International Journal of Innovative Research in Computer Science and Technology (IJIRCST)
ISSN (Online): 2347-5552, Volume-13, Issue-5, September 2025
DOI: https:/doi.org/10.55524/ijircst.2025.13.5.6
Article ID IRP-1682, Pages 38-43
www.ijircst.org

Emotion and Confidence Classifier for Mock Interviews Using Artificial Intelligence

Shalini Bhaskar Bajaj

Professor, Department of Computer Science and Technology, Amity University, Manser, Haryana, India

Correspondence should be addressed to Shalini Bhaskar Bajaj; shalinivimal@gmail.com

Received 27 August 2025; Revised 10 September 2025; Accepted 23 September 2025

ABSTRACT- As a steppingstone towards a candidate's career, a job interview is perhaps the most important occasion in which their distinct set of skills meets an expected opportunity. They are perhaps one of the most essential elements in educational system and corporate world, which when used properly can yield competent candidates to hire, and are based on testing of their skills. Mock interviews enhance communication skills, as well as confidence, which translates into improved interview performance. We present a new simulation system for practicing interviews, powered by artificial intelligence, designed to narrow the preparation-performance gap. Our system measures performance in two critical areas: empathy and confidence. For evaluating emotions, we apply sophisticated approaches to deep learning, particularly classifying facial expressions into seven basic emotions using a convolutional neural network (CNN). In analyzing confidence, we employ voice recognition and natural language processing (NLP) and LSTM (Long Short-Term Memory) models to process and identify speech correctly. The system helps attendees overcome stressful pre-interview situations, improves their self-perception and self-efficacy, and prepares them for real-life interviews.

KEYWORDS- Artificial Intelligence, Deep Learning, Convolutional Neural Network (CNN), Voice Recognition, Natural Language Processing (NLP)

I. INTRODUCTION

Artificial Intelligence (AI) has transformed various fields such as healthcare, finance, and education. AI in recruiting offers creative methods of evaluating applicants, improving the conventional hiring procedure. As regarded in the recruitment process, interviews are both defining moments in an individual's professional journey and a deciding factor for organizations seeking the most suitable candidates. For candidates, interviews are opportunities to showcase their skills, but the stress that comes with interviews tends to hamper their actual performance. The long-standing practice of preparing for interviews has been to conduct dummy sessions. However, the recent inclusion of artificial intelligence in interview preparation and practice has transformed the entire ecosystem by offering innovative data-driven solutions. With the application of artificial intelligence technologies, interview preparation has transformed remarkably. Inadequate interviewing techniques included rehearsing with peers, advisors, or mentors. Although these approaches aid in 'basic' preparation, they have systemic shortcomings which limit their widespread use such as lacking standardization, scalability, and most importantly providing objective feedback. AI-driven technologies that provide systematic and structured preparation are revolutionizing students' ability to configure their work-readiness. These platforms allow comprehensive analysis of interviews from different sectors and encompass virtually all scenarios and industry contexts. Traditional practice methods are enhanced using these AI technologies which transform objective review methodologies into vivid, dynamic learning experiences.

The goal of this concept is to leverage AI in identifying suitable candidates by analyzing facial and vocal cues. For example, Su Yusheng et al. research from 2021, presented a real-time AI agent that uses facial analysis and support vector machines (SVM) and convolutional neural networks (CNN) to forecast job applicants' potential behaviour. Likewise, Hung-Yue Suen and associates (2020) suggested an intelligent decision-making system for asynchronous video interviews that can forecast behaviour and communication from video recordings. Even with these improvements, there are still issues with current systems, such as time consumption, the requirement for manual data entry, and challenges when assessing behaviour across various people.

In order to address these problems, a smart conversational system that combines voice and facial analysis will be created in order to assess and comprehend candidates' personalities more thoroughly. Combining deep learning models, data processing methods, and advanced data communication will further improve the precision and effectiveness of candidate assessments. The ultimate goal is to use AI to transform the interview process and develop a more impartial, dependable, and efficient system.

