Journal of Information Systems Engineering and Management
2025, 10(42s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

An AI-Driven Approach to Enhance Interview Performance through Voice and Response Analysis

Ria Mary Sunil[1], Tessa Soji Cherian[2], Divya James[3], Ann Mariya Joy[4], Paul Dins[5]
[1,2,4,5]*Department of Computer Science and Business Systems, Rajagiri School of Engineering & Technology, Kochi, Kerala, India*
riamsunil11@gmail.com, tessasoji@gmail.com, annmariyaalukkal@gmail.com, pauldins@gmail.com
[3]*Associate Professor, Department of Computer Science and Business Systems, Rajagiri School of Engineering & Technology, Kochi, Kerala, India*
divyaj@rajagiritech.edu.in

ARTICLE INFO

Received: 28 Dec 2024

Revised: 18 Feb 2025

Accepted: 26 Feb 2025

ABSTRACT

In today's competitive job market, interview prepa ration is a daunting issue for candidates due to a lack of proper and personal practice opportunities. The conventional approach through self-practice and generic mock interviews cannot ensure an immersive and personalized experience, leaving candidates poorly prepared and nervous. To solve this problem, this pa per suggests an AI-based virtual interview system, delivering immersive, role-playing mock interviews that are customized to meet the requirements of different jobs. The system involves dynamic question selection, speech fluency assessment, confidence evaluation via facial emotion recognition, and answer relevance assessment via natural language processing methods. Questions are dynamically chosen according to job descriptions, skills, and levels of experience to provide a systematic and varied inter view process. Fluency in speech is evaluated through hesitation patterns, filler words, and pause detection, while confidence measurement is carried out in real-time using facial expression analysis based on a convolutional neural network . Relevance of the answers is determined based on Cosine Similarity as well as BERT , and the results have been shown to offer higher accuracy for evaluating contextual meaning with BERT. The candidates are given information about their strengths and weaknesses, allowing them to improve their answers, increase confidence, and enhance communication overall. Through the imitation of actual interview situations, the system is an all-around resource for job applicants, increasing their likelihood of success in business interviews.

Keywords: Electronic Learning, Academic perfor mance, Engineering Education, Mock Interview, Student Learning, Educational Environment, Activity Learning, Facial Emotion Recognition, Confidence Analysis, Speech-to-Text, Answer Relevance, Filler Words, Fluency Assessment, Communication Skills, Job Preparation, Interview Readiness, Tailored Feedback

Interviews are essential to the hiring process because they evaluate a candidate's technical skills as well as their con fidence, communication abilities, and capacity for handling pressure. Candidates are frequently unprepared for actual interviews due to a lack of constructive criticism and simulated practice, even with prior self-practice or coaching. Voice-based interview preparation is more interactive in comparison to traditional methods, enabling candidates to respond verbally and receive structured feedback. In contrast to text-based methods, it mimics actual interviews, assisting candidates in reducing stress and enhancing verbal communication.

An existing AI-powered tool, such as Yoodli, ana lyzes speech fluency, filler words, and pacing to help users improve their speaking abilities. However, as it does not tailor questions according to candidate's chosen job role or evaluate the relevance of an answer, lacks flexibility for structured interview preparation. This restricts its capacity to replicate actual interview situations and offer focused feedback that is necessary for a successful interview. In order to fill these gaps, our system replaces generic prompts with a customized question set that dynamically corresponds with the candidate's chosen job role, guaranteeing relevant and structured practice. Additionally, the system checks for answer relevance, assisting candidates in improving their answers to match industry stan dards. The system also assesses speech fluency by identifying filler words and provides confidence analysis. The scope of this research is to create a customized inter view preparation tool that assesses the relevance of responses, customizes question sets, and offers

Journal of Information Systems Engineering and Management
2025, 10(42s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

feedback on speech fluency and confidence. The motivation stems from the fact that many candidates still struggle in real interviews despite their prior preparation as existing approaches lack structured feedback, adaptability, and real-world simulation.

Candidates struggle in actual inter views as as traditional methods fail to replicate live interview pressure and interac tion. They often face difficulties in maintaining confidence and articulating structured response This research aims to over come these issues by developing an AI-powered voice-based interview preparation system that offers dynamic feedback to improve candidates' confidence and readiness for real-world interviews. In the following contents, Section II. conducts a thorough literatuve review, discussing existing interview preparation methods, AI evaluation models, and methodologies. Section III. introduces proposed methodolo gy, including dynamic question fetching, confidence and speech fluency analysis and answer relevance checking. Section IV covers the experimental setup and user interface design. Section V presents performance analysis of CNN and a comparison of BERT and Cosine Similarity. Section VI concludes with key findings and future scope.

LITERATURE SURVEY

A thorough literature review is carried out in four main areas: question formulation for interviews, confidence evaluation, detection of disfluency in speeches, and answer relevance assessment. Current approaches are examined to assess their benefits, drawbacks, and possible applications in diverse fields. This evaluations gives a clear understanding of the effectiveness, limitations and scope in real life applications.

