https://doi.org/10.46335/IJIES.2025.10.6.6 **e-ISSN: 2456-3463**
Vol. 10, No. 6, 2025, PP. 28-32

International Journal of Innovations in Engineering and Science, www.ijies.net

From Interviews to Insights: A Review of Multimodal Personality Assessment in Recruitment

K. R. Rajput[1], A. J. Kharade[2], A. P. Pawar[3], T. S. Wakhare[4], Prof. D. D. Ahir[5]

[1,2,3,4]UG student
Department of Computer Engineering MESWCOE, Pune, India, 411005,
[5]Professor : 0000-0001-8081-8571
Department of Computer Engineering MESWCOE, Pune, India, 411005,

Email of Corresponding Author: ***deepali.ahir@mescoepune.org***

Received on: *29 April, 2025* ***Revised on:*** *02 June ,2025* ***Published on:*** *06 June,2025*

Abstract – *The recruitment landscape has traditionally relied on manual assessments, often prone to biases and inefficiencies, making it challenging to evaluate candidates objectively. While modern recruitment has embraced asynchronous video interviews (AVI) for convenience, they lack the personal interaction of traditional methods, leading to gaps in assessing soft skills like personality traits. The Automatic Personality Recognition (APR) system bridges this gap by leveraging multimodal data—text, audio, and visuals—to evaluate candidates based on the Big Five personality traits. Using advanced deep learning techniques, the system processes recorded interviews to generate personality scores, enabling unbiased and scalable assessments. The APR system maintains candidate confidentiality and empowers recruiters with data-driven insights, enhancing decision-making while addressing the limitations of both traditional and modern recruitment processes.*

Keywords - *Big Five Personality Traits, Multimodal Data (MD), Computer Vision (CV), Asynchronous Video Interviews (AVI), Natural Language Processing (NLP)*

INTRODUCTION

Asynchronous Video Interviews (AVIs) are increasingly gaining popularity in modern recruitment processes due to their convenience, flexibility and scalability. However, traditional personality assessments used during interviews are subject to human judgment and are therefore prone to bias and inconsistency. While AVIs address logistical challenges, they neglect the essential face-to-face interaction required for the kind of all-rounded assessment. In an attempt to overcome these limitations, the current research thus proposes the building of an APR system that impartially assesses the Big Five personality traits of applicants—Extraversion, Agreeableness, Conscientiousness, Neuroticism, and Openness. This proposed system will rely on a multimodal approach, aggregating all the textual, auditory, and visual data for the recruitment agencies to perform a relatively deeper and unbiased assessment. In the system, sophisticated techniques in NLP, computer vision and speech analysis are used to ensure highly data-driven, scalable, and effective personality evaluations.

The APR system supports automation in the extraction and analysis of multimodal data stemming from video interviews, tackling effectively the issue of modality balance, bias reduction, and the attainment of high accuracy. This approach mitigates limitations inherent with legacy approaches while also providing recruiters with concrete understanding in personality. The approach, therefore, serves to establish a foundation for revolutionizing the contemporary recruitment processes with data-driven decision-making and communicating the concepts of increased fairness and scalability. This method further advances the qualities of both fairness and scalability while establishing the backbone of transforming modern recruitment processes into data-driven decision-making.

LITERATURE REVIEW

A. Multimodal Approach

https://doi.org/10.46335/IJIES.2025.10.6.6 **e-ISSN: 2456-3463**
Vol. 10, No. 6, 2025, PP. 28-32

This study [21] focuses on developing an automatic interview evaluation system to enhance the accuracy and dependability of human evaluation. It incorporates standardized video interview protocols and multimodal features, such as verbal content, visual cues, and personality traits, to enhance prediction accuracy. Data were collected from 36 participants who completed 419 mock interview responses and 68 presentations. Video and audio recordings were captured using Kinect and HD webcams, with the MultiSense framework synchronizing data streams from various sensors. Human evaluation rubrics assessed verbal content and personality, while feature extraction methods like Doc2Vec, LIWC, and visual word clustering provided fixed-sized feature vectors for machine learning models. These models predicted personality traits and performance scores, demonstrating improved human scoring accuracy through automated scoring representations.

Building on advancements in APR, new technologies like asynchronous video interviews enhance recruitment by predicting candidate hireability. The HireNet model in [16] employs hierarchical attention mechanisms and processes over 7,000 real candidate interviews, combining audio, text, and video features through multimodal fusion to surpass mono-modal approaches in accuracy. The interviews featured job-specific questions posed to candidates, and the data was divided into train, validation, and test sets in a ratio of 80/10/10. Sequential modeling using Bidirectional GRUs and feature extraction with eGeMAPS via OpenSmile enhanced analysis, while early and late fusion methods improved multimodal integration. HireNet outperformed vote-based baselines and sequential models, achieving superior F1-scores. Audio and text modalities demonstrated stronger performance compared to video models, with multimodal fusion further enhancing results. Future work aims to address ethical considerations and biases in automated evaluations.

