Ensofy

FUTURE ROLE OF AI AND VOICE TECHNOLOGY IN MANAGING MENTAL HEALTH DISORDERS

WHITEPAPER

HTTPS://WWW.ENSOFY.AI/

CONTENTS

EXECUTIVE SUMMARY

Depression is a global burden that severely affects individuals' mental health. Recently, its prevalence grew significantly due to the pandemic. Gaps in the healthcare system and barriers to mental health care access impede individuals' growing inquiries for mental health services. On top of that, a major issue constitutes the inability and difficulty of primary care physicians in identifying depression. Due to lack of time and resources, **primary care physicians are unable to identify more than half of depression cases.** As a result, depression remains one of the most under-diagnosed and under-treated disease.

The established correlational relationship between human voice and depression has led to the emergence of speech analysis via artificial intelligence and machine learning models that can identify specific vocal features **(vocal biomarkers)** in individuals. Such technologies have the potential to innovate and transform the management of mental health disorders.

An emerging application of vocal biomarkers technology is within the domain of telehealth. **Integration of voice technology solutions provides value to the patients, physicians, and telehealth service providers.** Benefits include: improvement of patient health outcomes via more optimal management of mental health disorders, reduction of physicians' workload and streamlining the patient consultations, increasing user engagement and offering novel decision-support and monitoring tools.

Our team developed a **proprietary vocal biomarkers machine learning model that identifies depression in human voice with 30 seconds of speech, called VoiceAI.** In contrast to the existing models, VoiceAI does not analyze the content of the speech - preserving the confidentiality and privacy of the users. Our model has demonstrated comparable performance to the existing state-of-art AI models, with a combined accuracy rate of **74%**.

Current developments focus on validating vocal biomarkers technology through clinical trials. These developments will drive the adoption of voice technology and AI in routine clinical use and establish them as an integral part for the management of mental health disorders.

GLOBAL TOLL OF DEPRESSION: HUMAN & ECONOMIC COSTS

Depression is the most common mental illness that ranges from mild, episodic mental disturbances to long-term, severe depressive disorders known as clinical depression (WHO, 2017). It is characterized by various lasting symptoms, from feelings of sadness, emptiness, and frustration to loss of interest in daily activities and suicidal thoughts (Hall-Flavin, 2017).

According to the World Health Organization (WHO, 2017; WHO, n.d.), approximately **300 million people worldwide suffer from depression, and up to 75% never receive any care.** The emergence of the recent pandemic caused a surge in depression and further increased its prevalence by **25%** worldwide (WHO, 2022). As a result, demands for mental health services grew exponentially. But the resources for such inquiries became increasingly limited which in turn created major gaps in the health-care system for the diagnosis, treatment, and support of patients.

While inquiries for professional mental health services have significantly increased over the years, barriers to mental health care access remain, hindering people from seeking the help they need. The most commonly encountered barriers include financial burden, lack of well-trained mental health professionals, and social stigma (Low et al., 2020).

Aside from gaps in the healthcare system, depression has significant economic implications for individuals and employers, involving direct costs (e.g., psychotherapy services) and indirect costs (e.g., low employee productivity and turnover) (Grazier, 2019). As stated by the Lancet global health (2020), **mental health disorders cost 1 trillion dollars annually to the global economy.**

HOW IS DEPRESSION TREATED TODAY? - CURRENT APPROACHES & LIMITATIONS

Primary care is the de facto mental health care in most developed markets. The first contact point for the patients is the primary care physician (aka GP) who is responsible for doing initial patient health assessment and referral (Faghri et al., 2010).

Various studies report on depression remaining under-treated due to the inability of GPs to recognize the symptoms. It is estimated that only 1 out of 5 GPs records the mental health state of their patients in the electronic health records (Mitchell et al., 2009).

To address the public health challenges related to depression, in 2016 US Preventive Services Task Force (USPSTF) published an official recommendation that all adults in primary care settings should be screened for depression. USPSTF identified that combining various screening tools with the respective therapeutic support systems, such as anti-depressant treatment and psychotherapy, led to vastly improved clinical outcomes for the patients. USPSTF's recommendation was a significant milestone in establishing routine screening for depression in primary care settings (Siu et al., 2016).

