A PRACTICAL APPROACH TO PREDICTING DEPRESSION: VERBAL AND NON-VERBAL INSIGHTS WITH MACHINE LEARNING

Sanskar Rahul Nanegaonkar and Kotaro Ando

I’m beside you Inc, Japan

ABSTRACT

While global standards have been established for diagnosing depression, the reliance on expert judgement and observation remains a challenge. This study delves into a potential approach of efficient data collection to increase the practicability of machine learning models in accurately predicting depression based on a comprehensive analysis of verbal and non-verbal cues exhibited by individuals.

KEYWORDS

Depression, Machine Learning, Free-Talk Sessions, Multimodal Cues, Depression Prediction, Mental Health, Diagnostic Tools

1. INTRODUCTION

Depression is a significant global health concern, impacting an estimated 5% of adults worldwide. Women are more commonly affected than men (For details, refer to: https://www.who.int/news-room/fact-sheets/detail/depression). The prevalence and impact of depression vary greatly across different regions. For instance, countries like Ukraine, the United States, Australia, and Estonia have some of the highest reported rates of depression, with prevalence rates around 6% (For details, refer to: https://worldpopulationreview.com/country-rankings/depression-rates-by-country).

Access to treatment for depression is highly disparate globally. In low- and middle-income countries, over 75% of individuals with mental disorders do not receive the necessary treatment (For details, refer to: https://www.who.int/news-room/fact-sheets/detail/depression). This lack of access is attributed to factors such as insufficient investment in mental health care, a shortage of trained health-care providers, and prevalent social stigma around mental health issues. In contrast, more developed nations have higher rates of diagnosis and treatment, although many cases still remain unreported or untreated due to stigma and lack of awareness.

The World Health Organization emphasises that there are effective treatments for depression, ranging from psychological treatments to medications (For details, refer to: https://www.who.int/news-room/fact-sheets/detail/depression). However, the accessibility and type of treatment vary significantly worldwide. The global impact of depression underscores the need for comprehensive public health strategies to address this condition, which significantly affects personal well-being, relationships, work performance, and broader societal aspects.

DOI : 10.5121/hiij.2024.13201

The burden of depression on global mental health has reached alarming levels, with the World Health Organization estimating approximately 300 million people suffering from this condition (For details, refer to: https://www.who.int/news-room/fact-sheets/detail/depression). Characterised by persistent sadness, loss of interest, and a range of emotional and physical problems, depression significantly impairs an individual's daily life. Traditionally, the diagnosis of depression has relied heavily on clinical assessments and patient self-reports. Standard diagnostic tools include structured interviews like the Diagnostic and Statistical Manual of Mental Disorders (DSM) [1] and assessment scales such as the Hamilton Depression Rating Scale (HDRS) [2]. While these methods are foundational in diagnosing depression, they often hinge on the subjective interpretation of symptoms and patient self-disclosure, which can lead to variability in diagnosis accuracy.

The integration of machine learning (ML) into healthcare, particularly in the realm of mental health diagnosis, heralds a groundbreaking paradigm shift with immense promise. Machine learning algorithms, with their unparalleled capability to sift through and make sense of vast, diverse datasets, are poised to significantly refine the accuracy of diagnoses. Through sophisticated techniques, ML not only enriches our understanding of depressive patterns but also establishes itself as an indispensable asset in the field of mental health predictive analytics.

Existing research in the field [3], has investigated the potential of Vocal Acoustic Features as Biomarkers for Depression. This particular study delved into the differences in acoustic characteristics between individuals diagnosed with depression and healthy controls, when reading positive, negative and neutral texts. The findings of this research concluded that certain acoustic features significantly differ between these two groups, underscoring the potential of these features in distinguishing depressed patients from healthy individuals.

Similarly in another study [4], researchers focused on the fast and accurate assessment of depression based on voice acoustic features. This research aimed to identify voice acoustic features that can effectively predict the severity of depression, employing an artificial neural network trained with voice features correlated to depression scores. Furthermore, the study included a longitudinal analysis, examining changes in these acoustic features following an Internet-based cognitive-behavioural therapy (ICBT) program. The results demonstrated that certain acoustic features significantly changed post-treatment, suggesting a potential correlation with specific treatment options and notable improvement in depression symptoms. This study underscores the potential of using voice acoustic features as a low-cost and efficient method for large-scale depression screening and monitoring the efficacy of treatment.