V. Kumar et al. [1] proposes a reinforcement learning-based framework for question generation (QG) that incorporates standard rewards (BLEU, GLEU, ROUGE-L, DAS) and novel QG-specific quality rewards (QSS+ANSS), enhancing syntactic, semantic, and relevance aspects of generated questions. The methodology builds upon pointer-generator networks with copy and coverage mechanisms, and evaluates models through both automatic metrics and human judgment. The advantages include significant improvements over state of-the-art baselines like L2A and AutoQG in both evaluation types, especially when QG-specific rewards are added. However, limitations include challenges in comparing with methods like NQGLC due to differing input requirements, lack of publicly available code for full reproduction, and only moderate inter-rater agreement in human evaluations.

M. Srivastava and N. Goodman [2] propose a methodology that fine-tunes pre-trained language models for knowledge tracing (LM-KT) to generate personalized, difficulty calibrated questions for adaptive education. Their work not only is able to make predictions about what a student might do on previously unseen questions, but also personally customizes learning to each person's needs. This approach excels at personalizing content, correctly matching question difficulty, and accommodating new question types. That said, it does have a few drawbacks—it requires large amounts of computational resources, relies strongly on rich student data, and hasn't been thoroughly tested for anything other than language translation problems yet.

P. Babakhani et al. [3] introduce Opinerium, a system to generate opinion-based, subjective questions using large language models (LLMs). This process employs fine-tuned LLMs to take input text as input and generate automatically interesting, non-factoid questions that stimulate critical thinking and open-ended discussion. The key advantages include high-quality question generation without manual rule-crafting and adaptability across domains. However, limitations involve potential biases inherited from the training data, challenges in ensuring relevance and coherence, and difficulties in controlling question style or tone consistently.

Jaiswal et al. [4] propose a deep learning-based methodology for facial emotion detection, utilizing convolutional neural networks (CNNs) to classify human emotions from facial expressions. Their approach involves training CNN models on facial image datasets to accurately recognize and categorize emotions such as happiness, anger, or sadness. The main advantages of this method are its high accuracy, scalability, and ability to learn complex features automatically from raw image data. However, limitations include sensitivity to variations in lighting, occlusions, and facial orientations, as well as potential bias if the training data lacks diversity.

Deshmukh et al. [5] propose a facial emotion recognition system using traditional machine learning techniques, where facial features are first extracted and then classified into emotional categories using algorithms like Support Vector Machines (SVM) and k-Nearest Neighbors (k-NN). The approach is computationally efficient and easier to interpret compared to deep learning methods, making it suitable for systems with limited processing power.

Journal of Information Systems Engineering and Management
2025, 10(42s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

However, its limitations include lower accuracy on complex datasets and reduced performance in handling diverse facial expressions, lighting conditions, and noise compared to more advanced deep learning models.

Pandimurugan et al. [6] presents a facial emotion recognition system tailored for monitoring students using machine learning techniques, aiming to assess their emotional engagement in educational settings. The proposed methodology leverages image preprocessing and feature extraction followed by classification using machine learning algorithms such as SVM and Random Forest. The key advantage lies in its potential to enhance adaptive learning by providing real-time emotional feedback. However, limitations include sensitivity to variations in lighting, pose, and facial occlusions, as well as the potential for biased predictions due to limited or imbalanced training data.

Shah et al. [7] propose a real-time facial emotion recognition system using CNN, focusing on accurately identifying emotions from live video feeds. The methodology involves preprocessing facial images, feeding them into a deep CNN model trained on standard emotion datasets, and deploying the system for real-time emotion classification. The advantages include high accuracy and fast processing suitable for live applications. However, limitations include the model's dependency on large, diverse datasets for generalization and challenges in handling extreme facial variations, occlusions, and low-resolution inputs.

Bhogan et al. [8] present a facial emotion detection system that combines machine learning and deep learning algorithms to enhance accuracy and robustness. When it is compared to Random Forest, CNN, and KNN, CNN showed superior accuracy. The system demonstrated significant improvements over the state-of-the-art methods, demonstrating its usefulness in psychology and human-computer interaction research. The difficulties lie in real-time processing and generalization across a range of facial expressions.

W. Mateo et al. [9] discuss speech disfluency detection through open-source software for automatic oral presentation training feedback. Their technique efficiently detects pauses, repetitions, and filler words to measure speech fluency. Although their method is useful in educational contexts, it has limitations in distinguishing between natural pauses and nervous speech patterns. Incorporating their disfluency detection model into speech emotion recognition would enable a more complete analysis of user speech.

D. S. Touretzky and C. Gardner-McCune [10] analyze Google Speech Recognition's understanding of language. They lead students in exploring bias and limitations in automatic transcription services. Although their research increases awareness of speech recognition issues, it does not have a competing approach to reducing biases. Their research may assist in optimizing bias mitigation in speech-based AI systems.

J. Liu et al. [11] suggest an effective pause extraction and encoding approach for Alzheimer's disease detection from spontaneous speech. Their approach is based only on acoustic features, minimizing reliance on linguistic information. Although their method is promising for early diagnosis, its efficacy across various languages and accents is not tested. Their pause extraction methods can be applied to speech emotion recognition models for identifying cognitive and emotional states.

J. Devlin et al. [12] present BERT, a deep bidirectional transformer model for language comprehension. BERT dramatically enhances NLP tasks, such as sentiment and emotion analysis. Although the model is highly accurate, it needs enormous computational power, which makes it difficult to apply in real-time. Nevertheless, fine-tuned BERT models might improve speech-based emotion recognition systems.

