

ORIGINAL

Advancements and challenges in emotion recognition technologies

Avances y retos en las tecnologías de reconocimiento de emociones

Anamika Kumari¹ , Smita Pallavi¹

¹Department of Computer Science, Birla Institute of Technology. Patna-800014, India.

Cite as: Kumari A, Pallavi S. Advancements and Challenges In Emotion Recognition Technologies. Gamification and Augmented Reality. 2025; 3:88. <https://doi.org/10.56294/gr202588>

Submitted: 12-03-2024

Revised: 19-08-2024

Accepted: 23-03-2025

Published: 24-03-2025

Editor: Dr. Adrián Alejandro Vitón Castillo 

Corresponding Author: Anamika Kumari 

ABSTRACT

Introduction: emotion recognition is a transformative technology that enhances human-computer interaction by enabling systems to interpret and respond to human emotions effectively.

Objective: this paper investigates the current landscape of emotion recognition technologies, emphasizing the diverse sources of emotional data, including facial expressions, voice, physiological signals, and textual content.

Method: we explore the methodologies and algorithms employed for emotion classification, ranging from traditional machine learning techniques to advanced deep learning models such as Convolutional Neural Networks (CNNs) and Recurrent Neural Networks (RNNs).

Results: this study provides a comparative analysis of various emotion recognition approaches, evaluating their accuracy, robustness, and computational efficiency.

Conclusions: this paper contributes to the ongoing discourse on emotion recognition by offering a comprehensive overview of current trends, challenges, and opportunities for advancing the field.

Keywords: Emotion Recognition; Human-Computer Interaction; Machine Learning; Deep Learning; Algorithms.

RESUMEN

Introducción: el reconocimiento de emociones es una tecnología transformadora que mejora la interacción persona-ordenador al permitir que los sistemas interpreten y respondan a las emociones humanas con eficacia.

Objetivo: este artículo investiga el panorama actual de las tecnologías de reconocimiento de emociones, haciendo hincapié en las diversas fuentes de datos emocionales, como las expresiones faciales, la voz, las señales fisiológicas y el contenido textual.

Método: exploramos las metodologías y algoritmos empleados para la clasificación de emociones, que van desde técnicas tradicionales de aprendizaje automático hasta modelos avanzados de aprendizaje profundo, como las redes neuronales convolucionales (CNN) y las redes neuronales recurrentes (RNN).

Resultados: este estudio proporciona un análisis comparativo de varios enfoques de reconocimiento de emociones, evaluando su precisión, robustez y eficiencia computacional.

Conclusiones: este artículo contribuye al debate en curso sobre el reconocimiento de emociones, ofreciendo una visión global de las tendencias actuales, los retos y las oportunidades para avanzar en este campo.

Palabras clave: Reconocimiento de Emociones; Interacción Persona-Ordenador; Aprendizaje Automático; Aprendizaje Profundo; Algoritmos.

INTRODUCTION

Emotion recognition is a critical component of affective computing, which seeks to bridge the gap between human emotions and computational systems. The ability to detect and interpret emotions can significantly enhance user experience in various applications, including virtual assistants, mental health diagnostics, and security systems. Humans express a wide range of emotions through multiple channels, such as facial expressions, voice, and physiological signals. However, accurately classifying these emotions remains a complex challenge due to their variability and subtlety.⁽¹⁾ This paper reviews the current technologies used for emotion detection, focusing on the sources of emotional data and the classification techniques employed. We also discuss the effectiveness of different algorithms and highlight the limitations of existing approaches. By providing a comprehensive analysis, we aim to contribute to the ongoing development of more robust and accurate emotion recognition systems.

Sources of Emotional Data

Facial Expressions

Facial expressions are one of the most studied sources of emotional data. The Facial Action Coding System (FACS) developed by Paul Ekman is a widely used method for analyzing facial movements. Recent advancements in computer vision and deep learning, particularly Convolutional Neural Networks (CNNs), have significantly improved the accuracy of facial emotion recognition systems.⁽²⁾

Voice and Speech

The tone, pitch, and rhythm of speech provide critical cues for emotion recognition. Speech emotion recognition (SER) systems analyze these acoustic features using techniques like Mel-Frequency Cepstral Coefficients (MFCCs) and machine learning algorithms. Deep learning models, such as Recurrent Neural Networks (RNNs) and Long Short-Term Memory (LSTM) networks, have been employed to capture temporal dependencies in speech.⁽³⁾

Physiological Signals

Physiological signals such as heart rate, skin conductance, and brain activity offer a more objective measure of emotions. Wearable devices and biosensors are increasingly used to capture these signals. Machine learning models, particularly Support Vector Machines (SVMs) and Neural Networks, are commonly employed to classify emotions based on physiological data.

