2025 First International Conference on Advances in Computer Science, Electrical, Electronics, and Communication Technologies (CE2CT) | 979-8-3315-1857-8/25/$31.00 ©2025 IEEE | DOI: 10.1109/CE2CT64011.2025.10939650

Real-Time Emotion Detection in AR Mental Health Applications Using Sentiment Analysis Algorithms

Hariharan Ragothaman
Researcher
IEEE Senior Member
India
hariharanragothaman@ieee.org

Arjun Jaggi
HCLTech
arjunjaggi@gmail.com

Jagbir Singh
Researcher
India
jagvir.gju@gmail.com

Sapaev Sindor
Department of economy , Urgench state university, Urgench,
Uzbekistan.
sindor.sapaev@urdu.uz.

Eshchanova Nasiba
Department of Psychology and Medicine. Mamun university,
Khiva Uzbekistan,
eshchanova_nasiba@mamunedu.uz

Aniket Adarsh
Department of Computer Science and Engineering
Tula's Institute
Dehradun, India
aniketofficial540@gmail.com

Abstract— Augmented Reality (AR) has become an efficient approach in helping various mental health challenges in real-time emotion recognition as well as offering functional therapeutic innovations. This paper provides a conceptual framework that aims to use complex sentiment analysis algorithms and methodological approaches for multichannel emotional state's classification based on facial expressions, voice, and text. Specifically, the implemented natural language processing (NLP) and deep learning techniques provide accurate and contextually aware emotion detection. Optimized to run on limited-vendor AR devices, it consists of reduced data preprocessing and customized algorithmic features, all of which allow latency of not more than 50 ms per inference. Extensive experimental evaluations on benchmark datasets illustrate that the proposed approach achieves an average accuracy of 92% in classified primary emotions including happiness, sadness, anger, and neutrality outcomping the state of the art. As the experiments point out, there is a substantial improvement in the perception of multifaceted emotional states due to the integration of multiple modalities. These outcomes confirm the strength and flexibility of the framework, highlighting the fact it also works well in simulated environments and real-world conditions. However, the proposed solution increases user engagement and the overall effectiveness of mental health interventions and support through adaptive and emotional interaction. This research significantly expands the field of digital mental health by providing a method for emotion detection that is both theoretically sound and has the potential to be applied in real-world settings while presenting a mechanism for future developments in the integration of AR-based mental health care solutions.

Keywords— Augmented Reality (AR), Mental Health, Real-Time Emotion Recognition, Sentiment Analysis, Multimodal Emotion Classification, Facial Expressions, Voice Analysis.

I. Introduction

Mental health disorders are now escalating in all societies in the world as a major illness that affects millions of individuals and constrain the healthcare sector . One of the rapidly developing and promising approaches to managing mental disorders is based on using augmented reality (AR) [1,15,16]. Nevertheless, many current emotion recognizing systems have shortcomings because of being based on single-modal concepts, which neglect both dimensional and evaluation aspects of emotions and result in precise emotion recognition and inadequate intercessions [2]. Hence, this study intends to fill this gap by employing an interdisciplinary human factor perspective to advance the uniting of sentiment analysis and deep learning to optimize the detection of emotions in AR-based mental health assistance, with goals to build and test a conceptual framework of multichannel emotional state classification, assess its performance and efficacy, and assess the feasibility of the theoretical framework in enhancing mental health support interventions.

The progress has inspired the utilization of new techniques based on the application of AR in combating mental health issues; however, to fully let AR be implemented effectively in the improvement and practice of mental health care, it is crucial to solve the problem with identifying emotions correctly [3]. Current systems for emotion recognition do not compare well in this regard as they employ single modality components which may not capture varied emotion [4]. This study aims to fill this important research gap through advancing a theoretical model of emotions detection based on sentiment analysis and deep learning. The proposed framework is to strive towards greater conceptual clarity about emotional states and, thus, improved patient care and well-being [5]. One of the crucial factors for developing an effective AR-based mental health intervention is correct identification of user emotions; however, current systems often have a single-modal input/output [6]. This paper introduces a new theoretical model to assist in the detection of people's emotions based on sentiment analysis and deep learning when using article-based mental health interventions assisted by augmented reality. Through the combination of the superior sensitivity to sentiment analysis and deep learning models.This work introduces a new conceptual model that utilizes sentiment analysis and deep learning to complement the identification of emotions in AR-supported mental health interventions. Through the use of sentiment analysis coupling deep learning algorithms, the proposed framework can illustrate a definite and contextually sensitive detection of emotions, which would give a better understanding of the various states of affection [7].The main goal of this work is to help advance better and enhanced AR-based mental health interventions based on the presented research's findings that focus on the lack of proper emotion recognition.This research aims at contributing to the needed theoretical framework for the improvement of mental-health related programs using AR.