R. Krithika and J. Narayanan [13] examine machine learning approaches for grading short answer responses. Their approach assesses responses semantically and structurally, enhancing the consistency of grading. Though their system makes the automated grading process better, domain-specific training data is necessary for generalizing over subjects effectively. Their study could help towards developing AI based education assessment systems with speech-based tests.

Munikar et al. [14] suggested a fine-grained sentiment classification with BERT (Bidirectional Encoder Representations from Transformers) in order to increase the accuracy of sentiment analysis. In contrast to conventional models aiming at binary sentiment classification, the method captures subtle variations in sentiment. Experiments showed that BERT performs better than other widely used models in sentiment classification without even necessitating complex architectures. Additionally, the study showed how it can be effective transfer learning is

Journal of Information Systems Engineering and Management
2025, 10(42s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

at natural language processing tasks, which can be a comprehensive solution for sentiment analysis across a variety of domains. Some challenges lie in computational overhead and dependence on datasets.

M. C. Wijaya [15] proposes an automatic grading system for Indonesian short answers using BERT. Although their system's accuracy is dependent on the quality of the dataset and linguistic similarity, it lowers the accuracy error when compared to keyword-based matching. The key advantage is the system's ability to capture contextual meaning in natural language, leading to more human-like evaluations. However, limitations include the need for large annotated datasets for f ine-tuning and potential biases inherited from pre-trained language models.

Mosaed et al. [16] proposed a BERT-based model for reading comprehension question answering, leveraging the transformer architecture to understand and extract relevant answers from textual passages. The methodology involves f ine-tuning BERT on QA datasets to improve contextual understanding and answer precision. Its advantages include strong performance in understanding complex queries and producing accurate answers. However, limitations involve high computational requirements and sensitivity to ambiguous or poorly phrased questions.

T. Zhang & R. Zhang [17] investigated the strength of BERT in text sentiment classification, that overcomes the shortcoming of existing sentiment analysis models that cannot extract contextual semantics. Their study developed BERT-based models and conducted experiments on the IMDB dataset, demonstrating that, in comparison to the best-performing models currently in use, even a linear layer with BERT increases the F1-score by 2.01 percentage. The study highlights how BERT can dynamically view contextual information to achieve more accurate sentiment classification, but it also confirms that training time and computational complexity remain significant obstacles.

Abdollahnejad et al. [18] also introduced a deep learning methodology using BERT to improve person job fit in talent hiring in order to hire talented candidates. Their method improves the accuracy of matching processes by using BERT's contextualized knowledge to match resumes with job descriptions. The work was presented at the 2021 International Conference on Computational Science and Computational Intelligence (CSCI), which was held in Las Vegas, Nevada, USA, from December 15–17, 2021.

Tawil and Alqaraleh [19] created a BERT-based topic specific crawler to improve web crawling efficiency through targeting specific topics. The methodology improves the relevance and accuracy of collected data compared to traditional keyword-based crawlers. Its main advantages include enhanced semantic understanding and improved precision in topic relevance. However, limitations include increased computational cost and dependency on large labeled datasets for fine-tuning BERT.

Ajagbe and Zhao [20] explored retraining a BERT model for transfer learning in requirements engineering, aiming to improve the understanding and classification of software requirements. Their methodology involved fine-tuning a pre-trained BERT model on domain-specific datasets to adapt it for tasks like requirements categorization and intent detection. BERT4RE transferability was shown in the study by fine-tuning it for the purposes of identifying key domain concepts with better performance compared to the generic BERT-based models. Challenges involve the requirement for large domain-specific data and computing resources for retraining. The literature review offers important insights into speech disfluency detection methods, FER models for emotion detection, adaptive question generation techniques, and the influence of BERT in NLP applications. While BERT based models improve sentiment analysis and contextual language comprehension, FER developments allow for more precise emotion recognition for real-time human-computer interaction.

PROPOSED METHODOLOGY

The proposed AI-based interview preparation system fol lows a structured workflow to simulate a real-life inter view experience. The methodology consists of dynamic ques tion retrieval, speech-to-text transcription, confidence analysis through facial emotion recognition, speech fluency analysis based on frequency of filler words, and answer relevance validation. The system assesses users based on multiple factors to provide constructive feedback and improve their interview performance.

Journal of Information Systems Engineering and Management
2025, 10(42s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

