IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 8, March 2026
Impact Factor: 8.2

PrepWise: AI Interview Assistant

Dr. Renuka Deshpande[1], Harshal Gawai[2], Advait Jadhav[3], Bhushan Sutar[4]
Department of Artificial Intelligence and Machine Learning[1-4]
Shivajirao S. Jondhale College of Engineering, Dombivli (E), Maharashtra, India

Abstract: *Evaluating candidate performance during interviews is a complex task that traditionally depends on human judgment, which can be inconsistent, time-consuming, and subjective. With the rapid advancement of Artificial Intelligence (AI), modern systems can now analyze verbal and non-verbal cues to provide objective and data-driven assessments. This paper presents a comprehensive review of recent deep learning and multimodal AI techniques used for automated interview evaluation. It focuses on the integration of Computer Vision, Natural Language Processing (NLP), and Speech Emotion Recognition for assessing facial expressions, tone of voice, and textual content of responses. Various approaches such as Convolutional Neural Networks (CNNs), Transformer-based architectures, and multimodal fusion models are discussed. The review highlights key research gaps including dataset scarcity, real-time processing challenges, bias and fairness issues, and the limited interpretability of model decisions. The paper also outlines future directions toward developing efficient, explainable, and scalable systems like* ***PrepWise****, an AI-powered interview evaluation assistant that supports consistent and unbiased candidate assessment*

Keywords: Interview Evaluation, Deep Learning, Natural Language Processing, Emotion Recognition, Multimodal Analysis, Explainable AI, Transformer Models

I. INTRODUCTION

The interview process is a critical component in recruitment, education, and professional evaluation, as it helps determine a candidate's communication ability, confidence, and personality traits [1]. Traditional interview assessments rely heavily on human judgment, which often leads to subjectivity, inconsistency, and bias [2], [3]. Different evaluators may interpret the same candidate's behavior differently, making the process unreliable and time-consuming. As organizations increasingly emphasize fairness and efficiency, the demand for automated and data-driven evaluation systems has grown substantially [4].

Recent advances in Artificial Intelligence (AI) and Machine Learning (ML) have enabled systems capable of analyzing human behavior objectively by leveraging Computer Vision (CV), Natural Language Processing (NLP), and Speech Emotion Recognition (SER) techniques [5], [6]. These technologies allow for the automatic extraction of facial expressions, voice tone, and textual content to assess communication quality and emotional state. For instance, CNN-based visual models and transformer-based language architectures such as BERT and RoBERTa have demonstrated remarkable success in detecting emotions, sentiments, and personality cues [7]–[9]. Similarly, hybrid and multimodal architectures that combine facial, textual, and audio features have achieved improved accuracy in understanding complex human behavior [10]–[12].

A growing body of research focuses on integrating these multimodal learning techniques into intelligent interview-evaluation frameworks [13]–[16]. Such systems aim to provide consistent feedback, minimize bias, and offer explainable decision-making to ensure transparency [17], [18]. However, several challenges remain, including limited availability of publicly annotated datasets [19], [20], computational complexity for real-time deployment [21], [22], and ethical concerns surrounding privacy and fairness [23], [24].

This review paper presents a comprehensive study of existing deep-learning and multimodal AI techniques applied to automated interview assessment. It categorizes research contributions across visual, textual, and fusion-based approaches and highlights existing gaps such as dataset scarcity, model interpretability, and bias mitigation. The

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 8, March 2026
Impact Factor: 8.2

insights derived from this analysis also support the design of ***PrepWise*** — an AI-powered multimodal interview assistant developed to provide unbiased, explainable, and scalable candidate evaluations.

II. REVIEW METHODOLOGY

This section describes the systematic procedure followed to analyze and categorize existing research works related to AI-based interview performance assessment. The methodology adopted for this review ensures comprehensive coverage, structured evaluation, and objective comparison of previously published systems.

A. Literature Identification and Selection

Relevant research papers were collected from reputed digital libraries including IEEE Xplore, ScienceDirect, SpringerLink, ACM Digital Library, and Google Scholar. Keywords such as automated interview evaluation, multimodal emotion recognition, speech emotion analysis, facial expression recognition, transformer-based sentiment analysis, and AI in recruitment were used to retrieve articles published between 2020 and 2024.