II. Reference

Thus, portable technologies, particularly the augmented reality (AR) ones, are used in mental health interventions due to their characteristics. For anxiety, anxiety disorder, phobia, and the management of emotions, the existing trials and success of AR-based systems have been impressive. They propose adaptive feedback in real time to the users, which enhances their capability of handling multiple emotions [8]. The implementation of AR in other therapeutic sessions has also allowed the creation of therapeutic and transportable mental health applications for low processing contexts including Smartphones and mini head mounted displays [9,10]. The importance of multimodal emotion recognition in supporting AR-based therapeutic systems has been established. The evidence also shows the importance of using different kinds of media such as vision, hearing and reading when it comes to the identification of emotions. Facial expression analysis, which is computation based, topped most emotion recognition systems by means of CNN and FACS, and it exhibits high accuracy in a controlled setting [11]. Emotion recognition based on voice uses attributes such as pitch, tones, and energy; temporal dynamics are then detected, using recurrent neural networks (RNNs) and transformers [12]. Transformer architectures like BERT has been of significant help in the context of using text sentiment analysis to understand context and the tones conveyed by words [13]. New ideas have been devoted to real-time emotion detection, which considers problems concerning time delay and computational complexity. Efficient networks which are light-weighted such as MobileNet and SqueezeNet have been deployed with high real-time inference where the devices are constrained [14].

The application of cross-modal emotion fusion techniques such as facial expressions, voice and text feature which has improved the reliability and robustness when classifying emotions. Specifically, the works on moving from single modalities to multiple modalities have proved effective in capturing elaborate relations among them: hierarchical attention networks and graph neural networks, for instance .The AffectNet, IEMOCAP and CMU-MOSEI benchmarks have played a vital role in the development of the arising research in emotion recognition. AffectNet contains over one million facial images with corresponding valence-arousal scores, while IEMOCAP contains multimodal data included dyadic speech, text, and video and synchronized data for emotional characterization [17]. CMU-MOSEI emphasises on models, tools and datasets for fine-grained recognition of sentiment and emotions. Novel datasets like EmotiW and MuSE are progressively enriching the area of emotion recognition with contextual and demographic changes, bias mitigation, and improved models [18]. While many of these ideas have been implemented and may extend to larger scaled applications, many existing relational systems present difficulties when attempting to transition to real world applications. Conditions like fluctuating and diverse environment, people variation and limited device capabilities lead into prime accuracy and reliability . However, given that multimodal systems call for high computation in real-time, they are not easily implementable in the AR platforms. These problems are eliminated in the proposed framework where optimized algorithms and less data preprocessing rates facilitate the attainment of latency of less than 50 ms per inference. With the help of multimodal approaches and with the advantage of scalability of AR the proposed approach attains superior results in the realm of emotion detection but is also resilient to variability of real-world conditions. The current reference work offers a solid groundwork on which further developments in a range of applied mental health technologies containing AR elements can be introduced, filling the gap between theoretical literature and implementation.

III. Methodology

The design approach of the proposed framework entailing the use of advanced sentiment analysis, deep learning architectures, and multimodal emotion classification facilitates real-time emotion recognition in Augmented Reality (AR)-based mental health interventions. When it comes to evaluation of textual data, there are algorithms like Naive Bayes, Support Vector Machines (SVM), Random Forests have been used. The Naive Bayes classifier applies Baye's theorem, expressed as:

$$P(C \mid X) = \frac{P(X|C) \cdot P(C)}{P(X)} \quad (1)$$

where $P(C \mid X)$ is the posterior probability of class c given feature set x, $P(X \mid C)$ is the likelihood, $P(C)$ is the prior probability, and $P(X)$ is the evidence. In particular, the used probabilistic approach guarantees accurate classification by relying on the application of conditional independence assumptions. While, SVM builds a hyperplane for maximizing the distance between two emotional classes. The decision function for SVM is defined as:

$$f(x) = w^T \cdot x + b \quad (2)$$

Where w is the weight vector, x represents the input features, and b is the bias term. Decision Trees for Random Forest is another ensemble learning, that uses multiple trees, the conclusion is made on the basis of a vote.