Extending the above applications, [24] explores the detection of Big Five personality traits during self-presentations. Using data from 89 participants, including employees, students, and external individuals (46 males, 43 females; 47 young, 42 adults), personality traits were assessed through short presentations lasting 30 to 120 seconds. Conscientiousness and Emotional Stability were the most accurately detected traits, particularly with the SVM-RBF model, while Extraversion and Agreeableness showed lower detection rates. A total of 29 features—17 visual, 9 acoustic, and 3 speech metrics—were extracted from webcam recordings and audio captured via clip-on microphones. The study employed Naive Bayes and SVM algorithms, with backward linear regression linking features to traits. Results highlight the influence of situational and nonverbal cues on personality expression, supporting the use of automated systems to enhance interview evaluations. Participants also completed the Italian Big Five questionnaire, providing a validated framework for trait analysis. This approach holds promise for improving job performance insights for both interviewers and candidates.

Recent research on multimodal models for personality recognition demonstrates significant advancements through the integration of visual, audio, and textual features using deep learning architectures. These models leverage datasets like ChaLearn, UDIVAv0.5, and First Impressions, containing thousands of annotated video clips [18]. Techniques such as ViT, Swin Transformers, BiLSTM, and VGGish process spatiotemporal and audio features [2] [11], while tools like librosa and FFmpeg extract advanced acoustic metrics [5]. Multimodal fusion consistently improves accuracy, with superior performance in personality dimensions like Extraversion, Openness, and Conscientiousness. The use of pretrained embeddings, hierarchical temporal aggregation, and AutoKeras hyperparameter tuning further optimizes these ML systems, achieving top-tier performance and high Pearson correlations across traits.

B. Approach based on Textual Features

Complementing advancements in the field, alternative approaches like the projective method in [12] challenge traditional inventories such as the Big Five, commented on for susceptibility to faking and response biases. By employing Z-tests, a projective testing method, this approach delves into unconscious personality traits, overcoming limitations of conventional inventories and achieving promising accuracy (AUC-ROC of 0.85). Meanwhile, machine learning models continue to transform APR systems, as seen in [6] where RNN, LSTM, and BERT models were applied to Spanish-language educational settings. The BERT-based model led with a 72% accuracy rate, underscoring the adaptability of Transformer-based models for job interviews by leveraging rich textual data from video responses. Similarly, [15] introduced SBERT, optimizing semantic text analysis with reduced

https://doi.org/10.46335/IJIES.2025.10.6.6 **e-ISSN: 2456-3463**
Vol. 10, No. 6, 2025, PP. 28-32

International Journal of Innovations in Engineering and Science, www.ijies.net

computational costs while maintaining high accuracy. This capability positions SBERT as a vital tool for real-time candidate evaluation in asynchronous video interviews, enabling efficient large-scale recruitment processes.

C. Approach based on Visual Features

Expanding the scope of APR, visual features have emerged as critical components, particularly in asynchronous video-based interviews. [23] explores the relationship between OCEAN traits and facial expressions, using the Computer Expression Recognition Toolbox (CERT) to analyze micro-expressions such as joy, anger, and surprise. The study found strong correlations between expressive facial behaviors and traits like Extraversion and Agreeableness, emphasizing the role of continuous facial activity as an indicator of personality. Further, deep learning models like FaceNet, introduced in [22] provide robust 128-dimensional embeddings that are invariant to lighting and pose, enabling precise extraction of personality-relevant visual features. By integrating these embeddings into hybrid models, combining text and visual inputs, advanced personality prediction systems can achieve enhanced accuracy, offering comprehensive insights for large-scale recruitment processes. By integrating these embeddings into hybrid models, combining text and visual inputs, advanced personality prediction systems can achieve enhanced accuracy, offering comprehensive insights for large-scale recruitment processes. Additionally, research on XLNet with refined highway and switching modules has further expanded APR capabilities, enhancing contextual understanding and feature integration for tasks like chatbot-based personality prediction [4].