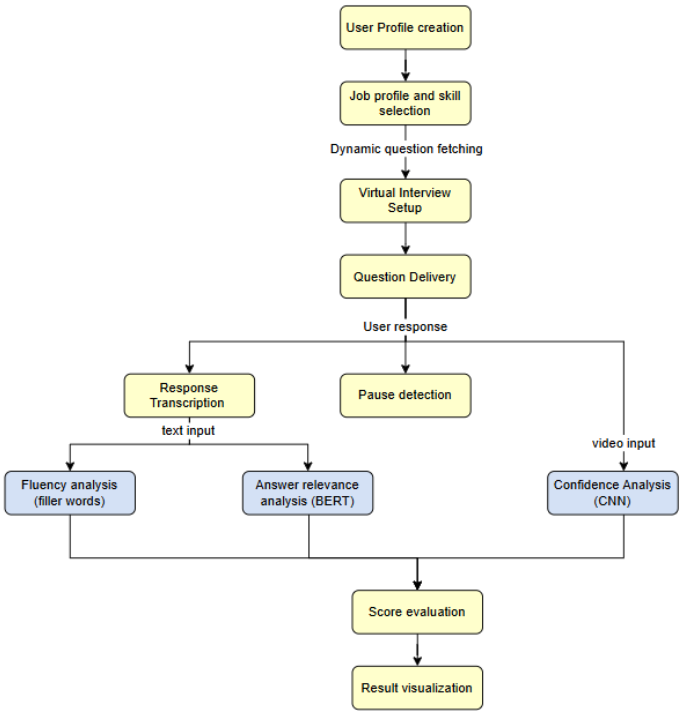

Figure. 1 Flow Diagram

The system architetcure represents a virtual interview plat form intended to assess user responses using different criterias. To personalize the interview process, it starts with the creation of the user profile and moves on to the job profile and skill selection. Using the chosen criteria, suitable questions are dynamically fetched and presented in a virtual interview environment. Pause detection is applied to the user's spoken response to evaluate hesitations, while confidence analysis is carried out using visual input. After the response has been transcribed into text, the relevance of the response is assessed in comparison to the expected response, and speech fluency is examined by identifying filler words. These evaluations are combined and a performance score is generated. This score is displayed as a detailed result visualization, allowing users to identify strength and areas of improvement.

A. Dynamic Question Retrieval:
Dynamic question retrieval is a crucial component of the proposed system, which guarantees that users receive ques tions that are relevant to their chosen job profile (e.g.,software engineering), skill (e.g.,Java), and experience level. This guar antee a valid and customized interview experience. The system distribute questions dynamically using a Round Robin tech nique, which enable the delivery of questions in a structured and varied manner. To achieve balanced representation, the Round Robin algorithm cycles through different skill sections which includes different skills, ensuring that each skill area contributes one question before the cycle restarts. This method effectively avoids the repetition of questions from the same skill and prevents overrepresentation of any single area, thereby sustaining a well-rounded distribution. Furthermore, questions that have already been answered are excluded from the pool, preserving both the relevance and diversity of the evaluation process.

B. Confidence Analysis:
Confidence analysis is a key component that leverages facial expressions to categorize a user's confidence level as high, medium, or low. This process utilizes a CNN model which is also called as Convolutional Neural Network to detect emotional cues linked to confidence by analyzing facial features in real time. The CNN model identifies seven core emotion: happiness, neutrality, sadness, fear, surprise, disgust, and anger. Confidence is estimated using these identified emotion: High confidence can be indicated by neutral and happy, medium confidence by sad, fear, and surprise, and low confidence by disgust and anger. To give complete evaluations of the user's confidence during the virtual interview, the system inspect multiple frames

Journal of Information Systems Engineering and Management
2025, 10(42s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

per second and calculates the average detected emotion captured using the frame detected over a given time window since facial expressions fluctuate during the interview.

C. Speech Fluency Analysis:
Speech fluency analysis improves verbal communication and feedback by identifying pauses and filler words . The B. User Interface of the Virtual Interview System system looks for hesitation patterns (filler words like ”actu ally” , ”so”) in transcripts using Regular Expressions, then determines their percentage of total words. While using fewer f illers shows good communication, using more fillers suggests poor fluency. Fluency is rated as excellent, good, moderate, or poor. In order to guarantee continuous speech, pause detection also detects silent periods that surpass a predetermined thresh old (measured in seconds) and sounds an alert. Through the elimination of unnecessary fillers and delays, this method helps candidates improve their speaking abilities, which eventually boosts their confidence and clarity in actual interviews.

D. Answer Relevance Checking:
The answer relevance is implemented using two differ ent approaches: cosine similarity and BERT-based NLP. The BERT model compare the transcribed responses against the expected responses which is stored in the database after analyzing sentence structures and contextual meaning to de termine semantic alignment. It provides a thorough linguistic assessment by evaluating coherence, accuracy, and completeness. Cosine similarity converts both of the user responses and expected answers into numerical vectors and measures to compute the similarity between them. A higher cosine score indicate more relevance. While cosine similarity emphasize how similar texts are and BERT concentrates on contextual understanding, both approaches separately offer insights into answer accuracy.

RESULTS AND DISCUSSIONS

A virtual interview system was developed to replicate the actual interview experience through an interface which includes interaction. It was built and tested using standard hardware and software tools. The system presents profile creation to interview simulation, aiming to ensure a smooth and engaging interaction.

4.1 Experimental Setup:

The experiments were conducted on a system consisting of the specified hardware and software requirements:

4.1.1. Hardware:

Processor: Intel Core i3 or higher RAM: 4 GB or more Storage: 512 GB or more.

