Kalmora: A Voice-Based Journaling App for Real-Time Emotion Detection and Sustainable Mental Well-Being

Kalasi Kavundhya
School of Computing
SLIIT City Uni
Kalasikavundya@gmail.com

D.R.C.G.K. Dampallessa
School of Computing
SLIIT City Uni
charu14.d@gmail.com

Abstract **- The current tools for journaling depend on personal self-reporting which fails to match accurately with how people genuinely feel, and emotional states affect sustainable societal development. This research introduces Kalmora, which stands as a mobile voice-journaling application which utilizes Wav2Vec2 speech emotion recognition model that identifies seven basic emotions (happiness, sadness, anger, fear, disgust, neutral, surprise) in real time. Kalmora's secure dual frontend backend framework consisting of Flutter and Flask and Firebase elements performs time-based emotion assessment and individual wellness guidance. The model evaluated using controlled TESS data reached 99.8% accuracy which surpassed CNN-LSTM benchmark models at 94.1% accuracy. User testing involved observing real users interacting with the app to evaluate the ease of voice journaling, accuracy of emotion detection, and overall user experience, leading to improvements based on their feedback. Through its combination of objective emotional knowledge and practical tips Kalmora brings new possibilities to digital mental healthcare that enable sustainable emotional self-care practices.**

Keywords **- Speech Emotion Recognition (SER), Mental Health, Voice Journaling, Wav2Vec2, Personalized Recommendations, Affective Computing**

I. INTRODUCTION

The ongoing mental health problems constrain sustainable development because they reduce both personal health and work efficiency. Text-based journals help people reflect about themselves yet lack the ability to detect authentic emotions through vocal expressions [1]. Kalmora functions as a mobile journaling application through voice-based entries to detect emotions in real-time while offering personalized guidance in order to support emotional health sustainably. Depression affects over 280 million people globally to the estimation of the World Health Organization.

Due to the lack of clinical care, mobile mental health apps take on more convenient forms available to anyone with interest, focusing on emotional well-being.

II. BACKGROUND AND RELATED WORK

A. Voice Emotion Detection Technology

Recent advancements in deep learning, particularly transformer-based models like Wav2Vec2, have significantly improved the accuracy of SER systems. Yang *et al.* [2] demonstrate that fine-tuned Wav2Vec2.0 models achieve state-of-the-art performance across multiple speech emotion recognition benchmarks. Kalmora builds on these findings, utilizing transfer learning to adapt a pre-trained Wav2Vec2 model for emotion detection. The choice of Wav2Vec2 to be used in this study was since it has been proven to outperform in embedding rich emotional features observed in voice recordings using transfer learning. Wav2Vec2 was retrained over TESS dataset, i.e., a collection of 2,800 annotated emotional samples. After fine-tuning, the accuracy of the model successfulness was 99.8%. Besides, data augmentation strategies, such as noise injection and time-stretching, have been proposed to support robustness and reduce overfitting. Such interventions strengthened the capability of the system in accommodating the actual speech input on mobile devices.

B. Existing Journaling Applications

The existing mental health apps including Daylio and Reflectly need users to manually track moods, but this method delivers inconsistent and unreliable results. Value-based applications Mumble Journal and Echo Journal do

https://doi.org/10.54389/HVNM9177

not provide sophisticated emotion recognition features. Kalmora bridges the existing gap through its combination of SER with relevant recommendations enabling users to gain objective and meaningful insights from their experiences. Unlike existing applications such as Lid or Mumble Journal which use basic sentiment detection, Kalmora offers deeper voice-based emotional analysis across seven categories. While Woebot and MoodMirror incorporate AI and emotion tracking respectively, Kalmora uniquely combines accurate emotion recognition with wellness recommendations based on cognitive behavioral therapy. This makes it more proactive, private (voice-only), and adaptable for day-to-day emotional self-care.

III. METHODOLOGY

A. System Architecture

Kalmora follows a modular architecture:

1. **Frontend (Flutter):** A user-friendly interface for voice recording and journal management.
2. **Backend (Flask API):** Processes audio files and performs emotion detection.
3. **Database (Firebase):** Securely stores user data and journal entries.