II. LITERATURE REVIEW

Artificial intelligence these days is playing a vital role in the interview process and has revolutionized the mock interview process wherein both the candidate and the recruiters can identify the best fit candidate for the job roles being published by the recruiters. The artificial intelligence software can read the facial expressions of the candidates during the interview process using deep learning algorithms. Convolutional neural networks (CNNs) can

identify emotional expressions from the images and are quite good at translating them into hidden feelings. There are several datasets available such as FER-2013, CK+, JAFFE which the used by the researchers to train their deep leaning model on various emotions like anger, sadness, fear, surprise and many more, thus, enhances the interpretation of non-verbal clues that plays very important role during the interview process.

During the interview process, it's not only the video or image analysis is considered by the recruiters but also analysis of the audio is becoming increasingly popular to understand the psychological and emotional characteristics. This is done by analyzing the speech. Prosodic feature-based techniques such as speech rate, change of pitch, vocal intensity are analyzed for evaluating the confidence of the speaker. In order to extract useful features from the audio signal, tools such as Pydub or Librosa are being used. Semantic and syntactic characteristics of the audio signal are tested using natural language processing (NLP) to check the appropriateness of the audio content. During interviews, NLP is employed to check the grammatical accuracy of the audio content, structure of the spoken sentence and much more. These days platforms like HireVue, MyInterview and other such platforms are available for automated interview evaluation. These platforms use machine learning algorithms to examine video and audio answers, sometimes even delivering ratings for communication, personality, and possible fit. Criticisms have been raised on account of their black-boxed scoring systems, possible model bias during training, and the absence of in-depth feedback for candidates. Moreover, many existing systems do not integrate multi-modal inputs effectively, leaving a gap in providing holistic evaluations.

Scholarly research in the field continues to delve into more transparent, ethical, and flexible systems. Certain research has identified the need for multi-modal fusion methods, where audio, visual, and text data are fused together to enhance prediction accuracy and stability. Others aim to develop interpretable AI systems that offer human-readable explanations of scores or classifications, promoting higher levels of user trust.

In spite of these advancements, there is a notable deficiency in systems providing real-time, transparent, and structured interview assessment incorporating emotional expression, speech confidence, and content correctness. The system under proposal works towards fulfilling all these requirements by utilizing CNN for detecting facial expression, speech and prosodic feature analysis through Pydub, and content assessment using NLP on the semantic, syntactic, and factual basis. In contrast to most other tools, this platform includes a detailed and balanced scoring mechanism that assigns a fair weight to each dimension, thus providing candidates with tailored insights and individual areas of improvement.

Given the rapid progress with AI technologies, there are even greater possibilities for using interview preparation systems and tools. Technology can be considered a major advancement in equalizing access to quality interview training while providing a platform to help candidates learn the skills needed to feel comfortable by inviting an interviewer into their lived experience while obtaining fact-based, and complete feedback on their interview skills. Because the system adapts to the needs of the individual while conforming to standardized rating frameworks, it can be a powerful tool in the professional development of job seekers and bridge the gaps between preparation and interview performance. Table 1 tabutales comparison between traditional Methods and AI-Powered Methods.

Table 1: Comparison between traditional Methods and AI-Powered Methods

ASPECT	TRADITIONAL METHODS	AI-POWERED METHODS
Feedback consistency	Varies based on human judgement may be subjective and inconsistent	Provides standardized, objective, and data-driven feedback
scalability	Limited to the availability of mentors, peers, or career advisors.	Highly scalable, accessible to unlimited users simultaneously
Emotional Analysis	Relies on subjective interpretation by interviewers.	Uses deep learning (CNN) for accurate facial emotion detection.
Confidence Evaluation	Subjective observation or self-assessment.	Uses speech recognition, NLP, and audio processing for confidence assessment.
Industry-Specific Preparation	Limited access to industry-specific interviewers.	AI can simulate diverse industry-based interview scenarios.
Adaptability	Fixed interview structure, does not adapt to individual progress.	Dynamically adjusts difficulty and questions based on performance.