Text Analysis

Emotion recognition from text involves Natural Language Processing (NLP) techniques to analyze written content. Sentiment analysis, a subset of text-based emotion recognition, classifies text into positive, negative, or neutral categories. More advanced models, such as BERT and GPT, have been applied to capture nuanced emotions in text.

Literature review

Ekman and Friesen's Facial Action Coding System (FACS) (1978)⁽⁴⁾ is a seminal work in the field of emotion recognition, providing a systematic framework for analyzing facial movements and their association with emotions. FACS categorizes facial expressions into Action Units (AUs), which represent specific muscle movements, enabling precise identification of emotional states such as happiness, sadness, and anger. This methodology has become a cornerstone in affective computing, offering a standardized approach for researchers to study facial expressions across cultures and contexts. FACS has been widely adopted in psychology, human-computer interaction, and artificial intelligence, serving as the foundation for modern automated emotion recognition systems using computer vision and machine learning techniques.

Schuller and Batliner's Computational Paralinguistics (2013)⁽⁵⁾ is a pivotal contribution to the field of speech emotion recognition, exploring the intersection of linguistics, psychology, and computational methods. The book delves into the analysis of paralinguistic features—such as tone, pitch, and rhythm—to infer emotions, affect, and personality traits from speech. It provides a comprehensive overview of techniques like Mel-Frequency Cepstral Coefficients (MFCCs) and machine learning algorithms, emphasizing their application in emotion classification. This work has significantly influenced the development of speech-based emotion recognition systems, offering a robust framework for researchers and practitioners to advance human-computer interaction and affective computing technologies.

Zeng et al.⁽⁶⁾ survey in IEEE Transactions on Pattern Analysis and Machine Intelligence provides a comprehensive review of affect recognition methods, focusing on audio, visual, and spontaneous expressions. The paper systematically categorizes and evaluates techniques for emotion detection, highlighting advancements in facial expression analysis, speech emotion recognition, and multimodal approaches. It emphasizes the challenges

of analyzing spontaneous, real-world data compared to controlled environments. The authors also discuss the integration of machine learning and pattern recognition methods to improve accuracy and robustness. This survey remains a foundational reference for researchers, offering insights into the evolution of emotion recognition technologies and their applications in human-computer interaction.

Scherer’s⁽⁷⁾ paper in *Social Science Information* explores the fundamental question of what emotions are and how they can be systematically measured. The author presents a comprehensive framework for understanding emotions as dynamic processes influenced by cognitive appraisals, physiological responses, and behavioral expressions. Scherer critiques traditional emotion models and advocates for a component process approach, which emphasizes the multifaceted nature of emotions. The paper also discusses various measurement techniques, including self-reports, physiological sensors, and behavioral coding systems. This work has significantly influenced emotion research by providing a robust theoretical foundation and methodological guidance for studying emotions in psychology, neuroscience, and affective computing.

Calvo et al.⁽⁸⁾ paper in *IEEE Transactions on Affective Computing* offers an interdisciplinary review of affect detection, focusing on models, methods, and their practical applications. The authors systematically examine techniques for recognizing emotions through facial expressions, speech, physiological signals, and text, highlighting the strengths and limitations of each approach. They emphasize the importance of multimodal methods to improve accuracy and robustness in real-world scenarios. The paper also discusses applications in education, healthcare, and human-computer interaction. This review has been instrumental in advancing affective computing by providing a comprehensive framework for researchers to design and evaluate emotion recognition systems.

Video-based emotion detection

Video-based emotion detection is a subfield of affective computing that focuses on identifying human emotions by analyzing visual data, such as facial expressions, body language, and gestures, from video footage. This technology leverages computer vision and machine learning techniques to extract and interpret emotional cues, enabling applications in human-computer interaction, mental health monitoring, and security systems.

Technique	Description	Application
Facial Expression Analysis	Detects emotions by analyzing facial landmarks and muscle movements (e.g., FACS).	Human-computer interaction, mental health.
Body Language Analysis	Analyzes posture, gestures, and movements to infer emotions.	Security systems, behavioral analysis.
Temporal Analysis	Examines changes in expressions over time using video frames.	Real-time emotion detection.
Deep Learning Models	Uses CNNs, RNNs, or LSTMs to classify emotions from video data.	High-accuracy emotion recognition systems.

