MULTIMODAL BASED INTELLIGENT INTERVIEW BOT SYSTEM

ABSTRACT:

The Interview Chatbot serves as an innovative, user-friendly solution to help individuals enhance their communication skills, build confidence, and prepare effectively for high-stakes interviews, setting a strong foundation for career success. Unlike conventional interview preparation tools, this system integrates natural language processing (NLP), confidence analysis, and emotion detection to provide personalized feedback. Existing interview chatbots focus primarily on user's response appropriateness and contextual understanding based on many AI assessment tools. These tools usually employ NLP models to analyse the textual responses and come up with adaptive questions. However, most solutions offer little more than the analysis of text but fail on the multi-modal aspects, such as real-time emotion recognition, speech, and confidence assessment. These parameters are also very important in assessing the overall performance of a candidate in addition to content relevance.

The proposed solution comprises of many high-level methodologies like Natural Language Processing (NLP), Natural Language Toolkit (NLTK) for interpreting human language, Mel Frequency Cepstral Coefficients (MFCC) for confidence analysis and Gemini-1.5-flash model for analysing users prompt and generating flow of questions. The chatbot interacts with users in a conversational format, posing frequently asked interview questions and recording the responses for overall analysis. The performance of the proposed system on the work done is 85% accuracy in emotion detection, 99% accuracy in confidence classification, and strong correlation between vocabulary analysis and interview success.

Keywords: Interview Chatbot, Speech Recognition, Emotion Detection, Natural Language Processing, Deep Learning.

1. INTRODUCTION:

In today's competitive job market, candidates often struggle to adequately prepare for interviews, resulting in a lack of confidence, poor communication skills, and limited vocabulary. Traditional interview preparation methods are often generic and fail to provide personalized feedback or simulate real interview scenarios effectively. To address these issues, there is a need for an intelligent, interactive solution that not only simulates realistic interview environments but also evaluates candidate's confidence, their emotional mindset and vocabulary through their spoken responses.

The Interview Chatbot creates an interactive environment where users can simulate real interview scenarios, practice answering questions, and evaluate their performance. After logging in through the website, users are guided step by step for recording their responses to AI generated interview questions, having their emotions analysed in real time with DeepFace, and getting their audio recordings assessed for confidence and vocabulary. The following questions are generated by tokenizing the transcribed text as input, transforms it through Gemini-1.5-flash model and formulate into a pertinent follow-up question. The chatbot transcribes carefully the user’s voice and analyses it, focusing on aspects like tone and modulation to assess the confidence level and the effectiveness of language usage. It also tracks user’s emotions during the session, giving valuable insights on their mental state. On the results page, there a detailed report with confidence score, vocabulary analysis, and most frequent emotion occurred during the interview. To make it even better, it offers YouTube video recommendations to help user improve. Beyond this, the chatbot can answer general questions about interviews, making the experience feel smooth and personal. This work is all about making interview preparation easy, affordable, and effective. With advanced AI working behind the scenes, it gives a personalized feedback and practical tips to build confidence and communication skills.

It can be used as a tool for practicing and enhancing the interview skills of job candidates, helping them prepare for real-life interviews by giving them personalized feedback. The chatbot can also be valuable to educational institutions in providing students with an interactive platform to practice interview scenarios before entering the job market. This technology will be integrated with the employee's development programs wherein professionals can use their interview techniques to sharpen either in internal promotion or job transfers.

The paper is divided into three main sections. The first section explains the methods we used, including how different types of input are collected and processed. The second section goes into how each part of the system works together at the same time. The final section shares the results we observed, along with the outputs and feedback from the model.

2. LITERATURE REVIEW:

Abhiram K R et al. (2024) developed an AI interview evaluator which employs machine learning, computer vision, and natural language processing on real-time to analyse the performance of candidates. It uses deep face and Haar-Cascade models to detect emotional states from facial expressions, Google Speech Recognition for converting speech to text for the assessment of communication skills. While the system provides insights to recruiters, candidates might not receive immediate feedback for improvement during the interview itself.

Kiran Waghmare et al. (2024) discussed about an AI interview preparation bot that uses ChatGPT API for real-life conversational interactions and the Whisper API for speech-to-text conversion. The responsive front-end runs on React and Vite, while the back-end is managed with FastAPI. But this approach lacks in analysing user confidence and vocabulary.

K. Kushal Jain et al. (2024) created AI Interviewer chatbot boasts Technical, Interpersonal, and Resume-Based interactions with candidates wherein one can test each for job-related knowledge, soft skills, and past experience. In this, feedback feature analyses performance for further improvement, with LangChain memory, vector embeddings, and Faiss relevant adaptive questions on hand. This primarily focuses on text-based interactions and not on voice analysis.

Amitha Caldera et al. (2023) proposed an Interview Bot methodology is based on speech and facial recognition and question categorization through a grading system. It transcribes responses, analyses emotional tone, and evaluates facial expressions. An updated question bank personalizes the interview by difficulty, and the grading system assesses response accuracy, relevancy, and emotional engagement through machine-learning algorithms.

The voice interviewing chatbot developed by Dr. T.Graceshalini et al. (2023) employs speech recognition (PyAudio, SpeechRecognition) to convert speech to text and includes NLP techniques (tokenization, vectorization) to analyse user intent. It uses LSTM-based Recurrent Neural Networks (RNNs) to manage sequential data and produce effective responses, using machine learning and deep learning to simulate interviews adaptively and effectively.

Gunjal Sumedha et al. (2023) the chatbot applies NLP (spaCy) and Deep Learning for tokenization, keyword extraction, and response analysis. It uses a User-Centered Design (UCD) approach with user feedback through interviews and task-based interactions. The system verifies candidate IDs and images, conducts interviews with dynamic questions, analyzes responses, and classifies results based on accuracy using keyword matching.