4.1.2. Software:

Operating System: Windows 10 or later ,IDE: Notepad++ ,Front-End: HTML, CSS, JavaScript, Back-End: MySQL, PHP, Python ,Toolkit: XAMPP

4.1.3. Dataset:

The FER-2013 dataset was used, which con sists of 35,887 grayscale images (48×48 pixels) categorized into seven emotion classes: Angry, Disgust, Fear, Happy, Neutral, Sad, and Surprise. The data was preprocessed to increase model precision through the use of normalization and data augmentation. The last dataset was divided into Training (28,709 images), Validation (3,589 images), and Testing (3,589 images) for handling balanced performance assessment.

4.2 User Interface of the Virtual Interview System

The system features a structured and interactive user in terface designed for seamless virtual interviews. It presents questions and captures user responses, ensuring a smooth and intuitive experience.

The homepage of the virtual interview assistant setup as shown in Figure. 2, acts as the main entry point for users to access important functions. It offers a methodical design that leads users through the interview procedure, guaranteeing a smooth transition from profile creation to interview simulation.

Journal of Information Systems Engineering and Management
2025, 10(42s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

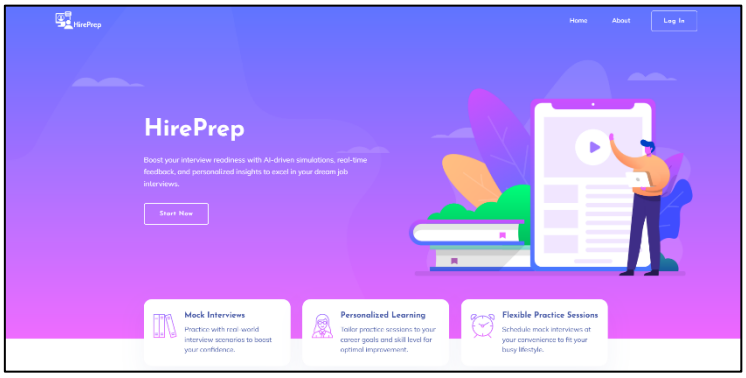

Figure. 2 Home page

The virtual interview screen is shown in Figure. 3, with a question prompt and a section for transcribed responses as part of the user interface. When no response is found, a pop up notification shows up, encouraging prompt interaction and increasing user engagement.

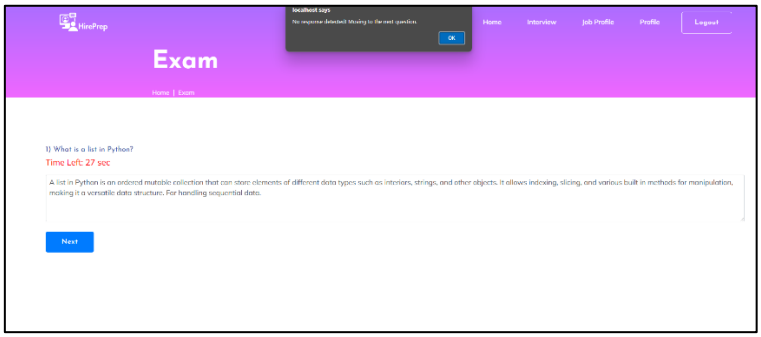

Figure. 3 Interview Simulation page

PERFORMANCE EVALUATION

Evaluating the performance of key components in the system is crucial to ensuring accuracy, reliability, and efficiency. The performance of the FER model is evaluated using training curves, confusion matrices, and performance metrics, which gives an indication of the accuracy of classification. Moreover, the performances of Cosine Similarity and BERT in identifying answer relevance during interviews are compared, thereby bringing out strengths and weaknesses.

5.1 Performance evaluation of Facial Emotion Recognition Model:

5.1.1 Training curves:

The Figure. 4 shows the training progress of a CNN used for Facial Expression Recognition (FER), with accuracy (left) and loss (right) plotted over 100 epochs:

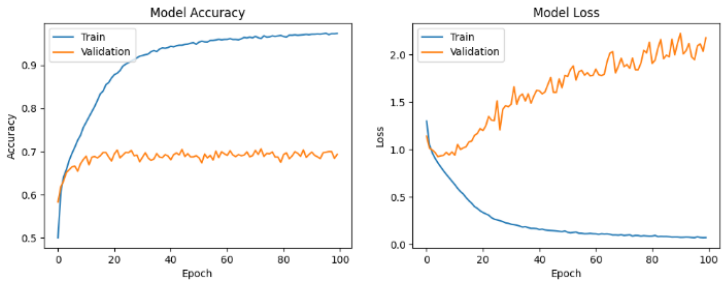

Figure. 4 Training curves of CNN

Journal of Information Systems Engineering and Management
2025, 10(42s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

The training curves in the figure show the learning trajectory of the CNN model for Facial Expression Recognition (FER). The accuracy curve also shows a steady upward trend, with training accuracy 97.56% and validation accuracy converging at 69.35%, showing that the model is well learning discrim inative facial features. At the same time, the loss curve shows a steady decrease, which also verifies efficient optimization and model convergence. Such stability of training guarantees the CNN model in being able to differentiate between many emotions efficiently and thus being ideally suited for FER applications in real-world environments where reliable and accurate emotion identification is crucial.

5.1.2. Confusion Matrix:

The confusion matrix in Figure. 5, showcases the classification performance of a CNN model for Facial Expression Recognition (FER).