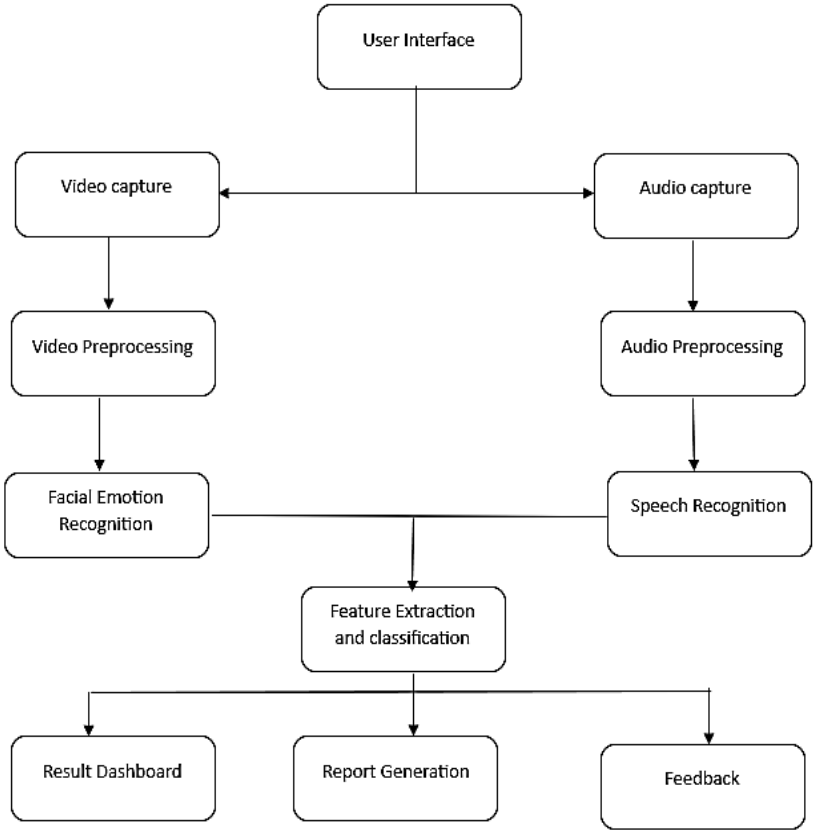

Figure 1: Proposed Model

III. PROPOSED MODEL

The planned emotion analysis model from video and audio data prescribes a powerful and comprehensive system with multiple connected components. It starts with the user interacting through a specific user interface, using which video or audio inputs are received. Such inputs are then preprocessed—video data are processed with tasks such as resizing, frame separation, and removal of noise, whereas audio data are processed through noise filtering and feature extraction. Deep learning algorithms that are advanced in nature are being employed to process the inputs. These algorithms extract features such as facial landmarks and audio spectrograms to identify different emotional states. These features are fed to the classification module that can effectively estimate the underlying emotions. It uses metrics like recall, precision and F1-score to measure the performance of the model. Finally, the user is shown the output on the dashboard in the form of a detailed report. The system also gives immediate feedback and performance information to support ongoing improvement and better interpretation of user feelings. Figure 1 gives flowchart for the proposed model.

IV. ALGORITHMS

The AI-based mock interview platform is structured around two core analytical modules—emotion analysis and confidence evaluation—each utilizing specialized algorithms .

CNNs are a type of deep learning model known for processing grid-like data. The proposed AI-powered mock interview platform thus uses two important modules, namely, emotional analysis and the other one is confidence evaluation of the candidate. Each module is implemented using specialized algorithm. The modules are designed to give suggestions to the candidate to improve their interview performance.

A. Convolutional Neural Network (CNNs) for Facial Recognition

CNNs specialize in image processing because it can identify complex features via layer extraction. Images are first taken based on the candidate's face during their interview, capturing images of the candidates throughout their interview. The images are then input into layer components of the CNN, which applies a filter to each image to detect relatively basic features of imagery such as edges, shape, and parts of the facial features (eyes, mouth, nose, etc.). As the data is processed layer by layer, CNN (Figure 2) learns to detect more complex features, such as facial expression and associated emotion (e.g., happy, angry, sad). Each convolutional operation is followed by an activation function (such as ReLU) to introduce non-linearity into the model, which allows for the capturing of more complex relationships between pixels during the creation of the image. After processing the images with convolutional operations, pooling is implemented (generally max pooling) to reduce the spatial dimension but retains key features of the imagery. In final layers, the CNN combines the feature information through fully connected layers, which delivers to classify the image into one of the emotion categories using the softmax activation function. The CNN will learn to optimize its internal parameters with the goal of reducing classification errors

via backpropagation during training. Subsequently, CNN can process new, real-time video frames to classify the candidate's emotion by analyzing their facial expressions and predicting one of the seven categorically emotion states.