Siddhant Dharmatti, et al. (2022) proposed an approach for an interview preparation chatbot which is AI-driven and utilizes the integration of PyAudio and Google Speech Recognition for accepting voice inputs, NLU for intent recognition and tokenization, and RNN with LSTM for recognizing long-term dependencies in a candidate's response. Detailed feedback is provided in real-time, evaluations are organized, and emotional insights are gleaned from responses to inform candidate preparations for interviews.

Vineet Agarwal et al. (2022) employed common Naïve Bayes classifier to check the technical skills by checking the words said by the candidate against a set of predetermined answers based on keyword matching. It further varies-tuning the difficulty of the question presented to the candidate, depending on their performance, and stores responses within a database, thus potentially leading to a rapid, automated interview process with progressive feedback.

Joko Siswanto at al. (2022) reviewed the literatures in the Indonesian frame of reference regarding the processes, gains, and shortcomings of the interview bot based on Behavioral Event Interview (BEI). This interview bot employs various NLP strategies for feature extraction, such as tokenization, stemming, part-of-speech tagging, and the term frequency-inverse document frequency. Bayesian inference is used for probabilistic modeling, while competency levels are judged based on the STAR framework (situation-task-action and result). This methodology thus allows for an efficient approach towards structured competency evaluation of candidates in a cost-effective way.

Sinung Suakanto et al. (2021) proposed a methodology for the development of an interview chatbot that will engage in structured conversations, which are distinctly different things from the usual information-based chatbots. This chatbot joins intent recognition, to evaluate candidate responses, and sequencing mechanisms, to generate topic-based and follow-ups, allowing deeper investigations into the interviewees' skills. Responses are interpreted either through a machine learning model trained on labelled datasets or through expert judgment, which ensures structured communication, scalability, and automation of repetitive tasks in recruitment.

Muhammad Laiq et al. (2020) proposed and developed an AI-enabled chatbot for IT interviews utilizes some aspects from machine learning, whereas in-depth text analysis and text to speech are conducted through natural language processing (with the use of spaCy). The methodology consists of candidate identification through unique IDs, a communication phase where the chatbot does targeted questioning, a response assessment phase through a pre-trained dataset, and an automated generation of results.

Sonali Rajguru et al. (2018) discussed the development of an AIML-based chatbot that provides answers to frequently asked questions, employing Artificial Intelligence Markup Language to generate preset responses using pattern matching. The effectiveness is thought to be based on the input collection, query processing, and response generation done by the use of the AIML chatbot. Limitations include dependency on static AIML scripts, which hinder adaptability to non-static and complex queries.

In general, the latest trends in AI-driven interview practice chatbots utilize machine learning, NLP, computer vision, and speech recognition to provide natural-sounding and engaging interview simulations. These technologies screen candidates based on facial expression analysis, speech-to-text, intent detection, and adaptive questioning. While some provide real-time feedback and structured scoring, others focused on user experience and conversation flow. Yet many do not possess a holistic strategy that incorporates emotion detection, confidence analysis, and in-depth vocabulary assessment. To address this challenge, a multimodal approach has been proposed.

3. PROPOSED METHOD:

An interactive multimodal interview chatbot is proposed by using audio and video input, is the purpose of this work. The system enables the users to log in, participate in an interview by responding to questions orally. The integrated Gemini-1.5-flash model via the Google API allows structured prompt with NLP techniques to ensure context-aware and concise questioning. The responses of the user are recorded and analysed for different aspects like confidence (using speech modulation), vocabulary (using language analysis), and emotional response (using DeepFace for emotional perception). At the end of the interview, the user will receive a report with a summary of the performance including confidence, vocabulary, and emotion scores. Additionally answer choices generated by the chat-box for most common questions asked during interviews can also be made specific to the user, providing an experienced and personalized touch. This work is aimed at preparing people for interviews through feedback on their speaking skills.

The block diagram of the proposed interactive chatbot is shown in Fig.1. The approach involves the following modules:

- Input acquisition
- Preprocessing
- Feature Extraction
- Classifying
- Feedback

The detailed explanation about the modules are provided in subsequent sections.

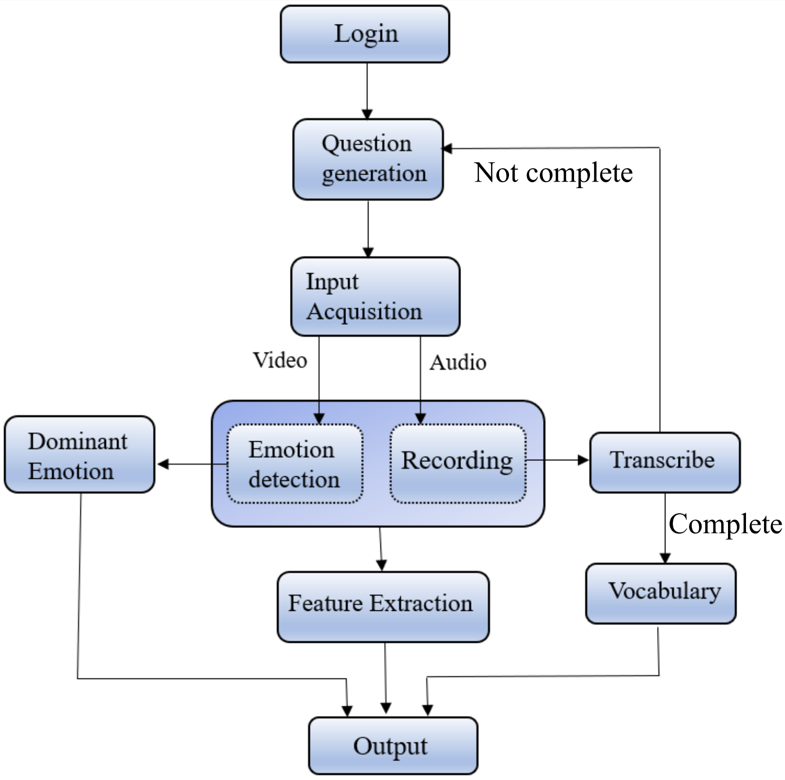

Fig.1 Flowchart of the proposed solution

3.1 Input Acquisition and Processing:

The user logs into the platform using their credentials, ensuring secure access. Upon successful login, the system displays instructions to the user, outlining how to navigate through the interview process. The input acquisitions are captured from the user, as audio and video through microphone and camera respectively.