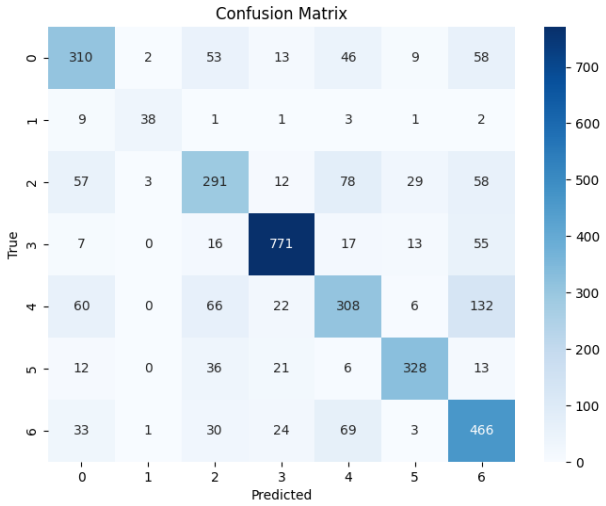

Figure. 5 Confusion matrix displaying absolute values for predictions on FER dataset

It demonstrates the CNN model's good capability for classification, as it shows high diagonal values, which represent accurate predictions. The matrix incorporates analysis across seven emotions: Angry (0), Disgust (1), Fear (2), Happy (3), Neutral (4), Sad (5), and Surprise (6). The model performs highly well with certain expressions, especially with class 3 (Happy) where there were 771 instances classified correctly and class 6 (Surprise) where there were 466 instances correctly classified, reflecting the high level of its capability in facial expression identification. This performance highlights the promise of CNN-based methods for FER, since they can well capture and analyze facial signals.

5.1.3. Precision, Recall and F1 score analysis:

The CNN model for Facial Expression Recognition (FER) demonstrates a balanced performance, with a precision of 70.37%, meaning most of its predictions are correct, and a recall of 69.99%, indicating it successfully identifies the majority of actual expressions. The F1-score of 69.96% confirms a good trade off between precision and recall, making it reliable for FER tasks. While improvements can be made, these metrics highlight model's effectiveness in accurately recognizing facial expressions.

Model	Precision	Recall	F1 Score
CNN	70.37	69.99	69.96

Table. 1 Performance metrics of CNN

Journal of Information Systems Engineering and Management
2025, 10(42s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

5.1.4 ROC Curve:

The Figure. 6 demonstrates the ROC curve of the trained CNN model. The ROC curve analyzes the performance of the CNN model by charting the True Positive Rate (TPR) against the False Positive Rate (FPR) for various emotion classes.

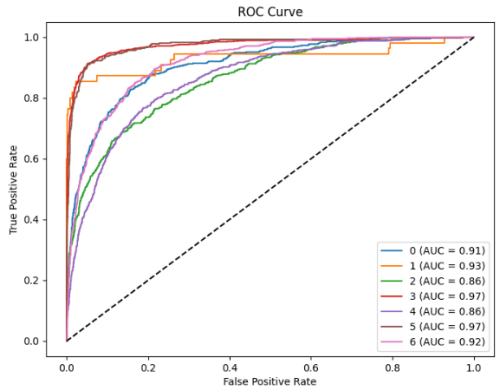

Figure. 6 ROC Curve of CNN model

The model has high classification strength, with AUC values between 0.86 and 0.97, reflecting good discrimination between facial expressions. Classes 3 and 5 reach an AUCof 0.97, indicating very close-to-perfect separability. The curves remaining high above the diagonal (random classifier) reflect the reliability of the model in reducing misclassification. This encouraging performance reflects the robustness of the CNN in identifying emotions and hence being very well placed for real-world applications like human-computer interaction, affective computing, and analytics driven by emotions.

5.2. Performance evaluation of BERT and Cosine similarity for Answer Relevance:

The effectiveness of BERT and Cosine Similarity was assessed using a standardized set of interview questions and matching database responses in determining answer relevance. Each approach produced a similarity score for each response after comparing the interviewee's recorded responses to pre defined database answers. Based on the calculated scores, the efficiency of each strategy was examined, demonstrating the advantages and disadvantages of each method.

5.2.1 Cosine Similarity:

The Figure. 7 showcases the score generated for answer relevance by assessing the similarity between user responses and database answers using Cosine Similarity.

```
Question ID  | Similarity Score   | Score (Out of 10)
-----------------------------------------------------------------
1            | 0.54               | 5.36/10
2            | 0.50               | 4.95/10
3            | 0.35               | 3.46/10
4            | 0.47               | 4.67/10
5            | 0.52               | 5.21/10
6            | 0.32               | 3.16/10
7            | 0.37               | 3.75/10
8            | 0.27               | 2.71/10
9            | 0.52               | 5.24/10
10           | 0.45               | 4.54/10

Final User Scores:
------------------------------
User 1 Total Score = 43.26%
```

Figure. 7 Answer Relevance Checking using Cosine Similarity

Journal of Information Systems Engineering and Management
2025, 10(42s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

Cosine Similarity assesses textual similarity by transforming the database response and the interviewee's response into numerical vector representations. The cosine of the angle between these vectors is used to calculate the similarity score, which ranges from 0 (totally different) to 1 (totally identical in wording). When responses contain exact word matches, this method works well and is computationally efficient. However, the results demonstrate how Cosine Similarity struggles with synonyms, reworded responses, and contextual variations, ac cording to the results. The final generated score was 43.26%, showcasing its limitations in capturing contextual meaning beyond lexical overlap.