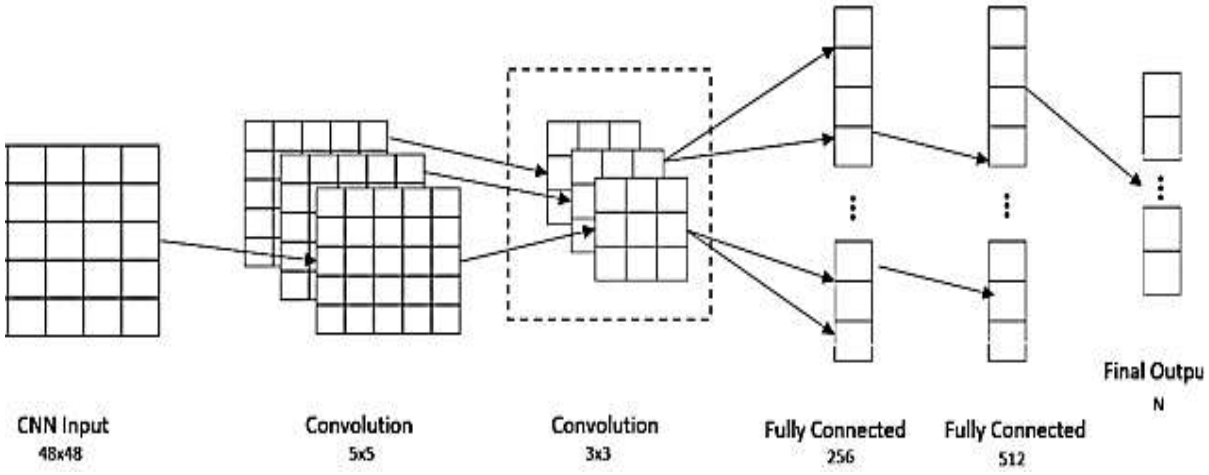

Figure 2: Working of CNN Algorithm

B. Natural Language Preprocessing (NLP) for Speech Recognition

Natural Language Preprocessing (NLP) is a field of artificial Intelligence that focusses on the interaction between computers and human language. NLP tools, supports analysis of both vocal and verbal characteristics. Within the arena of speech recognition, NLP methods are utilized to convert spoken language to text. In order to properly transcript a language, NLP is used as it can provide the context off the language as well as comprehension. Thus, we can make speech recognition system more robust, flexible to different styles of speaking or different dialects, more accurate, if we could harness the capabilities of NPL algorithms.

C. Long Short-Term Memory (LSTM)

To identify and process speech correctly and thus for enhancing the capabilities of the system, Long Short-Term Memory algorithms can be deployed. Speech data is, by nature, sequential in the sense that one word or sound relies on other words before it becomes meaningful. LSTM is a subclass of recurrent neural network (RNN) with specialized skills to handle such sequential data based on learning long-term dependencies. The most outstanding aspect of it currently exists. structure, whereby it can maintain significant information in the long term while eliminating useless data through its input, forget, and output gates. This process helps LSTM to cope with the temporal aspect of speech efficiently, such that past information is considered when processing present input. Speech input usually also differs in length, and LSTM is appropriate for variable-length sequences since it can handle input whose size is not fixed. By using its property of remembering context over long sequences, LSTM assists in enhancing the recognition of longer sentences and complex speech patterns. It results in more accurate output in the entire system than in traditional models, which can get stuck with holding context in longer sequences. Thus, the addition of LSTM in your speech recognition system greatly improves its performance so it can interpret and recognize speech in everyday situations better.