Taking it one step forward, this study [5] focuses on developing a depression detection model by utilising both audio and visual features from YouTube vlogs. It addresses the gap in existing literature by concentrating on video content rather than text or images from social media for depression detection. The study collected and annotated YouTube vlogs, then extracted audio and visual features to analyse differences between depression and non-depression content. A machine learning model, achieving over 75% accuracy, was built to detect depression, highlighting the potential of video logs as a valuable resource for detection of depression in individuals.

Moreover, as discussed in the study reviews by [6], [7], [8], and other research papers like [9], various studies have examined the connection between multi-modal data and mental health. However, most of them focus on using sensing data, heart rate monitoring, social media analysis, or synthetic data generation through tasks or stimuli like reading sentences. Although certain studies incorporate clinical assessments, their direct applicability across different scenarios remains constrained. From a socio-economic perspective, these studies appear promising, but from an individual standpoint, a more practical approach is necessary.

The findings from the aforementioned studies indicate a promising avenue for enhancing the practicability of machine learning models for depression prediction. By further expanding these studies to incorporate a broader array of verbal/acoustic and non-verbal features, and integrating data that closely corresponds to real-life scenarios, there is potential to significantly refine the precision with which depression, including its severity, can be detected and assessed. The approach discussed hereafter suggests that a more comprehensive analysis of these features along with real-life scenario data could lead to more nuanced and accurate diagnostic tools in mental health care.

2. MATERIALS AND METHODS

For this study, we had requested cooperation via the internet from a pool of participants who had a track record of applying for counselling and the participant pool is composed of mainly Japanese individuals without a history of depression and those who had previously been diagnosed with depression and thus a total of over 60 participants were enrolled. An experienced psychiatrist from i2medical (See: https://i2medical.co.jp) was appointed to conduct a series of four distinct sessions with the participants. These sessions included Free-talk interactions and structured assessments, which encompassed the use of several standardised tools, namely the Semi-Structured Clinical Interview for DSM-5 (SCID-5) [1] the Montgomery-Åsberg Depression Rating Scale (MADRS) [10], and the Patient Health Questionnaire-9 (PHQ-9) [11]. Initially, Free-talk sessions of around 10-20 minutes, followed by SCID and MADRS sessions of around 20-25 minutes and finally PHQ-9 of around 5 minutes, all of which are conducted on the same day.

All sessions were conducted and recorded using the Zoom video conferencing tool and it was made sure that their personal identity like name, date of birth, place, etc. is anonymised and they were advised to be alone in the session with their face clearly visible in the camera and be in a quiet environment. An example of the recorded video is illustrated in Figure 1 (For privacy considerations, facial identities have been blurred in the example while the original visual data remains unaltered). The recorded videos were then processed through the code pipelines to extract low-level attributes of the patient such as blink rate (0 when eyes are open and 1 when closed), eye-offset ([12], [13]) which is the angle the eyes w.r.t the camera (values range from -180° to 180°), gaze which is derived from eye-offset based on the threshold of 25° (1 when the eye-offset angle is less than 25° else 0) and facial expressions ([14] and [15]) (values range from 0 to 100) which includes anger, disgust, fear, happy, sad, surprise, neutral and negative/positive affect at each timestamp or second. Additionally, pitch and root mean square (RMS) or volume were extracted from separated patient audio. Due to technical issues and limitations such as some video recordings being unavailable due to technical glitches on Zoom and some with no audio, feature extraction was successfully performed on data from only 56 participants (35 females and 21 males) with distribution as shown in Figure 2 and Figure 3.