Only peer-reviewed journal articles and conference papers focusing on deep learning, multimodal fusion, explainable AI, and fairness-aware recruitment systems were considered. Studies unrelated to interview or behavioral assessment were excluded to maintain domain relevance.

B. Categorization of Existing Systems

The selected studies were organized based on the primary methodology employed:

Visual-Based Approaches – Systems that rely on facial expression recognition using CNNs, Vision Transformers, or hybrid deep architectures.

Audio-Based Approaches – Methods focusing on speech emotion recognition using spectral feature extraction, CNN–RNN hybrids, or transformer-based acoustic models.

Text-Based Approaches – Research utilizing NLP techniques such as BERT, RoBERTa, sentiment analysis, and personality prediction models.

Multimodal Approaches – Integrated frameworks combining visual, audio, and textual modalities through early fusion, late fusion, or attention-based fusion strategies.

This categorization enabled structured comparison of performance metrics, datasets, and architectural choices.

C. Comparative Evaluation Criteria

To analyze the effectiveness of earlier systems, common evaluation parameters were identified:

Model Architecture: CNN, RNN, BiLSTM, Transformer, Vision Transformer, or hybrid frameworks.

Dataset Used: Public datasets such as FER-2013, IEMOCAP, RAVDESS, CMU-MOSI, CMU-MOSEI, AffectNet, or custom interview datasets.

Performance Metrics: Accuracy, F1-score, precision, recall, and robustness under noisy or real-world conditions.

Computational Complexity: Feasibility of real-time deployment and hardware requirements.

Explainability: Use of XAI methods such as Grad-CAM, SHAP, or LIME.

Fairness Considerations: Bias detection or mitigation mechanisms.

These parameters were used to construct comparative tables and identify strengths and limitations of each technique.

D. Analysis of Fusion Strategies

For multimodal systems, particular attention was given to the fusion methodology adopted. Existing works typically implement:

Early Fusion: Combining raw features from different modalities before classification.

Late Fusion: Merging predictions from independent modality-specific models.

Hybrid or Attention-Based Fusion: Dynamically weighting modality importance using attention mechanisms.

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 8, March 2026
Impact Factor: 8.2

The review evaluates how these fusion strategies impact performance, scalability, and interpretability in interview evaluation scenarios.

E. Gap Identification Process

After analyzing architectural designs, datasets, and performance outcomes, recurring limitations across studies were identified. These include dataset imbalance, lack of real-world validation, high computational requirements, limited transparency, and ethical concerns. By systematically examining these aspects, research gaps were derived to highlight areas requiring further investigation.

III. ANALYSIS OF EXISTING TECHNIQUES

The field of automated interview evaluation has evolved rapidly with advancements in deep learning, natural language processing, and computer vision. Early research primarily focused on using single-modality approaches such as facial expression analysis or speech-based assessment to evaluate candidate behavior. However, recent studies demonstrate that combining multiple modalities significantly improves performance and reliability in emotion understanding and behavioral prediction [4], [5].

A. Visual Analysis Techniques

Initial works on visual-based candidate evaluation utilized Convolutional Neural Networks (CNNs) for recognizing facial expressions and micro-expressions that indicate emotions such as confidence, anxiety, or attentiveness [6], [7]. CNN architectures like VGGNet, ResNet, and MobileNet were widely employed for feature extraction from facial regions, achieving accuracy levels between 85% and 95% on benchmark datasets such as FER-2013 and AffectNet. Recent studies also explored Vision Transformers (ViT), which capture spatial dependencies more effectively and outperform traditional CNNs in recognizing subtle expressions [8]. Explainable AI methods like Grad-CAM and LIME have been incorporated to highlight the specific regions of the face influencing classification decisions, improving interpretability and trust in the evaluation process [10], [12].

B. Speech and Audio-Based Techniques

Speech emotion recognition (SER) has become another crucial component of interview performance assessment. Researchers have used hybrid deep architectures such as CNN–BiLSTM and Transformer-based acoustic models to analyze tone, pitch, and energy variations in speech [13]. These models effectively capture temporal and spectral features, providing insights into emotional cues like nervousness, confidence, or enthusiasm. Studies have reported accuracies ranging from 82% to 90% on datasets such as RAVDESS and IEMOCAP. To enable deployment on real-time platforms, lightweight CNN and MobileNet variants have been developed to process audio data efficiently on edge devices without significant loss of accuracy.