One of the most commonly used screening tools for depression is the "Patient Health Questionnaire (PHQ)."

PHQ is a standardized, paper-based questionnaire developed in the 1990s by Pfizer Inc. Various versions of PHQ have been validated in clinical and non-clinical settings. In North America and other parts of the world, it remains the most commonly used screening tool in primary care settings (PHQ screeners, n.d.).

Despite being well-validated in clinical settings, standardized questionnaires, such as PHQ, come with significant limitations which hinder their broad adoption in primary care. These limitations include social desirability bias, reliance on the subject's retrospective recall, and significant time commitment from the physician (Ben-Zeev et al., 2009).

A physician and a patient fill-out the questionnaire together during the consultation. Completing such test can take up to 10 minutes or more. The healthcare systems have witnessed increasing complexity over the last decades resulting in a significant reduction in the time doctor spends with a patient. Nowadays an average doctor-patient visit lasts only 10-15 minutes and is further declining (Topol, 2019). Thus in the existing circumstances, utilization of time-intensive paper questionnaires is severely limited. This compounds the problem of under-recognition and under-treatment of depression in primary care.

HUMAN VOICE: A WINDOW INTO THE HUMAN MIND

Human voice and depression are inherently correlated. As early as 1921, German psychiatrist and the founder of modern psychiatry - Emil Kraepelin - described depressed patients as having a lower vocal pitch, monotonous speech, low speech intensity, and rate, in addition to exhibiting more hesitations, stuttering, and whispering (Low et al., 2020).

Nowadays, we know that neurophysiological and cognitive changes in depressed individuals impact the motor control functions that drive vocal output.

Individuals affected by clinical depression suffer from psychomotor retardation, which directly affects the acoustic, linguistic, and prosodic features of the voice. Other depression-associated cognitive declines include speech planning, ability to select words, and respiratory rate (Low et al., 2020).

VOCAL BIOMARKERS FOR PSYCHIATRIC DISORDERS

AI-enabled speech analysis and natural language processing techniques have shown great promise in detecting vocal changes related to psychiatric disorders.

Machine learning models can identify specific prosodic (i.e., rhythm, pitch, shimmer, jitter) and linguistic features in the voice with relatively high accuracy. These features are currently being researched as potential biomarkers for various psychiatric disorders, including major depressive disorder (Low et al., 2020).

To identify vocal biomarkers for diagnosis, human-emitted sounds are classified into three categories of verbal, syllable, and non-verbal vocalizations. Predictive models utilize audio features to detect and classify psychiatric disorders. Segmental features, such as the Mel Frequency Cepstral Coefficients[1], along with spectrogram[2] and Geneva Minimalistic Acoustic Parameterization Set (eGeMAPS)[3] enable predictive models to extract complex representations from voice and correlate it to the mental health of the subject (Low et al., 2020; Fagherazzi et al., 2021).

Once the features are selected, predictive models are trained and tested on separate datasets, to ensure that model is able to generalize on previously unseen data. Cross-validation is used to train model and obtain reliable estimates. The accurate selection of performance metrics is crucial as it affects the interpretation of the results. Specificity and sensitivity are one of the most commonly used performance metrics in relation to the diagnostic systems.

The future potential of vocal biomarkers technology is multitude. Amongst them are:

A. Population screening of at-risk individuals for psychiatric disorders before access to the mental health or primary care system. For the individuals outside the system, these technologies enable remote mental health assessments and the potential, via collecting data on the same individuals over time, to offer personalized treatment alternatives.

B. Adjunct screening and decision-supporting tools in the clinician's arsenal. These technologies can increase the assessment accuracy of patients by physicians, who often face difficulties in correctly identifying and diagnosing certain disorders due to their episodic nature and high comorbidity rates. Furthermore, vocal biomarkers can potentially classify diagnoses that will allow predicting the most appropriate treatment and identifying risk-factor patients.

C. Supporting therapy and interventions as a monitoring tool in between consultations with the healthcare professional. Real-time remote monitoring enables individuals and physicians to closely assess mental health conditions and observe the necessity to seek help.

