gender is a noteworthy avenue for future exploration. Integrating these methods into our framework could provide a more comprehensive and nuanced understanding of depression, contributing to the refinement and expansion of our diagnostic capabilities.

REFERENCES

[1] World Health Organization: WHO & World Health Organization: WHO. (2023a, March 31). Depressive disorder (depression). https://www.who.int/news-room/fact-sheets/detail/depression (Accessed 2023 November 08)

[2] J. Rybakowski, S. Pużyński, J. Wciórka: Psychiatria. T. 2. Wrocław: Elsevier Urban & Partner, 2010, s. 305–375. ISBN 978-83-7609-102-0

[3] American Psychiatric Association, DSM-5 Task Force, Diagnostic and statistical manual of mental disorders: DSM-5™ (5th ed.). American Psychiatric Publishing, Inc.. 2013. https://doi.org/10.1176/appi.books.9780890425596

[4] American Association for Suicidology. (2009, June 23). Some Facts About Suicide and Depression. https://suicidology.org. https://www.cga.ct.gov/asaferconnecticut/tmy/0129/Some%20Facts%20About%20Suicide%20and%20Depression%20-%20Article.pdf (Accessed 2023 November 08)

[5] Suicide statistics. (2023, July 10). American Foundation for Suicide Prevention. https://afsp.org/suicide-statistics/ (Accessed 2023 November 08)

[6] Gove W.: Gender differences in mental and physical illness: The effects of fixed roles and nurturant roles. Soc. Sci. Med., 1984, 19(2), 77-84

[7] Frąckowiak-Sochańska M.: Rodzinne i społeczno-kulturowe uwarunkowania zaburzeń psychicznych – analiza z perspektywy płci społeczno-kulturowej. Rocz. Socjol. Rodz., 2010, tom XX, s. 153-185.

[8] Shin C, Lee SH, Han KM, Yoon HK, Han C. Comparison of the Usefulness of the PHQ-8 and PHQ-9 for Screening for Major Depressive Disorder: Analysis of Psychiatric Outpatient Data. Psychiatry Investig. 2019;16(4):300-305. https://doi.org/10.30773/pi.2019.02.01

[9] Sun Y, Fu Z, Bo Q, Mao Z, Ma X, Wang C. The reliability and validity of PHQ-9 in patients with major depressive disorder in psychiatric hospital. BMC Psychiatry. 2020;20(1):474. Published 2020 Sep 29. https://doi.org/10.1186/s12888-020-02885-6

[10] Gábor, K., Vicsi, K., Mono- and multi-lingual depression prediction based on speech processing. International Journal of Speech Technology, 2017. https://doi.org/20.10.1007/s10772-017-9455-8

[11] Cannizzaro M, Harel B, Reilly N, Chappell P, Snyder PJ. Voice acoustical measurement of the severity of major depression. Brain Cogn. 2004;56(1):30-35. https://doi.org/10.1016/j.bandc.2004.05.003

[12] Flint AJ, Black SE, Campbell-Taylor I, Gailey GF, Levinton C. Abnormal speech articulation, psychomotor retardation, and subcortical dysfunction in major depression. J Psychiatr Res. 1993;27(3):309-319. https://doi.org/10.1016/0022-3956(93)90041-y

[13] 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 Neurolinguistics. 2007;20(1):50-64. https://doi.org/10.1016/j.jneuroling.2006.04.001

[14] Yamamoto M, Takamiya A, Sawada K, et al. Using speech recognition technology to investigate the association between timing-related speech features and depression severity. PLoS One. 2020;15(9):e0238726. Published 2020 Sep 11. https://doi.org/10.1371/journal.pone.0238726

[15] B. Stasak, D. Joachim and J. Epps, "Breaking Age Barriers With Automatic Voice-Based Depression Detection," in IEEE Pervasive Computing, vol. 21, no. 2, pp. 10-19, 1 April-June 2022, doi: 10.1109/MPRV.2022.3163656.

[16] Amiri S. Unemployment associated with major depression disorder and depressive symptoms: a systematic review and meta-analysis. Int J Occup Saf Ergon. 2022;28(4):2080-2092. doi:10.1080/10803548.2021.1954793

[17] Karol Chlasta, Krzysztof Wołk, Izabela Krejtz, Automated speech-based screening of depression using deep convolutional neural networks, Procedia Computer Science, Volume 164, 2019, Pages 618-628, ISSN 1877-0509, https://doi.org/10.1016/j.procs.2019.12.228.

[18] Liu, C.H.; Zhang, E.; Wong, G.T.F.; Hyun, S. Factors associated with depression, anxiety, and PTSD symptomatology during the COVID-19 pandemic: Clinical implications for US young adult mental health. Psychiatry Res. 2020, 290, 113172

[19] Karmolińska-Jagodzik, Ewa. (2019). Zaburzenia depresyjne wśród studentek I studentów w Polsce. Studia Edukacyjne. 139-162. 10.14746/se.2019.55.9.

[20] Depression statistics. Depression and Bipolar Support Alliance. (2019, July 12). https://www.dbsalliance.org/education/depression/statistics/. Accessed 6/06/2023.

[21] Szpitalak, Malwina, and Katarzyna Prochwicz. "Płeć psychologiczna osób z depresją kliniczną. Doniesienia wstępne.". Psychiatria Polska 47 no. 1 (2013): 53-64.

[22] Marcus SM, Kerber KB, Rush AJ, Wisniewski SR, Nierenberg A, Balasubramani GK, Ritz L, Kornstein S, Young EA, Trivedi MH. Sex differences in depression symptoms in treatment-seeking adults: confirmatory analyses from the Sequenced Treatment Alternatives to Relieve Depression study. Compr Psychiatry. 2008 May-Jun;49(3):238-46. Doi: 10.1016/j.comppsych.2007.06.012. Epub 2008 Mar 4. PMID: 18396182; PMCID: PMC2759282.