C. Text and NLP-Based Techniques

Textual analysis plays a key role in understanding candidates' linguistic style, sentiment, and domain knowledge. Transformer-based language models such as BERT, RoBERTa, and DistilBERT have achieved state-of-the-art results in sentiment and personality prediction tasks [15]. These models extract semantic embeddings from candidate responses, allowing systems to evaluate clarity, tone, and confidence in verbal communication. Studies have also explored the use of contextual emotion recognition and semantic coherence scoring to predict candidate suitability and engagement during interviews.

D. Multimodal Fusion Approaches

To achieve a holistic evaluation, several studies have proposed multimodal systems that combine visual, audio, and textual data [16]. Fusion techniques are typically categorized as **early fusion** (combining raw features) or **late fusion** (combining individual model outputs). Approaches such as M3ER and Multimodal Transformers have demonstrated

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429 Volume 6, Issue 8, March 2026 Impact Factor: 8.2

improved emotion recognition accuracy and better generalization to real-world interview conditions [17]. Furthermore, recent frameworks integrate Explainable AI mechanisms to identify modality-specific contributions and ensure transparency in decision-making [17], [19]. While these multimodal systems provide robust analysis, they require large annotated datasets and high computational resources, which limit their practical scalability.

E. Challenges and Observations

Although current models have achieved promising results, several limitations remain unresolved. Most approaches depend heavily on constrained datasets that lack diversity in culture, accent, lighting, and background conditions [19], [22]. Real-world interviews often involve noisy audio, varied camera angles, and limited labeled data, which degrade model accuracy. Additionally, fairness and privacy issues have emerged as major concerns, as AI systems may unintentionally inherit demographic biases from training data [23], [24]. Recent work in fairness-aware multimodal learning and ethical AI highlights the need for interpretable and bias-resistant systems to ensure responsible deployment in recruitment scenarios.

Overall, existing studies confirm that integrating multimodal deep learning and explainable AI frameworks can significantly improve the accuracy, transparency, and fairness of interview evaluations. However, further research is needed to create publicly available, balanced datasets and lightweight yet interpretable models suitable for real-time applications such as **PrepWise**.

Sr. No.	Author(s) / Year	Technique / Model Used	Dataset(s)	Modality	Accuracy / Result	Key Contribution
1	A. Kumar et al., 2023 [1]	CNN + Facial Feature Extraction	FER-2013	Visual	91.2 %	Automated behavioral analysis for interview evaluation
2	R. Sharma & M. Bhatia, 2023 [2]	Random Forest, Logistic Regression	Custom Interview Data	Textual	85.6 %	Bias reduction in candidate evaluation
3	Y. Zhao & L. Chen, 2022 [8]	CNN–BiLSTM Hybrid	RAVDESS, IEMOCAP	Audio	88.4 %	Robust speech emotion recognition under noise
4	S. Li & H. Zhang, 2023 [5]	Multimodal Transformer	CMU-MOSEI	Text + Visual	90.1 %	Vision-language fusion for multimodal sentiment analysis
5	T. Mittal et al., 2022 [7]	M3ER Framework	CMU-MOSI	Visual + Audio + Text	92.3 %	Multiplicative fusion for emotion recognition
6	H. Kim & J. Choi, 2023 [14]	Vision Transformer (ViT)	AffectNet	Visual	93.5 %	Real-time facial expression recognition
7	N. Deshmukh & R. Kulkarni, 2023 [9]	BERT + Personality Trait Classification	Essays Dataset	Textual	87.4 %	Text-based personality prediction
8	M. Reddy & P. Varma,	Audio-Visual Fusion CNN	Custom Dataset	Audio + Visual	89.8 %	Confidence estimation using

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429 Volume 6, Issue 8, March 2026 Impact Factor: 8.2

	2023 [16]					multimodal cues
9	A. Sahu et al., 2023 [15]	Multimodal Deep Learning	IEMOCAP + Interview Clips	Audio + Visual + Text	91.0 %	Interview performance prediction using deep fusion
10	A. Ghosh & M. Ray, 2024 [21]	Fairness-Aware Transformer	Balanced Multimodal Data	Multimodal	88.7 %	Bias mitigation in multimodal AI systems

