http://dx.doi.org/10.21622/ACE.2025.05.1.1317

AN AI-BASED FRAMEWORK FOR IMPROVING EFFICIENCY AND FAIRNESS IN THE INTERVIEW PROCESS

Mohannad Taman[1,3], Yahia Khaled[2,3], Dalia Sobhy[3]

[1] AI Engineer, Obeikan Digital Solutions, Cairo, Egypt.

[2] SOC Analyst, Banque Misr, Cairo, Egypt.

[3] Arab Academy of Science and Technology and Maritime Transport, Alexandria, Egypt.

Emails: {m.taman@obeikan.com.sa, yahiakhaled49@gmail.com, dalia.sobhi@aast.edu }

Received on, 27 April 2025 - Accepted on, 25 May 2025 - Published on, 18 June 2025

ABSTRACT

Artificial intelligence (AI) technologies have advanced to the point where they can assist human resource specialists, such as recruiters, by automating major aspects of the hiring process and efficiently filtering candidate pools. However, limited research has evaluated the effectiveness of AI systems in virtual interviews. This paper presents InstaJob, an AI-powered framework designed to enhance both efficiency and fairness in the hiring pipeline. The system integrates multiple deep-learning components to analyze candidate responses during interviews. Facial emotion recognition is performed using a convolutional neural network (CNN) trained on the FER2013 dataset, achieving a validation accuracy of 77% and outperforming several state-of-the-art approaches. For speech processing, IBM Watson is used to convert spoken responses into text. The transcribed text is then analyzed using EmoRoBERTa, a transformer-based model, to detect emotional signals from verbal content. In addition, IBM Watson is employed to detect filler words and assess speech fluency. These components collectively enable InstaJob to assess candidates' soft skills in a structured and unbiased manner, offering a comprehensive and data-driven evaluation of interview performance.

Index words: Artificial intelligence, virtual interviews, Facial emotion recognition, speech processing, Deep learning applications.

1. INTRODUCTION

Emotion is a complex and dynamic state that arises in response to various stimuli, including experiences, thoughts, or social interactions. It includes subjective feelings, cognitive processing, behavioral responses, physiological changes, and communication cues. Therefore, the ability to identify emotions is essential in a variety of domains, including marketing, human-robot interaction, mental health evaluation, and employment interviews [1, 2]. Using virtual interviews in employment processes has become more common recently.

Virtual video interviews offer several advantages to interviewers and interviewees alike: they enable HR personnel to assess a large number of job applications; (2) they enable HR personnel to review and make decisions offline, and (3) facilitate cost-effective long-distance interviews. However, virtual video interviews pose several challenges, particularly in areas like assessing non-verbal communication, building rapport, and ensuring consistency across candidates. Studies have indicated that

http://dx.doi.org/10.21622/ACE.2025.05.1.1317

in virtual environments, biases related to race, gender, and socioeconomic class May develop [3, 4]. In order to minimize bias by standardizing the interview process and Delivering data-driven insights for more equitable decisions, there is growing interest in the usage of AI-driven tools that objectively analyze applicant responses and non-verbal cues. These technologies can help make virtual interview recruitment fairer by ensuring that hiring decisions are based on qualifications rather than opinions. According to recent surveys [5, 6], AI in recruiting is becoming more widely accepted, with 60% of HR professionals believing it will have a major impact. The global market for AI recruiting is expected to reach $7.1 billion by 2025, highlighting its growing influence in the field. HR professionals can save up to 80% of their time during the hiring process by implementing AI, while candidates who practice using AI-powered interview tools have a higher chance of securing employment opportunities [5, 6]. In this context, the use of automated systems for assessing video interviews to evaluate applicants have grown in popularity in recent years across a range of industries [7].

Hiring managers can save time with these tools, and candidates can schedule interviews at their convenience [8]. Nowadays, automatic video interview assessment systems are in use all over the world. More than 80 million interviews have been assessed according to HireVue is a well-known startup in the automated hiring sector [9]. Emerging applications have made use of social signal processing (SSP) [10] to assist in the study and assessment of interview performance. Nevertheless, emotional cues The voice and face have not been fully investigated in prior studies. This paper addresses the following research question: ***RQ:*** How can artificial intelligence (AI) be applied to ensure the fairness of the applicant's evaluation and the efficiency of the interview process using facial and speech cues?

The remainder of the paper is structured as follows: The related work section presents the literature review related to AI in recruitment and emotion detection techniques. The proposed model section describes the proposed AI-driven framework. The experimental evaluation section provides the set of experiments conducted. Finally, the conclusion section provides conclusions and future work.

RELATED WORK

AI-based recruitment techniques can promote better hiring results, decrease bias, and accelerate the hiring process [11–13]. However, the effectiveness of AI in hiring depends on on the approaches adopted and the implementation environment. This section presents an overview of research on AI-based recruitment practices.

AI IN RECRUITMENT

The use of AI in recruitment has grown significantly over the years. Various studies have explored automated systems for CV parsing, skill assessment, and candidate filtering, with most approaches focusing on text analysis [14, 15]. However, few work have focused on emotional detection from video interviews. In recent years, several recruitment platforms have been proposed [16–18]. For example, a multi-fine-grained method based on recurrent neural networks (RNN) and keyword-question attention mechanisms are proposed for interviews [16]. This method scores different personality characteristics of interviewees, obtaining a comprehensive interview score. The approach leverages a two-stage model learning mechanism and a keyword-question-level attention mechanism to reliably predict personality traits after removing words and phrases that are associated with those traits. This work has focused on determining personality traits based on speech without considering facial emotion recognition. Further, "vRecruit" [17], a machine learning-based web application for virtual recruitment was proposed, which incorporates a client-specific interview process and a text-based sentiment analysis engine. In [18], the authors used CNN to classify a candidate's emotions based on images and nervousness from blink rate during a virtual interview process, with an accuracy of 60%. However, the focus of these works was on

http://dx.doi.org/10.21622/ACE.2025.05.1.1317

evaluating candidates using still images, which may introduce significant bias compared to video interview analysis.