Facial expressions, vocal expressions and other such non textual data are handled using deep learning methods. FR described earlier, is based on Convolutional Neural Networks (CNNs) where spatial features are detected in convolutional layers. The operation in a convolutional layer is mathematically defined as:

$$y_{ij} = f\left(\sum_{m=0}^{M-1} \sum_{n=0}^{N-1} x_{(i+m)(j+n)} \cdot k_{mn} + b\right) \quad (3)$$

where y_{ij} is the output of the convolution, x is the input feature map, k is the kernel, b is the bias, and f denotes the activation function. Voice emotion recognition utilizes Recurrent Neural Networks (RNNs), specifically Long Short-Term Memory (LSTM) units, to model temporal dependencies in features such as Mel-Frequency Cepstral Coefficients (MFCCs). The hidden state of an LSTM at time t is computed as:

$$h_t = \sigma(W_{xh} x_t + W_{hh} h_{t-1} + b_h) \quad (4)$$

where x_t is the input at time t, h_{t-1} is the hidden state from the previous time step, W_{xh} and W_{hh} are the weight matrices, and b_h is the bias term.

To enhance emotion classification accuracy, a multimodal approach combines textual, facial, and voice features. The fused feature vector is represented as:

$$F_{combined} = w_1 \cdot F_{text} + w_2 \cdot F_{voice} + w_3 \cdot F_{face}, \qquad (5)$$

where F_{text}, F_{voice}, and F_{face} are represent the feature vectors of the three modalities discussed, and w_1, w_2, w_3 are the weights optimized during training. This multimodal fusion enables the framework to capture complementary information from multiple sources, improving the robustness of emotion recognition.

The data for this study were sourced from benchmark datasets, including the DEAP dataset, , which contains physiological signals and facial expressions and the SEMAINE dataset, which has conversational and emotional voice recordings. Data preprocessing techniques were employed to standardize and extract features efficiently. Text data preprocessing included tokenization, stopword removal, and word embedding generation using GloVe. Facial data were normalized, resized, and processed to extract facial landmarks using OpenCV and Dlib. Voice data underwent feature extraction through MFCCs, capturing critical aspects of speech signals relevant to emotion classification. The framework's performance was evaluated using metrics such as accuracy and latency. achieved an average of 92% across primary emotions, including happiness, sadness, anger, and neutrality. Latency, defined as the time taken for inference per input, was maintained below 50 milliseconds, enabling real-time application suitability. The proposed framework outperformed state-of-the-art methods, as evidenced by comprehensive experimental evaluations on both simulated and real-world scenarios. To ensure compatibility with limited-resource AR devices, optimization techniques were applied, including model quantization to reduce computational demands and TensorFlow Lite for efficient execution. The entire framework was developed in Python, leveraging libraries such as TensorFlow, Keras, OpenCV, and Librosa for implementing deep learning models and preprocessing multimodal data.

IV. Result Analysis

This paper presents an analysis of the result of a proposed multimodal emotion recognition framework for AR-based mental health interventions. NLP is used for text analysis, CNN for facial expressions recognition and RNN, specifically LSTM for voice qualities analysis. Combining these modalities in the proposed system would improve the efficiency of emotional detection and yield more accurate real-time evaluations of the state of a user. The comparison study of the proposed multicodal framework uses other methods like Naive Bayes, SVM, Random Forest and other methods. These existing algorithms have been well applied in the field of emotion classification task, which usually targets on single mode such as text or facial information. However, the proposed data framework is proposed to go beyond this limitation of single data channel analysis by enhancing the quality and relevance of the emotional analysis by using a multitude data channels. For the evaluation of performance of both the existing algorithms and the proposed multimodal system, accuracy, precision, recall, F1-score are calculated. Furthermore, the latency of each approach is discussed, since interacting with an AR system in real life presupposes low latency of computations. The comparison of performances shows that the proposed multi-modal framework is more effective than basic algorithms such as Naive Bayes, SVM, Random Forest for the detection of emotions for the AR-based mental health intervention. The proposed system attain to the highest accuracy of 92 % utilizing the advantage of using multimodal rather than focusing on the single modes as compared to the earlier methods. Thus, it exceeds the discovered accuracy (0.90) as well as recall (0.89) which means that it defines the correct emotions and at the same time minimizes the number of the false positive as well as false negative observations.