The overall system architecture as shown in Fig 1.

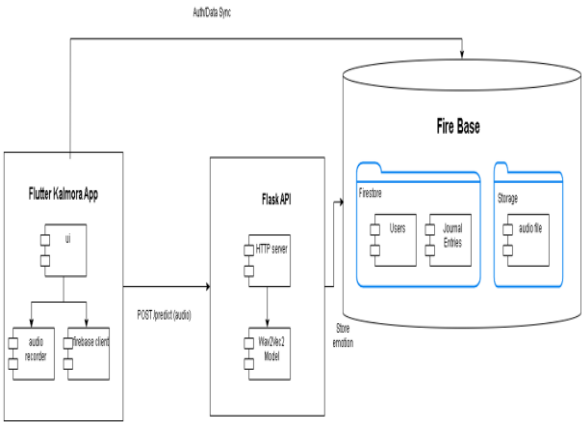

Fig 1: System architecture

B. Emotion Detection Model

Kalmora brought a Wav2Vec2 model trained on the Toronto Emotional Speech Set (TESS), a databank that collects 2,800 labelled audio files into seven emotional groups: happiness, sadness, anger, fear, disgust, neutral, and pleasant surprise. The material was also resampled to 16 kHz and normalized to unity gain thus providing equal input scaling. Additional processes of data augmentation, such as adding Gaussian noise and time-stretching, were used, focusing on less common classes of emotion, whose representation was made robust by means of methods. The split of 80-20 was introduced stratified to train and test. The training was conducted within 10 epochs using the Adam optimizer, learning rate of 1-5 and a batch size of 16, which was selected to reduce overfitting and improve generalization. In this setup, the model attained test-set accuracy of 99.8 %, so it was better compared with CNN-LSTM baseline (accuracy = 94.1 %) [2].

C. Personalized Recommendations

A rule-based system translates identified emotions to appropriate interventions which use cognitive behavioral therapy methods including anxiety breathing exercises combined with gratitude prompts [3].

D. Security and Privacy

Each of the users the study incorporated in the user-testing of the Kalmora application gave consent to participate in the study willingly before the research continued. The study conformed to ethics authorized under institutional research regulations though it did not formally solicit institutional review board (IRB) approval because the app was not intended to apply to a clinical setting. In order to preserve privacy and data safety, the voice entries of users were encoded (end to end) and any sensitive information was anonymized prior to complete analysis. Firebase Authentication was used to protect access to journals and emotional data and restrict access to them only to authorized people. No personally identifiable information (PII) had been gathered, and a participant could revoke at any time. Together, these protocols were developed in an attempt not only to continue to generate user trust, but also to align with the ethical norms governing within the sphere of digital mental-health software.

IV. RESULTS AND EVALUATION

A. Model Performance

Wav2Vec2 after fine-tuning delivered exceptional results based on precision exceeding 98%, recall exceeding 98% and F1-score values over 98%, and test run accuracy reached 99.8%. Emotion classification performance is detailed in the confusion matrix (Figure 2) The recorded data came from standard-speaking professionals who used amplified emotional cues while

working with a small voice range because of their limited presenter diversity. Through the architecturally optimized transformer model and 960 hours of self-supervised audio pretraining Wav2Vec2 processed prosodic features accurately from a small dataset consisting of 2,800 samples. The evaluation method included a stratified 80:20 split of the data for training and testing purposes and manual verification of randomized audio trials to validate performance consistently. The Kalmora model demonstrates superior performance, as shown in [8], when compared to CNN-LSTM models on the TESS dataset. It also achieves benchmark-level results across multiple controlled datasets, validated through randomized audio trials and manual verification.