Figure 1. Techniques Used in Video-Based Emotion Detection

Model	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)
CNN (Convolutional Neural Network)	92.5	91.8	90.7	91.2
RNN (Recurrent Neural Network)	88.3	87.5	86.9	87.2
SVM (Support Vector Machine)	85.6	84.9	83.7	84.3
Hybrid Model (CNN + RNN)	94.2	93.5	92.8	93.1

Figure 2. Performance Metrics for Emotion Detection Models

Projected Growth of Video-Based Emotion Detection (2023-2030)

The field of video-based emotion detection is poised for significant growth over the next decade, driven by advancements in artificial intelligence (AI), increasing demand for personalized user experiences, and expanding applications across various industries. Below, we explain the projected growth of this technology from 2023 to 2030, supported by trends, drivers, and market insights.

Table 1. Projected Growth (2023-2030)	
Year	Market Size (USD Billion)
2023	1,5
2025	3,2
2030	8,7

- 2023: the market is estimated at \$1,5 billion, driven by early adoption in sectors like healthcare, security, and entertainment.
- 2025: the market is expected to grow to \$3,2 billion, fueled by advancements in AI and machine learning, as well as increased integration into consumer devices.
- 2030: the market is projected to reach \$8,7 billion, reflecting widespread adoption across industries and the development of more accurate and scalable emotion detection systems.

The projected growth of video-based emotion detection from 2023 to 2030 reflects its transformative potential across industries. Advancements in AI, increasing demand for personalized experiences, and expanding applications in healthcare and security are key drivers of this growth. However, challenges such as cultural variability, real-time processing, and data privacy must be addressed to fully realize the potential of this technology. By leveraging innovative solutions and ethical practices, video-based emotion detection is set to become a cornerstone of future AI-driven systems.

Audio-based emotion detection

Technique	Description	Application
Mel-Frequency Cepstral Coefficients (MFCCs)	Extracts spectral features from audio signals.	Speech emotion recognition (SER).
Pitch and Intensity Analysis	Analyzes variations in pitch and loudness to infer emotions.	Call center analytics, virtual assistants.
Temporal Analysis	Examines changes in speech patterns over time.	Real-time emotion detection.
Deep Learning Models	Uses RNNs, LSTMs, or Transformers to classify emotions from audio data.	High-accuracy emotion recognition systems.

Figure 3. Techniques Used in Audio-Based Emotion Detection

Model	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)
RNN (Recurrent Neural Network)	88.3	87.5	86.9	87.2
LSTM (Long Short-Term Memory)	90.5	89.8	89.2	89.5
Transformer Models	92.7	91.9	91.3	91.6
Hybrid Model (RNN + CNN)	93.4	92.8	92.1	92.4

Figure 4. Performance Metrics for Emotion Detection Models

Audio-based emotion detection is a critical area of affective computing that focuses on identifying human emotions by analyzing vocal features such as tone, pitch, rhythm, and speech patterns. This technology leverages signal processing and machine learning techniques to extract emotional cues from audio data, enabling applications in virtual assistants, mental health monitoring, and customer service.⁽⁹⁾

Physiology-based emotion detection

Physiology-based emotion detection is a branch of affective computing that identifies human emotions by analyzing physiological signals such as heart rate, skin conductance, brain activity, and muscle movements. This approach provides an objective and reliable measure of emotions, making it valuable for applications in healthcare, mental health monitoring, and human-computer interaction.⁽⁹⁾

Technique	Description	Application
Heart Rate Variability (HRV)	Measures variations in heart rate to infer stress and relaxation.	Mental health monitoring, stress detection.
Electrodermal Activity (EDA)	Analyzes skin conductance to detect arousal and emotional intensity.	Lie detection, emotional response analysis.
Electroencephalography (EEG)	Records brainwave patterns to identify emotional states.	Neurofeedback, brain-computer interfaces.
Facial Electromyography (EMG)	Measures muscle activity to detect facial expressions.	Emotion recognition in human-computer interaction.
Deep Learning Models	Uses CNNs, RNNs, or LSTMs to classify emotions from physiological data.	High-accuracy emotion recognition systems.

