iJRASET
International Journal For Research in Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 14 **Issue:** II **Month of publication:** February 2026

DOI: https://doi.org/10.22214/ijraset.2026.77631

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 14 Issue II Feb 2026- Available at www.ijraset.com

A Survey on Real-time Multimodal Emotion Detection using Deep Learning

Alan Prakash[1], Rishikesh S[2], Shriram A[3], Dr. R. Punithavathi[4]
[1, 2, 3]*UG Scholar, Department of Computer Science and Engineering Chettinad College of Engineering and Technology, Puliyur, Karur, Tamil Nadu, India*
[4]*Professor and Head, Department of Computer Science and Engineering Chettinad College of Engineering and Technology, Puliyur, Karur, Tamil Nadu, India*

Abstract: ***While human-to-human communication relies naturally on the interpretation of facial expressions and vocal nuances, detecting emotional states in Human-Computer Interaction (HCI) presents significant technical challenges. This survey paper explores the development of an integrated Emotion Recognition System (ERS) that leverages both Speech Emotion Recognition (SER) and Facial Emotion Recognition (FER) to bridge this gap. We analyze the effectiveness of deep learning architectures, specifically a hybrid CNN+BiLSTM model, in processing multimodal inputs for real-time applications. The study reviews the role of data augmentation techniques—such as noise addition and spectrogram shifting—in improving model robustness across benchmark datasets including TESS, EmoDB, and RAVDESS. By synthesizing facial and vocal features, the proposed framework aims to enhance the naturalness of human-machine communication and provide a foundation for intelligent systems in mental health, customer service, and robotics.***

Keywords: Multimodal Emotion Detection, Speech Emotion Recognition(SER), Facial Emotion Recognition (FER), Deep Learning, CNN-BiLSTM, Human-Computer Interaction (HCI), Affective Computing.

I. INTRODUCTION

In the evolving landscape of digital innovation, the interface between humans and machines is increasingly focused on achieving natural and effective communication. Human-Computer Interaction (HCI) now prioritizes user-centric systems that can adapt to the user's psychological state to improve learning, support development, and enhance overall user experience. Because emotions are a fundamental pillar of human expression, the ability of a machine to recognize emotional cues is essential for the next generation of robotics and intelligent assistants.

Emotional states are traditionally expressed through a combination of facial expressions, body language, and speech patterns. This research specifically focuses on a multimodal approach, integrating SER—which extracts features from vocal intonations—and FER—which interprets visual facial cues. By employing advanced feature extraction methods like Mel-Frequency Cepstral Coefficients (MFCC), Chroma, and Mel Spectrograms, and utilizing hybrid deep learning models such as CNN combined with Bidirectional Long Short-Term Memory (BiLSTM), this study addresses the limitations of unimodal systems that often fail in complex, real-world environments

A. Core Themes of This Survey

This paper is organized around four primary thematic areas:

1) Deep Learning & Hybrid Architectures: A review of neural network models, including MLP, CNN, and BiLSTM, used for high-accuracy emotion classification
2) Multimodal Feature Extraction: Analysis of speech signals (MFCC, RMS, ZCR) and visual markers (Haar Cascades, HOG) for comprehensive emotion capture
3) Data Augmentation & Robustness: Study of techniques like noise addition and image rotation to prevent model overfitting and improve performance in diverse environments.
4) Societal Impact & SDG Mapping: Evaluation of how real-time emotion detection contributes to mental health monitoring (SDG 3) and industrial innovation (SDG 9).

By consolidating existing research and proposing an integrated risk-aware routing framework, this survey bridges the gap between predictive models and practical delivery operations, providing a foundation for future advancements in multimodal emotion detection using deep learning.

©IJRASET: All Rights are Reserved

II. REVIEW OF EXISTING RESEARCH PAPERS

Recent advancements in affective computing have shifted the focus toward hybrid models that can process the complexities of human emotion in real time. Research in this field primarily examines the integration of facial and vocal data to overcome the limitations of single-modality systems (Song, Y., & Zhang, L.). Studies have shown that Facial Emotion Recognition (FER) frequently utilizes Transfer Learning and data augmentation to handle variations in lighting and orientation. In parallel, Speech Emotion Recognition (SER) research has evolved from basic statistical models to deep learning architectures like CNN-LSTM, which effectively capture temporal dependencies in audio signals. Key findings indicate that models without robust augmentation often suffer from overfitting when applied to diverse real-world datasets (Zhang, L., et al).