The bar chart, Figure 2 illustrates the performance comparison between the CNN-LSTM and Kalmora (Wav2Vec2) models in terms of emotion detection accuracy using the TESS dataset. The CNN-LSTM model achieved an accuracy of 94.1%, while the Kalmora (Wav2Vec2) model significantly outperformed it with a remarkable 99.8% accuracy. This highlights the superior capability of the Wav2Vec2-based Kalmora model in capturing and interpreting emotional cues in speech data.

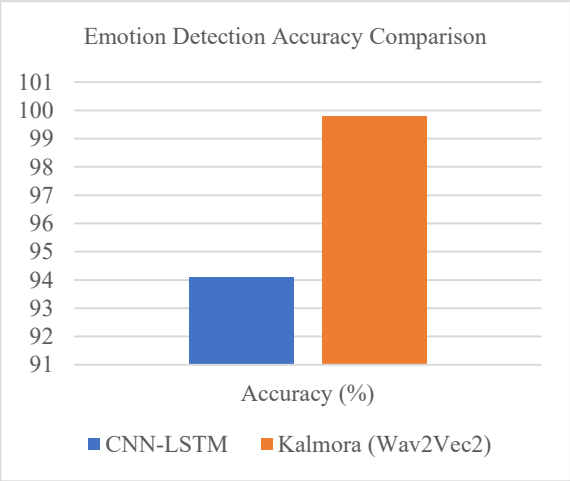

Fig 2: Emotion Detection Accuracy Comparison

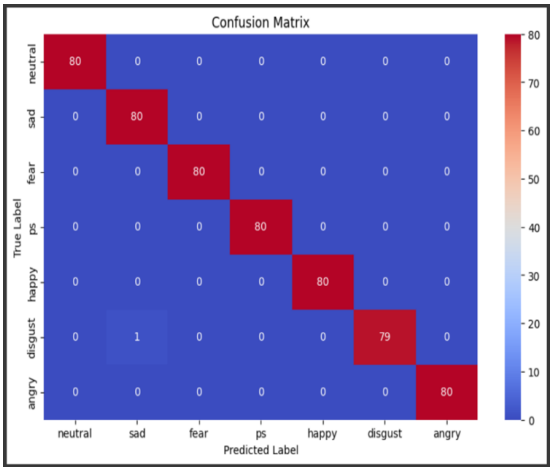

Fig 3: The confusion matrix

B. User Feedback

The Kalmora system underwent user testing on Android smartphones involving a sample of 10 participants selected through convenience sampling. The group included university students and young adults aged 18 - 28, with varying levels of technical experience. Users were instructed to record emotional voice entries and interact with features including voice journaling, playback, and real-time emotion recognition. Feedback was collected using semi-structured interviews and a 5-point Likert scale. Participants rated voice journaling experience an average of 4.6/5, and emotion prediction accuracy 4.4/5. Most users responded positively to the interface and emotional feedback features. Recommendations were made to reduce processing time for emotion detection and to improve button visibility in the UI, both of which were incorporated in subsequent design iterations. The User Acceptance Testing (UAT) phase confirmed that all core functionalities worked as expected and the application was ready for real-world use.

V. DISCUSSION

Kalmora establishes the possibility of running real-time speech emotion recognition through mobile devices through its high precision alongside low latency performance. Kalmora connects automatic emotion detection to tailored programmatic interventions because it resolves the requirement for adaptable mental wellness tools that move beyond basic mood tracking functionality. The method supports emotional resistance over time because it constitutes a central requirement for sustaining societies [4]. The “Digital Innovation for a Sustainable Future” program benefits from Kalmora because it

improves person wellness which creates sustainable community environments.

While the TESS dataset provides clean and clearly labeled audio samples, it is composed of acted emotional expressions by professional speakers in controlled environments. This may not fully represent the nuances of spontaneous, real-life emotional speech, potentially affecting real-world accuracy. To address this limitation, future work will involve testing the model on more naturalistic datasets such as RAVDESS or CREMA-D, which include greater speaker diversity, varying emotion intensities, and real-life acoustic conditions. This will help improve the model's generalizability to uncontrolled, user-generated voice inputs.