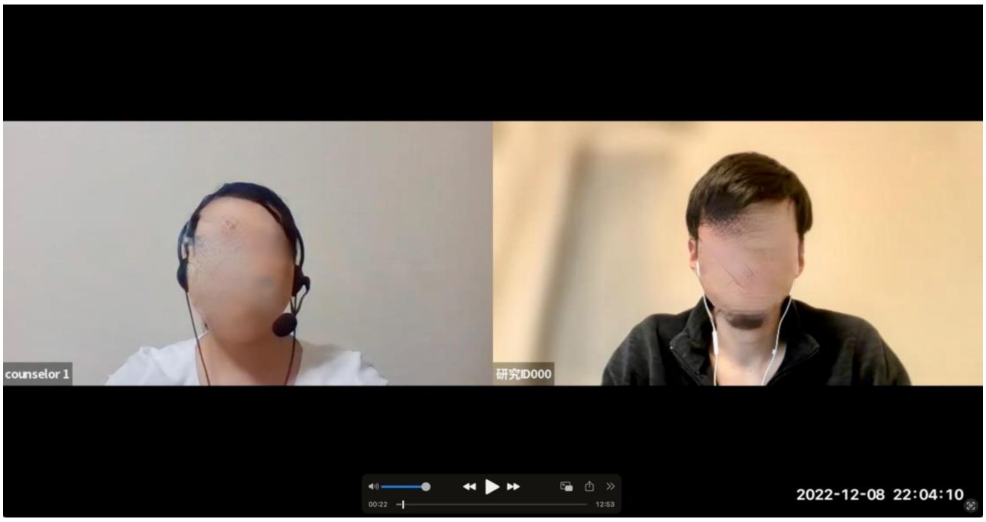

Figure 1. An example of Zoom recording

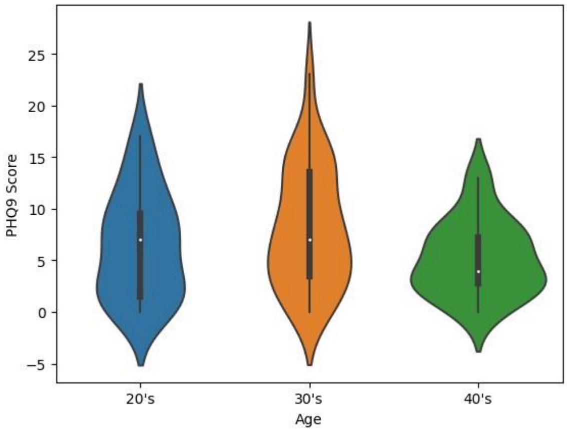

Figure 2. Violin plot showing the distribution of PHQ9 Score for each Age range

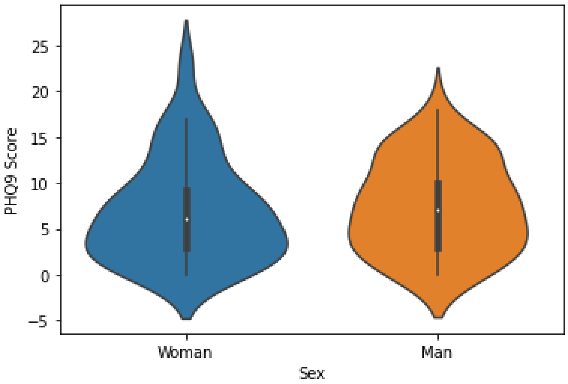

Figure 3. Violin plot showing the distribution of PHQ9 Score for each Sex

From these low-level features, we computed high-level statistics, including mean (characterising the average value, represented as _m), standard deviation (characterising the fluctuation, represented as _s), and standard deviation of standard deviation (characterising the instability, represented as _ss). Notably, the standard deviation of standard deviation was calculated by first computing the standard deviation for each 30-second interval and then calculating the standard deviation of these interval-level standard deviation values. However, in the final feature set, we omitted the mean value derived from facial features to mitigate potential bias associated with individual differences in facial expressions and orientation. For instance, individuals may naturally exhibit varying degrees of neutrality or expressiveness in their resting facial expressions. By excluding the mean based on facial features, we aimed to ensure that the relative changes in facial expressions were assessed independently of individual differences in baseline expressions and orientation. The _s based features can be understood as an indication of how the features are deviating or fluctuating over time and the _ss based features can be understood as an indication of how the features are stable or unstable over time. For example, one might have a natural behaviour of constantly moving their head, so _s will help us to gauge it and then _ss will give us an understanding of how stable/unstable their movement is over time.

So, our feature set ultimately comprised of RMS (Volume), pitch_m, pitch_s, blink_m, angry_s, fear_s, happy_s, sad_s, neutral_s, negative_s, positive_s, blink_s, gaze_s, eye_offset_s, angry_ss, fear_ss, happy_ss, sad_ss, neutral_ss, negative_ss, positive_ss, blink_ss, gaze_ss, and eye_offset_ss.