A. Comparison of Past Methodologies

Current methodologies for emotion detection can be categorized into three primary technical approaches:

1) Rule-Based and Statistical Models: Early systems relied on manual feature engineering and predefined thresholds, which struggle with the non-linear nature of emotional expression.
2) Single-Modality Deep Learning: These approaches focus exclusively on either FER or SER. While effective in controlled environments, they are highly sensitive to background noise in audio or lighting variations in video
3) Hybrid Multimodal Frameworks: Modern systems, such as the CNN+BiLSTM model, integrate features from multiple sources to provide a more holistic emotional profile. These frameworks utilize advanced extraction techniques like MFCC for speech and HOG for facial analysis to achieve higher classification accuracy.

III. STRENGTHS AND WEAKNESSES OF EXISTING APPROACHES

A. Strengths

1) Hybrid Deep Learning Models (CNN + BiLSTM): These models are highly effective in capturing both spatial features from facial expressions and temporal dependencies from speech signals. The integration of CNN for visual processing and BiLSTM for sequential audio data provides a more robust analysis than single-architecture systems.
2) Data Augmentation Techniques: The application of noise addition, spectrogram shifting, and image rotation significantly improves model robustness against overfitting. These techniques allow the system to maintain accuracy even when dealing with diverse environmental conditions or varied datasets (Li et al.).
3) Multimodal Integration: Combining facial and vocal cues enables the system to provide a comprehensive emotional profile, addressing the limitations of unimodal systems that may fail in complex interactions.
4) Comprehensive Feature Extraction: Utilizing advanced features such as MFCC, Chroma, and Mel Spectrograms allows the system to identify subtle vocal intonations and pitch characteristics essential for precise emotion classification.

B. Deep Learning and Multimodal Approaches

Traditional and modern deep learning algorithms have been extensively applied in emotion recognition due to their superior performance in handling complex, unstructured data such as audio and video. While Multi-Layer Perceptron (MLP) models serve as a strong baseline for structured features, Convolutional Neural Networks (CNN) are preferred for automated feature extraction from facial regions and audio spectrograms. Hybrid models, particularly the CNN+BiLSTM architecture, further enhance accuracy by modeling the long-term temporal relationships found in human speech.

Despite these advancements, existing deep learning approaches face certain limitations. Many models rely heavily on large, high-quality benchmark datasets like TESS or RAVDESS and may struggle to generalize in real-world environments with significant background noise or varied lighting. Furthermore, real-time inference requires optimized hardware to maintain high frame rates and low latency during live data processing. Nevertheless, deep learning remains the foundation for intelligent Human-Computer Interaction (HCI) and continues to drive innovation in affective computing.

IV. MACHINE LEARNING DEEP -BASED MODELS

Deep learning-based models are utilized to analyze multimodal audio and visual data to recognize human emotional states in real-time. Acoustic features such as Mel-Frequency Cepstral Coefficients (MFCC), Chroma, and Mel Spectrograms, along with facial landmarks, serve as the primary input vectors for the training process. Supervised learning techniques are preferred in this framework due to their high performance in complex classification tasks involving unstructured media. Architectures such as the Multi-Layer Perceptron (MLP), Convolutional Neural Networks (CNN), and hybrid CNN-BiLSTM models are implemented to

categorize emotional states into distinct classes. The MLP provides a fundamental and computationally efficient baseline for classification using extracted audio features. Convolutional Neural Networks are employed to capture spatial patterns within spectrograms and facial images, while the integration of BiLSTM layers enables the system to learn the temporal dynamics of speech and expressions. These deep learning approaches are widely applicable in Human-Computer Interaction (HCI), mental health monitoring, and interactive robotics (Ouyang, Q., et al).