KEY FINDINGS

The literature survey highlights significant advancements in automated personality recognition, leveraging multimodal approaches and deep learning techniques. The OCEAN personality model provides a robust theoretical basis for mapping video cues to the Big Five personality traits. Sequential modeling methods like LSTM and RNN capture temporal dependencies in video and audio data, enhancing accuracy. In natural language processing, SBERT enables deep linguistic feature extraction, improving personality trait predictions in text-based inputs. Bi-LSTM further strengthens context understanding by analyzing input sequences bidirectionally, accommodating complex dependencies. A deep ensemble approach, integrating CNNs, LSTMs, and transformer models, demonstrates improved accuracy and robustness by combining diverse strengths, offering a comprehensive framework for automated personality recognition in asynchronous video interviews.

OUTCOMES

A. Comparison of preferred Datasets

The ChaLearn's First Impressions V2 dataset offers significant advantages for personality assessment in AVIs. It provides multimodal data (audio, visual, and text) with pre-labeled annotations for the Ocean personality model, reducing the need for manual labeling. Unlike UDIVA, which focuses on dyadic interactions, or MIT Interviews, which lack comprehensive behavioral annotations, ChaLearn V2 is specifically designed for single-speaker scenarios, making it more aligned with real-world asynchronous interviews. Its diverse demographics and rich annotations capture subtle cues like micro-expressions, gestures, and tonal variations, ensuring better training data for accurate and robust personality prediction models.

Table 1- Dataset Characteristics

Sr. no	Name	Size	Interview type	Metadata provided
1	First Impressions v2	~10,000 clips of 5-10 seconds, across 2,000 participants	Monologues	Ethnicity, gender, age annotations
2	UDIVA	~188 sessions, 90.5 hours of interactions	Dyadic interactions	Participant demographics, Personality traits, Session details
3	MIT Interview Dataset	~2,900 interviews across 1,100 participants	Dyadic interactions	Facial expressions, Speech features, Language use

B. Model Evaluation for Effective Personality Trait Prediction

https://doi.org/10.46335/IJIES.2025.10.6.6 **e-ISSN: 2456-3463**
Vol. 10, No. 6, 2025, PP. 28-32

International Journal of Innovations in Engineering and Science, www.ijies.net

BiLSTM (Bidirectional Long Short-Term Memory) stands out as an effective model for personality recognition tasks when compared to alternatives like SVM, RNN, CNN, and Swin Transformer. Unlike SVM, that acts as a linear classifier, is insufficient in handling sequential as well as multimodal data, whereas BiLSTM can handle temporal dependencies, and hence makes it suitable for the time-series analysis of video interviews. RNNs can learn sequential patterns; however, they incur vanishing gradient problems, meaning that they can only learn short-term dependencies and not long-term dependencies.

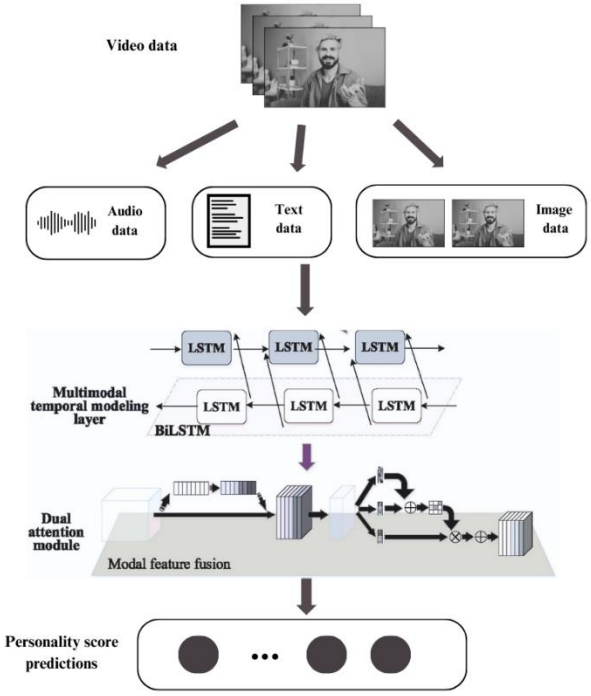

Fig. 1- BiLSTM for multimodal video analysis

BiLSTM manages to overcome these issues by processing the sequences in both forward and backward directions, so it becomes easier to acquire extensive contextual relationships across audio, visual, and textual data. Unlike CNNs, which are good at extracting spatial feature but fall short in catching temporal dynamics, BiLSTMs are capable of extracting meaningful temporal and sequential characteristics. Furthermore, while architectures like Swin Transformers show incredible accuracy in visual tasks, they require an immense amount of computational resources and are not well adapted to medium-sized datasets such as ChaLearn V2, especially when they need real-time processing. Instead, BiLSTMs strike a good balance because they are computationally inexpensive, noisy resistant, and strong at multimodal integration, so it is the most appropriate choice for this application.