Kalmora stands out by using Wav2Vec2-based voice-only emotion detection, ensuring both emotional depth and user privacy, while generating wellness tips aligned with detected emotional states. This voice-first approach makes Kalmora more intuitive, private, and context-aware than many existing tools.

VI. FUTURE WORK

Model Validation: Test the emotion recognition model using varied and authentic speech datasets such as **CREMA-D** and **RAVDESS** for improved accuracy and robustness.

Hybrid Emotion Analysis: Combine **voice emotion detection** with **text sentiment analysis** to increase precision, especially for detecting faint emotions like *fear* and *disgust*.

Multilingual Support: Enable the system to recognize emotional expressions in **multiple languages**, improving accessibility and cultural adaptability.

Personalized Recommendations: The current recommendation engine maps detected emotions to cognitive behavioral therapy (CBT) techniques using a static rule-based system. While effective for initial deployment, this approach does not adapt to user behavior over time. Future iterations of Kalmora aim to integrate a machine learning-based recommendation system that learns from user interactions, emotional history, and feedback. This would enable the app to deliver adaptive, personalized wellness suggestions based on long-term usage patterns rather than fixed mappings.

User Feedback Loop: Implement a feedback system where users can rate the relevance of recommendations, allowing for continuous personalization improvement.

On-Device Emotion Detection: Develop lightweight emotion detection features that can run directly on the user's device, reducing dependence on cloud services and improving offline access.

Mental Health Support Features: Expand content to include **mindfulness training** and **breathing techniques** tailored to the user's current emotional state and feedback history.

Interactive Calendar: Add a feature that lets users **browse and reflect on past voice journals** through a calendar view, helping track emotional patterns over time.

User Testing: The aim of the present user testing loop became the creation of preliminary proof of concept. The next stages will be larger, real-life testing with a more heterogeneous population, including participants of different ages, potentially speaking a different language and having diverse cultural backgrounds, to determine the accuracy of emotional recognition and the overall user involvement of a high magnitude.

VII. CONCLUSION

Kalmora represents a significant advancement in digital mental health tools by combining voice journaling with real-time emotion detection. Its scalable architecture, high accuracy, and user-centric design make it a promising solution for enhancing emotional awareness. Future developments will further refine their capabilities, ensuring broader accessibility and impact.

ACKNOWLEDGMENT

I extend my gratitude to Ms. Charunika Dampalessa for her invaluable guidance and to all participants who contributed feedback during the testing and requirement gathering.

REFERENCES

[1] M. B. Akçay and K. Oğuz, "Speech emotion recognition: Emotional models, databases, features…," *Speech Communication*, vol. 116, pp. 56–76, 2020.

[2] J. Yang, A. S. Hasani, A. Ali, and E. M. Provost, "Fine-tuning Wav2Vec2 for robust cross-corpus speech emotion recognition," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process. (ICASSP)*, 2023.

[3] J. Firth, S. Torous, J. Nicholas, A. Carney, S. Pratap, and J. E. Rosenbaum, "Efficacy of smartphone-based mental health interventions," *World Psychiatry*, vol. 16, no. 3, pp. 287–298, 2017.

[4] A. Poria, E. Cambria, D. Hazarika, and N. Mazumder, "Multimodal emotion recognition: Trends and perspectives," *IEEE Trans. Affective Comput.*, Early Access, 2023.

[5] K. Zhang and H. Wang, "Contrastive domain adaptation for speech emotion recognition," presented at the *Emotion AI Conf.*, 2024.

[6] K. K. Fitzpatrick, A. Darcy, and M. Vierhile, "Delivering cognitive behavior therapy to young adults with symptoms of depression and anxiety using a fully automated conversational agent (Woebot): A randomized controlled trial," *JMIR Mental Health*, vol. 4, no. 2, e19, 2017.

[7] J. M. Garcia-Garcia, V. M. Penichet, and M. D. Lozano, "Emotion detection," in *Proc. XVIII Int. Conf. Human-Computer Interaction*, 2017, pp. 1–8.