V. HYBRID MODELS COMBINING MULTIPLE TECHNIQUES

Hybrid models in intelligent emotion recognition systems combine spatial feature extraction with temporal sequence modeling to improve detection accuracy and reliability. In this approach, deep learning architectures such as Convolutional Neural Networks (CNN) and Bidirectional Long Short-Term Memory (BiLSTM) are integrated to process multimodal data, including speech signals and facial expressions. The spatial human emotional states over time. Hybrid models offer improved precision, real-time adaptability, and comprehensive analysis in Human-Computer Interaction. However, they introduce increased computational demands and require significant preprocessing of multimodal datasets for effective performance. Despite these challenges, such hybrid approaches have proven effective in mental health monitoring, customer service analytics, and the development of responsive robotic systems.

VI. MULTIMODAL FEATURE ANALYSIS AND EMOTION SYSTEM MODELING

Multimodal feature analysis and system modeling play a fundamental role in understanding and improving the accuracy of real-time emotion recognition systems. Multimodal analysis evaluates emotional states based on integrated factors such as vocal intonation, speech features, and facial expressions. System modeling focuses on how these diverse data streams are collected, synchronized, and processed—from raw live inputs to final emotional classification—considering computational constraints and real-time response priorities (Pillalamarri & Shanmugam).

These modeling techniques enable the identification of subtle emotional cues and high-intensity expressions, supporting more natural Human-Computer Interaction. However, their effectiveness depends on the availability and accuracy of multimodal data, such as benchmark audio-visual datasets (TESS, RAVDESS, EmoDB) and precise facial landmark localization.

Computational complexity may increase when processing simultaneous high-definition video and audio streams with varying temporal constraints. Despite these challenges, multimodal analysis and system modeling have been successfully applied in mental health monitoring, customer service optimization, and real-time social robotics.

Emotion recognition approaches have evolved from basic single-modality rule-based systems to intelligent hybrid deep learning frameworks. While traditional models like MLP or simple CNNs offer efficiency for basic classification, hybrid models that integrate spatial CNN features with temporal BiLSTM layers provide significantly improved reliability in real-world scenarios. Future research should focus on enhancing real-time adaptability across diverse environments, improving model efficiency for low-end hardware, and integrating Explainable AI (XAI) to provide deeper insights into the decision-making process of emotion detection systems.

VII. ADVANTAGES AND LIMITATIONS OF VARIOUS APPROACHES

A. Multi-Layer Perceptron (MLP)

- Advantages: Computationally efficient for classifying pre-extracted audio features and serves as a highly interpretable baseline for emotion detection.
- Limitations: Struggles to capture complex spatial patterns in images or temporal dependencies in long audio sequences

B. Convolutional Neural Network (CNN)

- Advantages: Highly effective at automated spatial feature extraction from facial images and audio spectrograms.
- Limitations: Primarily focused on static spatial data and may overlook the evolving temporal context of emotional expressions

C. Bidirectional Long Short-Term Memory (BiLSTM)

- Advantages: Capable of capturing long-term temporal dependencies by processing data sequences in both forward and backward directions
- Limitations: Requires significant computational memory and is prone to longer training times due to its complex recurrent structure

D. Data Augmentation (Noise Addition, Shifting, Rotation)

- Advantages: Enhances model generalization and reduces overfitting by creating synthetic variations of training data, improving performance in noisy environments
- Limitations: Requires careful tuning to ensure the augmented data remains representative of real-world emotional cues without introducing excessive distortion.

E. Hybrid Models (Multimodal Deep Learning + Feature Fusion)

- Advantages: Combines visual, audio, and textual emotional cues, improves robustness and accuracy, captures cross-modal emotional dependencies, and enhances real-time emotion recognition performance
- Limitations: Higher system increased computational overhead, synchronization issues between modalities, and challenging model deployment.

TABLE I: Comparison of Different Smart Parcel Delivery Techniques

Technique	Strengths	Weaknesses	Performance Metrics	Applications
Traditional MLP Classifier	Fast and efficient for processing pre-extracted MFCC features.	Cannot process raw spatial images or long temporal sequences effectively	Accuracy: Moderate Latency: Low	Basic audio emotion classification and baseline benchmarking
Convolutional Neural Network (CNN)	Excellent at extracting spatial features from facial images and audio spectrograms	Limited ability to understand the temporal flow of speech and facial changes over time	SpatialPrecision:High, F1-Score: Strong.	Facial Emotion Recognition (FER) and spectrogram-based audio analysis.
BiLSTM / LSTM Networks	Captures long-term temporal dependencies in speech patterns and vocal intonations	Computationally intensive and requires large datasets to avoid vanishing gradients.	Temporal Accuracy: High Recall: Strong.	Sequence-based speech emotion recognition and time-series analysis
Transfer Learning	Utilizes pre-trained weights to achieve high accuracy with smaller custom datasets	Can be sensitive to domain shifts and background noise if not fine-tuned properly.	Transfer Efficiency: High Precision: High.	Rapid deployment of facial recognition in variable lighting conditions