---

[1] MFCCs: Set of features with 10-20 coefficients that delineate the shape of the spectrum.
[2] Spectrogram: Spectrum of audio frequencies visually represented.
[3] eGeMAPS: Consensus set of 88 human-emitted sound features for emotion recognition

VOICE TECHNOLOGY IN TELEMEDICINE: BENEFITS FOR PROVIDERS, DOCTORS, AND PATIENTS

An up-and-coming application of vocal biomarkers technology is the automatic screening and monitoring of mental health disorders on the telehealth service provider platforms. In contrast to using standardized questionnaires, integration of such decision-support tools offer the following benefits:

- Under-detection and under-treatment of mental health disorders is reduced - leading to the improvement of patient health outcomes.
- Physicians' workload related to mental health screening and monitoring is also reduced, enabling physicians to streamline the patient visits and save time and resources.
- Providers can longitudinally monitor the efficacy of their services and objectively measure patient health outcomes.
- Provide innovative decision-support tools to the physicians and improve their engagement and retention on the provider platforms.

Vocal biomarkers technology can be integrated into virtually any type of telehealth platform. Digital nature of such platforms enables instant data exchange via API and information analysis, meaning that the physician receives the patient's mental health assessment almost instantly.

Several groups, amongst them ours, are already working in this direction - implementing voice technology solutions in various telehealth and remote monitoring platforms. We believe that vocal biomarkers technology will play an increasing role in telemedicine over the coming years and deliver true value to the patients, physicians, and providers.

Ensofy VoiceAI

Over the last six months, our team has developed a proprietary machine learning model based on the deep neural networks, called VoiceAI. VoiceAI identifies depression in human voice using 30 seconds of free-form speech.

One of the barriers to adopting vocal biomarkers technology in clinical use is the privacy concern of the physicians and the patients. Our market research showed us that the physicians strongly prefer the conversation during a patient-visit to remain private and anonymous - given the sensitive nature of the topic. To address this concern, we decided to focus only on the non-linguistic/prosodic features of the voice analysis, thus preserving the privacy of our users.

In partnership with certified psychotherapists, our team conducted a data collection trial to collect high-quality vocal and clinical data from the randomly selected subjects. The resulting proprietary vocal biomarkers dataset was used for the development of VoiceAI.

In total, data from 209 subjects with approximately 2000 minutes of speech were used for training and development purposes. The following table provides the demographic and clinical information of the subjects.

	NO. OF SUBJECTS
FEMALE	99 (47%)
MALE	110 (53%)
DEPRESSED	47 (22%)
NON-DEPRESSED	162 (78%)
TOTAL	209 (100%)

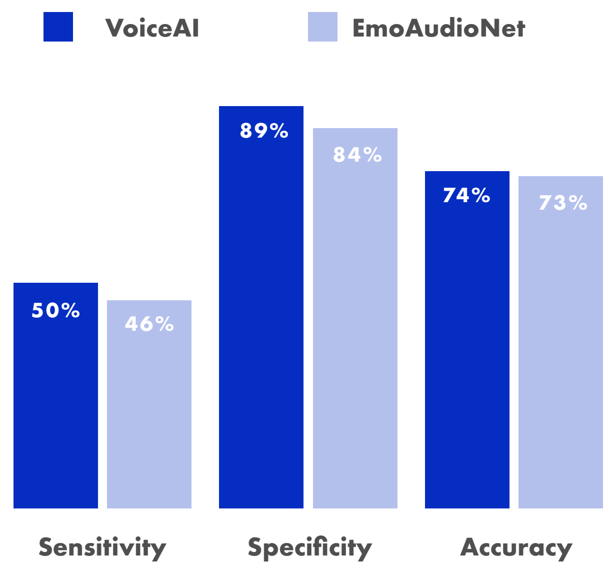
* Demographic & clinical profile of the subjects.

Our study design follows the established protocol as set out in the “Audio/Visual Emotion Challenge and Workshop 2019”: Psychotherapists conducted 1-on-1 sessions with the participants and recorded the session audio. Participants filled out PHQ before or after the session. The conducted trial complied with all relevant ethical, data protection, and safety standards and was carried out by qualified healthcare professionals.