V. METHODOLOGY

The AI-powered interview simulation system conceived in this paper is intended to evaluate candidate performance using multimodal input analysis through facial expression identification, speech understanding, and knowledge-based analysis. The methodology comprises several interconnected stages to ensure a comprehensive and objective assessment process.

A. Data Acquisition

The system records real-time audio-visual information from subjects in simulated interview contexts using webcam and microphone interfaces. Video streams are processed to recognize facial emotions, and audio inputs are analysed for speech analysis and confidence assessment. This two-channel data acquisition supports concurrent evaluation of verbal and non-verbal cues.

B. Emotion Analysis

CNN trained on facial expression datasets can be effectively used for the recognition of different facial expressions and the meaning associated with them. The model can be trained on seven different emotions such as fear, anger, sadness, happiness, feeling of surprise, feeling of disgust, neutral (no emotion). The process of training involves giving a score to each emotion from 1 to 7. This is done to standardize the results. The final score can be computed by giving weightage of 20% to the emotional expression, thus highlighting that expressions are important in any communication.

C. Analysis of Speech

NLP can be used to analyze the response of the candidate. The weightage given to this component is again 20%. Features like pronunciation clarity, speech rate, use of filler

words etc. can be used to rate the final score of the candidate. The model thus highlights the importance of speech articulation, verbal fluency, coherence etc.

D. Confidence Assessment

Confidence assessment is performed by examining prosodic characteristics of the audio input, including pitch variability, voice volume, and the rate of hesitations. Pydub library is utilized for streamlined audio segmentation and preprocessing. Confidence is measured in percentage form, with the Confidence Score accounting for 10% of the overall performance assessment. This light-weight scoring is meant to reflect confidence importance without significantly overlapping with speech and emotional analysis.

E. Knowledge Base Assessment

The content of candidate responses is critically assessed through NLP models on three essential dimensions: Semantic Correctness (10%) – assessing the logical appropriateness of the response. Syntax Correctness (10%) – analyzing grammatical structure and sentence form. Answer Correctness (30%) – analyzing the factual veracity and completeness of the response. Combined, the subcomponents constitute the Knowledge Base Score with 50% of the overall score, emphasizing the highest significance of domain knowledge and answer quality in interviews.

F. Scoring Framework

The final score can be obtained by providing weighted sum of all the components that include, emotion score (20%), speech score (20%), confidence score (10%), knowledge score (50%), knowledge score being the highest and the most useful is given weightage of 50 percent.

Weightage allocation is the representation of actual-world significance of every dimension in interview performance. Knowledge is of the highest priority, then speech and emotional engagement (communication skills), and then individual confidence. This systematic, balanced scoring scheme provides an objective, holistic, and actionable assessment framework, thus improving the process of preparation for actual-world interviews.

VI. RESULTS

The model harnessing AI for emotion analysis and confidence assessment enables multimodal input instances of video, audio, and textual data to provide a comprehensive representation of emotional state and confidence. The emotion analysis mechanism, which uses a CNN, has provided highest success in classifying different emotions as listed in the above section that include the emotion of surprise, anger, sadness, happiness, feeling of disgust, feeling of being fearful, neutral (no emotion). The model could detect positive emotions more efficiently than negative emotions. The confidence assessment mechanism employs speech recognition, techniques using NLP, and audio processing to obtain verbal and non-verbal assessments, also identifying markers of confidence levels such as pitch variability, volume consistency and speech rate. Visualizations, such as pie charts and bar charts, provided a visual understanding of the relationship between emotional states and confidence, in addition to supplementing prior research and categorization demonstrating that positive emotional states, such as joy and love, were associated with greater confidence, while negative emotional states of sadness and fear were associated with lower confidence. The analysis nuanced findings regarding emotional states such as anger where confidence was mixed based on the podcaster and perspective.