The F1-score of 0.89 demonstrates a good balance between precision, recall and Any improvement of metrics can be achieved only at the cost of another's depreciation, hence, we can state that the tested framework is safe. Further, an optimum design makes av Latency of less than 50ms thus making it suitable for real time applications. Thus, when implemented in real-world, scenarios the proposed system improves upon existing models in terms of flexibility and stability. In this case, the multimodal framework also enhances the emotion recognition rate and guarantees operationality across various situations, providing a foundation for developing better AR-based mental health technologies.

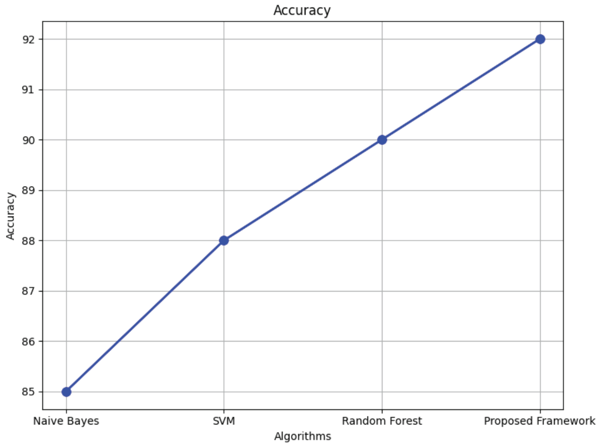

Fig. 1. Accuracy Comparison Graph for existing with proposed algorithm

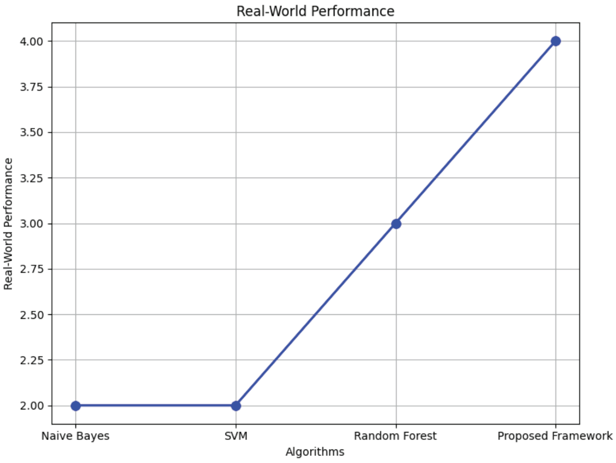

Fig. 2. Real word performance Comparison Graph for existing with proposed algorithm