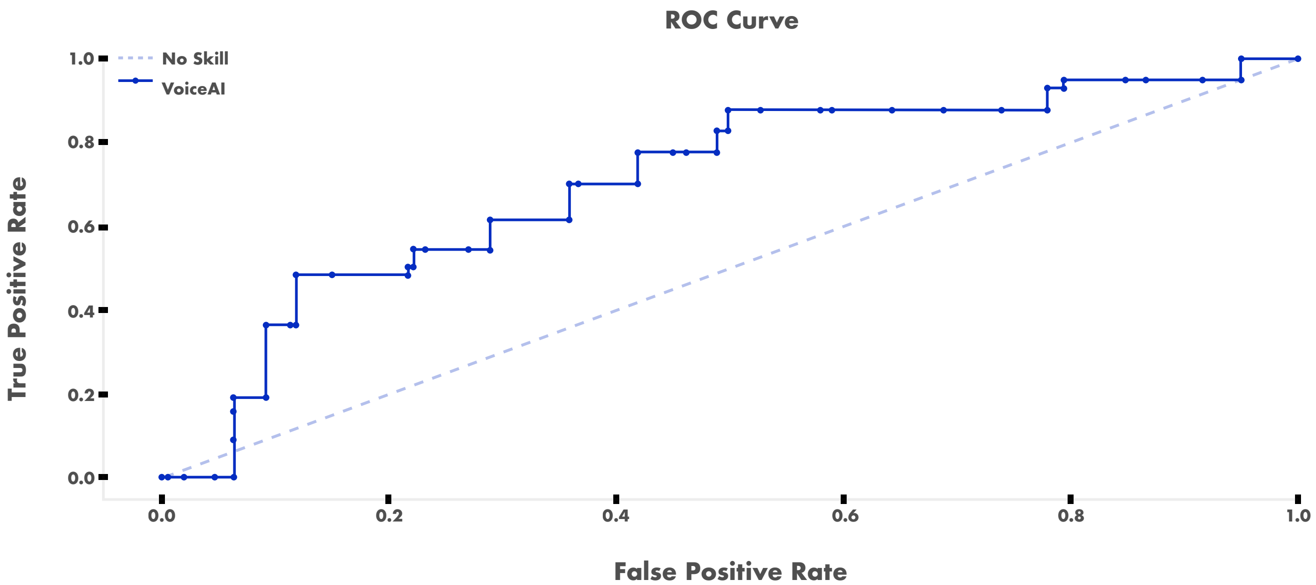
The accuracy of VoiceAI in correctly identifying depressed and non-depressed samples was tested on an existing research dataset and compared to the state-of-art AI model - “EmoAudioNet”. As described by Rejaibi et al. (2019), “EmoAudioNet outperforms state-of-art approaches of automatic depression recognition” and is one of the best performing models in the field.

Ensofy VoiceAI

Following **FIGURE 1** demonstrates performance of VoiceAI against EmoAudioNet on the same test dataset. As seen VoiceAI has comparable performance as EmoAudioNet and even shows minor performance improvements. **FIGURE 2** demonstrates detailed performance of VoiceAI on the ROC curve[4].



* **FIGURE 1** VoiceAI vs. EmoAudioNet



* **FIGURE 2** VoiceAI ROC curve.

Our team continues the development of VoiceAI to achieve higher accuracy rates by collecting and incorporating additional high-quality data in the machine learning model.

---

[4] **Receiver operating characteristic (ROC) curve:** Commonly used graphical plot used to visualize diagnostic performance of a system. It is created by plotting true-positive rate (i.e. sensitivity) against false-positive rate (i.e. 1 minus specificity).

ONGOING DEVELOPMENTS

One of the limitations of the vocal biomarkers technology is the use of PHQ or other standardized questionnaires as the comparator (i.e., ground truth for assessing depression). Despite strong validation of such questionnaires in clinical settings, they remain inferior to the "gold standard" of diagnosing mental health disorders that is structured or semi-structured interview by a qualified psychiatrist. When compared to the "gold standard", PHQ achieves accuracy rates of up to 85% (Levis et al., 2019). Thus using PHQ as the ground truth for the development of vocal biomarkers technology compounds the limited accuracy of the vocal biomarkers models.