3.1.1 Video processing:

This process performs real-time emotion detection using the DeepFace library and OpenCV. The webcam captures video frames, and then every frame is analysed for emotions. By starting with face detection using the Multi-task Cascaded Convolutional Networks (MTCNN model) because it has high precision and can reliably detect faces in real-time even in challenging conditions like varying angle and light. The MTCNN algorithm uses PNet to find the candidate area of the face, and uses RNet and Onet to further select the face candidate area with the highest probability When a face is detected, the VGG 16-Face model is used for emotion analysis. VGG 16-Face is chosen because of being a deep convolutional neural network trained on a huge number of faces, which allows stable and accurate recognition of emotions from different expressions. It develops a combination of deep learning with CNNs to perform feature extraction and pattern recognition. This model uses transfer learning to fine-tune a pre-trained VGG-16 to work on emotion datasets by replacing the output layer with one that works with emotion classes. The proposed system uses image preprocessing techniques such as resizing and normalization, softmax activation function for probability predictions, and training using optimizers such as Adam together with cross-entropy loss. These elements altogether provide a data-efficient architecture for efficiently detecting emotion from facial expression as illustrated in Fig.2.

It also tracks emotions across frames through a counter. When the user stops detecting, the most frequent emotion registered during that session displays. This effectively stitches video processing in real-time with the analysis of facial emotions, with MTCNN and VGG 16-Face considered both reliable detectors and recognizers of faces and emotions for interactive applications.

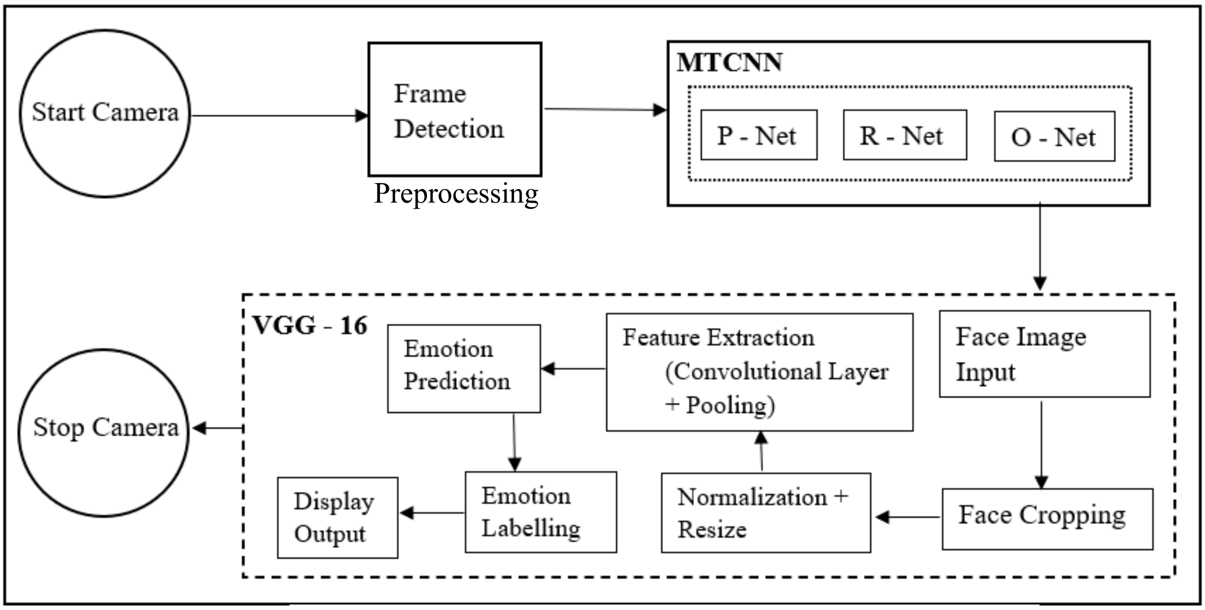

Fig.2 Flow diagram of Face and Emotion Module

3.1.2 Audio Transcribing:

This process is intended for recording verbal responses to interview questions. The transcription begins using OpenAI's Whisper, a highly accurate Automatic Speech Recognition (ASR) model capable of transcribing and later translating that audio in multiple languages. Whisper is a robust speech-to-text system that listens to audio and translates it into text accurately. In this case, the so-called "medium" model of Whisper is used to process recorded files of audio, MP3, WAV, M4A, and FLAC file formats. It starts by dividing the audio into small pieces and converting them into spectrograms, which are like visual representations of sound. It then examines these spectrograms to learn about the language and the speaker's way of speaking. If the language is unknown, Whisper identifies it automatically. Then, the system employs speech recognition to determine the most probable words being uttered. If necessary, it can even translate the text into a different language. Lastly, it refines the output to enhance accuracy and readability, making the transcription as clear as possible (Fig.3).

There is an iteration through the audio files stored in a previously defined folder. For each audio file, Whisper provides a transcription that is saved as a text file encoded in UTF-8. The entire process goes through an automation using the os module in Python, which handles files. After the transcription is done as shown in Fig.8, the text will be kept for further analysis to enable progression through the next stage of interviews. The transcribed words subsequently act as a basis for evaluating user’s responses, allowing the system to evaluate their vocabularies and confidence levels.