3. Results

Our analysis revealed a strong correlation between the PHQ-9 and MADRS assessments, as depicted in Figure 4, generated using [15]. This finding is consistent with previous research results [16]. Additionally, we observed a high correlation between the extracted features from Free-talk and MADRS sessions as shown in Figure 5. This strong correlation, along the diagonal, indicates a resemblance in participant behaviour across these contexts.

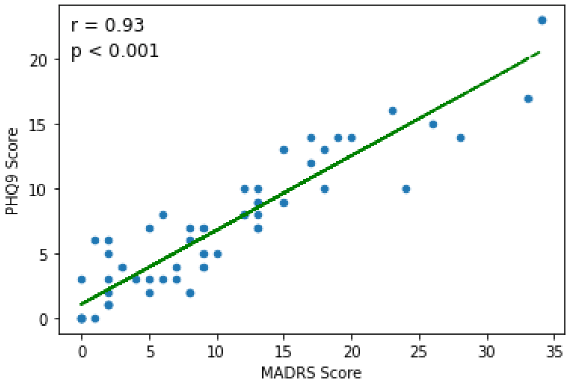

Figure 4. Scatter plot between MADRS and PHQ9 Scores

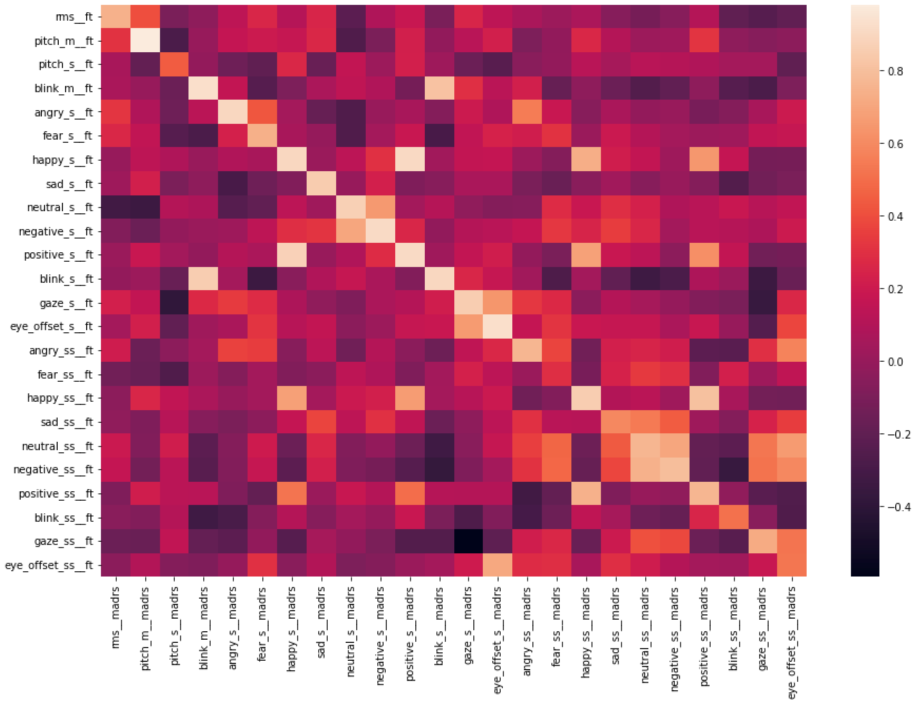

Figure 5. Heat-map showing the correlation between features extracted from Free-talk (on y-axis) and MADRS (on x-axis) session

Consequently, we chose to employ Free-talk sessions to predict PHQ-9 scores, recognizing the richness of data captured during these sessions. The unstructured nature of Free-talk sessions allows for spontaneous expressions of thoughts and emotions, offering deeper insights into an individual's mental state compared to standardised assessments alone, refining it for greater practical application.

PHQ-9 scores were categorised into two classes based on a threshold of 10 [11] in order to distinguish between mild and severely affected individuals. Scores greater than or equal to 10 were considered indicative of depression, while scores less than 10 indicated a healthy state. We divided the dataset in a stratified manner into three sets: 60% for training (33 samples, 24 healthy, and 9 affected), 20% for validation (11 samples, 8 healthy, and 3 affected), and 20% for testing (12 samples, 8 healthy, and 4 affected). Min-Max Scaler was employed to normalise the data.