EMOTION DETECTION IN AI

Emotion detection has been a significant area of AI research, especially with the use of deep learning models such as convolutional neural networks (CNN) for facial expression analysis. The 2013 Facial Expression Recognition (FER) Challenge data set (FER-2013 dataset) [19], a widely used benchmark in this field. It provides a diverse range of facial expressions necessary for training and evaluating emotion detection models. The first model, the Local Learning Bag of Words (BOW), was designed for competition application [20]. This was later improved by the Local Learning Deep+BOW model, which combined deep learning features with traditional BOW and was applied to mental disorder detection [21]. In 2020, Ensemble ResMaskingNet, an ensemble learning approach demonstrated the effectiveness of combining multiple CNNs to improve CNN performance [22]. Another CNN-based model from the same year, aimed at enhancing human-computer interaction [23]. By 2022, efforts to improve emotion detection methods using data augmentation with the VGG16 model was introduced, which was used to improve the quality of images in the FER-2013 dataset [24]. However, it achieved a low accuracy as compared to the to previous approaches. In 2023, custom architectures and other methods achieved moderate improvements, with one such custom architecture reaching a considerable accuracy, though the application details were unspecified [25].

Other studies from 2023, reported proposed some models with applications remaining unspecified [26] [27]. To the best of our knowledge, none of these approaches have used facial emotion detection to enhance the recruitment process. In this context, we aim to introduce an novel deep learning model specifically tailored for this purpose. In parallel with facial emotion recognition, text-based emotion recognition has evolved significantly. Models like BERT and EmoRoBERTa have been employed for analyze sentiment and emotional tone in textual data. These models have been particularly useful in assessing emotional intelligence in scenarios such as interviews, where the analysis of both facial expressions and text plays a pivotal role. The evolution of emotion detection models from traditional handcrafted Features extraction techniques to advanced deep learning and ensemble methods underscore the continual advancements in this field. The shift towards deeper architectures and hybrid approaches, along with the integration of multiple datasets, represents a significant leap forward in the accuracy and reliability of emotion detection systems. Ensemble models, in particular, have proven highly effective, combining diverse architectures to achieve superior performance. As emotion detection continues to mature, it will remain a essential component in applications ranging from human-computer interaction (HCI) to personalized content recommendation systems.

2. PROPOSED FRAMEWORK

FULL SYSTEM ARCHITECTURE

This system analyzes video interviews through two parallel paths (Fig. 1). The left path focuses on facial analysis, using HaarCascade for face detection and CNN techniques to assess candidate expressions. The right path processes audio, transcribing it with IBM Watson, then analyzing it for fillers and emotions using IBM Watson and EmoRoBERTa. Each analysis produces a score, which is combined to generate a comprehensive assessment of the candidate's performance.

http://dx.doi.org/10.21622/ACE.2025.05.1.1317

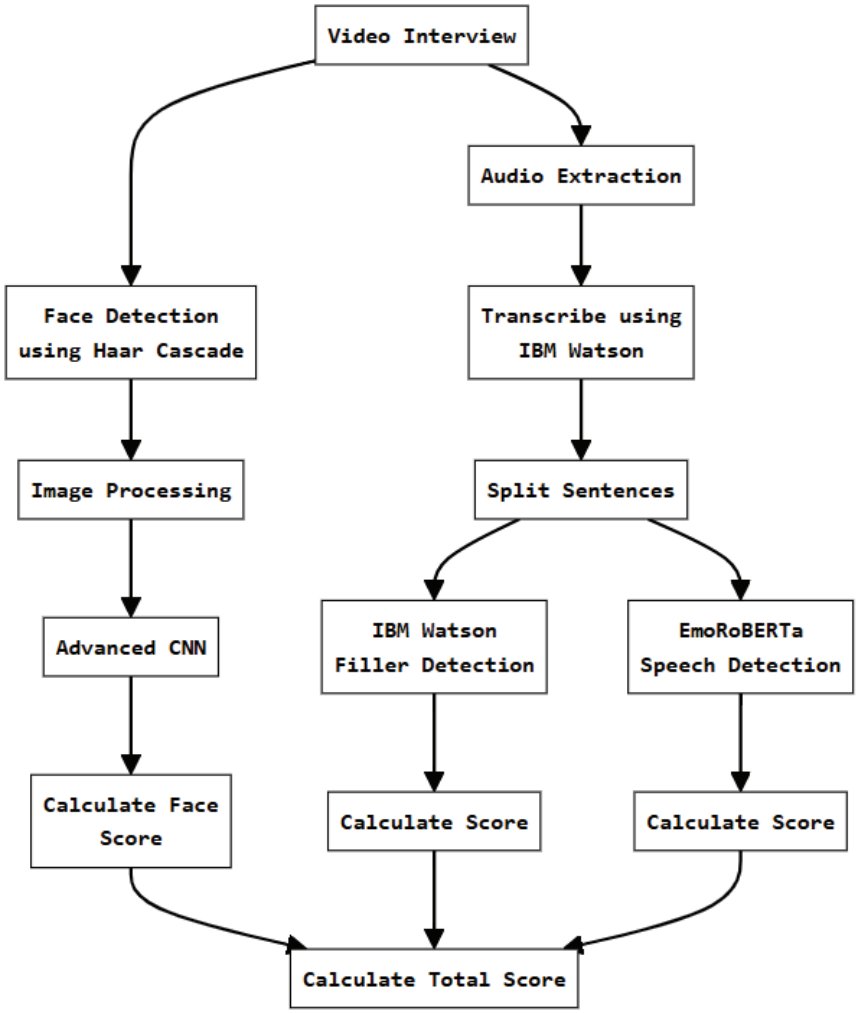

Fig 1. Interview process system Architecture.