CONCLUSION

This research works as a significant advancement in Recruitment and selection processes can be streamlined by the creation of an APR system. By leveraging multimodal data—integrating audio, visual, and textual inputs—and employing advanced machine learning techniques, the proposed system demonstrates the potential to deliver accurate and unbiased personality assessments. The comparison of various algorithms and the selection of BiLSTM highlights the importance of using models that balance predictive accuracy and computational efficiency for large-scale applications.

The study addresses critical gaps in traditional personality assessment methods by minimizing human biases and introducing scalable, automated tools that align with the demands of modern recruitment. While it also acknowledges the ongoing challenges in optimizing multimodal integration, ensuring fairness, and improving interpretability, establishing a foundation for future exploration in this domain, paving the way for more robust, ethical, and impactful AI-driven solutions in recruitment and beyond.

REFERENCES

[1] *Bounab, Y., Oussalah, M., Arhab, N., Bekhouche, S. (2024). Towards job screening and personality traits estimation from video transcriptions. Expert Systems with Applications, 238(D), 122016. https://doi.org/10.1016/j.eswa.2023.122016.*

[2] *X. Duan, H. Li, F. Yang, B. Chen, J. Dong, and Y. Wang, "Multimodal Automatic Personality Perception Using ViT, BiLSTM and VGGish," 2024 5th International Conference on Computer Engineering and Application (ICCEA), Hangzhou, China, 2024, pp. 549-553, doi: 10.1109/ICCEA62105.2024.10604109.*

[3] *K. A. Dnyaneshwar and G. Poonam, "AI-Driven Insights: Personality Evaluation in Asynchronous Video Interviews for Informed Hiring Decisions," 2024 IEEE International Conference on Information Technology, Electronics and Intelligent Communication Systems (ICITEICS), Bangalore, India, 2024, pp. 1-6, doi: 10.1109/ICITEICS61368.2024.10625601.*

[4] *O. T.-C. Chen, C.-H. Tsai, and M.-H. Ha, "Automatic Personality Recognition via XLNet with Refined Highway and Switching Module for Chatbot," 2024 IEEE International Symposium on Circuits and Systems (ISCAS), Singapore, Singapore, 2024, pp. 1-5, doi: 10.1109/ISCAS58744.2024.10558116.*

[5] *S. Ghassemi et al., "Unsupervised Multimodal Learning for Dependency-Free Personality Recognition," IEEE Transactions on Affective Computing, vol. 15, no. 03, pp. 1053-1066, July-Sept. 2024, doi: 10.1109/TAFFC.2023.3318367.*

[6] *Zatarain Cabada, Ramón Barron Estrada, Maria Bátiz Beltrán, Víctor Sapien, Ramón Ruiz, Gerardo. (2023).*

Sentiment Analysis of Spanish Text for Automatic Personality Recognition in Intelligent Learning Environments. pp. 1-4. doi: 10.1109/ENC60556.2023.10508699.

[7] *D. Nagajyothi, S. A. Ali, P. H. Sree, and P. Chinthapalli, "Automatic Personality Recognition In Interviews Using CNN," 2023 4th IEEE Global Conference for Advancement in Technology (GCAT), Bangalore, India, 2023, pp. 1-7, doi: 10.1109/GCAT59970.2023.10353423.*

[8] *Holtrop, Djurre, Oostrom, Janneke, Breda, Ward, Koutsoumpis, Antonis, and de Vries, Reinout. (2022). Exploring the application of a text-to-personality technique in job interviews. European Journal of Work and Organizational Psychology, 31, 1-18. doi: 10.1080/1359432X.2022.2051484.*

[9] *J. R. Lima, H. J. Escalante, and L. V. Pineda, "Sequential Models for Automatic Personality Recognition from Multimodal Information in Social Interactions," 2022 IEEE International Autumn Meeting on Power, Electronics and Computing (ROPEC), Ixtapa, Mexico, 2022, pp. 1-6, doi: 10.1109/ROPEC55836.2022.10018711.*

[10] *X. Duan, Y. Yu, Y. Du, H. Liu, and Y. Wang, "Personality Recognition Method Based on Facial Appearance," 2022 3rd International Conference on Computer Vision, Image and Deep Learning (CVIDL ICCEA), Changchun, China, 2022, pp. 710-715, doi: 10.1109/CVIDLICCEA56201.2022.9824658.*

[11] *X. Duan, Q. Zhan, S. Zhan, Y. Yu, L. Chang, and Y. Wang, "Multimodal Apparent Personality Traits Analysis of Short Video Using Swin Transformer and Bi-directional Long Short-Term Memory Network," 2022 4th International Conference on Frontiers Technology of Information and Computer (ICFTIC), Qingdao, China, 2022, pp. 1003-1008, doi: 10.1109/ICFTIC57696.2022.10075178.*