5.2.2 Bidirectional Encoder Representation from Transformers(BERT):

The Figure. 8 demonstrates the result generated using BERT model for answer relevance by assessing the similarity between user responses and database answers.

```
Question ID  | Similarity Score  | Score (Out of 10)
-----------------------------------------------------------------
1            | 0.93              | 9.33/10
2            | 0.95              | 9.52/10
3            | 0.93              | 9.26/10
4            | 0.86              | 8.55/10
5            | 0.96              | 9.57/10
6            | 0.91              | 9.14/10
7            | 0.79              | 7.89/10
8            | 0.73              | 7.32/10
9            | 0.90              | 8.97/10
10           | 0.92              | 9.2/10

Final User Score:
------------------------------
Total Score = 88.75/100
```

Figure. 8 Answer Relevance Checking using BERT

BERT analyses responses based on context and meaning rather than precise word matching by utilising deep learning and bidirectional processing. It creates semantic embeddings, tokenises text, and uses deep contextual alignment to calculate similarity scores. The findings show that BERT effectively detects synonym-based variations and paraphrased responses, resulting in noticeably higher relevance scores. A significant improvement in accuracy over Cosine Similarity was demonstrated by the final calculated total score of 88.75% using BERT. This demonstrates how well BERT assesses the relevance of answers in interview situations.

5.2.3 Comparative analysis between Cosine Similarity and BERT:

The table below presents a compararison of Cosine Similarity and BERT in evaluating answer relevance, alongside confidence scores, speech fluency scores, and overall interview performance.

Method	Confidence	Fluency	Relevance	Overall
Cosine Similarity	90	95.50	43.26	76.25
BERT	90	95.50	88.75	91.42

Table. 2 Comparison between Cosine similarity and BERT for Answer Relevance Checking

The comparison between Cosine similarity and BERT highlights that Cosine similarity is computationally efficient while it lacks semantic understanding which actually makes it less effective for real-world assessments. This is reflected in its low answer relevance score of 43.26%, which significantly impacts its overall interview score of 76.25%. On the other hand, BERT's contextual awarenesse helps it to capture meaning beyond lexical similarity, resulting in a significantly higher answer relevance score of 88.75% and an overall interview score of 91.42%. BERT is the recommended option for interview evaluations due to its accuracy, even though it has a higher computational cost.

The results of the evaluation illustrate the strengths and limitations of the different methods employed in our system. The FER model exhibits high classification accuracy, clearly separating diverse emotions. Due to its deep contextual comprehension, BERT surpasses Cosine Similarity in performing answer relevance. When combined, these models offer detailed analysis and feedback, creating a well-performing and efficient approach to enhancing interview preparedness.

CONCLUSION AND FUTURE SCOPE

The proposed AI-based virtual interview system combines confidence analysis, speech fluency analysis, and answer relevance checking to deliver an interview preparation assissting system. CNN-based facial expression analysis accurately concludes confidence level, while dynamic question retrieval ensures balanced coverage across skills. Speech fluency assessment helps identify pauses and hesitation, aiding verbal communication improvement. BERT outperforms Cosine similarity in answer evaluation, offering more accurate insights despite higher computational demands. Overall, the system provides a scalable, AI-driven solution to enhance interview readiness.

Future enhancements include multilingual support, advanced sentiment and emotion analysis, adaptive learning for personalized feedback, and real-time AI guidance. Integration with VR simulations will offer realistic practice environments.

ACKNOWLEDGEMENTS

We also would like to express our heartfelt gratitude to Dr. Divya James for her invaluable guidance, support, and encouragement throughout the course of this research. Her expertise and insightful feedback have played a significant role in shaping this work.

Finally, we extend our thanks to Rajagiri School of Engineering & Technology for providing an excellent academic environment, resources, and infrastructure to carry out this work.

REFERENCES

[1] V. Kumar, G. Ramakrishnan, and Y. F. Li, “Putting the horse before the cart: A generator-evaluator framework for question generation from text,” Proceedings of the Annual Meeting of the Association for Com putational Linguistics, pp. 812–820, 2020. https://aclanthology.org/K19-1076

[2] M. Srivastava and N. Goodman, “Question generation for adaptive education,” Journal of Educational Technology, 2021. https://aclanthology.org/2021.acl-short.88/

[3] P. Babakhani, A. Lommatzsch, T. Brodt, D. Sacker, F. Sivrikaya, and S. Albayrak, “Opinerium: Subjective question generation using large language models,” pp. 105–110, 2024. https://ieeexplore.ieee.org/document/10522667

[4] A. Jaiswal, A. K. Raju, and S. Deb, “Facial emotion detection using deep learning,” in Proceedings of the 2020 International Conference for Emerging Technology (INCET), Belgaum, India, 2020. https://www.researchgate.net/publication/343414057_Facial_Emotion_Detection_Using_Deep_Learning