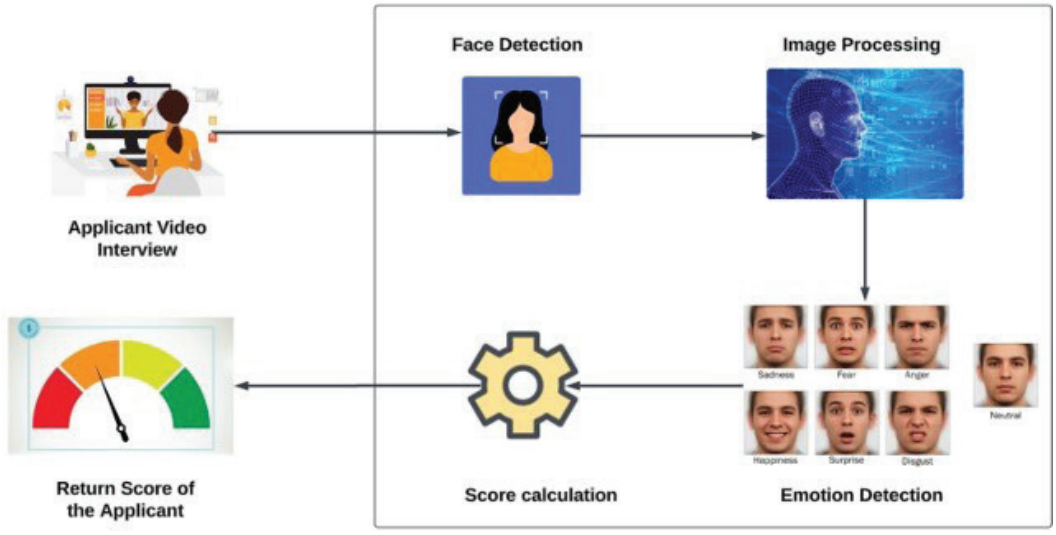

Fig 2. Face Emotion Detection System Architecture.

http://dx.doi.org/10.21622/ACE.2025.05.1.1317

3. FACE EMOTION DETECTION

SYSTEM ARCHITECTURE

The system architecture is shown in Fig. 2. The Face Emotion Detection module aims to analyze facial expressions to detect emotions such as happiness, sadness, anger, and neutrality. It provides users with feedback on their emotional responses during video interviews or interactions, enhancing self-awareness and interview performance. This

The analysis is performed offline after the candidate uploads the video. Finally, the recruiter receives a score indicating the candidate's performance.

HAARCASCADE FOR FACE DETECTION

In our video processing pipeline, we utilize haar cascade frontal face default.xml from the OpenCV library for face detection. This pre-trained HaarCascade classifier is specifically designed to identify frontal faces within images or video frames (Fig. 3).

The classifier operates by evaluating patterns of pixel intensities that correspond to the characteristic features of a human face, such as the arrangement of eyes, nose, and mouth. Upon detection, the classifier marks the location of each identified face with a bounding box, effectively isolating facial regions from the rest of the image or frame.

This bounding box ensures that only the essential facial features are extracted as inputs for further processing, such as emotion detection using machine learning models. Integrating HaarCascade face detection enhances the precision and efficiency of our system in capturing and analyzing facial expressions within video streams.

EMOTION DETECTION

After detecting faces using haar cascade frontal face default.xml in the video processing pipeline, the extracted facial regions (bounded by boxes) are passed to the emotion detection model. These regions, containing key facial features (Fig. 4), are preprocessed through resizing and pixel normalization to ensure consistent input formatting. We developed a Convolutional Neural Network (CNN) model for face emotion detection.

The CNN consists of multiple convolutional and pooling layers followed by a fully

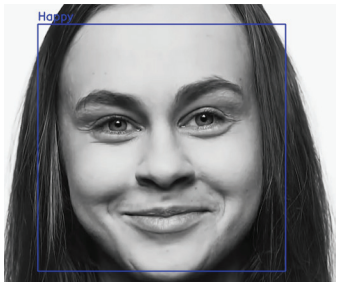

Fig 3. Sample of Face Detection using Haar Cascade.

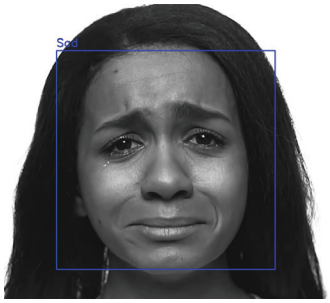

Fig 4. Face Detection Detection.

http://dx.doi.org/10.21622/ACE.2025.05.1.1317

Connected layers. The input shape of images is 48x48 pixels, grayscale. The architecture of our CNN model is shown in Fig. 5. The deep learning model then classifies emotions—such as happiness, sadness, anger, surprise, and neutrality—based on these features. By combining Haar Cascade face detection with emotion recognition model, the system efficiently captures and interprets emotional cues from video streams, enabling meaningful real-time insights.

SPEECH ANALYSIS

It consists of three modules. First, the Speech-to-Text Conversion module, which is implemented using the IBM Watson Speech-to-Text API. It inputs the candidate's speech performs some preprocessing (e.g., sentence segmentation using punctuation (.,?)), and then provides the transcribed text. Second, the Speech Emotion Detection the module uses the pre-trained EmoRoBERTa (28 emotion classes). It performs three main processes: contraction expansion (e.g., "can't" → "cannot"), sentence-level emotion classification, and dominant emotion flagging. The third module is Filler Word Detection, which uses regex-based pattern matching. It searches for some fillers (e.g., 'uh,' 'um,' 'like,' etc.) and provides the timestamped filler occurrences. At the end, it calculates a score for speech detection and filler detection.