[12] *Camati, Ricardo Enembreck, Fabrício. (2020). Text-Based Automatic Personality Recognition: a Projective Approach. pp. 218-225. doi: 10.1109/SMC42975.2020.9282859.*

[13] *Z. Su, Z. Lin, J. Ai, and H. Li, "Rating Prediction in Recommender Systems Based on User Behavior Probability and Complex Network Modeling," IEEE Access, vol. 9, pp. 30739-30749, 2021, doi: 10.1109/ACCESS.2021.3060016.*

[14] *K.Yesu, K., Shandilya, S., Rekharaj, N., Ankit, K., and Sairam, P. S. (2021). Big Five Personality Traits Inference from Five Facial Shapes Using CNN. IEEE 4th International Conference on Computing, Power and Communication Technologies (GUCON), Kuala Lumpur, Malaysia, pp. 1-6. doi: 10.1109/GUCON50781.2021.9573895.*

[15] *Reimers, N., Gurevych, I. (2019). Sentence-BERT: Sentence Embeddings using Siamese BERT-Networks. Conference on Empirical Methods in Natural Language Processing. doi: 10.48550/arXiv.1908.10084.*

[16] *L´eo Hemamou, Ghazi Felhi, Vincent Vandenbussche, Jean-Claude Martin, and Chlo´e Clavel. 2019. HireNet: a hierarchical attention model for the automatic analysis of asynchronous video job interviews. In Proceedings of the Thirty-Third AAAI Conference on Artificial Intelligence and Thirty-First Innovative Applications of Artificial Intelligence Conference and Ninth AAAI Symposium on Educational Advances in Artificial Intelligence (AAAI'19/IAAI'19/EAAI'19). AAAI Press, Article 71, 573–581.https://doi.org/10.1609/aaai.v33i01.3301573.*

[17] *S. Zhang, S. Zhang, T. Huang, W. Gao, and Q. Tian, "Learning Affective Features With a Hybrid Deep Model for Audio–Visual Emotion Recognition," IEEE Transactions on Circuits and Systems for Video Technology, vol. 28, no. 10, pp. 3030-3043, Oct. 2018, doi: 10.1109/TCSVT.2017.2719043.*

[18] *J. Gorbova, E. Avots, I. Lüsi, M. Fishel, S. Escalera, and G. Anbarjafari, "Integrating Vision and Language for First-Impression Personality Analysis," IEEE MultiMedia, vol. 25, no. 2, pp. 24-33, Apr.-Jun. 2018, doi: 10.1109/MMUL.2018.023121162.*

[19] *T. Baltrusaitis, A. Zadeh, Y. C. Lim, and L.-P. Morency, "OpenFace 2.0: Facial Behavior Analysis Toolkit," 2018 13th IEEE International Conference on Automatic Face & Gesture Recognition (FG 2018), Xi'an, China, pp. 59-66, doi: 10.1109/FG.2018.00019.*

[20] *L. Chen, R. Zhao, C. W. Leong, B. Lehman, G. Feng, and M. E. Hoque, "Automated video interview judgment on a large-sized corpus collected online," 2017 Seventh International Conference on Affective Computing and Intelligent Interaction (ACII), San Antonio, TX, USA, 2017, pp. 504-509, doi: 10.1109/ACII.2017.8273646.*

[21] *Chen, Lei, Feng, Gary, Leong, Chee Wee, Lehman, Blair, Martin-Raugh, Michelle, Kell, Harrison, Lee, Chong Min, Yoon, Su-Youn. (2016). Automated Scoring of Interview Videos using Doc2Vec Multimodal Feature Extraction Paradigm. doi: 10.1145/2993148.2993203.*

[22] *Schroff, Florian, Kalenichenko, Dmitry, Philbin, James. (2015). FaceNet: A Unified Embedding for Face Recognition and Clustering. Proceedings of CVPR.*

[23] *Biel, J., Teijeiro-Mosquera, L., & Gatica-Pérez, D. (2012). FaceTube: predicting personality from facial expressions of emotion in online conversational video. International Conference on Multimodal Interaction.*

[24] *Batrinca, Ligia Maria, Mana, Nadia, Lepri, Bruno, Pianesi, Fabio, and Sebe, Nicu. (2011). Please, tell me about yourself: automatic personality assessment using short self-presentations. Proceedings of the 13th International Conference on Multimodal Interfaces. doi: 10.1145/2070481.2070528.*