To address this problem, our team is initiating a pilot clinical trial where the diagnosis by qualified psychiatrists are used as a comparator. Handful of academic and commercial organizations are also pursuing similar pilot trials. The outcomes of these trials, if positive, will accelerate the ongoing adoption of vocal biomarkers technology in routine clinical use and establish voice technology and AI as an integral part of managing mental health disorders.

REFERENCES

Ben-Zeev, D., Young, M. A., & Madsen, J. W. (2009). Retrospective recall of affect in clinically depressed individuals and controls. Cognition and Emotion, 23(5), 1021-1040. https://doi.org/10.1080/02699930802607937

Fagherazzi, G., Fischer, A., Ismael, M., & Despotovic, V. (2021). Voice for Health: The Use of Vocal Biomarkers from Research to Clinical Practice. Digital Biomarkers, 5(1), 78–88. https://doi.org/10.1159/000515346

Faghri, N. M. A., Boisvert, C. M., & Faghri, S. (2010). Understanding the expanding role of primary care physicians (PCPs) to primary psychiatric care physicians (PPCPs): enhancing the assessment and treatment of psychiatric conditions. Mental health in family medicine, 7(1), 17-25.

Grazier, K. L. (2019). The economic impact of depression in the workplace. Mental Health in the Workplace (pp. 17-26). Springer, Cham.

Hall-Flavin, D. K. (2017). Clinical depression: What does that mean? Mayo Clinic. https://www.mayoclinic.org/diseases-conditions/depression/ep-ert-answers/clinical-depression/faq-20057770

Levis, B., Benedetti, A., & Thombs, B. D. (2019). Accuracy of Patient Health Questionnaire-9 (PHQ-9) for screening to detect major depression: individual participant data meta-analysis. bmj, 365.

Low, D. M., Bentley, K. H., & Ghosh, S. S. (2020). Automated assessment of psychiatric disorders using speech: A systematic review. Laryngoscope Investigative Otolaryngology, 5(1), 96–116. https://doi.org/10.1002/lio2.354

Mitchell, A. J., Vaze, A., & Rao, S. (2009). Clinical diagnosis of depression in primary care: a meta-analysis. The Lancet, 374(9690), 609-619.

Patient Health Questionnaire (PHQ) Screeners. Screener Overview. (n.d.). Phqscreeners. https://www.phqscreeners.com/select-screene

Rejaibi, E., Kadoch, D., Bentounes, K., Alfred, R., Daoudi, M., Hadid, A., & Othmani, A. (2019). Clinical depression and affect recognition with emoaudionet. arXiv preprint arXiv:1911.00310.

Siu, A. L., Bibbins-Domingo, K., Grossman, D. C., Baumann, L. C., Davidson, K. W., Ebell, M., ... & US Preventive Services Task Force. (2016). Screening for depression in adults: US Preventive Services Task Force recommendation statement. Jama, 315(4), 380-387. doi:10.1001/jama.2015.18392

The Lancet Global Health. (2020). Mental health matters. The Lancet Global Health, 8(11), e1352. https://doi.org/10.1016/s2214-109x(20)30432-0

Topol, E. (2019). Deep Medicine. Drerictopol.
https://drerictopol.com/portfolio/deep-medicine/
Wikipedia contributors. (2022, June 21). Receiver operating characteristic. Wikipedia. Retrieved June 21, 2022, from
https://en.wikipedia.org/wiki/Receiver_operating_characteristic

World Health Organization. (n.d.). WHO Special Initiative for Mental Health.
https://www.who.int/initiatives/who-special-initiative-for-mental-health

World Health Organization. (2017). Depression and other common mental disorders: global health estimates (No. WHO/MSD/MER/2017.2). World Health Organization.

World Health Organization. (2022). COVID-19 pandemic triggers 25% increase in prevalence of anxiety and depression worldwide.
https://www.who.int/news/item/02-03-2022-covid-19-pandemic-triggers-25-increase-in-prevalence-of-anxiety-and-depression-worldwide

DISCLAIMER

For more information contact: irakli@ensofy.ai
The project is co-sponsored by "Georgia's Innovation and Technology Agency".