VIII. DISCUSSION ON REAL-WORLD APPLICABILITY

A. Real-Time Multimodal Emotion Recognition

Machine learning–based delivery risk prediction models combined with route optimization techniques have strong applicability in real-world last-mile logistics operations. Predictive models such as Logistic Regression, Decision Trees, and Random Forest can be used to identify high-risk deliveries based on factors like customer availability, traffic conditions, and delivery time windows. These predictions help logistics managers proactively plan delivery schedules and reduce failed delivery attempts (Pillalamarri & Shanmugam).

B. Mental Health Monitoring and Well-Being

Real-time emotion detection plays a vital role in healthcare by enabling the continuous monitoring of mental health states, including stress and anxiety. These systems can be integrated into psychological counseling platforms to provide early intervention during emotional distress or crises. By promoting early support through real-time monitoring, the technology contributes directly to global health and well-being goals.

C. Industrial Innovation and Workspace Safety

In industrial settings, emotion recognition contributes to digital innovation by enhancing Human-Computer Interaction (HCI) and intelligent communication. In workplaces, the system can improve productivity and human performance by monitoring employee well-being and streamlining human-machine communication. Furthermore, it supports safety in critical environments, such as aviation and customer service, by optimizing human performance based on emotional analytics.

©IJRASET: All Rights are Reserved | SJ Impact Factor 7.538 | ISRA Journal Impact Factor 7.894 |

D. Enhancing public safety and Smart Communities

The deployment of emotion detection technology in shared spaces can promote safer environments and improve public safety through smart monitoring. By identifying emotional cues in real-time video streams, these systems can encourage secure mobility and support the development of sustainable, secure communities (Li et al).

IX. CURRENT LIMITATIONS IN THE FIELD

A. Real-Time Data Integration

Deep learning-based emotion recognition models have demonstrated significant effectiveness in identifying emotional states using benchmark datasets. However, integrating real-time data such as live audio streams and dynamic video inputs remains a substantial challenge. Delays in processing or incomplete data frames during active monitoring can reduce the precision of emotion detection during real-world interactions.

B. User-Friendly Applications

Many multimodal emotion detection models remain confined to experimental or analytical environments and lack intuitive interfaces for practical daily use. The absence of accessible platforms makes it difficult for practitioners in fields like healthcare or customer service to interpret emotional analytics effectively. Without clear visualization and seamless user interaction, the widespread adoption of intelligent emotion recognition systems is limited.

C. Data Scarcity and Quality

High-quality and well-structured multimodal data is essential for accurate emotion prediction. Factors such as background noise in audio recordings, varied accents, and poor lighting conditions in video streams can significantly degrade system performance. Furthermore, many existing models rely heavily on curated benchmark datasets, which may not always reflect the diversity of real-world human expressions.

This dependency often leads to a performance gap when models move from laboratory settings to uncontrolled, everyday environments. The scarcity of naturalistic data where emotions are subtle and overlapping rather than exaggerated remains a major hurdle for training truly resilient systems (Li et al).

D. Model Generalization

Emotion detection models trained on specific datasets often struggle to generalize across different cultural groups or varied social environments.

For instance, a model optimized for a specific language or demographic may perform poorly when applied to diverse linguistic patterns or different facial features. This highlights the ongoing need for transferable emotion recognition frameworks that are applicable across diverse global contexts.

E. Computational Complexity

Hybrid models, such as the CNN+BiLSTM architecture, require significant computational resources, especially as the demand for high-resolution real-time monitoring increases. This can make real-time emotion prediction and live video inference challenging on low-end hardware or resource-constrained mobile devices.

X. PROBABLE RESEARCH OPPORTUNITIES

A. Real-Time Emotion Monitoring Applications

Developing real-time, web-based applications for emotion prediction and visualization can bridge the gap between analytical models and real-world deployment. Features such as emotion risk dashboards and live sentiment tracking can significantly improve operational usability in clinical and professional settings.