[8] S. Chamishka, S. Subhashini, and R. Rodrigo, "A voice-based real-time emotion detection technique using recurrent neural network empowered feature modelling," *Multimedia Tools Appl.*, vol. 81, no. 24, pp. 35173–35194, 2022.

[9] Lid, "Lid: AI-Powered Voice Journaling." [Online]. Available: https://www.getlid.co/. [Accessed: 2-Apr-2025].

[10] G. Marques and S. Brissos, "Mental health mobile apps for depression diagnosis and monitoring: A systematic review," *J. Med. Syst.*, vol. 45, no. 7, p. 70, 2021.

[11] N. Martinezmartin and K. Kreitmair, "Ethical issues for direct-to-consumer digital psychotherapy apps: Addressing accountability, data protection, and consent," *JMIR Mental Health*, vol. 5, no. 2, e32, 2018.

[12] D. C. Mohr, D. Tomasino, S. Lattie, and S. Schueller, "IntelliCare: An eclectic, skills-based app suite for the treatment of depression and anxiety," *J. Med. Internet Res.*, vol. 19, no. 1, e10, 2017.

[13] Moodfit, "Tools & Insights for your mental health." [Online]. Available: https://www.getmoodfit.com/. [Accessed: 2-Apr-2025].

[14] MoodMirror, "MoodMirror: Multimodal Emotion Recognition for Mental Health." [Online]. Available: https://apps.apple.com/us/app/moodmirror-journal/id6739550398. [Accessed: 3-Apr-2025].

[15] Mumble, "Mumble Journal: Voice Recognition for Mental Health." [Online]. Available: https://apps.apple.com/us/app/mumble-journal/id6639614466. [Accessed: 3-Apr-2025].

[16] C. Nebeker, N. Harlow, C. Espinoza Giacinto, and K. Weibel, "Patient-centered research ethics review: Integrating ethics into your methods," *Ethics Human Res.*, vol. 41, no. 3, pp. 37–40, 2019.

[17] J. Nicholas, M. E. Larsen, A. Proudfoot, and H. Christensen, "Mobile apps for bipolar disorder: A systematic review of features and content quality," *J. Med. Internet Res.*, vol. 19, no. 4, e105, 2017.

[18] J. W. Pennebaker, "Expressive writing in psychological science," *Perspect. Psychol. Sci.*, vol. 13, no. 2, pp. 226–229, 2018.

[19] ReflectAI, "ReflectAI: Voice-to-Text Journaling with Sentiment Analysis." [Online]. Available: https://reflectaiapp.com/. [Accessed: 2-Apr-2025].

[20] Reflectly, "Reflectly: Journal and AI Mental Health Companion." [Online]. Available: https://reflectly.app/. [Accessed: 2-Apr-2025].

[21] S. M. Schueller, A. Aguilera, and D. C. Mohr, "Ecological momentary interventions for depression and anxiety," *Depress. Anxiety*, vol. 34, no. 6, pp. 540–545, 2017.

[22] J. W. Shin, J.-H. Chang, and N. S. Kim, "Voice activity detection based on statistical models and machine learning approaches," *Comput. Speech Lang.*, vol. 24, no. 3, pp. 515–530, 2010.

[23] J. Torous, K. Firth, J. Huckvale, and H. Christensen, "Clinical review of user engagement with mental health smartphone apps: Evidence, theory and improvements," *Evid. -Based Ment. Health*, vol. 21, no. 3, pp. 116–119, 2018.

[24] A. Aguilera and R. F. Muñoz, "Text messaging as an adjunct to CBT in low-income populations: A usability and feasibility pilot study," *Prof. Psychol.: Res. Pract.*, vol. 42, no. 6, pp. 472–478, 2011.

[25] M. A. Azad, J. Arshad, S. M. A. Akmal, F. Riaz, S. Abdullah, M. Imran, and F. Ahmad, "A comprehensive analysis of the adoption of security and privacy measures in mobile mental health applications," *IEEE Access*, vol. 9, pp. 35622–35637, 2021.