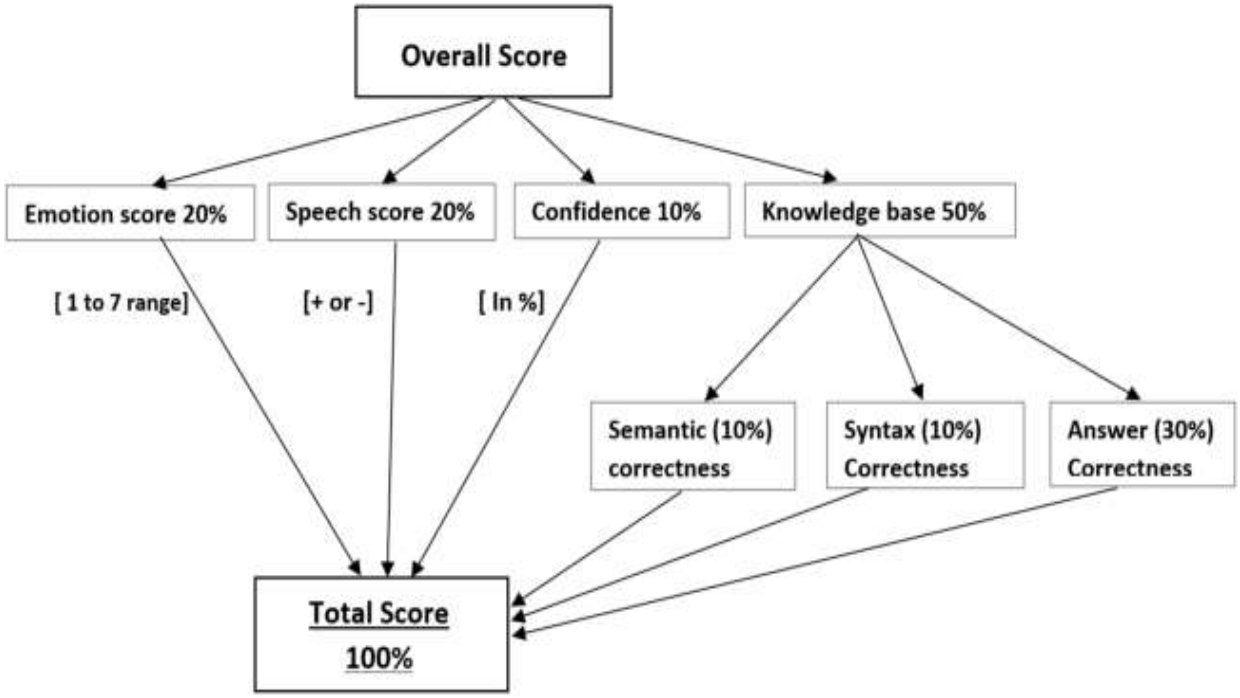

Figure 3: Scoring Module

Figure 3 gives a systematic assessment scheme for assessing the performance of candidates during the interview process. The interview outcome is judged on four different parameters such as emotion, confidence, speech, and knowledge. The emotion score (20%) is rated on a scale of 1 to 7 and likely considers facial expressions and engagement. The speech score (20%) assesses fluency and clarity in speech while also making plus or minus adjustments based on pronunciation and articulation in speech. Confidence (10%) would be assigned as a percentage and would likely be determined based on modulation of voice, stability of tone, and body language analysis of a candidate. The knowledge base (50%), which is the most important component to evaluate, is made up of three dimensions, which account for a subcategory of aspects within a candidate's interview performance: (1) semantic correctness (10%), which would assess logical consistency; (2) syntax correctness (10%), which would assess grammar usage; and (3) answer correctness (30%). The answer correctness dimension is focused on whether factual statements are factually correct and relevant to the question. To have a standardized, data driven evaluation of the candidates, scores recorded on the above four parameters are combined based on the percentage specified. The above-mentioned mechanism needs to be followed in a systematic manner so that the proper evaluation of the candidates can be done. The proposed system provides an additional level of objectivity and reliability to interview evaluations, allowing for more precise preparation for candidates and analysis on the interview performance of candidates.

VII. CONCLUSION

This research illustrates the strong associations between emotional states and levels of confidence, leveraging a multimodal approach with AI. The positive emotions of joy and love are consistently associated with increased confidence, while the negative emotions of sadness and fear are associated with decreased confidence. The addition of video, audio, and textual data allows us to undertake a more comprehensive analysis, while the visualizations show clarity over these behavioral features. The research findings demonstrate the ability of this AI-based system to have practical applications for mock interviews, mental health assessments, and personal development applications. This current research has put forward a basis for further research in a quest to understand and analyze human emotions and confidence and makes great addition to AI and human behavior analysis.