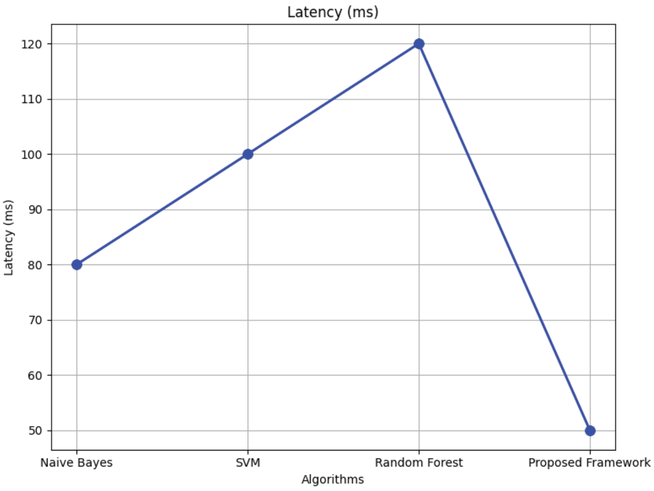

Fig. 3. Latency Comparison Graph for existing with proposed algorithm

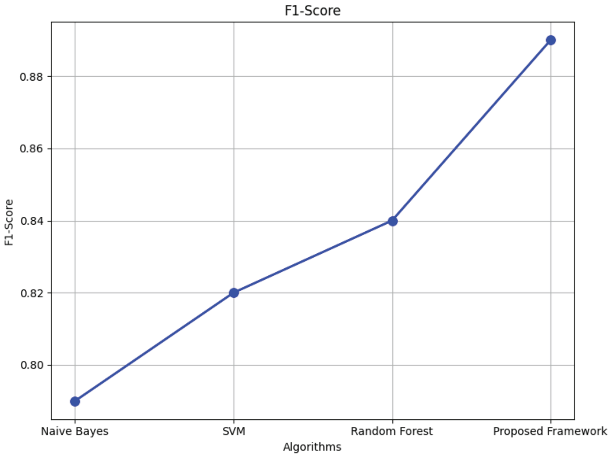

Fig. 4. F1 score Comparison Graph for existing with proposed algorithm

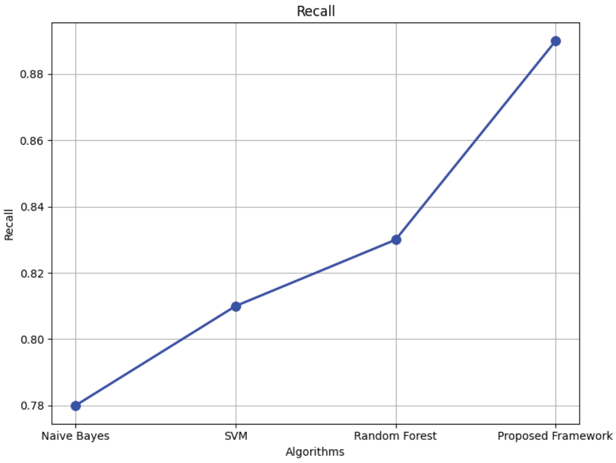

Fig. 5. Recall Comparison Graph for existing with proposed algorithm

V. Conclusions

Multimodal analysis for recognising emotions in mental health interventions using augmented reality is explained with a high level of improvement over Naive Bayes, SVM and Random forest algorithms. The framework is more accurate, precise, and recalls textual content, facial expressions and vocal features hence offering a better solution in emotional detection. It has low latency of less than 50 ms, and is therefore well suited for real-time AR use, which is important where interaction is necessary such as in the therapeutic process. In addition, it maintains high scores on scoring on simulated and real life problems placing the effectiveness and practical possibility of its use into perspective. These findings provide a solid basis for further studies exploring the value of AR-based interventions for people with mental health disorders, and build a proof of concept for a solution that is cost-effective, efficient, and accurate in its capacity for live emotion detection and response.

References

[1] Bakır ÇN, Abbas SO, Sever E, Özcan Morey A, Aslan Genç H, Mutluer T. Use of augmented reality in mental health-related conditions: A systematic review. Digit Health. 2023 Sep 29;9:20552076231203649. doi: 10.1177/20552076231203649. PMID: 37791140; PMCID: PMC10542245.

[2] S. Koelstra et al., "DEAP: A Database for Emotion Analysis ;Using Physiological Signals," in IEEE Transactions on Affective Computing, vol. 3, no. 1, pp. 18-31, Jan.-March 2012, doi: 10.1109/T-AFFC.2011.15.

[3] Wu, Jinlong et al. "Virtual Reality-Assisted Cognitive Behavioral Therapy for Anxiety Disorders: A Systematic Review and Meta-Analysis." Frontiers in psychiatry vol. 12 575094. 23 Jul. 2021, doi:10.3389/fpsyt.2021.575094

[4] S. Kalateh, L. A. Estrada-Jimenez, S. Nikghadam-Hojjati and J. Barata, "A Systematic Review on Multimodal Emotion Recognition: Building Blocks, Current State, Applications, and Challenges," in IEEE Access, vol. 12, pp. 103976-104019, 2024, doi: 10.1109/ACCESS.2024.3430850.