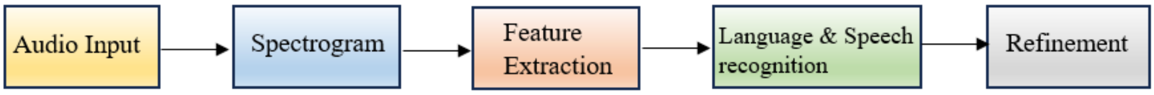

Fig.3 Steps involved in Transcribing

3.2 Feature Extraction:

The subsequent section explains the feature extraction process for question prompting, speech confidence and vocabulary analysis.

3.2.1 Questions prompting:

The interview chatbot questioning process is based on Gemini-1.5-flash, a highly capable AI model. It is employed to create dynamic and context-sensitive questions from user answers. Instead of relying on a pre-defined list of questions, Gemini-1.5-flash model examines the transcribed audio to understand patterns, coherence, and important topics, so that the interview becomes more interactive and responsive. This approach enhances engagement by mimicking real interview scenarios, improving a candidate's ability to think on the spot. By integrating NLP and AI-driven question generation, the system personalizes the interview process, making it more realistic and effective for skill development.

3.2.2 Confidence Analysis:

The Speech Confidence is computed based on audio classification using feature extraction, dataset management, and deep learning. The audio signal studied is one of the most widely used feature extraction methods, namely MFCC (Mel Frequency Cepstral Coefficients). MFCC was introduced into the field of audio analysis during the 1970s. The approach can be said to capture distinctive characteristics within a speech signal, defining the first few coefficients of the power spectrum in a manner representative of human auditory perception.

The working of MFCC can be summarized in a few steps, each involving mathematical transformations to convert raw audio into features that represent its perceptual characteristics:

- **Pre-Emphasis:** The signal is pre-emphasized to amplify high frequencies using equation (1) which apply a high-pass filter:

$$y[n]=x[n]-\alpha x[n-1] \quad \textbf{(1)}$$

  where, α is typically 0.95, and $x[n]$ is the input signal.
- **Framing**: The continuous signal is divided into small overlapping frames of size N samples, typically 20-40 ms, with 50% overlap using equation (2):

$$N=sampling\ rate \times frame\ duration. \quad \textbf{(2)}$$

- **Windowing**: A window function, such as the Hamming window, is applied to each frame to smooth the edges with equation (3):

$$y[n]=x[n].w[n] \quad \textbf{(3)}$$

  where $w[n]$ is the window function.

- **Fast Fourier Transform (FFT)**: Each windowed frame is transformed to the frequency domain using FFT, which converts the signal from time domain to frequency domain as demonstrated in equation (4):

$$X[k] = \sum_{n=0}^{N-1} x[n]e^{-j2\pi kn/N} \quad \textbf{(4)}$$

where $X[k]$ represents the frequency spectrum.

- **Mel Filter Bank**: The frequency spectrum is passed through a Mel filter bank, which is designed to simulate the human ear's frequency resolution. The Mel scale is a perceptual scale, and the transformation is defined as equation (5):

$$m = 2596. log_{10}(1 + \frac{f}{700}) \quad \textbf{(5)}$$

where m is the Mel frequency and f is the actual frequency. The output is the Mel-filtered spectrum.

- **Logarithm**: To approximate the way human perceive loudness, the logarithm of the Mel-filtered power spectrum is taken as equation (6):

$$Log\ Energy[k] = \log\left(\sum\nolimits_i |X_i|^2\right) \quad \textbf{(6)}$$

where X_i are the frequency bins after filtering.

- **Discrete Cosine Transform (DCT)**: Finally, in the equation (7) the log Mel spectrum is transformed using DCT to extract a set of coefficients (MFCCs), reducing dimensionality and capturing the most important features:

$$c[n] = \sum_{k=0}^{K-1} Log\ Energy[k]\ .\cos(\frac{\pi n(k + .05)}{K}) \quad \textbf{(7)}$$

The output consists of a small number of coefficients (usually 12-13 MFCCs) for each frame, which capture the essential characteristics of the audio, useful for tasks like speech recognition or audio classification. The MFCC representation of a sample audio is shown in (Fig.4).

The use of PyTorch's Dataset and DataLoader to effectively manage data efficiently, allowing for batch processing with shuffling in order to increase training efficiency and avoid overfitting. It employs fully connected neural network for multi-class classification with dropout layers, along with non-linear and probabilistic functions, such as ReLU and softmax, respectively. The loss function used is the cross-entropy loss, with the Adam optimizer in place to allow for effective, speedy training. It provides a scalable mechanism for audio classification, spanning the pipeline from feature extraction to model building.