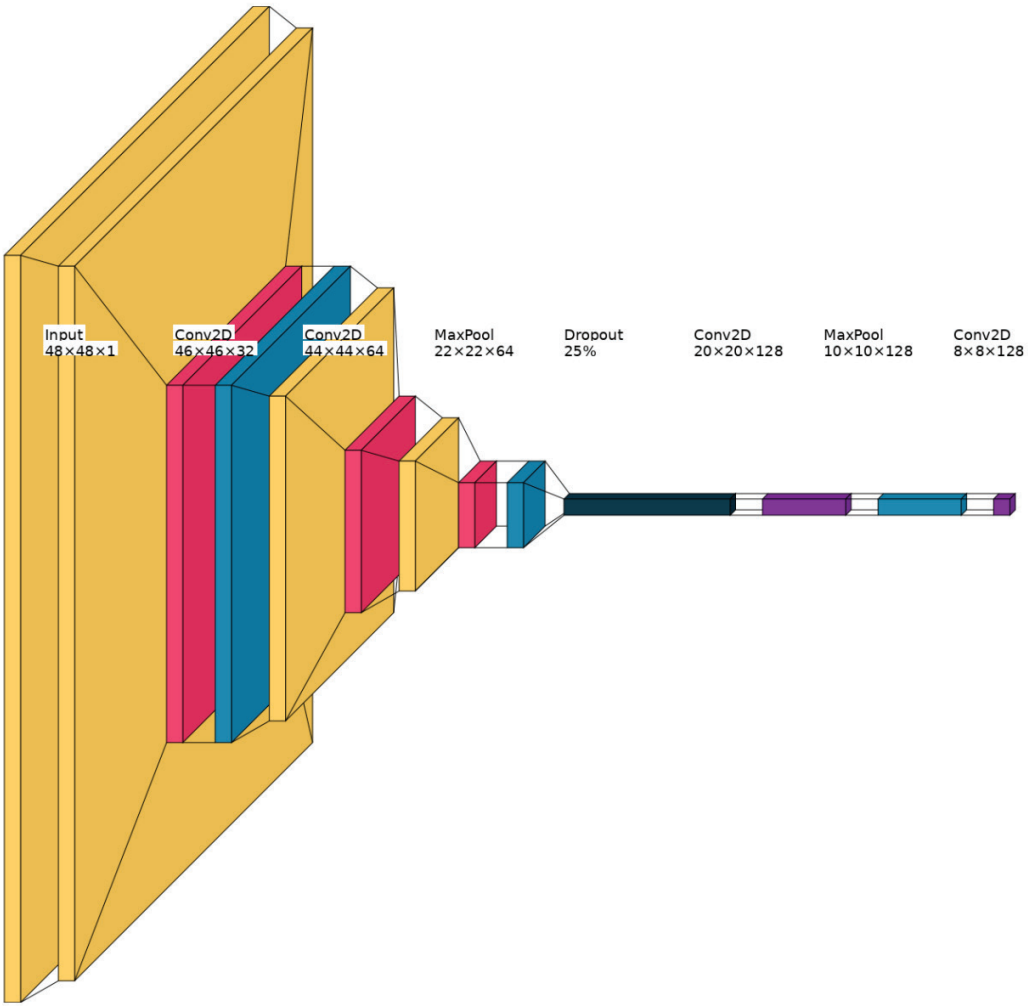

Fig 5. The CNN model for the face emotion detection feature.

http://dx.doi.org/10.21622/ACE.2025.05.1.1317

SCORING METHODOLOGY

The score S is calculated as a weighted sum of negative emotion instances (E) and filler word occurrences (F) from three independent models ($S = \sum_{i=1}^{3} w_i \cdot T_i$). T_i representsthe count of triggers (emotions/filler words) for model i , whereas w_i is the weight per 159 models (inverse of triggers needed to increment the score by 1). Table 1 shows the scoring o f 160 rules per model.

Table 1. Scoring rules per model.

Model	Trigger	Score Increment
Face Emotions Detection	4 negative emotions	+1
Speech-to-Text Emotions AI	6 filler words	+1
Text Emotions AI	2 negative emotions	+1

FACE EMOTION DETECTION EXPERIMENTAL EVALUATION

In this section, we will discuss the experimental evaluation.

4. DATASET

The dataset consists of facial expressions obtained from the Facial Expression Recognition 2013 (FER2013) dataset [28], which was provided by Kaggle at the International Conference on Machine Learning (ICML). The FER-2013 dataset contains a set of 35,887 grayscale face images, all standardized to 48×48 pixels. Each image is annotated with one of seven emotion categories: Angry, Disgust, Fear, Happy, Sad, Surprise, or Neutral (Fig. 6). These emotions cover a wide spectrum of affective states that are essential for examining human reactions in interview situations, where nonverbal clues such as facial expressions offer crucial information about applicants' emotional reactions, degrees of engagement, and behavioral indicators.

Fig 6. FER2013 Seven Emotions [28].

The dataset is divided into training, validation, and test sets, facilitating Model development and performance evaluation across different stages. There are 28,709 images in the training set and 3,589 images in the test set. In the training set, there are 4,953 images of anger, 547 images of disgust, 5,121 images of fear, 8,989 images of happiness, 6,077 images of sadness, 4,002 images of surprise, and 6,198 images of neutrality. This imbalance presents an additional challenge in training models for perform accurately across all classes. The dataset's low resolution and grayscale format add further complexity, as models must learn to identify subtle emotional cues from limited visual data.

