DESIGN AND DEVELOPMENT OF A SPEECH-BASED DIARY FOR DEPRESSION RELAPSE PREVENTION

Kyra Kannen[1], Zohre Azimi[1], Muhammad Saif-Ur-Rehman[1], Sonja Dana Roelen[2], Sebastian Schnieder[2], Steffen Holsteg[3], André Karger[3], Philip Mildner[4], Jens Piesk[4], & Jarek Krajewski[1]
[1]*Institute of experimental Psychophysiology (Germany)*
[2]*University of Applied Sciences for Media Communication and Business Cologne (Germany)*
[3]*Clinical Institute of Psychosomatic Medicine and Psychotherapy, Heinrich-Heine-University Düsseldorf (Germany)*
[4]*Nuromedia GmbH (Germany)*

Abstract

Background. Relapse is a significant risk among people treated for depression, with approximately 50% of patients experiencing a recurrence after their initial depressive episode. This risk increases with subsequent episodes. This alarming statistic underscores the urgent need for innovative, reliable tools to identify individuals at high risk of relapse for timely intervention. Leveraging the capabilities of machine learning to analyze speech patterns offers a novel pathway for early detection of changes in depression status to prevent relapse.
Methods. We trained two machine learning models on speech data from patients with depression. The first, a Depression Detection System (DDS), is trained on speech data from patients with four levels of depression, as assessed by the Beck Depression Inventory (BDI) score (from non-depressed to severely depressed). The second, Speech Emotion Recognition (SER) model, employs deep learning techniques to identify seven basic emotions. Additionally, we developed a speech-based diary, where patients reflect on their week and address topics related to their condition. The recorded diary entries provide a rich dataset for our models to analyze depression-related voice parameters. This approach offers the opportunity to monitor changes in patients' symptoms and enables professionals to take timely action and prevent relapse.
Results. The performance of models achieved a satisfactory level, as measured by the $F1$-score (DDS: 81% averaged across four classes, and SER: 62% averaged across seven emotion classes). These results highlight the potential efficacy of our approach in the nuanced task of speech-based emotion and depression detection.
Conclusion. This innovative approach may serve as a reliable tool for relapse prevention in depression. Empowering both therapists and patients with valuable insights from weekly speech-based self-reflection, this method holds potential for early intervention. As a next step, clinical trials will be necessary to validate the effectiveness of this speech-based diary in preventing depression relapse.

Keywords: *Depression, relapse prevention, acoustic speech-analysis, machine learning, speech emotion recognition.*

1. Introduction

Depression, affecting approximately 6.38% in Europe (Arias-de la Torre et al., 2021), poses a significant challenge due to its high recurrence rate post-recovery, escalating up to 85% within fifteen years (Kanai et al., 2003; Melartin et al., 2004, Mueller et al., 1999). Key predictors of relapse include the number of prior episodes and subclinical residual symptoms (Hardeveld et al., 2010). Beyond individual suffering, depression burdens society, notably increasing treatment costs by approximately 158% (König et al., 2019). Despite the need for professional therapy after relapse, the limited therapy slots do not allow access to proper treatment of such patients (Moitra et al., 2022). To address this gap, developing new monitoring tools for early detection of relapse and symptoms onset is essential. This study explores the use of machine learning for early detection of depression status, by analyzing speech patterns. Depression can affect speech, resulting in changes such as decreased speech volume, a slower speaking rate, or increased jitter (Sahu & Espy-Wilson, 2014; Silva et al., 2021). Similarly, emotions can influence a person's speech

p-ISSN: 2184-2205 e-ISSN: 2184-3414 ISBN: 978-989-35106-6-7 © 2024

production, leading to changes in prosody, spectral features, and voice quality (Akçay & Oğuz, 2020). Thus, we aimed to develop a speech-based diary and two machine learning models to assess depression severity and emotional states, offering a novel pathway for timely intervention and relapse prevention.

2. Methods

2.1. Design of a speech-based diary

The design process for a speech-based diary began with the development of a set of guided questions to explore the emotional responses and coping mechanisms of depressive patients. It is assumed that individuals with depression use maladaptive coping strategies, such as avoidance or rumination, while non-depressed individuals use adaptive methods like problem-solving and seeking social support (Joormann & Gotlib, 2010). In collaboration with mental health professionals and treatment providers, guided questions were developed that reflect emotional regulation and coping.