B. Adaptive Multimodal Optimization

Further research can explore dynamic system optimization that continuously updates detection thresholds based on environmental factors like background noise or lighting. Integrating multimodal emotion prediction with adaptive, lightweight neural network variants can enhance real-time efficiency on mobile devices

C. Explainable Emotion Prediction (XAI)

Improving model interpretability through explainable machine learning (XAI) techniques can increase trust and adoption among healthcare providers and educators. Visualizing key emotional factors, such as specific facial landmarks or vocal pitch characteristics, improves decision-making and transparency.

D. Scalable and Transferable Emotion Frameworks

Developing scalable emotion recognition frameworks that perform efficiently across varying cultural demographics and languages remains a critical research direction. Such frameworks can support deployment in both urban and rural environments where data might be sparse.

E. Intelligent Decision-Support Systems for Healthcare

Integrating delivery risk insights into decision-support systems can assist logistics planners in optimizing scheduling, workforce allocation, and service-level strategies based on predicted delivery outcomes

XI. EMERGING TECHNOLOGIES THAT COULD IMPROVE EXISTING METHODS

A. Real-Time Web-Based Processing

Deploying emotion detection models within web-based platforms enables near real-time decision-making for remote healthcare and interactive systems. Lightweight models integrated into these applications can process incoming audio and video data quickly, reducing response time for emotional feedback.

B. Advanced Hybrid Architectures

Enhancements in hybrid deep learning variants, such as combining Attention mechanisms with CNN-BiLSTM, can further optimize the detection of subtle emotional shifts. These techniques can handle multiple objectives, including minimizing latency and maximizing classification accuracy across diverse datasets. .(Ouyang, Q., et al).

C. Secure Data Handling and Transparency

Improving data handling mechanisms ensures reliable and consistent emotion prediction across sensitive applications. Structured data storage and secure data flow between the detection models and healthcare applications help maintain data integrity and user privacy.

D. Intelligent Decision-Support Systems

Integrating deep learning-based emotion prediction with decision-support modules can enhance mental health monitoring and crisis intervention. Such systems provide actionable insights through dashboards, enabling professionals to make informed support and counseling decisions efficiently

XII. REAL-TIME APPLICATIONS AND PRACTICAL IMPLEMENTATION ISSUES

A. Real-Time Emotion Monitoring and Visualization

A web-based application can be utilized to monitor emotional status, predicted sentiment levels, and real-time detection results. This interface provides a visual representation of emotional shifts as they occur during human-computer interaction.(Kumar, A., & Raj, S).

B. Emotional Alerts and Decision Support

Early alert mechanisms can assist healthcare providers or practitioners in responding to signs of emotional distress or crisis. These systems provide actionable insights through dashboards, enabling professionals to make informed support and counseling decisions.

C. Scalability and Accessibility

Scaling the emotion recognition system to handle simultaneous multimodal streams from multiple users while maintaining high prediction accuracy is a significant challenge.(Ouyang, Q., et al)

ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 14 Issue II Feb 2026- Available at www.ijraset.com

XIII. SUMMARY OF KEY FINDINGS

This survey paper examined the integration of Deep Learning (DL) and Artificial Intelligence (AI) techniques for real-time multimodal emotion detection using speech and facial features. The key findings are summarized as follows:

1) Deep Learning Models, Hybrid architectures, particularly the CNN + BiLSTM model, were found to be effective in recognizing emotions by analyzing facial expressions and vocal intonations. Their ability to process complex, multi-dimensional audio-visual data makes them highly suitable for real-time emotional state classification

2) Multimodal Integration, The system demonstrates that combining Speech Emotion Recognition (SER) and Facial Emotion Recognition (FER) provides a more comprehensive understanding of human feelings compared to single-modality systems.(Kumar, A., & Raj, S).

3) Data Augmentation & Feature Extraction, Techniques such as noise addition, spectrogram shifting, and image rotation significantly improve model robustness and prevent overfitting. Extracting advanced features like MFCC, Chroma, and Mel Spectrograms is essential for capturing precise vocal characteristics.