DATASET PREPROCESSING AND AUGMENTATION

To fit the interview context, data cleaning was essential. Label inaccuracies, particularly in the "happy" class, where images were mislabeled as "angry," were identified, and

http://dx.doi.org/10.21622/ACE.2025.05.1.1317

removed. Additionally, emotions such as fear, surprise, and disgust were excluded, focusing on four expressions: angry, sad, happy, and neutral (Fig. 7). After cleaning, the dataset showed an imbalance in emotional state distribution, with "happy" having the most instances, followed by "neutral" and "sad," while "angry" had the least (Fig. 8). This imbalance could bias the model towards recognizing dominant emotions like happiness and neutrality, potentially affecting its ability to accurately identify less common emotions like anger. Addressing this imbalance is crucial for the model's performance in real interview contexts.

The train data generator is an instance of ImageDataGenerator configured to augment image data during the training phase of our machine learning model. This configuration includes several augmentation techniques aimed at enhancing the model's ability to generalize from the training data. The images are first rescaled to a range of 0 to 1 to normalize the pixel values. During training, the generator randomly applies a rotation of up to 15 degrees, zooms images by up to 10%, performs horizontal flips, and shifts images horizontally and vertically by up to 5% of the total width and height, respectively, using a wrapping fill mode to handle gaps. These augmentations not only help expose the model to a wider variety of image variations, reducing overfitting and improving its ability to learn robust features from the data, but also effectively address the initial class imbalance by generating synthetic samples for under-represented classes like "angry." This approach ensures a more balanced distribution of training instances across different emotional states, thereby enhancing the model's overall performance and generalization capability.

Fig 7. Cleaned FER2013.

http://dx.doi.org/10.21622/ACE.2025.05.1.1317

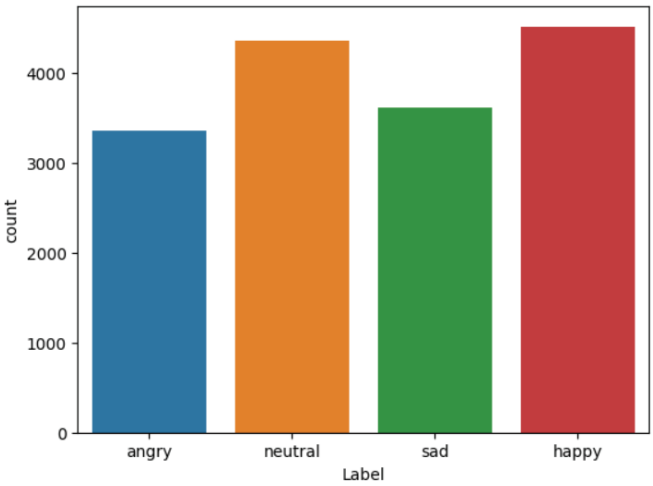

Fig 8. The cleaned FER2013 class distribution.

Fig 9. Examples of the augmented dataset.

5. EXPERIMENTAL EVALUATION

The optimization of the neural network architecture for the FER2013 dataset involved a systematic exploration of hyperparameters and training configurations to maximize model performance. Initial experimental iterations employed traditional optimization algorithms, including Stochastic Gradient Descent (SGD), Adagrad, and RMSprop, coupled with relatively shallow dense layers and limited dropout regularization. These baseline models yielded moderate accuracies, ranging from 52% to 62%, highlighting the necessity for further architectural and training refinements. Subsequently, the focus

http://dx.doi.org/10.21622/ACE.2025.05.1.1317

shifted towards the Adam optimizer due to its adaptive learning rate capabilities and superior convergence properties, which are well-documented in deep learning literature. Progressive adjustments were made by increasing the dense layer capacity to 1024 neurons, incorporating dropout layers with rates of 0.25 and 0.5 to mitigate overfitting and expand the convolutional stack to include four layers with filter sizes of (32, 64, 128, 128), all utilizing ReLU activation functions to introduce non-linearity and improve gradient flow. The training was conducted with a batch size of 64 over 100 epochs, a configuration empirically determined to balance convergence speed and generalization.

This rigorous tuning process culminated in the final model, which achieved a peak accuracy of 77%, representing a significant improvement over the earlier attempts. Additional post-final experiments, which explored alternative activation functions such as LeakyReLU and alternative optimizers like RMSprop at comparable training lengths failed to surpass the established performance threshold, thereby validating the robustness and optimality of the selected hyperparameter set. This comprehensive experimentation underscores the critical role of optimizer selection, network depth, dropout regularization, and training regimen in enhancing the predictive

Table 2. Progressive Model Tuning Experiments for FER2013 Dataset: Optimizer, Layers, and Hyperparameters.

Edition	Neurons	Dropout	Conv Layers (Filter Size)	Training Params	Accuracy
Exp 1	Dense: 64	None	(32), (3,3)	SGD, 64 batch, 10 epochs, ReLU	52%
Exp 2	Dense: 128	0.1	(32, 64), (3,3)	SGD, 64 batch, 20 epochs, ReLU	57%
Exp 3	Dense: 256	0.25	(32, 64), (3,3)	Adagrad, 64 batch, 30 epochs, ReLU	60%
Exp 4	Dense: 256	0.25	(32, 64), (3,3)	RMSprop, 64 batch, 40 epochs, ReLU	62%
Exp 5	Dense: 256	0.25	(32, 64), (3,3)	RMSprop, 32 batch, 60 epochs, Tanh	61%
Initial Adam Model	Dense: 256	None	(32, 64), (3,3)	Adam, 64 batch, 20 epochs, ReLU	65%
Edition 1	Dense: 512	0.25	(32, 64), (3,3)	Adam, 64 batch, 40 epochs, ReLU	68%
Edition 2	Dense: 512	0.25, 0.5	(32, 64, 128), (3,3)	Adam, 64 batch, 60 epochs, ReLU	71%
Edition 3	Dense: 1024	0.25	(32, 64, 128), (3,3)	Adam, 32 batch, 80 epochs, ReLU	73%
Edition 4	Dense: 1024	0.25, 0.5	(32, 64, 128), (3,3)	Adam, 64 batch, 90 epochs, ReLU	75%
Final Model	**Dense: 1024**	**0.25, 0.5**	**(32, 64, 128, 128), (3,3)**	**Adam, 64 batch, 100 epochs, ReLU**	**77%**
Post-Exp 1	Dense: 1024	0.25, 0.5	(32, 64, 128, 128), (3,3)	RMSprop, 64 batch, 100 epochs, ReLU	75%
Post-Exp 2	Dense: 1024	0.25, 0.5	(32, 64, 128, 128), (3,3)	Adam, 64 batch, 100 epochs, LeakyReLU	74%