[23] Mendoza, E., Valencia, N., Muñoz, J., & Trujillo, H. (1996). Differences in voice quality between men and women: use of the long-term average spectrum (LTAS). Journal of voice : official journal of the Voice Foundation, 10(1), 59–66. https://doi.org/10.1016/s0892-1997(96)80019-1

[24] Simpson, Adrian. (2009). Breathiness differences in male and female speech. Is H1-H2 an appropriate measure?.

[25] Li, S., Scherer, R. C., Wan, M., Wang, S., & Song, B. (2020). Intraglottal Pressure: A Comparison Between Male and Female Larynxes. Journal of voice : official journal of the Voice Foundation, 34(6), 813–822. https://doi.org/10.1016/j.jvoice.2019.06.005

[26] Interspeech 2017, 18th Annual Conference of the International Speech Communication Association, Stockholm, Sweden, August 20-24, 2017. Wydawca: ISCA. Redaktor: Francisco Lacerda. DOI: 10.21437/Interspeech.2017. URL: https://doi.org/10.21437/Interspeech.2017

[27] Kroenke, K., Strine, T. W., Spitzer, R. L., Williams, J. B., Berry, J. T., & Mokdad, A. H. (2009). The PHQ-8 as a measure of current depression in the general population. Journal of affective disorders, 114(1-3), 163–173. https://doi.org/10.1016/j.jad.2008.06.026

[28] Shen, Y., Yang, H., & Lin, L. (2022). Automatic Depression Detection: An Emotional Audio-Textual Corpus and a GRU/BiLSTM-based Model. arXiv [Eess.AS]. Retrieved from http://arxiv.org/abs/2202.08210

[29] Jokelainen, J., Timonen, M., Keinänen-Kiukaanniemi, S., Härkönen, P., Jurvelin, H., & Suija, K. (2019). Validation of the Zung self-rating depression scale (SDS) in older adults. Scandinavian journal of primary health care, 37(3), 353–357. https://doi.org/10.1080/02813432.2019.1639923

[30] Zung, W. W., & Zung, E. M. (1986). Use of the Zung Self-rating Depression Scale in the elderly. Clinical Gerontologist: The Journal of Aging and Mental Health, 5(1-2), 137–148. https://doi.org/10.1300/J018v05n01_07

[31] Campo-Arias, Adalberto & Diaz Martinez, Luis & Eduardo, German & Rueda-Jaimes, German & Cadena, Laura & Hernández, Nubia. (2006). Validation of Zung Self-Rating Depression Scale among the Colombian general population. Social Behavior and Personality: an international journal. 34. 87-94. 10.2224/sbp.2006.34.1.87.

[32] Yoon, J., Kang, C., Kim, S., & Han, J. (2022). D-vlog: Multimodal Vlog Dataset for Depression Detection. Proceedings of the AAAI Conference on Artificial Intelligence, 36(11), 12226-12234. https://doi.org/10.1609/aaai.v36i11.21483

[33] Livingstone, S. R. (2019, January 19). Ravdess emotional speech audio. Kaggle. https://www.kaggle.com/datasets/uwrfkaggler/ravdess-emotional-speech-audio

[34] Home - daic. WOZ. (2022, March 7). https://dcapswoz.ict.usc.edu/

[35] Google. (n.d.). DVLOG-dataset.zip. Google Drive. https://drive.google.com/file/d/1IPBs-tLWCpRDg8ZNPS3iUeWGBCEh895g/view?usp=sharing

[36] Tlachac, M. L., Toto, E., Lovering, J., Kayastha, R., Taurich, N., & Rundensteiner, E. (2021). EMU: Early Mental Health Uncovering Framework and Dataset. 20th IEEE International Conference of Machine Learning Applications (ICMLA) Special Session Machine Learning in Health.

[37] Speechandlanguageprocessing. (n.d.). Speechandlanguageprocessing/ICASSP2022-depression: Automatic depression detection: A GRU/ BILSTM-based model and an emotional audio-textual corpus. GitHub. https://github.com/speechandlanguageprocessing/ICASSP2022-Depression

[38] Gratch, J., Artstein, R., Lucas, G., Stratou, G., Scherer, S., Nazarian, A., … Morency, L.-P. (2014, May). The Distress Analysis Interview Corpus of human and computer interviews. In N. Calzolari, K. Choukri, T. Declerck, H. Loftsson, B. Maegaard, J. Mariani, … S. Piperidis (Eds.), Proceedings of the Ninth International Conference on Language Resources and Evaluation (LREC'14) (pp. 3123–3128). Retrieved from http://www.lrec-conf.org/proceedings/lrec2014/pdf/508_Paper.pdf

[39] M. Tlachac, E. Toto, J. Lovering, R. Kayastha, N. Taurich and E. Rundensteiner, "EMU: Early Mental Health Uncovering Framework and Dataset," 2021 20th IEEE International Conference on Machine Learning and Applications (ICMLA), Pasadena, CA, USA, 2021, pp. 1311-1318, doi: 10.1109/ICMLA52953.2021.00213.

[40] Robert L. Spitzer, Janet B.W. Williams, & Kurt Kroenke. (n.d.). Patient Health Questionaire and general anxiety disorder (PHQ-9 and gad-7). https://uhs.fsu.edu/sites/g/files/upcbnu1651/files/docs/PHQ-9%20and%20GAD-7%20Form_a.pdf