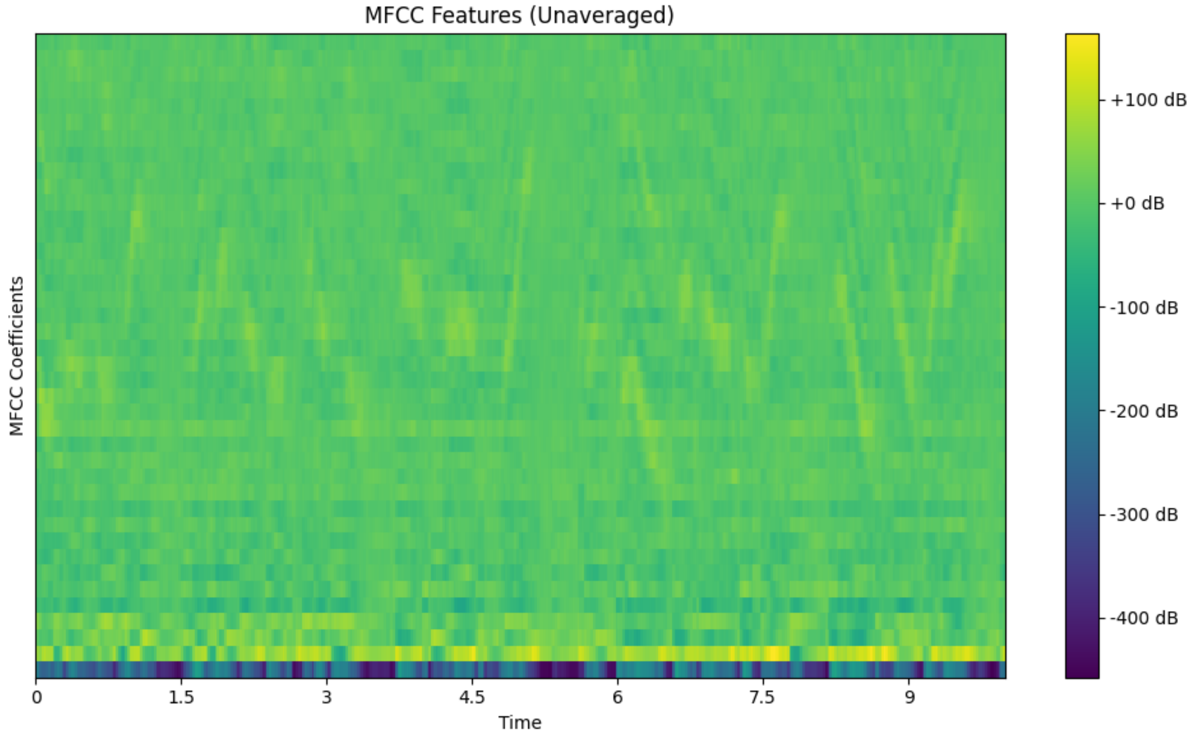

Fig.4 A Sample MFCC graph for a Audio

(3.3.c) Vocabulary Analysis:

The vocabulary evaluation program that ultimately evaluates the user's vocabulary in the transcribed text comparing depth, richness, and diversity with regard to complexity. The vocabulary assessment tool evaluates the depth, richness, and variety of a user's vocabulary in transcribed answers through analysis of complexity and word usage. Developed using Python-based NLP libraries, it uses Natural Language Toolkit (NLTK) for basic text preprocessing operations like

- tokenization
- stop-word elimination
- stemming
- part-of-speech tagging

enabling extensive linguistic analysis. It also uses TextBlob, a high-level NLP library that makes tasks such as sentiment analysis, noun phrase extraction, and text translation easy.

The program crosses user response to individual words with a dataset of professional vocabulary, determining overlaps and gaps in word use. Through the integration of NLTK's rich text processing with TextBlob's easy-to-use sentiment analysis, the system offers an all-around assessment of the user's vocabulary. This provides for personalized feedback on word complexity, variety, and sentiment, allowing users to hone their language skills for improved interview performance as depicted in Fig.5.

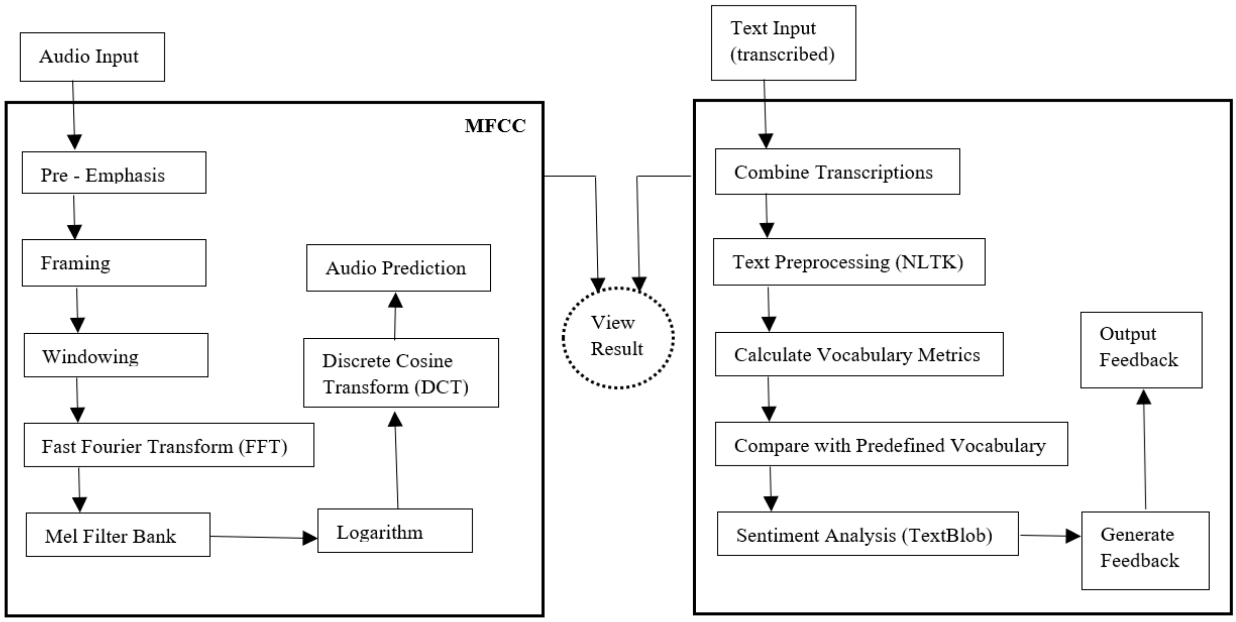

Fig. 5 Workflow of MFCC and Transcribing

3.3 Performance review:

Once the analysis is complete, the results are displayed on a dedicated result page whereby the user can view feedback on three cardinal aspects: Vocabulary, Confidence, and Emotion.