Accuracy of deep neural networks on the FER2013 dataset.

http://dx.doi.org/10.21622/ACE.2025.05.1.1317

6. RESULTS AND DISCUSSION

To evaluate the effectiveness and robustness of our face emotion detection model, we conducted an extensive series of simulations utilizing the FER-2013 dataset. The FER-2013 dataset is well-regarded in the field for its diverse range of facial expressions and comprehensive annotations, making it an ideal benchmark for assessing Model performance. Our simulations were designed to rigorously test the model's ability to accurately identify and classify a wide array of facial emotions across various conditions, including different lighting, angles, and occlusions. Through these simulations, we aimed to visualize and analyze the model's predictions in comparison to the actual ground truth labels. This process not only helps in quantifying the model's accuracy but also in identifying specific instances where the Model excels or struggles. By showcasing these visualizations, we provide a detailed insight into the operational dynamics of our face emotion detection system.

Fig. 10 illustrates sample results from our simulations, demonstrating the model's capability to interpret and categorize facial emotions. Each image is accompanied by the predicted emotion label and the corresponding ground truth, allowing for a clear comparison. These visualizations highlight the practical applicability of our model i real-world scenarios, underscoring its potential for integration into various applications such as AI-powered interviews, mental health assessments, and human-computer interaction systems.

In addition to evaluating the model's accuracy, these simulations also shed light on its limitations and areas for improvement. For instance, certain emotions may be more than challenging to detect under low-light conditions or when the face is partially obscured.

Understanding these limitations is crucial for guiding future enhancements and ensuring that the model performs reliably across all intended use cases. Our model achieved a validation accuracy of 77%, a validation loss of 0.6, a validation precision of 0.82 and a validation recall of 0.71. Table 3 shows that our the model outperforms the other state-of-art approaches.

Table 3. Summary of the state-of-the-art approaches and their performance as compared to the proposed model on the FER2013 dataset.

Study	Year	Method	Accuracy	Application
[20]	2014	Local Learning BOW	67.48%	Competition
[21]	2018	Local Learning Deep+BOW	75.42%	Mental Disorder Detection
[22]	2020	Ensemble ResMask- ingNet	76%	Boost the performance of CNNs
[23]	2020	CNN	70.14%	Enhance Human Computer Interaction
[24]	2022	VGG16	58.6%	Improve Quality of Images in FER2013
[27]	2023	N/A	61.88%	N/A
[25]	2023	Custom Architecture	66.6%	N/A
[29]	2024	Deep CNN	65.68%	VR, Robotics, Marketing and Mental Health
Ours	**2025**	**Advanced CNN**	**77%**	**AI Recruitment**

http://dx.doi.org/10.21622/ACE.2025.05.1.1317

While validation accuracy provides a general measure of the model's performance, precision and recall offer more nuanced insights into its behavior. Precision, the ratio of correctly predicted positive observations to the total predicted positives is crucial in scenarios where the cost of false positives is high. Recall that the ratio of correctly predicted positive observations to all actual positives is essential in situations where it is important to capture as many true positives as possible. These metrics help us understand the trade-offs between different types of errors, guiding us to optimize the model for its intended applications.

Hence, to provide a balanced measure of the model's performance, especially when dealing with imbalanced datasets or when the costs of false positives and false negatives