REFERENCES

1. Yang Li, Constantinos Papayiannis, Viktor Rozgic, Elizabeth Shriberg, and Chao Wang, "Confidence estimation for speech emotion recognition based on the relationship between emotion categories and primitives," IEEE, 2023. Available from: https://ieeexplore.ieee.org/abstract/document/9746930
2. S. Sridhar, S. Mootha, and S. Kolagati, "A University Admission Prediction System using Stacked Ensemble Learning," 2020 Advanced Computing and Communication Technologies for High Performance Applications (ACCTHPA), pp. 162–167, 2020. Available from: https://ieeexplore.ieee.org/abstract/document/9213205
3. Sivasangari, V. Shivani, Y. Bindhu, D. Deepa, and R. Vignesh, "Prediction Probability of Getting an Admission into a University using Machine Learning," 2021 5th International Conference on Computing Methodologies and Communication (ICCMC), pp. 1706–1709, 2021. Available from: https://ieeexplore.ieee.org/abstract/document/9418279
4. Dulmini Yashodha Dissanayake, Venuri Amalya, Raveen Dissanayaka, Lahiru Lakshan, Pradeepa Samarasinghe, Madhuka Nadeeshani, et al., "AI-based Behavioural Analyser for Interviews/Viva," IEEE 16th International Conference on Industrial and Information Systems (ICIIS), 2021. Available from: https://ieeexplore.ieee.org/abstract/document/9660757
5. R. N., P. Karlapati, M. R. Mulagondla, P. Amaranayani, and A. P. Toram, "An Innovative Emotion Recognition and Solution Recommendation Chatbot," 2022 8th International Conference on Advanced Computing and Communication Systems (ICACCS), Coimbatore, India, 2022, pp. 1100–1105. Available from: https://tinyurl.com/tpptm4ba
6. Zelin Chen, Guoxin Qiu, Xiangyu Li, Caixia Li, Kexin Yang, Zhuanggui Chen, et al., "Exploring the relationship between children's facial characteristics emotion and processing speech communication ability using deep learning on eye tracking and speech performance measures," IEEE, 2022.
7. R. Chinmayi, N. Sreeja, A. S. Nair, M. K. Jayakumar, R. Gowri, and A. Jaiswal, "Emotion Classification Using Deep Learning," 2020 Third International Conference on Smart Systems and Inventive Technology (ICSSIT), Tirunelveli, India, 2020, pp. 1063–1068. Available from: https://ieeexplore.ieee.org/abstract/document/9214103
8. R. Pathar, A. Adivarekar, A. Mishra, and A. Deshmukh, "Human Emotion Recognition using Convolutional Neural Network in Real Time," 2019 1st International Conference on Innovations in Information and Communication Technology (ICIICT), Chennai, India, 2019, pp. 1–7. Available from: https://ieeexplore.ieee.org/abstract/document/8741491
9. J. D. Berrios and Y. S. Lee, "Transforming Talent Acquisition: The Role of AI Technologies," Journal of Business Research, 2024. Available from: https://doi.org/10.4324/9781003512813
10. Chen Zhu, Hengshu Zhu, Hui Xiong, Chao Ma, Fang Xie, Pengliang Ding, and Pan Li, "Person-Job Fit: Adapting the Right Talent for the Right Job with Joint Representation Learning," ACM Transactions on Management Information Systems (TMIS), vol. 9, no. 3, pp. 1–12, 2018. Available from: https://dl.acm.org/doi/abs/10.1145/3234465

ABOUT THE AUTHOR

Dr. Shalini Bhaskar Bajaj is working as Professor in the department of Computer Science and Engineering at Amity University Haryana. She has more than 25 years of experience in the field of teaching, research and industry. She completed her B. E. (CSE) from Murthal Engineering College, Sonipat, M.E. (CTA) from Delhi College of Engineering, Delhi University and PhD from Amarnath School of IT, IIT Delhi.