- The system analyses the vocabulary of the user, by first filtering for unique words, considering the lexical diversity, and comparing this to a stored set of professional vocabulary to draw feedback on any missed or underused words. This generates insightful feedback on language complexity, which assists the user in developing the vocabulary for the next interview.
- The confidence feedback deals with the user, exploring their confidence level, cessation pipe of speech such as tone, pace, and hesitation for words. In turn, this analysis detects such patterns from a given speech to determine whether a speech presented certainty or hesitation.
- The emotional feedback illustrates the most dominant feeling that prevailed during the interview concerning the emotional tint in tone that the user expressed in reply.
- In addition, these suggestions on the result page leads the user to several YouTube videos addressing interview practice, vocabulary development, or emotional intelligence based on their interested aspects. These recommendations are, therefore, personalized to help the user to improve in their respective areas of concern and further develop their conversational skills for the next interview.

3.4 Dedicated Chatbot for Question and Answer:

The chatbot embedded in the results page finds answers to general queries related to interview and inform candidates about their feedback. This module uses JavaScript (ES6) for dynamic programming and interaction with conversation states, sending messages typed by the user, and updating the chat interface in real-time. HTML and CSS structure allows forming an attractive and responsive design with chat bubbles, an input field, and buttons. The chat interface is made dynamic by adding the inputs into the chatbot with the help of DOM Manipulation. Back-end communication is via Fetch API to send members' queries as POST requests to the Flask server's endpoint, which processes these inputs and returns responses. This asynchronous capability allows perfect handling of communication, so this is made possible via CORS settings to allow secure cross-origin requests. The design employs modern patterns to make the application responsive and friendly. With the use of these components, an interactive chatbot can be constructed to provide instant responses to interview preparation concerns.

4. RESULT AND ANALYSIS:

The experimental setup comprises hardware and software parts required for efficient operation of the system. In hardware, a computer system with Intel i5/i7 processor, 8GB RAM, and NVIDIA GPU is utilized to execute deep learning tasks efficiently. A high-resolution web camera is needed for real-time emotion detection, whereas a high-quality microphone is utilized to capture clear audio recordings for speech analysis.

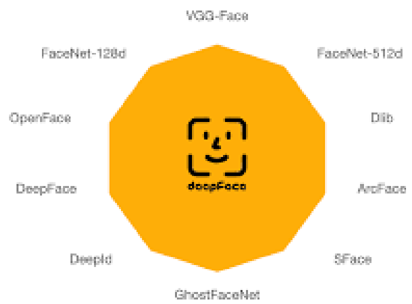

Fig. 6 DeepFace

On the software front, the system is developed within a Python 3.8+ environment, incorporating robust libraries for different tasks. DeepFace (Fig.6) is employed for emotion detection, and OpenCV facilitates real-time video processing. Depending on the implementation, either TensorFlow or PyTorch is used to support deep learning models. Whisper is used for speech-to-text transcription, and NLTK and spaCy offer natural language processing functionality. Also, SpeechRecognition and PyAudio make voice input handling accurate. The whole system is hosted with the help of Flask, providing a smooth web-based interface for users to interact with the application in a timely manner.