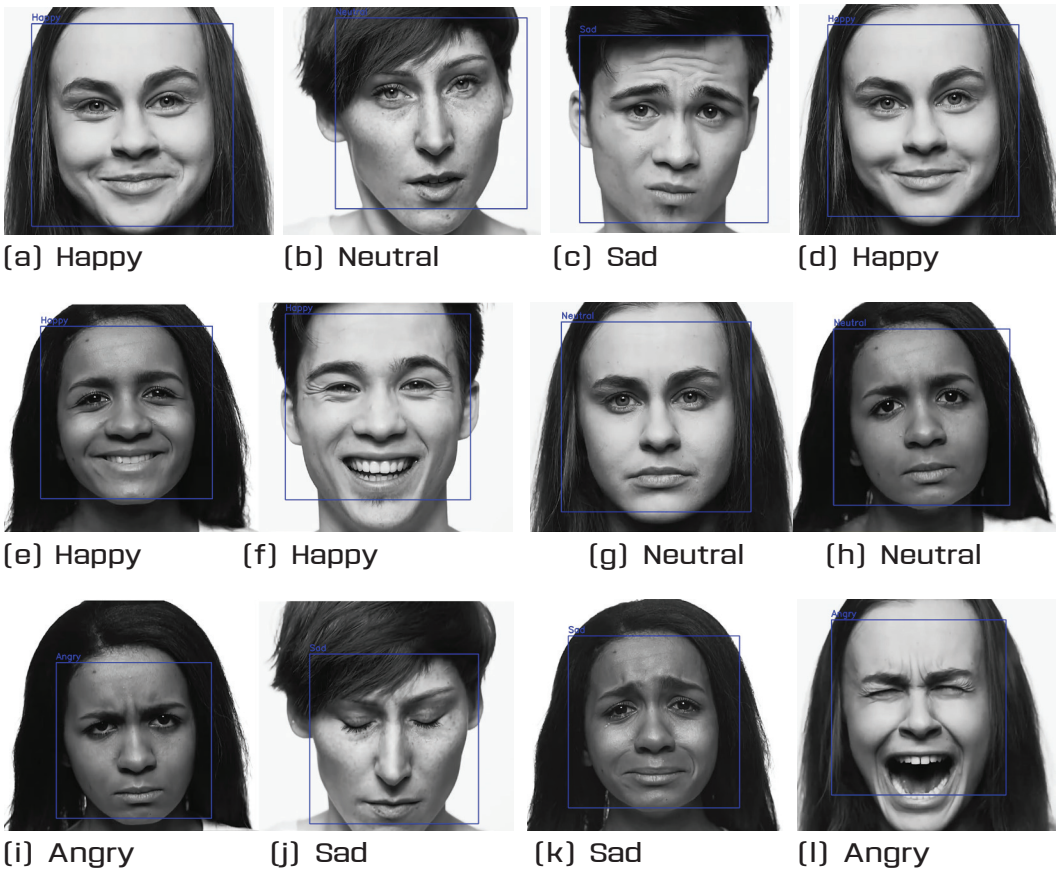

Fig 10. Face Emotion model examples.

Are different, we compute the F1 score. The F1 score is the harmonic mean of precision and recall, calculated as

$$F1 = \frac{2 \cdot \text{precision} \cdot \text{recall}}{\text{precision} + \text{recall}} = 0.761$$

The F1 score balances precision and recall, offering a single metric that accounts for both types of errors. It is especially important in applications like emotion detection, where both false positives and false negatives can have significant implications. For instance, incorrectly identifying a negative emotion when there is none (false positive) can cause unnecessary concern while missing a genuine negative emotion (false negative) can overlook someone in need of help. Thus, the F1 score provides a reliable metric for assessing the overall effectiveness of our face emotion detection model.

http://dx.doi.org/10.21622/ACE.2025.05.1.1317

IMPLEMENTATION EXAMPLE ON SCORING METHODOLOGY

For a session with eight negative emotions (Face), 12 filler words (Speech), and three negative emotions (Text), the score is computed as follows (with respect to Table 1):

$$S = \left\lfloor \frac{8}{4} \right\rfloor + \left\lfloor \frac{12}{6} \right\rfloor + \left\lfloor \frac{3}{2} \right\rfloor = 2 + 2 + 1 = 5$$

Generally, higher scores reflect cumulative performance issues. The main limitation is that it assumes uniform impact of triggers; calibration may vary by context, where other advanced scoring methodologies will be addressed in the future.

7. CONCLUSION

This paper presents InstaJob, an AI-powered framework that enhances fairness and efficiency in the recruitment process by integrating facial emotion detection, speech analysis, and AI-based technical assessments. The system utilizes datasets such as FER-2013 and GoEmotions to power its models and offer a scalable solution for conducting initial candidate evaluations.

While the platform demonstrates promising results, several limitations must be acknowledged. The variability in video and audio quality can affect the accuracy of the models, and there is a risk of biased outcomes due to cultural differences not fully represented in the training data. Additionally, the virtual interview format may not be perfectly replicate in-person interactions, potentially impacting candidate behavior and model performance.

Future work will focus on improving model generalizability, expanding dataset diversity to reduce cultural and linguistic bias, and testing the platform further in real-world scenarios across various industries. It is also essential to address the ethic considerations of using AI in recruitment by ensuring transparency, fairness, and inclusivity in the decision-making process.

REFERENCES

[1] D. Jekauc et al., "Recognizing affective states from the expressive behavior of tennis players using convolutional neural networks," Knowl Based Syst, vol. 295, p. 111856, Jul. 2024, doi: 10.1016/j.knosys.2024.111856.

[2] S. K. Khare, V. Blanes-Vidal, E. S. Nadimi, and U. R. Acharya, "Emotion recognition and artificial intelligence: A systematic review (2014–2023) and research recommendations," Information Fusion, vol. 102, p. 102019, Feb. 2024, doi: 10.1016/j.inffus.2023.102019.

[3] C. Dando, D. A. Taylor, A. Caso, Z. Nahouli, and C. Adam, "Interviewing in virtual environments: Towards understanding the impact of rapport-building behaviours and retrieval context on eyewitness memory," Mem Cognit, vol. 51, no. 2, pp. 404–421, Feb. 2023, doi: 10.3758/s13421-022-01362-7.

[4] V. Coleman et al., "Lessons Learned From Conducting Virtual Multiple Mini Interviews During the COVID-19 Pandemic," The Journal of Physician Assistant Education, vol. 35, no. 3, pp. 287–292, Sep. 2024, doi: 10.1097/JPA.0000000000000606.

[5] "Artificial Intelligence Insights & Articles | QuantumBlack | McKinsey & Company," 2023. [Online]. Available: https://www.mckinsey.com/capabilities/quantumblack/our-insights/

[6] S. Yadav and S. Kapoor, "RETRACTED ARTICLE: Adopting artificial intelligence (AI) for employee recruitment: the influence of contextual factors," International Journal of System Assurance Engineering and Management, vol. 15, no. 5, pp. 1828–1840, May

2024, doi: 10.1007/s13198-023-02163-0.

[7] C. Fernández-Martínez and A. Fernández, "AI and recruiting software: Ethical and legal implications," Paladyn, vol. 11, no. 1, pp. 199–216, May 2020, doi: 10.1515/pjbr-2020-0030.

[8] E.-R. Lukacik, J. S. Bourdage, and N. Roulin, "Into the void: A conceptual model and research agenda for the design and use of asynchronous video interviews," Human Resource Management Review, vol. 32, no. 1, p. 100789, Mar. 2022, doi: 10.1016/j.hrmr.2020.100789.