4) Integrated Recognition Frameworks, The proposed system bridges the gap between theoretical AI and practical human-computer interaction by combining real-time feature extraction with high-accuracy hybrid models.

XIV. FINAL THOUGHTS ON ADVANCEMENTS IN THE FIELD

The field of intelligent Human-Computer Interaction (HCI) and real-time emotion recognition has witnessed notable advancements in recent years, driven by the adoption of Artificial Intelligence (AI), Deep Learning (DL), and multimodal data processing techniques.

These advancements have enabled improved emotion detection accuracy, natural communication interfaces, and data-driven psychological insights in complex interaction environments. AI-based models have demonstrated strong potential in addressing key challenges such as interpreting human feelings through speech and facial expressions, automating mental health monitoring, and improving workplace well-being. However, issues related to real-time data integration, environmental noise, and computational scalability continue to pose challenges for widespread deployment in live systems. The development of integrated multimodal emotion detection frameworks represents a significant step forward in overcoming these limitations by combining facial feature extraction with sequential speech modeling.

XV. FUTURE PERSPECTIVES AND POSSIBLE RESEARCH DIRECTIONS

A. Enhancing Real-Time Capabilities

Future research should focus on improving the real-time adaptability of emotion recognition systems by integrating live audio-visual streams and high-speed data processing. This would enable dynamic adjustments to emotional classification and more accurate mood assessment during active human-computer interactions.

B. Explainable AI (XAI) for Affective Decision-Making

The development of explainable AI (XAI) techniques for emotion detection models can enhance transparency and trust among clinicians and users. Visual explanations, such as feature importance in speech or specific facial landmark activations, can support informed and confident decision-making in therapeutic and professional settings.

C. Integration of Environmental and Temporal Factors

Incorporating environmental conditions such as background noise levels, lighting variations, and long-term behavioral patterns can improve the accuracy of emotion prediction. Future studies should explore how these external factors influence emotional expression and integrate them into real-time monitoring systems.

D. Scalable and Transferable Emotion Models

Designing scalable and transferable emotion recognition models that can be applied across different demographics and languages is a key research direction.

Pre-trained models can be fine-tuned for new cultural contexts or specific user groups with minimal data, improving adaptability and deployment efficiency

©IJRASET: All Rights are Reserved

E. User-Friendly Decision Support Applications

Future work should emphasize the development of intuitive and user-friendly decision support tools for mental health professionals and educators. Interactive dashboards and visualization interfaces can simplify the interpretation of emotional data and enhance operational efficiency in support systems

F. Integration of Behavioral and Socioeconomic Factors

Incorporating user behavior patterns, personality traits, and socioeconomic indicators can help identify individuals at higher risk of emotional distress.