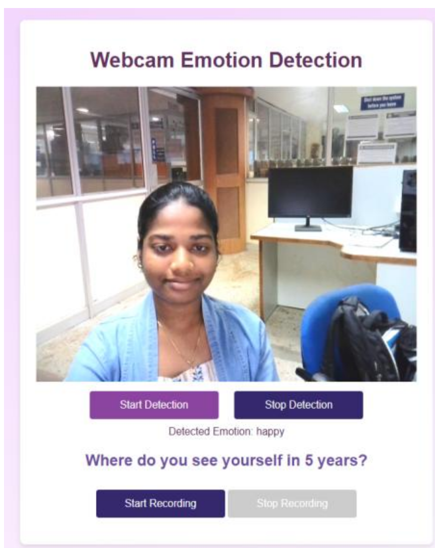

Fig. 7 Sample Frame of Emotion detection and Audio Recording

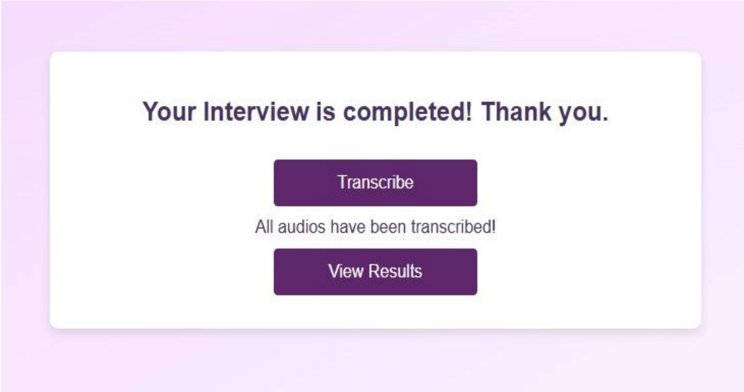

Fig. 8 Transcribing and Generating Results

Fig. 9 Result of the proposed system

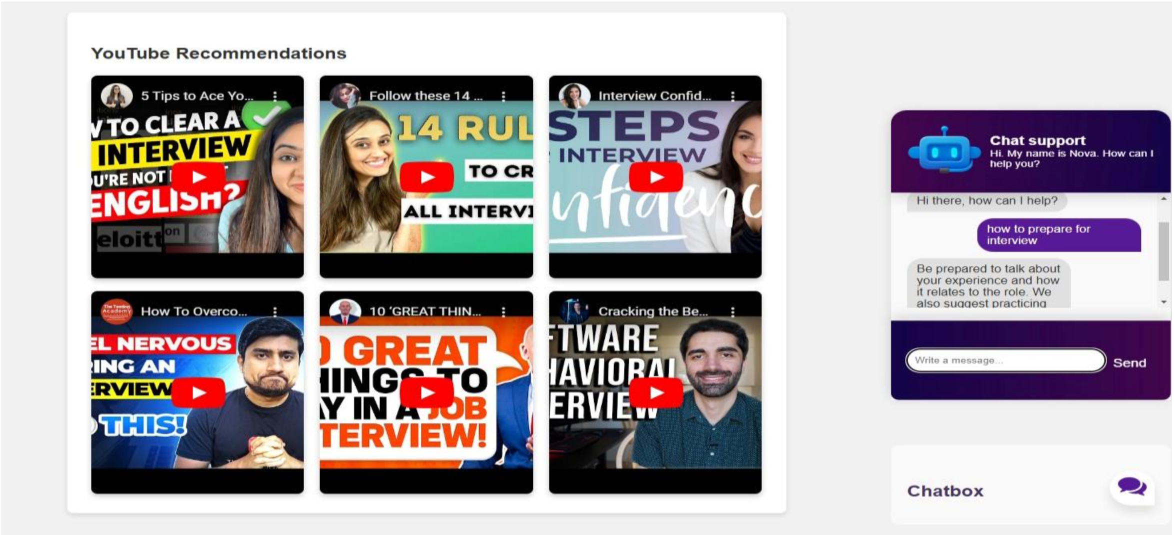

Fig. 10 YouTube Recommendation and Chatbot

Figures 7 to 10 illustrate various stages and features of the proposed interview evaluation system. A sample frame from the interview is shown in Fig.7, where the webcam records the candidate's audio responses and their facial expressions for real-time emotion detection. The system automatically transcribes the audio after the interview is over and allows result generation, as seen in Fig.8. A confidence level based on linguistic patterns, the dominant emotion identified during the interview, and a vocabulary analysis displaying frequently and infrequently used words are all included in the final analysis results, which are shown in Fig.9. Lastly, Fig.10 illustrates the post-analysis support features, which include a chatbot interface to help users with common interview-related questions and feedback, as well as personalised YouTube video recommendations to enhance interview skills.

4.1 Performance analysis:

In this study, two different audio feature extraction techniques were explored to evaluate their effectiveness in assessing candidate responses: **Mel Spectrogram** and **MFCC (Mel Frequency Cepstral Coefficients)**. Table 1 presents a side-by-side comparison of the Mel Spectrogram and MFCC on how each method contributes to performance.

Table 1 General analysis of audio feature extraction methods

PARAMETERS	Mel Spectrogram	MFCC
No of Datasets	Confident – 80 Unconfident - 150	Confident – 500 Unconfident - 500
Trained using	CNN & Mel - Spectrogram	CNN & MFCC
Result	Low Training and Validation Accuracy	High Training and Validation Accuracy
Reason	Imbalance Dataset and Complex Model	Balanced data and adaptive model

Table 2 shows the accuracy achieved by both models during evaluation. The Mel Spectrogram-based model and the MFCC-based model were tested on the test dataset under similar conditions

Table 2 Accuracy of the 2 approaches

Accuracy	Mel Spectrogram	MFCC
Training Accuracy	53.92%	98.62 %
Validation Accuracy	56.87%	99.50 %

Table 3 provides a detailed performance analysis, including metrics such as precision, recall, F1-score, and inference time for proposed solution using MFCC.