[9] HireVue, "About the Company | Leadership & CEO | HireVue." [Online]. Available: https://www.hirevue.com/about

[10] A. Vinciarelli et al., "Bridging the Gap between Social Animal and Unsocial Machine: A Survey of Social Signal Processing," IEEE Trans Affect Comput, vol. 3, no. 1, pp. 69–87, Jan. 2012, doi: 10.1109/T-AFFC.2011.27.

[11] C. Signore, B. Della Piana, and F. Di Vincenzo, "Digital Job Searching and Recruitment Platforms: A Semi-systematic Literature Review," 2023, pp. 313–322. doi: 10.1007/978-3-031-42134-1_31.

[12] A. Fabris et al., "Fairness and Bias in Algorithmic Hiring: A Multidisciplinary Survey," ACM Trans Intell Syst Technol, vol. 16, no. 1, pp. 1–54, Feb. 2025, doi: 10.1145/3696457.

[13] C. Qin, L. Zhang, R. Zha, and D. Shen, "A Comprehensive Survey of Artificial Intelligence Techniques for Talent Analytics," arXiv preprint , 2023, doi: http://dx.doi.org/10.48550/arXiv.2307.03195.

[14] E. T. Albert, "AI in talent acquisition: a review of AI-applications used in recruitment and selection," Strategic HR Review, vol. 18, no. 5, pp. 215–221, Oct. 2019, doi: 10.1108/SHR-04-2019-0024.

[15] M. Kathiravan, M. Madhurani, S. Kalyan, R. Raj, and S. Jayan, "A modern online interview platform for recruitment system," Mater Today Proc, vol. 80, pp. 3022–3027, 2023, doi: 10.1016/j.matpr.2021.06.459.

[16] I. Naim, Md. I. Tanveer, D. Gildea, and M. E. Hoque, "Automated Analysis and Prediction of Job Interview Performance," IEEE Trans Affect Comput, vol. 9, no. 2, pp. 191–204, Apr. 2018, doi: 10.1109/TAFFC.2016.2614299.

[17] S. Mhadgut, N. Koppikar, N. Chouhan, P. Dharadhar, and P. Mehta, "vRecruit: An Automated Smart Recruitment Webapp using Machine Learning," in 2022 International Conference on Innovative Trends in Information Technology (ICITIIT), IEEE, Feb. 2022, pp. 1–6. doi: 10.1109/ICITIIT54346.2022.9744135.

[18] A. K. A, A. H, N. P. Nair, V. A, and A. T, "Interview Performance Analysis using Emotion Detection," in 2022 4th International Conference on Inventive Research in Computing Applications (ICIRCA), IEEE, Sep. 2022, pp. 1424–1427. doi: 10.1109/ICIRCA54612.2022.9985667.

[19] I. J. Goodfellow et al., "Challenges in Representation Learning: A Report on Three Machine Learning Contests," 2013, pp. 117–124. doi: 10.1007/978-3-642-42051-1_16.

[20] I. J. Goodfellow et al., "Challenges in Representation Learning: A report on three machine learning contests," arXiv:1307.0414 [cs, stat], May 2013, [Online]. Available: https://arxiv.org/abs/1307.0414

[21] M.-I. Georgescu, R. T. Ionescu, and M. Popescu, "Local Learning With Deep and Handcrafted Features for Facial Expression Recognition," IEEE Access, vol. 7, pp. 64827–64836, 2019, doi: 10.1109/ACCESS.2019.2917266.

[22] L. Pham, T. H. Vu, and T. A. Tran, "Facial Expression Recognition Using Residual Masking Network," in 2020 25th International Conference on Pattern Recognition (ICPR), IEEE, Jan. 2021, pp. 4513–4519. doi: 10.1109/ICPR48806.2021.9411919.

[23] A. Jaiswal, A. Krishnama Raju, and S. Deb, "Facial Emotion Detection Using Deep Learning," in 2020 International Conference for Emerging Technology (INCET), IEEE, Jun. 2020, pp. 1–5. doi: 10.1109/INCET49848.2020.9154121.

[24] J. Luo, Z. Xie, F. Zhu, and X. Zhu, "Facial Expression Recognition using Machine Learning models in FER2013," in 2021 IEEE 3rd International Conference on Frontiers Technology

http://dx.doi.org/10.21622/ACE.2025.05.1.1317

of Information and Computer (ICFTIC), IEEE, Nov. 2021, pp. 231–235. doi: 10.1109/ICFTIC54370.2021.9647334.

[25] S. Hassan, M. Ullah, A. S. Imran, and F. A. Cheikh, "Attention-Guided Self-supervised Framework for Facial Emotion Recognition," 2024, pp. 294–306. doi: 10.1007/978-981-99-7025-4_26.

[26] Y. Wu, L. Zhang, Z. Gu, H. Lu, and S. Wan, "Edge-AI-Driven Framework with Efficient Mobile Network Design for Facial Expression Recognition," ACM Transactions on Embedded Computing Systems, vol. 22, no. 3, pp. 1–17, May 2023, doi: 10.1145/3587038.

[27] Y. Mao, "Optimization of Facial Expression Recgonition on ResNet-18 using Focal Loss and CosFace Loss," in 2022 International Symposium on Advances in Informatics, Electronics and Education (ISAIEE), IEEE, Dec. 2022, pp. 161–163. doi: 10.1109/ISAIEE57420.2022.00041.

[28] Dumitru, I. Goodfellow, W. Cukierski, and Y. Bengio, "Challenges in Representation Learning: Facial Expression Recognition Challenge," 2013. [Online]. Available: https://www.kaggle.com/c/challenges-in-representation-learning-facial-expression-recognition-challenge

[29] D. Bhagat, A. Vakil, R. K. Gupta, and A. Kumar, "Facial Emotion Recognition (FER) using Convolutional Neural Network (CNN)," Procedia Comput Sci, vol. 235, pp. 2079–2089, 2024, doi: 10.1016/j.procs.2024.04.197.