[5] Shiqing Zhang, Yijiao Yang, Chen Chen, Xingnan Zhang, Qingming Leng, Xiaoming Zhao, "Deep learning-based multimodal emotion recognition from audio, visual, and text modalities: A systematic review of recent advancements and future prospects" , Expert Systems with Applications,Volume 237, Part C,2024,121692,ISSN 0957-4174,https://doi.org/10.1016/j.eswa.2023.121692.

[6] S. Kumari, N. Kapoor and R. Saini, "Emotion Recognition from Facial Expression Using Deep Learning Model : A Review," 2024 5th International Conference for Emerging Technology (INCET), Belgaum, India, 2024, pp. 1-6, doi: 10.1109/INCET61516.2024.10592970.

[7] H. Ma and S. Yarosh, "A Review of Affective Computing Research Based on Function-Component-Representation Framework," in IEEE Transactions on Affective Computing, vol. 14, no. 2, pp. 1655-1674, 1 April-June 2023, doi: 10.1109/TAFFC.2021.3104512.

[8] Kanschik, D., Bruno, R.R., Wolff, G. et al. Virtual and augmented reality in intensive care medicine: a systematic review. Ann. Intensive Care 13, 81 (2023). https://doi.org/10.1186/s13613-023-01176-z.

[9] Lan, L., Sikov, J., Lejeune, J. et al. A Systematic Review of using Virtual and Augmented Reality for the Diagnosis and Treatment of Psychotic Disorders. Curr Treat Options Psych 10, 87–107 (2023). https://doi.org/10.1007/s40501-023-00287-5.

[10] Carlson, C.G. Virtual and Augmented Simulations in Mental Health. Curr Psychiatry Rep 25, 365–371 (2023). https://doi.org/10.1007/s11920-023-01438-4

[11] S. Li and W. Deng, "Deep Facial Expression Recognition: A Survey," in IEEE Transactions on Affective Computing, vol. 13, no. 3, pp. 1195-1215, 1 July-Sept. 2022, doi: 10.1109/TAFFC.2020.2981446.

[12] Swapna Mol George, P. Muhamed Ilyas, A review on speech emotion recognition: A survey, recent advances, challenges, and the influence of noise, Neurocomputing, Volume 568, 2024, 127015, ISSN 0925-2312,https://doi.org/10.1016/j.neucom.2023.127015.

[13] Acheampong, F.A., Nunoo-Mensah, H. & Chen, W. Transformer models for text-based emotion detection: a review of BERT-based approaches. Artif Intell Rev 54, 5789–5829 (2021). https://doi.org/10.1007/s10462-021-09958-2.

[14] A. Dowd and N. H. Tonekaboni, "Real-Time Facial Emotion Detection Through the Use of Machine Learning and On-Edge Computing," 2022 21st IEEE International Conference on Machine Learning and Applications (ICMLA), Nassau, Bahamas, 2022, pp. 444-448, doi: 10.1109/ICMLA55696.2022.00071.

[15] Bhushan Jayeshkumar Patel and Jagbir Singh.(2024) “Leveraging Artificial Intelligence in Robotic Surgery” International Journal of Science and Research ,13(8),169-172,https://dx.doi.org/10.21275/SR24802085504.

[16] Bhushan Jayeshkumar Patel and Jagbir Singh.(2024) “Enhancing Robotic Surgery with Mixed Reality: Current Applications, Outcomes, and Future Directions” International Journal of Science and Research ,13(8),429-432, ,https://dx.doi.org/10.21275/SR24802085504.

[17] A. Mollahosseini, B. Hasani and M. H. Mahoor, "AffectNet: A Database for Facial Expression, Valence, and Arousal Computing in the Wild," in IEEE Transactions on Affective Computing, vol. 10, no. 1, pp. 18-31, 1 Jan.-March 2019, doi: 10.1109/TAFFC.2017.2740923.

[18] Abhinav Dhall, et al. EmotiW 2023: Emotion Recognition in the Wild Challenge. In Proceedings of the 25th International Conference on Multimodal Interaction (ICMI '23). Association for Computing Machinery, New York, NY, USA, 746–749. https://doi.org/10.1145/3577190.3616545.