In our study, the evaluation of various machine learning models was meticulously carried out using an independent test dataset. We focused primarily on metrics such as overall accuracy, precision, recall, and F1-scores for each class. Among these models, the support vector machine [17] model with the hyper-parameters as (C: 1, kernel: 'rbf', class_weight: 'balanced'), implemented using the scikit-learn library [18], emerged as a standout performer. It achieved an overall accuracy of 83%, with accuracy for predicting healthy participants at 88% and for predicting affected participants at 75%, as illustrated in the confusion matrix in Figure 6. This high level of accuracy underscores the model's robustness and reliability in diagnosing depression.

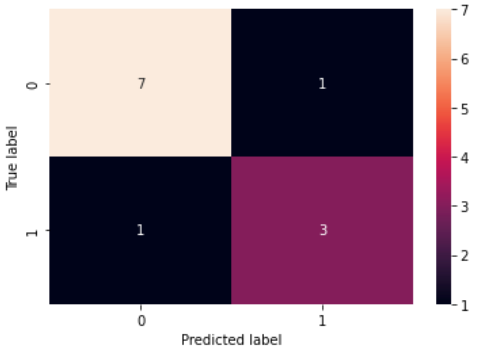

Figure 6. Confusion matrix of the test set prediction

Utilising SHAP (SHapley Additive exPlanations) [19], a violin plot, as shown in Figure 7, was generated to explain the influence of verbal and non-verbal features on the model's predictions. Notably, features such as blink_m (mean blink rate), positive_s (fluctuation in positive emotion), and fear_ss (instability in fear emotion) emerged as highly influential, indicating significant impacts on the model's output. Particularly, a higher value of blink_m was found to be associated with identifying depression. These insights are instrumental in understanding the underlying factors that the model deems significant in predicting depression, highlighting the complexity of the condition and the multifaceted nature of its expression. These insights were confirmed by the psychiatrist and a medical expert from Keio University to underscore the relevance of the identified features. They agree that such features are consistent with the critical cues they rely on for diagnosing depression, lending credibility to our findings.

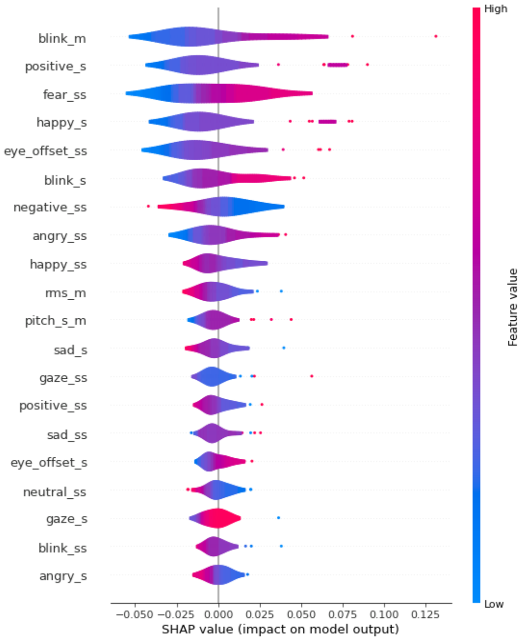

Figure 7. Violin plot showing the feature importance using SHAP

In conclusion, the results of our study indicate a promising potential for machine learning models in accurately predicting depression. These findings make a significant contribution to the field of mental health diagnostics, offering a new perspective on enhancing the accuracy and reliability of depression prediction tools, with a practical edge.

4. DISCUSSION

The efficacy of the machine learning algorithms in predicting depression, as indicated by an 83% accuracy rate, has notable implications for the field of mental health diagnostics. The application of SHapley Additive exPlanations (SHAP) values has provided us with a granular understanding of feature impact, revealing that not all features contribute equally to the model's predictive power. Our findings suggest that mean blink rate and emotional variability are potent indicators of depression, validating the nuanced complexity of psychiatric assessments and the potential for machine learning to augment them.

With that, the implications of our findings for clinical practice are considerable. The integration of machine learning models into diagnostic procedures could herald a new era of precision in mental health care, moving beyond the traditional reliance on self-reporting and clinician interpretation. Such an approach could lead to earlier and more accurate detection of depression, thereby improving patient outcomes through timely intervention.