Table 3 Performance Analysis of the proposed solution for various metrics

	Precision	F1 Score	Recall
Confident	99	100	99
Unconfident	100	99	100

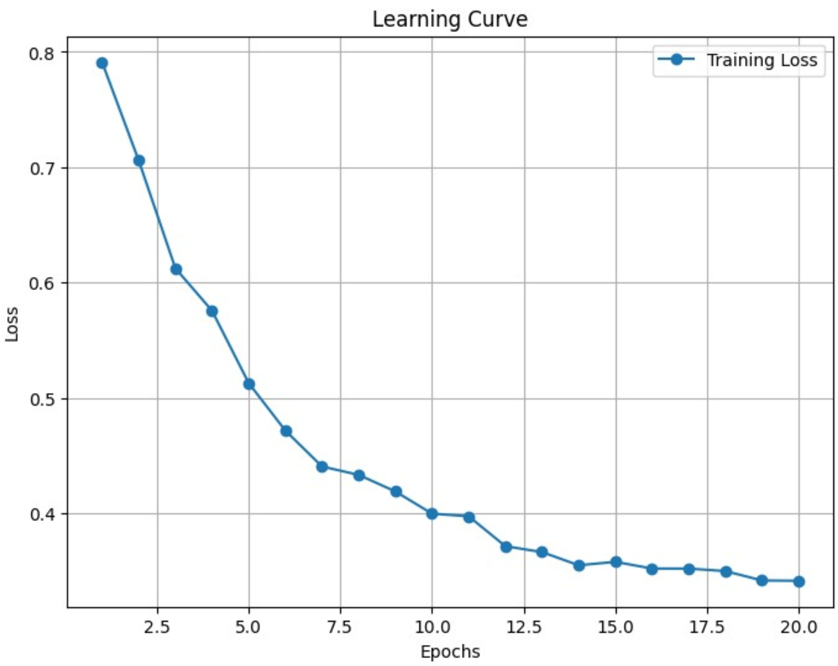

Fig. 11 Training Loss Graph

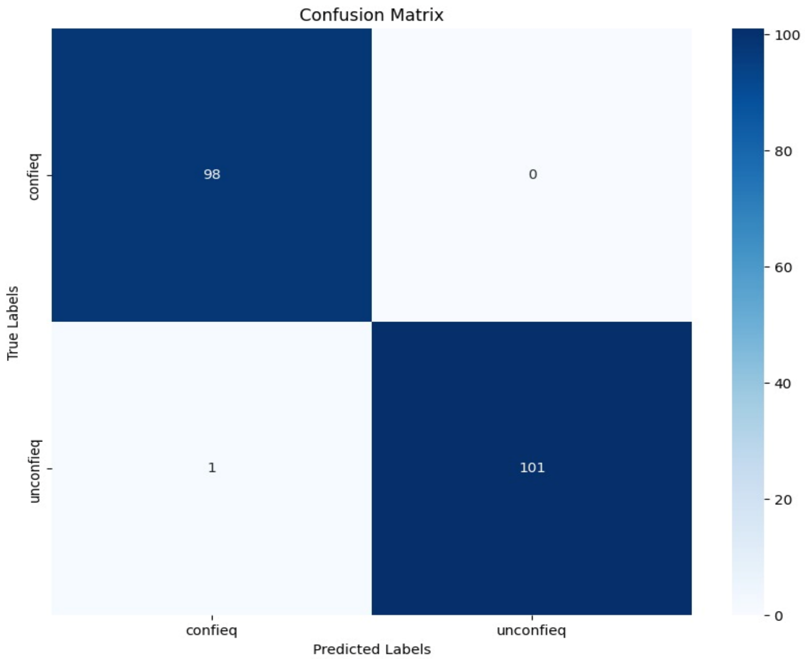

Fig. 12 Confusion Matrix

The training curve represented in Fig. 11 plots the model training process over 20 epochs in classifying confident and unconfident speech audio data. There is a clear and consistent decreasing trend in the training loss from around 0.79 to 0.34, depicting effective convergence. The lack of random fluctuations or premature plateaus reflects stable learning behavior and perfect parameter tuning. This indicates that the model is not just learning meaningful structures from the audio data but is also generalizing sufficiently without overfitting while training.

After the training period, classification performance of the model is tested using a confusion matrix presented in Fig. 12. The outcome is extremely promising, with 98 confident and 101 unconfident speech samples classified correctly and just one case of misclassification. This provides a very impressive accuracy of 99.5%, demonstrating the model's high ability to differentiate between confident and unconfident speech tones. The almost perfect classification performance verifies the success of the training process, as reflected by the loss curve, and highlights the model's reliability for downstream applications like real-time interview assessment and feedback generation.

5. CONCLUSION:

The developed Interview Chatbot project demonstrates the use of artificial intelligence to change how individuals prepare for interviews. In that regard, by incorporating real-time emotion detection, speech analysis, and natural language processing, the chatbot gives valuable insights to users regarding confidence, vocabulary, and emotional states. Its interactive design, together with personalized feedback and actionable recommendations, ensures users can practice, improve, and feel more confident when handling real interview scenarios. The preparation for an interview is now much more accessible and economical, while the entire experience is enhanced due to the realism offered and personalized guidance. Facilities such as YouTube recommendations and a built-in chatbot providing answers to general questions make the system learn from its mistakes continuously.

In future the proposed interview chatbot can be enhanced with multilingual support for users, allowing them the freedom to communicate in their language of choice, making the platform more inclusive and accessible.

7. REFERENCES:

1. Abhiram K R, et al. "AI Based Interview Evaluator: An Emotion and Confidence Classifier" International Journal of Advanced Research in Computer and Communication Engineering Impact Factor 8.102, Peer-reviewed & Refereed journal, Vol. 13, Issue 4, April 2024

2. Kiran Waghmare, et al. "AI Bot For Interview Preparation." Proceedings of the 2024 International Journal of Research Thoughts, Volume 12, Issue 1 January 2024 | ISSN: 2320-2882.

3. K. Kushal Jain, et al. "AI Interviewer Chatbot for Technical and Hr Brilliance: A Tool for Upskilling Candidate" International Journal of Engineering Research & Technology (IJERT) ISSN: 2278-0181 Volume 13, Issue 02, February

4. Amitha Caldera, et al. "Interview Bot: Automating Recruitment Process using Natural Language Processing and Machine Learning" International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395-0056 Volume: 10

5. Dr. T. Graceshalini et al., "Interview Practice–Voice-Based Chatbot" Grenze International Journal of Engineering and Technology, vol. 9, no. 2, pp. 1952–1955, 2023.

6. Gunjal Sumedha, et al. “Interview Bot Development with Natural Language Processing and Machine Learning AI Talk” International Journal of Creative Research Thoughts (IJCRT), Volume 11, Issue 12 December 2023 | ISSN: 2320-2882

7. Siddhant Dharmatti, et al. "Interview Practice – Voice-based Chatbot" International Journal of Advances in Engineering Research (IJAER) 2022, Vol. No. 23, Issue No. V, May e-ISSN: 2231-5152, p-ISSN: 2454-1796.

8. Vineet Agarwal, et al. "Chatbot for Interview." International Journal of Recent Technology and Engineering (IJRTE) ISSN: 2277-3878, Volume-11 Issue-2, July 2022.

9. Joko Siswanto, et al. " Interview Bot Development with Natural Language Processing and Machine Learning" International Journal of Technology 13(2) 274-285.

10. Sinung Suakanto, et al. "Interview Bot for Improving Human Resource Management." 2021 International Conference on ICT for Smart Society (ICISS) | DOI: 10.1109/ICISS53185.2021.9533248.

11. Muhammad Laiq, et al. "Chatbot-based Interview Simulator: A Feasible Approach to Train Novice Requirements Engineers." 2020 IEEE 10th International Workshop on Requirements Engineering Education and Training (REET) DOI 10.1109/REET51203.2020.00007.

12. Sonali Rajguru, et al. "Interview Preparation by Chatbots" SSRG International Journal of Computer Science and Engineering (SSRG – IJCSE) – Volume 5 Issue 12 – December 2018.