Table I. Comparison of Key Literature on AI-Based Interview Evaluation

IV. DISCUSSION AND IDENTIFIED RESEARCH GAPS

The reviewed studies demonstrate that artificial intelligence has significantly advanced the automation of interview evaluation. Multimodal systems combining facial, textual, and audio cues have outperformed single-modality methods in emotion recognition and behavioral analysis [7]. These approaches enhance the reliability of candidate assessment by capturing subtle indicators such as micro-expressions, vocal tone, and linguistic sentiment. However, despite these achievements, several challenges and gaps persist that must be addressed to build efficient, fair, and scalable AI-based evaluation systems.

A. Dataset Limitations

One of the major constraints in existing research is the lack of large-scale, diverse, and annotated datasets for real interview scenarios. Most models are trained on datasets like FER-2013, IEMOCAP, and RAVDESS, which were not originally designed for interview analysis [8]. As a result, models trained on such datasets often fail to generalize well to real-life environments involving varying lighting, background noise, and spontaneous expressions. Furthermore, publicly available multimodal datasets are limited in size, causing overfitting and bias toward specific demographic groups.

B. Real-Time Processing and Scalability

Another limitation arises from the computational complexity of multimodal deep learning models. Transformer-based and fusion architectures often require high-end GPUs and large memory resources, restricting their real-time usability on low-cost devices [9]. Deploying such models in real interview settings demands lightweight, optimized architectures capable of maintaining high accuracy while minimizing latency and power consumption.

C. Explainability and Interpretability

Although recent studies incorporate Explainable AI (XAI) techniques such as Grad-CAM, SHAP, and LIME to visualize feature importance, the interpretability of decisions in multimodal systems remains limited [10]. Most models function as black boxes, providing predictions without adequate justification. This lack of transparency makes it difficult for recruiters and candidates to understand or trust the evaluation outcomes. Therefore, developing interpretable and human-understandable AI systems is a key research priority.

D. Ethical and Fairness Concerns

Bias and fairness have emerged as critical concerns in AI-driven recruitment. Many studies highlight that training data may contain gender, racial, or cultural imbalances, leading to biased predictions [11], [13], [14]. Moreover, privacy issues arise when personal visual or audio data are stored or transmitted without proper encryption. Ethical frameworks for responsible AI usage in recruitment must be integrated into model design to ensure transparency, consent, and non-discrimination.

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429 Volume 6, Issue 8, March 2026 Impact Factor: 8.2

E. Integration with Real-World Applications

While several works achieve high accuracy in controlled environments, few systems have been validated in real-world corporate or academic interviews [15], [16]. Bridging the gap between research and deployment requires collaboration between AI researchers, HR professionals, and psychologists. Hybrid evaluation frameworks that combine automated scoring with minimal human supervision could provide an effective balance between efficiency and ethical accountability.

Summary

In summary, despite considerable progress in multimodal deep learning for candidate assessment, current systems face challenges related to data quality, interpretability, and fairness. Addressing these issues through explainable, unbiased, and resource-efficient models is crucial for enabling reliable real-world applications. These insights directly influence the development of ***PrepWise***, which aims to integrate multimodal learning with explainable and ethical AI to ensure transparent and trustworthy interview evaluations.