Our findings also underscores that relying solely on voice features is insufficient for accurate depression diagnosis. The significant contribution of facial and eye movements to prediction accuracy highlights the necessity of adopting a multimodal approach. While voice data is valuable, incorporating additional cues enhances diagnostic precision. Discussions with psychiatrists have reinforced the practicality of our multimodal methodology, affirming its importance in not overlooking critical indicators of depression.

Additionally, we observed correlation between features from free-talk and MADRS sessions emerged as a critical insight, suggesting that depression prediction might not solely depend on structured assessment sessions. This finding is significant as it points toward the potential of utilising free-talk analysis as an efficient screening tool for identifying depression, thereby offering a novel and practical approach that could streamline the preliminary assessment process. Such a method could significantly aid psychiatrists by prioritising client evaluations more effectively, without the need for additional, time-consuming diagnostic sessions. This possibility marks an important step forward in optimising mental health care delivery. These Free-talk sessions can be considered to be similar to video journaling and well-being apps where monologue videos are recorded and thus can be used to gauge the mental state of the person.

Our study furthermore highlights the significant correlation observed between the MADRS and the PHQ-9 assessments. This correlation not only validates the reliability of self-reported measures in mirroring clinician-administered evaluations but also emphasises their potential in enriching diagnostic approaches. The alignment between these two distinct methods of assessment indicates a broader applicability of self-administered tools in clinical settings, suggesting they could serve as effective complements especially in scenarios where direct clinician interaction is not feasible. This insight paves the way for further exploration into how diverse assessment tools can be harmonised within clinical practice, enhancing our understanding and approach to diagnosing depression.

However, the limitations of our study, including the sample size and the specificity of the dataset to a single cultural context, suggest that further research is warranted. Future studies should aim to replicate these findings across larger and more diverse populations, explore the integration of additional behavioural and physiological data, and examine the efficacy of other more complex machine learning algorithms.

In summary, our research contributes to the evolving landscape of psychiatric diagnostics, demonstrating the transformative potential of machine learning in detecting depression. As we advance, it is imperative to continue refining these models, ensuring their ethical application and their alignment with the complexities of human behaviour and mental health.

5. CONCLUSIONS

This study demonstrated the potential of using free-talk sessions and machine learning models, achieving an 83% accuracy rate in predicting depression through the analysis of verbal and non-verbal cues. Our findings suggest that machine learning can significantly enhance the objectivity and precision of mental health diagnostics. The integration of specific behavioural indicators, identified as crucial by our analysis, underscores the importance of multi-modal data in understanding complex psychiatric conditions.

With that, the observed correlation between PHQ-9 and MADRS assessments underscores the value of self-administered tools, especially when clinical interaction is impractical. Furthermore, the effective use of free-talk sessions to identify depression signifies a significant advancement, promising to impact future diagnostic practices positively and effectively. While acknowledging the limitations of our dataset's size and diversity, this research paves the way for future studies to build upon these findings, exploring broader applications and diversifying the analytical approaches in mental health diagnostics. Ultimately, our work contributes to the evolving landscape of psychiatric assessment, advocating for the incorporation of advanced computational methods to improve diagnostic accuracy and patient outcomes in mental health care.

6. FUTURE WORK

To enhance the depth and applicability of our research, future studies should aim to utilise a larger and more diverse dataset, addressing the current limitation of a small, culturally homogeneous group. This expansion will enable a more comprehensive understanding of depression indicators across different demographics. Additionally, incorporating a broader spectrum of features, could offer richer insights into depressive behaviours, enriching the model's diagnostic capabilities. Expanding the scope to include other mental health disorders alongside depression will also be crucial, as it could lead to the development of a more holistic diagnostic tool capable of addressing the complex spectrum of mental health issues, facilitating the creation of tailored treatment plans for a diverse patient population.

REFERENCES

[1] American Psychiatric Association, *Diagnostic and statistical manual of mental disorders (5th ed., text rev.)*, 2022.

[2] Hamilton M. A rating scale for depression. *J Neurol Neurosurg Psychiatry*, 1960.

[3] Zhao Q., Fan H.Z., Li Y.L., Liu L., Wu Y.X., Zhao Y.L., Tian Z.X., Wang Z.R., Tan Y.L., and Tan S.P. Vocal Acoustic Features as Potential Biomarkers for Identifying/Diagnosing Depression: A Cross-Sectional Study. *Front Psychiatry*, 2022.

