
Wellness Buddy: An AI-Enabled Emotion-Aware Companion for Digital Wellbeing

Supriya P. Jagtap, Arya Pawar, Samarth Bidkar, Nidhish Chincholkar, Aarush Sengupta and Adithya Sagri

Department of Information Technology, Vishwakarma Institute of Technology, Pune, Maharashtra, India, supriya.jagtap@vit.edu, arya.pawar24@vit.edu, samarth.bidkar24@vit.edu, nidhish.chincholkar24@vit.edu, aarush.sengupta24@vit.edu, adithya.sagri24@vit.edu

Abstract.

Existing music recommenders mainly use listening history and collaborative filtering, overlooking users' current emotions. Wellness Buddy integrates real-time facial emotion recognition, emotion-aware music recommendation, and digital wellbeing tools. A CNN trained on FER2013 classifies four emotions, while a content-based recommender uses valence, energy, tempo, and acousticness for music selection. The system also includes Pomodoro sessions, hydration reminders, journaling, and an LLaMA-3.1 wellness assistant. It achieves 93% accuracy, 0.92 precision/recall/F1, and 40 ms (~25 FPS) latency, supporting real-time interaction.

Keywords. Emotion recognition, digital wellbeing, music recommendation, convolutional neural networks, human-computer interaction.

1. INTRODUCTION

Advances in AI and deep learning have enabled practical emotion-aware computing for personalized digital wellness and content delivery. However, conventional music recommenders mainly rely on listening history and overlook users' current emotions. Wellness Buddy addresses this gap through real-time facial emotion recognition and emotion-based music recommendation, complemented by Pomodoro timers, reminders, and an AI conversational assistant to support productivity and wellbeing.

2. RELATED WORK

Studies show that multimodal emotion recognition using facial, speech, and physiological data can improve accuracy [1,2,6,9]. CNN and deep-learning models achieve strong performance in emotion and music recommendation but often require high computational resources and large datasets [8,17]. Existing systems also face challenges such as limited modalities, cold-start issues, privacy, and real-world validation [4,9–16]. Therefore, a scalable and privacy-aware framework integrating multiple data sources is needed.

Table 1 compares existing approaches with the proposed system.

Ref	Authors & Year	Key Findings	Limitations / Research Gap
[1]	Kim et al. (2010)	Multimodal input improves prediction of music preference	Limited personalisation; no real-time adaptation
[2]	Soleymani et al. (2017)	Multimodal fusion improves emotion understanding	Needs scalable handling of large heterogeneous data
[3]	Pichappan (2025)	Emotion improves recommendation relevance	Cold-start and sparse user data
[4]	Yousefian Jazi et al. (2019)	Interaction data improves recommendations	Few features; no multimodal data
[6]	Chakravathi et al.	Facial + speech raises recognition accuracy	No large-scale validation or physiological data
[8]	Van den Oord et al.	Predicts latent preferences from audio content	Ignores user emotion and multimodal signals
[17]	Bhagat et al.	82.56% training / 65.68% validation accuracy	Low validation accuracy; emotion confusion
[9]	Abdullah et al.	Multimodal outperforms unimodal systems	Real-time processing; dataset standardisation
-	Proposed (Wellness Buddy)	Unified emotion detection, recommendation and wellbeing tools; 93% accuracy at 40 ms	Single modality at present; multimodal extension planned

3. KEY CONTRIBUTION

The proposed framework integrates real-time emotion recognition, emotion-aware music recommendation, productivity tools, and wellbeing support into a single system. It combines CNN-based emotion detection, valence- and energy-based recommendation, and an LLaMA-3.1 chatbot to provide personalized wellness and emotional support.

4. METHOD

The system uses Haar cascade and CNN-based emotion detection to classify four emotions from 48×48 facial images. Detected emotions guide music recommendation using valence, energy, acousticness, and tempo. Additional features include Pomodoro, hydration reminders, journaling, scheduling, and an LLaMA-3.1 chatbot.

5. EXPERIMENTS

Each module—emotion detection, music recommendation, productivity tools, and chatbot—was tested individually and as an integrated system for accuracy, functionality, and responsiveness.

6. RESULTS

The emotion classifier achieved 93% accuracy, with 0.92 precision, recall, and F1-score. The average latency was 40 ms (~25 FPS), enabling real-time interaction. The music

recommender provided emotion-matched tracks, while productivity and chatbot modules functioned correctly.

Table 2. Performance evaluation metrics of the proposed system.

Metric	Formula	Experimental Value
Accuracy	$(TP + TN) / (TP + TN + FP + FN)$	93%
Precision	$TP / (TP + FP)$	0.92
Recall	$TP / (TP + FN)$	0.92
F1-score	$2PR / (P + R)$	0.92
Total latency	$T_{face} + T_{class} + T_{update}$	40 ms
Theoretical FPS	$1000 / T_{total}$	25 FPS

As shown in Table 2, the system sustains high classification performance while keeping latency low enough for real-time use.

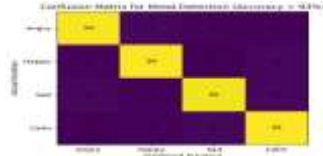


Fig.6.1. Confusion matrix for the four-class facial emotion recognition model.

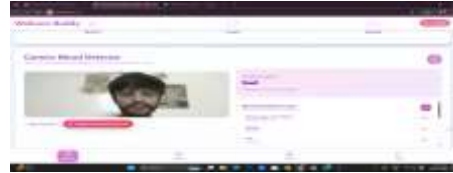


Fig. 6.2. Real-time emotion detection and music recommendation dashboard.