Sr. No.	Identified Research Gap	Description / Observation	Proposed Solution / Future Direction
1	Limited and Non-Diverse Datasets	Existing datasets (FER-2013, IEMOCAP, RAVDESS) lack diversity and real-interview scenarios	Develop large-scale, open, balanced datasets capturing diverse cultures and interview environments
2	Poor Generalization of Models	Models trained on lab data fail in real-world settings with lighting or noise variations	Use data augmentation and transfer learning to enhance model robustness
3	High Computational Complexity	Transformer and fusion models require heavy GPU resources and are hard to deploy in real-time	Design lightweight CNN/Transformer architectures and optimize for edge devices
4	Lack of Explainability	Deep models act as black boxes, providing no clarity on decision logic	Integrate Explainable AI (XAI) methods such as Grad-CAM, SHAP, and LIME for interpretability
5	Ethical and Bias Issues	Models inherit gender or cultural bias from training data	Implement fairness-aware learning and bias-mitigation algorithms with ethical AI principles
6	Privacy Concerns	Facial and voice data may expose sensitive personal information	Use data encryption, anonymization, and user consent frameworks
7	Lack of Real-World Validation	Most systems tested only in controlled environments	Collaborate with industries and academia to test in real interview conditions
8	Limited Fusion Techniques	Few studies compare fusion strategies comprehensively	Evaluate early, late, and hybrid fusion for optimal multimodal performance
9	Unbalanced Modality Importance	Some modalities dominate (overfitting to visual features)	Apply attention-based fusion and dynamic weighting mechanisms
10	Lack of Standard Benchmark Metrics	Inconsistent evaluation across datasets and tasks	Define standardized evaluation benchmarks for interview assessment systems

Table II. Identified Research Gaps and Proposed Solutions.

V. CONCLUSION AND FUTURE DIRECTIONS

The review highlights the rapid evolution of artificial intelligence in automating candidate evaluation through multimodal analysis of visual, audio, and textual cues. AI-driven systems have shown remarkable progress in

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429 Volume 6, Issue 8, March 2026 Impact Factor: 8.2

improving the objectivity and consistency of interview assessments compared to traditional manual methods. Techniques such as convolutional neural networks, transformer-based architectures, and multimodal fusion frameworks have significantly enhanced the accuracy of emotion, sentiment, and behavioral recognition.

However, despite these advancements, several limitations persist. Most studies rely on limited or domain-specific datasets, leading to poor generalization in real-world scenarios. Current models often operate as black boxes, offering limited transparency regarding their decision-making process. Additionally, issues related to computational cost, ethical fairness, and data privacy remain major challenges that restrict large-scale deployment of these systems.

Future research should focus on creating open-source, diverse, and balanced multimodal datasets specifically designed for interview environments. The development of lightweight, resource-efficient architectures capable of functioning in real-time applications is also essential. Moreover, integrating explainable AI techniques will help make predictions interpretable and trustworthy for both recruiters and candidates. Ethical frameworks and fairness-aware learning algorithms must be embedded into model design to prevent discrimination and bias.

The insights gathered from this review directly support the ongoing development of ***PrepWise***, an AI-based interview assistant designed to provide automated, unbiased, and explainable feedback to candidates. By combining multimodal deep learning, interpretability, and ethical AI principles, ***PrepWise*** aims to bridge the gap between academic research and real-world application — paving the way toward intelligent, transparent, and equitable recruitment systems.

Acknowledgement

We would like to express our sincere gratitude to our guide, to Dr. Renuka Deshpande (Project Guide), for her invaluable guidance, continuous support, and constructive suggestions throughout the course of this research work. Her insightful feedback, encouragement, and constant motivation played a vital role in the successful completion of this project.

We are deeply thankful to Shivajirao S. Jondhale College of Engineering (SSJCOE), for providing the necessary academic support, resources, and motivation required to carry out this work effectively and for offering a conducive learning environment and the required facilities that enabled us to successfully complete this project.

Finally, we express our sincere thanks to all faculty members, staff, and peers who directly or indirectly contributed to the successful completion of this work.

We would like to place on record our heartfelt thanks to our project guide for constant guidance, encouragement, and support, which have been of great assistance in organizing our thoughts and keeping us on the correct path with our objectives. We would like to place on record our heartfelt thanks to all the Shivajirao Jondhale College of Engineering teaching staff, non-teaching staff, administrative staff, and all support staff members who have assisted in creating an environment for learning and research. We thank them for the opportunity to work under their supervision and hope to get their constant support as we continue with our project.

REFERENCES

[1] A. Kumar and S. Mehta, "AI-Based Behavioral Analysis for Automated Interview Evaluation," *IEEE Access*, vol. 11, pp. 131245-131258,2023.

[2] R. Sharma and M. Bhatia, "Reducing Human Bias in Recruitment Evaluation Using Machine Learning," *Expert Systems with Applications*, vol. 213, pp. 118927,2023.

[3] V. Patel et al., "Deep Learning Approaches for Automated Candidate Assessment," *Procedia Computer Science*, vol. 218, pp. 205–214, 2023.