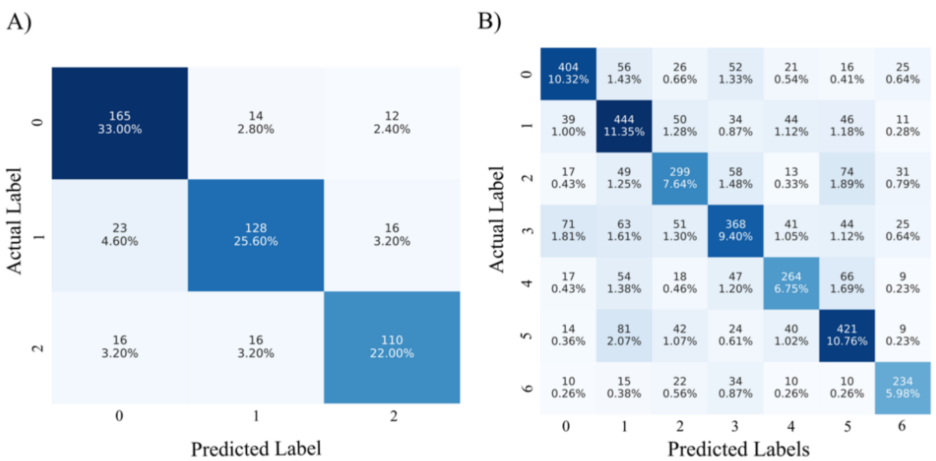
2.2. Integrating machine learning algorithms

Leveraging the capabilities of machine learning to analyze speech patterns offers a novel pathway for early detection of changes in depression status. To analyze the data recorded from the speech-based diary, we have developed two speech models. The first model, DDS model, employs an ensemble of machine learning classifiers – Support Vector Machine (SVM), K-Nearest Neighbors (K-NN), and Random Forest (RF) – to analyze audio features including Linear Predictive Coefficients (LPC), Fundamental Frequency, and Zero and Mean Crossing Rate. The DDS leverages OpenSmile (Eyben et al., 2010) for feature extraction and operates in two stages: initially differentiating *normal* from *depressed* states, then further classifying *depressed* signals into *mild*, *moderate*, or *severe*. The clinical dataset used for training and evaluation of the model was collected during a short story reading task and a free speech task and involves 7,250 audio data annotated with BDI (Beck et al., 2013) scores, which were assigned by expert psychiatrists. The second, SER model, utilizes deep learning to identify the emotional content of speech data. Trained on a diverse database of 25,596 emotional speech samples (acted/non-acted) in four different languages (English, German, Italian, and Japanese), SER can categorize speech data into one of the seven basic emotion classes (anger, disgust, fear, happiness, neutral, sadness, and surprise). The model is designed with a combined architecture of convolutional neural network and Long-Short-Term-Memory. Using these techniques, the model analyzes audio signals' Mel spectrogram and predicts the likely emotion class.

3. Results

Our evaluation shows that the models performed well in their respective tasks. The SER model attained an average F_1-score of 62% across seven emotion categories, which indicates a good capacity for emotion classification from speech. On the other hand, the DDS model achieved an impressive F_1-score of 81% in discerning among four levels of depression severity. These outcomes underscore the utility of our models in identifying crucial speech characteristics associated with emotional states and depression severity.

Figure 1. Confusion matrices for the (A) DDS and (B) SER models. Each matrix shows the number and percentage of samples from the test set that were correctly classified (diagonal) and misclassified (off-diagonal) by the models. The percentage is calculated as the number of samples in each category divided by the total number of samples. Darker shades represent higher percentages, illustrating the model's predictive accuracy for each class.



4. Discussion, outlook and conclusion

Given the high relapse rate of people with depression, we trained two models (SER and DDS) to monitor changes in symptoms. The robust performance of the models indicates their ability to detect fluctuations in emotion and depression levels by analyzing specific voice parameters in audio recordings, indicating the potential for continuous symptom monitoring to facilitate early relapse detection. As a next step, we will obtain audio data from a clinical trial using the developed diary to evaluate the model's performance on novel test data and to fine-tune the models. Additionally, we are assessing the diary's usability and user experience within the framework of human-centered design to determine its suitability for constant symptom monitoring.