7. DISCUSSION

The proposed system provides accurate, real-time emotion recognition and emotion-based recommendations, supporting personalized wellness and productivity. However, performance may decline under poor lighting, occlusion, and subtle expressions due to its reliance on facial input.

8. CONCLUSIONS

Wellness Buddy combines CNN-based emotion recognition, emotion-aware music recommendation, productivity tools, and an LLaMA-3.1 chatbot. It achieved 93% accuracy, approximately 0.92 precision, recall, and F1-score, and 40 ms latency (~25 FPS). Future work will focus on multimodal inputs, advanced models, and personalized recommendations.

9. REFERENCES

- [1] Kim, Y.E., Schmidt, E.M., Migneco, R., Morton, B.G., Richardson, P., Scott, J., Speck, J.A., Turnbull, D.: Music emotion recognition: a state-of-the-art review. In: Proc. 11th Int. Society for Music Information Retrieval Conf. (ISMIR), pp. 255-266 (2010).
- [2] Soleymani, M., Garcia, D., Jou, B., Schuller, B., Chang, S.-F., Pantic, M.: A survey of multimodal sentiment analysis. *Image and Vision Computing* 65, 3-14 (2017).
- [3] Pichappan, P.: A review of the emotion-induced music recommendation systems. *Journal of Digital Information Management* 23(2) (2025). ISSN 0972-7272.

- [4] Yousefian Jazi, S., Kaedi, M., Fatemi, A.: An emotion-aware music recommender system: bridging the user's interaction and music recommendation. *Multimedia Tools and Applications* 80, 13559-13574 (2021).
- [5] Singh, A.K., Mandal, D., Kumar, S., Venkat, D.: Emotion-based music recommendation system using deep learning. In: 2025 3rd IEEE Int. Conf. on Industrial Electronics: Developments & Applications (ICIDeA), Bhubaneswar, pp. 1-6 (2025).
- [6] Chakravathi, K., Tejashwini, G., SriVibhavana, G., Shirisha, C.: Context-aware music recommendation using multimodal emotion recognition. *Int. J. of Bioinformatics and Biomedical Engineering Research (IJBAR)*. ISSN 2249-3352.
- [7] Talaghat, M.A., Parvinnia, E., Mehrabi, M., Boostani, R.: A multimodal deep network for music emotion recognition using audio chorus and lyrics. *IEEE Access* 13, 177509-177519 (2025).
- [8] Van den Oord, A., Dieleman, S., Schrauwen, B.: Deep content-based music recommendation. In: *Advances in Neural Information Processing Systems (NIPS) Workshop on Deep Learning and Representation Learning* (2013).
- [9] Saleem, S.M., Ameen, S.Y., Sadeeq, M.A.M., Zeebaree, S.R.M.: Multimodal emotion recognition using deep learning. *Journal of Applied Science and Technology Trends* 2(2), 52-58 (2021).
- [10] Sunitha, M., Adilakshmi, T., Chitturi, S.S., Kapila, P.: Deep learning-based emotion-aware music recommendation system. In: *Proc. 2nd Int. Conf. on Cognitive and Intelligent Computing (ICCIC 2022)*, Cognitive Science and Technology. Springer, Singapore (2023).
- [11] Sathya, A.R., Varma, A.R.: Emotion-based song recommendation system. In: *Intelligent Data Engineering and Analytics (FICTA 2023)*, Smart Innovation, Systems and Technologies, vol. 371. Springer, Singapore (2023).
- [12] Regulwar, G.B., Kathirisetty, N.: Predictive music based on mood. *Int. J. of Scientific Research in Science, Engineering and Technology (IJSRSET)*. ISSN 2395-1990.
- [13] Mollahosseini, A., Hasani, B., Mahoor, M.H.: AffectNet: a database for facial expression, valence, and arousal computing in the wild. *IEEE Transactions on Affective Computing* 10(1), 18-31 (2019).
- [14] Li, S., Deng, W.: Deep facial expression recognition: a survey. *IEEE Transactions on Affective Computing* 13(3), 1195-1215 (2022).
- [15] Shah, D., Chavan, K., Shah, S., Kanani, P.: Real-time facial emotion recognition. In: 2021 2nd Global Conf. for Advancement in Technology (GCAT), Bangalore, pp. 1-4 (2021).
- [16] Zhou, N., Liang, R., Shi, W.: A lightweight convolutional neural network for real-time facial expression detection. *IEEE Access* 9, 5573-5584 (2021).
- [17] Bhagat, D., Vakil, A., Gupta, R.K., Kumar, A.: Facial emotion recognition (FER) using convolutional neural network (CNN). In: *ICMLDE 2023, Procedia Computer Science* 235, 2079-2089 (2024).

Biographies



Supriya P. Jagtap received the BE degree from SPPU in 2012 , ME degree from SPPU 2016 and Phd from Symbiosis International university in 2024



Arya Pawar Pursuing B.Tech degree from VIT Pune (TY)



Samarth Bidkar Pursuing B.Tech degree from VIT Pune (TY)



Nidhish Chincholkar Pursuing B.Tech degree from VIT Pune (TY)



Aarush Sengupta Pursuing B.Tech degree from VIT Pune (TY)



Adithya Sagri Pursuing B.Tech degree from VIT Pune (TY)