[5] R. S. Deshmukh, V. S. Jagtap, and S. S. Paygude, ”Facial emotion recognition system through machine learning approach,” in Proceedings of the 2017 International Conference on Intelligent Computing and Control Systems (ICICCS),Madurai,India,2017. https://www.researchgate.net/publication/316989659_Facial_Emotion_Recognition_System_through_Machine_Learning_approach

[6] P. Vellaisamy, “Facial emotion recognition for students using machine learning,” in Proc. Int. Conf. Comput. Commun. Informatics (ICCCI), Jan. 2023, doi: 10.1109/ICCCI56745.2023.10128425. https://www.researchgate.net/publication/371032293_Facial_Emotion_Recognition_for_Students_Using_Machine_Learning

[7] P. Kanani, A. Shah, R. Kumar, and H. Patel, "Real-Time Facial Emotion Recognition," in *Proceedings of the 2021 2nd Global Conference for Advancement in Technology (GCAT)*, Bangalore, India, Oct. 2021 . https://www.researchgate.net/publication/356075748_Real-Time_Facial_Emotion_Recognition

[8] S. Bhogan, K. Sawant, N. Gondalekar, R. Carvalho, V. Kalangutkar, and A. Mathew, ”Facial Emotion Detection using Machine Learning and Deep Learning Algorithms,” 2023 2nd International Conference on Edge

Computing and Applications (ICECAA), Namakkal, India, 2023. https://jglobal.jst.go.jp/en/detail?JGLOBAL_ID=202302220626068102

[9] W. Mateo, L. Eras, G. Carvajal, and F. Dom ınguez, "Detecting speech disfluencies using open-source tools in automatic feedback systems for oral presentation training," Proceedings of the International Conference on Speech and Language Processing, pp. 213–219, 2024. https://www.scitepress.org/PublishedPapers/2024/126221/

[10] D. S. Touretzky and C. Gardner-McCune, "Guiding students to investi gate what google speech recognition knows about language," Proceed ings of the AAAI Conference on Artificial Intelligence, vol. 37, no. 13, pp. 16 040–16 047, 2024. https://ojs.aaai.org/index.php/AAAI/article/view/26905

[11] J. Liu, F. Fu, L. Li, J. Yu, D. Zhong, S. Zhu, Y. Zhou, B. Liu, and J. Li, "Efficient pause extraction and encode strategy for alzheimer's disease detection using only acoustic features from spontaneous speech," Brain Sciences, vol. 13, no. 3, p. 477, 2023. https://www.mdpi.com/2076-3425/13/3/477

[12] J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova, "Bert: Pre-training of deep bidirectional transformers for language understanding," arXiv preprint arXiv:1810.04805v2, pp. 1–16, 2021. https://arxiv.org/abs/1810.04805

[13] R. Krithika and J. Narayanan, "Learning to grade short answers using machine learning techniques," 2023. https://www.amrita.edu/publication/learning-to-grade-short-answers-using-machine-learning-techniques/

[14] M. Munikar, S. Shakya, and A. Shrestha, "Fine-grained Sentiment Clas sification using BERT," 2019 Artificial Intelligence for Transforming Business and Society (AITB), Kathmandu, Nepal, 2019. https://arxiv.org/abs/1910.03474

[15] M. C. Wijaya, "Automatic short answer grading system in indonesian language using bert machine learning," Revue d'Intelligence Artificielle, 2021. https://iieta.org/journals/ria/paper/10.18280/ria.350609

[16] A. A. Mosaed, H. Hindy, and M. Aref, "BERT-Based Model for Reading Comprehension Question Answering," 2023 IEEE Conference on Artificial Intelligence(IEEE AI), 2023. https://www.researchgate.net/publication/377522193_BERT-Based_Model_for_Reading_Comprehension_Question_Answering

[17] T. Zhang and R. Zhang, "Revealing the Power of BERT for Text Sen timent Classification," 2021 IEEE 4th International Conference on Au tomation, Electronics and Electrical Engineering (AUTEEE), Shenyang, China, Nov. 19-21, 2021. https://www.researchgate.net/publication/357719476_Revealing_the_power_of_BERT_for_text_sentiment_classification

[18] E. Abdollahnejad, M. Kalman, and B. H. Far, "A Deep Learning BERT-Based Approach to Person-Job Fit in Talent Recruitment," 2021 International Conference on Computational Science and Computational Intelligence (CSCI), Las Vegas, NV, USA, Dec. 15-17, 2021. https://www.researchgate.net/publication/363289867_A_Deep_Learning_BERT-Based_Approach_to_Person-Job_Fit_in_Talent_Recruitment

[19] Y. Tawil and S. Alqaraleh, "BERT Based Topic-Specific Crawler," 2021 International Conference on Computational Science and Computational Intelligence (CSCI), Las Vegas, NV, USA, Dec. 2021. https://www.researchgate.net/publication/356372534_BERT_Based_Topic-Specific_Crawler

[20] Ajagbe, M., & Zhao, L. (2022). Retraining a BERT Model for Trans fer Learning in Requirements Engineering: A Preliminary Study. In Proceedings of the 30th IEEE International Requirements Engineering Conference (RE 2022). https://www.researchgate.net/publication/365112611_Retraining_a_BERT_Model_for_Transfer_Learning_in_Requirements_Engineering_A_Preliminary_Study