[4] J. Wang and K. Lee, "A Survey on Multimodal Emotion Recognition: Audio, Visual, and Textual Cues," *IEEE Transactions on Affective Computing*, vol. 14, no. 2, pp. 321-339,2023.

[5] S. Li and H. Zhang, "Multimodal Sentiment Analysis Using Vision and Language Transformers," *Pattern Recognition Letters*, vol. 165, pp. 44-52,2023.

[6] P.Gupta et al., "Facial Emotion Recognition Using Convolutional Neural Networks: A Comprehensive Review," *Multimedia Tools and Applications*, vol .82,pp. 17465-17489,2023.

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429 Volume 6, Issue 8, March 2026 Impact Factor: 8.2

[7] T. Mittal et al., "M3ER: Multiplicative Multimodal Emotion Recognition Using Facial, Text, and Speech Features," *Proceedings of AAAI*,pp.1351-1360,2022.
[8] Y. Zhao and L. Chen, "Speech Emotion Recognition Using CNN-BiLSTM Networks," *IEEE Signal Processing Letters*, vol. 29, pp. 1245-1249,2022.
[9] N. Deshmukh and R. Kulkarni, "Personality Prediction from Text Using Transformer-Based Models," *Applied Intelligence*, vol. 53, pp. 16615-16630,2023.
[10] A. Jaiswal et al., "Human Emotion Detection from Facial Images Using Deep Learning and Explainable AI," *Cognitive Computation*, vol. 15, pp. 844-861,2023.
[11] S. Tripathi and R. Singh, "A Comprehensive Survey on Multimodal Human Behaviour Analysis," *ACM Computing Surveys*, vol. 55, no. 7, pp. 1-35,2023.
[12] K. Bhattacharya and D. Chaudhuri, "Explainable Artificial Intelligence in Human Resource Analytics," *IEEE Access*, vol. 10, pp. 97200-97214,2022.
[13] B. Pandey et al., "Speech-to-Text Emotion Analysis Using Transformer-Based Acoustic Models," *Journal of Intelligent Systems*, vol. 32, no.4, pp. 489-500,2023.
[14] H. Kim and J. Choi, "Vision Transformer for Facial Expression Recognition in Real-Time Applications," *Sensors*, vol. 23, no. 4, p. 2056, 2023.
[15] A. Sahu et al., "Interview Performance Prediction Using Multimodal Deep Learning," *International Journal of Information Management Data Insights*, vol. 3, no. 2, p. 100118,2023.
[16] M. Reddy and P. Varma, "Fusion of Audio-Visual Features for Candidate Confidence Estimation," *IEEE Transactions on Multimedia*, vol. 25, pp. 4112-4123,2023.
[17] S. Kumar et al., "Transformer-Based Emotion and Sentiment Recognition for Conversational AI," *Neural Computing and Applications*, vol. 35, pp. 21467-21482,2023.
[18] L. Wu and G. He, "Benchmarking Multimodal Datasets for Human Behaviour Understanding," *Data in Brief*, vol. 46, p. 108901, 2023.
[19] R. Narayan and K. Tiwari, "Explainable Models for Recruitment Decision Support," *Knowledge-Based Systems*, vol. 273, p. 110584, 2023.
[20] S. Joshi et al., "Real-Time Emotion Recognition on Edge Devices Using Lightweight CNN Models," *IEEE Internet of Things Journal*, vol. 10, no.8, pp. 6578-6589,2023.
[21] A. Ghosh and M. Ray, "Bias and Fairness in Multimodal AI Systems: A Survey," *Information Fusion*, vol. 101, pp. 102021, 2024.
[22] E. Chandra et al., "A Review on Speech Emotion Recognition: Datasets, Features, and Methods," *Electronics*, vol. 13, no. 6, p. 1103, 2024.
[23] V. R. Patil and A. Mahajan, "DeepFake and Forgery Detection in Candidate Interviews Using Vision Transformers," *IEEE Access*, vol. 12, pp.18234-18247,2024.
[24] K. Srinivasan and P. Yadav, "Ethical and Privacy Considerations in AI-Driven Recruitment," *AI and Ethics*, vol. 4, pp. 231-247, 2024