[4] Wang Y., Liang L., Zhang Z., Xu X., Liu R., Fang H., Zhang R., Wei Y., Liu Z., Zhu R., Zhang X., Wang F. Fast and accurate assessment of depression based on voice acoustic features: a cross-sectional and longitudinal study. *Front Psychiatry*, 2023

[5] Min K., Yoon J., Kang M., Daeun L., Eunil P., and Jinyoung H. Detecting depression on video logs using audiovisual features. *Humanit Soc Sci Commun* 10, 788 (2023)

[6] T. T. Nguyen, V. H.-Q. Pham, D.-T. Le, X.-S. Vu, F. Deligianni, and H. D. Nguyen, Multimodal machine learning for mental disorder detection : a scoping review, *in 27th international conference on knowledge based and intelligent information and engineering systems (KES 2023)*, 2023, vol. 225, pp. 1458–1467.

[7] Enrique Garcia-Ceja, Michael Riegler, Tine Nordgreen, Petter Jakobsen, Ketil J. Oedegaard, and Jim Tørresen. Mental health monitoring with multimodal sensing and machine learning: A survey. *Pervasive and Mobile Computing*, 51:1-26, 2018.

[8] Khoo LS, Lim MK, Chong CY, McNaney R. Machine Learning for Multimodal Mental Health Detection: A Systematic Review of Passive Sensing Approaches. *Sensors*. 2024.

[9] Thati R.P., Dhadwal A. S., Kumar P., Sainaba P. A novel multi-modal depression detection approach based on mobile crowd sensing and task-based mechanisms. *Multimed Tools Appl* 82, 4787–4820 (2023).

[10] Lena C. Q., Jennifer J. R., Jean-Pierre R., Filip De F., Frédéric R., and R. Michael B. The structure of the Montgomery-Åsberg depression rating scale over the course of treatment for depression. *International journal of methods in psychiatric research*, 2013.

[11] Kurt K., Robert L S., and Janet B W W. The PHQ-9: validity of a brief depression severity measure. *Journal of general internal medicine*, 2001.

[12] Deng J., Guo J., Yuxiang Z., Jinke Y., Irene K., and Zafeiriou S. RetinaFace: Single-stage Dense Face Localisation in the Wild. *arxiv*, 2019.

[13] Wang Z., Chai J., and Xia S. Realtime and Accurate 3D Eye Gaze Capture with DCNN-based Iris and Pupil Segmentation. *IEEE transactions on visualisation and computer graphics*, 2019.

[14] Kaipeng Z., Zhanpeng Z., Zhifeng L., and Yu Q. Joint Face Detection and Alignment Using Multitask Cascaded Convolutional Networks. *IEEE Signal Processing Letters*, 2016.

[15] Serengil, Sefik I., and Ozpinar A. HyperExtended LightFace: A Facial Attribute Analysis Framework. 2021 International Conference on Engineering and Emerging Technologies (ICEET). 1-4 (2021).

[16] Hunter J. D. Matplotlib: A 2d graphics environment. *Computing in Science & Engineering*, 9(3):90–95, 2007.

[17] Shirin J.-O., Christine C., Erin F., Hongyan L., Kari C.-Z., Edna F.-C., Curtis T., Kristin A. C., Samden L., and Martha S. Assessing depression severity with a self-rated vs. rater-administered instrument in patients with epilepsy. *Science Direct*, 2018.

[18] Corinna C. and Vladimir V. Support-Vector Networks. 1995.

[19] Pedregosa F., Varoquaux G., Gramfort A., Michel V., Thirion B., Grisel O., Blondel M., Prettenhofer P., Weiss R., Dubourg V., Vanderplas J., Passos A., Cournapeau D., Brucher M., Perrot M., and Duchesnay E. Scikit-learn: Machine Learning in Python. Journal of Machine Learning Research. 12:2825–2830, 2011.

[20] Erik S., and Igor K. Explaining prediction models and individual predictions with feature contributions. Knowl Inf Syst 41, 2013.

AUTHORS

Sanskar Rahul Nanegaonkar, a Data Scientist at I’mbesideyou Inc., is keen on uncovering actionable insights from complex data. Beyond data, he is an avid backpacker, exploring the serenity of nature.

Kotaro Ando, a CSO at I’mbesideyou Inc., leverages AI insights to address societal challenges like mental health, education, and the environment. Passionate about positive change, he finds inspiration in hiking outdoors.