In summary, our models can effectively classify depression levels and emotion from audio data, showing promise in preventing depression relapse. This approach offers valuable insights for a proactive intervention through weekly speech-based introspection, facilitating early relapse detection. However, its efficacy and reliability have yet to be evaluated in clinical trials.

References

Akçay, M. B., & Oğuz, K. (2020). Speech emotion recognition: Emotional models, databases, features, preprocessing methods, supporting modalities, and classifiers. *Speech Communication*, *116*, 56-76. https://doi.org/10.1016/j.specom.2019.12.001

Arias-de la Torre, J., Vilagut, G., Ronaldson, A., Serrano-Blanco, A., Martín, V., Peters, M., ... & Alonso, J. (2021). Prevalence and variability of current depressive disorder in 27 European countries: a population-based study. *The Lancet Public Health*, *6*(10), e729-e738. https://doi.org/10.1016/S2468-2667(21)00047-5

Beck, A.T., Brown, K.G, Steer, R.A. (2013). *Beck-Depressions-Inventar-FS (BDI-FS). Manual*. Deutsche Bearbeitung von Sören Kliem & Elmar Brähler. Pearson Assessment, Frankfurt am Main.

Eyben, F., Wöllmer, M., & Schuller, B. (2010, October). Opensmile: the munich versatile and fast open-source audio feature extractor. In *Proceedings of the 18th ACM international conference on Multimedia* (pp. 1459-1462). https://doi.org/10.1145/1873951.1874246

Hardeveld, F., Spijker, J., De Graaf, R., Nolen, W. A., & Beekman, A. T. F. (2010). Prevalence and predictors of recurrence of major depressive disorder in the adult population. *Acta Psychiatrica Scandinavica*, *122*(3), 184-191. https://doi.org/10.1111/j.1600-0447.2009.01519.x

Joormann, J., & Gotlib, I. H. (2010). Emotion regulation in depression: Relation to cognitive inhibition. *Cognition and Emotion*, *24*(2), 281-298. https://doi.org/10.1080/02699930903407948

König, H., König, H. H., & Konnopka, A. (2029). The excess costs of depression: a systematic review and meta-analysis. *Epidemiology and Psychiatric Sciences*, *29*, e30. https://doi.org/10.1017/S2045796019000180

Kanai, T., Takeuchi, H., Furukawa, T. A., Yoshimura, R., Imaizumi, T., Kitamura, T., & Takahashi, K. (2003). Time to recurrence after recovery from major depressive episodes and its predictors. *Psychological Medicine*, *33*(5), 839-845. https://doi.org/10.1017/S0033291703007827

Melartin, T. K., Rytsala, H. J., Leskela, U. S., Lestela-Mielonen, P. S., Sokero, T. P., & Isometsa, E. T. (2004). Severity and comorbidity predict episode duration and recurrence of DSM-IV major depressive disorder. *Journal of Clinical Psychiatry*, *65*(6), 810-819. https://doi.org/10.4088/JCP.v65n0612

Moitra, M., Santomauro, D., Collins, P. Y., Vos, T., Whiteford, H., Saxena, S., & Ferrari, A. J. (2022). The global gap in treatment coverage for major depressive disorder in 84 countries from 2000–2019: A systematic review and Bayesian meta-regression analysis. *PLoS medicine*, *19*(2), e1003901. https://doi.org/10.1371/journal.pmed.1003901

Mueller, T. I., Leon, A. C., Keller, M. B., Solomon, D. A., Endicott, J., Coryell, W., ... & Maser, J. D. (1999). Recurrence after recovery from major depressive disorder during 15 years of observational follow-up. *American Journal of Psychiatry*, *156*(7), 1000-1006. https://doi.org/10.1176/ajp.156.7.1000

Sahu, S., & Espy-Wilson, C. (2014). Effects of depression on speech. *The Journal of the Acoustical Society of America*, *136*(4), 2312-2312. https://doi.org/10.1121/1.4900372

Silva, W. J., Lopes, L., Galdino, M. K. C., & Almeida, A. A. (2021). Voice acoustic parameters as predictors of depression. *Journal of Voice*, *38*(1), 77-85. https://doi.org/10.1016/j.jvoice.2021.06.018