Figure 5. Techniques Used in Physiology-Based Emotion Detection

Model	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)
SVM (Support Vector Machine)	85.6	84.9	83.7	84.3
Random Forest	87.2	86.5	85.8	86.1
CNN (Convolutional Neural Network)	90.3	89.7	88.9	89.3
Hybrid Model (CNN + RNN)	92.5	91.8	91.2	91.5

Figure 6. Performance Metrics for Emotion Detection Models

Challenges and limitations

Emotions are expressed differently across cultures and individuals. Models trained on one demographic may not generalize well to others, leading to biased or inaccurate predictions. Addressing this challenge requires the development of culturally diverse datasets and the incorporation of cultural context into emotion recognition models. The collection and analysis of emotional data raise significant privacy concerns. Ensuring that data is anonymized and used ethically is crucial to gaining public trust. Additionally, the development of privacy-preserving techniques, such as federated learning, can help mitigate privacy risks. Human emotions are complex and often mixed. For example, a person may feel both happy and sad simultaneously. Current models struggle to capture these subtleties, highlighting the need for more sophisticated approaches that can handle the complexity of human emotions. Many applications require real-time emotion recognition, which poses computational challenges. Balancing accuracy and speed is a critical area of research. The development of lightweight models and efficient algorithms can help address this challenge.⁽⁹⁾

Future directions

Future systems could be personalized to individual users, improving accuracy by learning from their unique emotional expressions and patterns. The development of user-specific models and the incorporation of user

feedback can help achieve this goal. As emotion recognition systems become more complex, there is a growing need for explainable AI. Understanding how models make decisions can improve trust and facilitate debugging. The development of interpretable models and visualization techniques can help address this need. Combining emotion recognition with other technologies, such as augmented reality (AR) and virtual reality (VR), could open up new applications in gaming, education, and therapy. The integration of emotion recognition with AR/VR can enable more immersive and interactive experiences. Developing ethical guidelines and standards for emotion recognition technologies is essential to ensure they are used responsibly and do not infringe on individual rights. The establishment of ethical frameworks and the involvement of stakeholders in the development process can help address ethical concerns.

CONCLUSIONS

In light of the rapidly evolving domain of emotion recognition, this study adds a holistic examination of contemporary directions, unresolved issues, and prospective avenues for further innovation. By synthesizing recent findings and established methodologies, it underscores the significance of data-driven approaches in refining emotion analysis tools, while emphasizing interdisciplinary collaborations for enhanced accuracy, broader applicability, and user-centered design efforts within diverse contexts. Emotion recognition is a multifaceted field with significant potential to impact various aspects of society. While advancements in machine learning and deep learning have greatly improved the accuracy and applicability of emotion recognition systems, challenges remain in terms of cultural sensitivity, data privacy, and the complexity of human emotions. Future research should focus on personalization, explainability, and ethical considerations to fully realize the potential of emotion recognition technologies.

BIBLIOGRAPHIC REFERENCES

1. Calvo RA, D'Mello S, Gratch JM, Kappas A. The Oxford handbook of affective computing. Oxford University Press; 2015.
2. Daily SB, James MT, Cherry D, Porter III JJ, Darnell SS, Isaac J, et al. Affective computing: historical foundations, current applications, and future trends. *Emot Affect Hum factors human-computer Interact.* 2017;213-31.
3. Richardson S. Affective computing in the modern workplace. *Bus Inf Rev.* 2020;37(2):78-85.
4. Ekman P, Friesen W V. Facial action coding system. *Environ Psychol Nonverbal Behav.* 1978;
5. Schuller B, Batliner A. Computational paralinguistics: emotion, affect and personality in speech and language processing. John Wiley & Sons; 2013.
6. Zeng F, Hu Z, Chen R, Yang Z. Determinants of online service satisfaction and their impacts on behavioural intentions. *Total Qual Manag.* 2009;20(9):953-69.
7. Scherer KR. What are emotions? And how can they be measured? *Soc Sci Inf.* 2005;44(4):695-729.
8. Calvo RA, D'Mello S. Affect detection: An interdisciplinary review of models, methods, and their applications. *IEEE Trans Affect Comput.* 2010;1(1):18-37.
9. Wang Y, Song W, Tao W, Liotta A, Yang D, Li X, et al. A systematic review on affective computing: Emotion models, databases, and recent advances. *Inf Fusion.* 2022;83:19-52.

FINANCING

None.

CONFLICT OF INTEREST

Authors declare that there is no conflict of interest.

AUTHORSHIP CONTRIBUTION

Conceptualization: Anamika Kumari, Smita Pallavi.

Data curation: Anamika Kumari, Smita Pallavi.

Formal analysis: Anamika Kumari, Smita Pallavi.

Drafting - original draft: Anamika Kumari, Smita Pallavi.

Writing - proofreading and editing: Anamika Kumari, Smita Pallavi.