G. Scalable Logistics Analytics

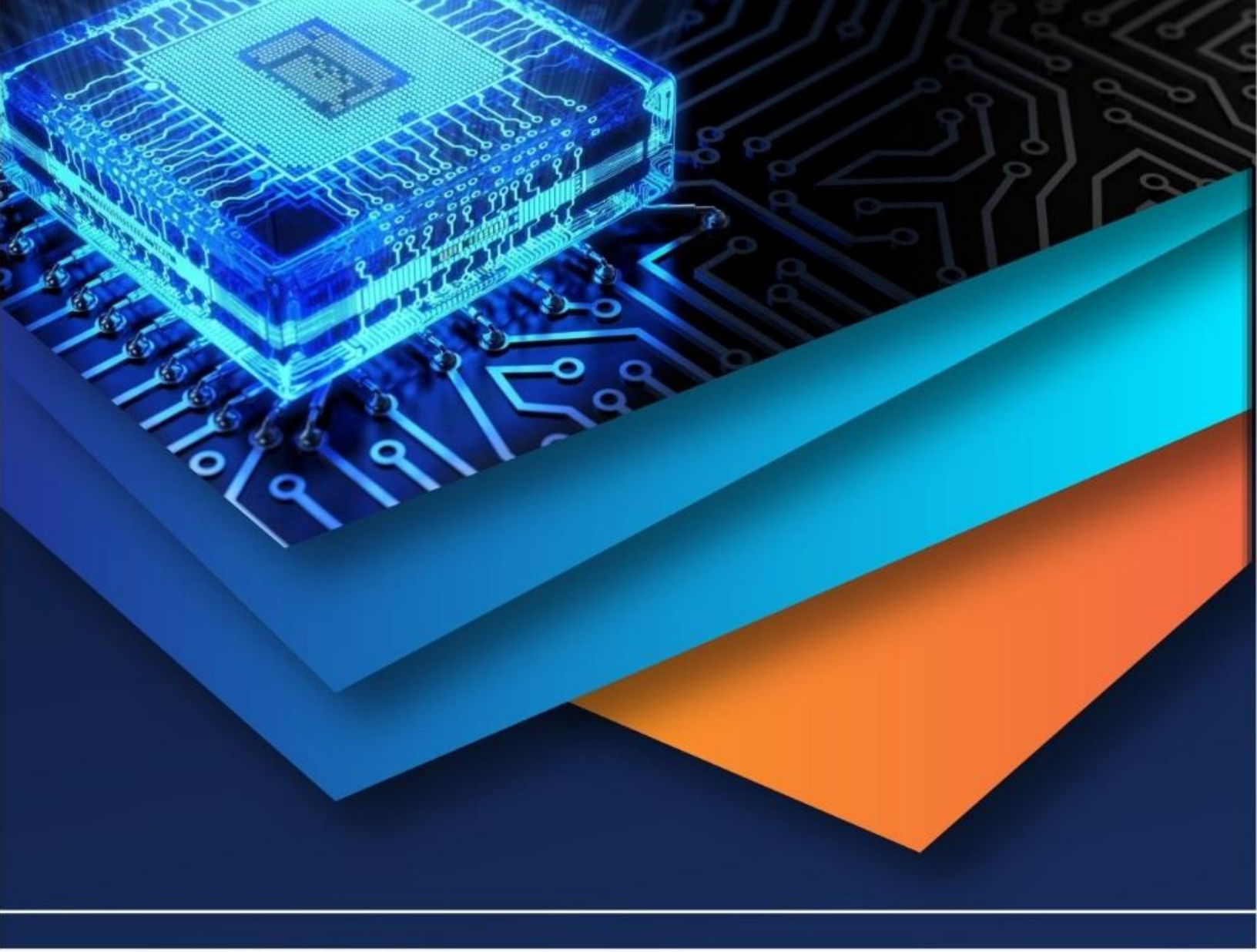
Future research should focus on building scalable frameworks that handle large datasets while maintaining real-time accuracy and responsiveness

XVI. CONCLUSION

The integration of deep learning-based emotion recognition with real-time multimodal analysis has the potential to significantly enhance human-computer interaction and mental health support. By addressing challenges such as multimodal synchronization and data augmentation, our project supports data-driven decision-making and helps provide early interventions for emotional distress. Our project provides a practical approach to smarter and more responsive communication by enabling effective emotional monitoring and execution. As digital systems continue to evolve, our project contributes toward building efficient and sustainable solutions for modern interactive networks.

REFERENCES

[1] Kumar, A., & Raj, S. (2024). Facial Emotion Detection with Data Augmentation Techniques. Journal of Computer Vision and Pattern Recognition.
[2] Li, Z., et al. (2025). "Enhancing Emotion Recognition Accuracy Through Data Augmentation and Deep Neural Networks." International Journal of Intelligent Systems and Applications.
[3] Li, Z., et al. (2025). "Enhancing Emotion Recognition Accuracy Through Data Augmentation and Deep Neural Networks." International Journal of Intelligent Systems and Applications.
[4] Ouyang, Q., et al. (2025). "Speech Emotion Detection Based on MFCC Features and CNN-LSTM Hybrid Model." arXiv preprint arXiv:2501.10666.
[5] Pillalamarri, R., & Shanmugam, U. A. (2025). "A Review on Multimodal Learning for Emotion Recognition Using EEG and Visual Signals." Artificial Intelligence Review (Springer).
[6] Song, Y., & Zhang, L. (2025). Facial and Speech-Based Emotion Recognition Using Deep Learning Electronics (MDPI).
[7] Zhang, L., et al. (2025). "Facial Emotion Recognition Using CNN and Transfer Learning for RealTime Human-Computer Interaction." IEEE Access.



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Call : 08813907089 (24*7 Support on Whatsapp)