INTERNATIONAL JOURNAL
FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Emo Confident Interviewer: An AI MockInterview Evaluator

Bharati Thawali D[1], Pranjal Kalal[2], Samarth Dugam[3], Harshada Lokhande[4]

Department of Electronics & Telecommunication JSPM's RSCOE Tathawade Pune, Savitribai Phule Pune University, Pune

Abstract: Interviews hold great significance for candidates as it's the moment when their hard work is put to the test in hopes of achieving their desired and successful life results. It plays a crucial role in our educational system and hiring process by helping to identify the best applicant based on the necessary abilities. We can do better by improving our communication and confidence skills through mock interviews. This article introduced new way to practice for interviews using AI-based mock interview platform. The three characteristics that our system will utilize to evaluate the user are emotions, confidence, and knowledge base. A deep learning CNN algorithm uses facial expressions to determine emotion classify the emotion into one of the seven categories, and the basis for evaluating confidence is voice recognition through the use Python modules for Pydub audio and natural language processing. A web scraping module will map keywords to internet resources by extracting them from incoming answers. Semantic analysis technique is utilized for knowledge assessment and keyword mapping. So, using this method will help the job candidate feel more confident and less stressed or anxious before the actual job interview.
Keywords: NPL, CNN, deep learning, Emotions, Confidence evaluation

I. INTRODUCTION

Interviews are crucial while recruiting. They assist the recruiting manager in selecting the most qualified candidate. They also assist the candidate in determining whether they are serious about the position. [1]. Usually, an interview lasts 45 to 90 minutes, or about 1.5 hours. Human psychology states that one has only seven seconds to make a lasting impression [2], and we are all aware of how important first impressions are. During this time, candidates worry a lot about looking well, trying to impress the recruiter, not seeming apprehensive, and keeping eye contact and confidence. Eye contact is important, according to recent statistical research on interviews 39% of candidates leave a bad impression on the recruiter due to their vocal quality, lack of confidence, or lack of grin, according to the results of 67% of recruiters [3][4]. Generally speaking, the present standard interview technique is very physical; that is, the interviewee answers after the interviewer asks the question [5]. The interviewer makes the decision to choose or reject the candidate based on the interviewee's answers, degree of confidence, and knowledge. It's crucial to practice interviews when you're about to go on one. "Do I really need to do practice interviews?" may be on your mind. But they're really beneficial, I promise. You become accustomed to the interview process via practice, particularly in the case of online interviews. Additionally, when the actual interview Additionally, because you've done this previously, you'll feel far more comfortable during the actual interview. Although there are a lot of online interview platforms available these days, most of them just take candidates' knowledge into account. Interviews, however, are about more than just information. They also concern an individual's personality, conduct, and character. Currently, candidates are frequently selected by AI-powered algorithms based on their recorded videos. More than 82% of recruiters utilize the internet for employment conversations, even in the midst of the pandemic. But some applicants discover Virtual interviews can be difficult to do well in because of social indications such head tilts, smiles, and nods. Recognize that remembering those specifics is harder. when we're under stress already. Practice, practice, and more practice is the only cure. This study employs artificial intelligence and a dynamic interview assessor as a mock interview evaluator to address the problem. Traits, and character. While there are numerous online platforms available to conduct interviews, the system only takes the candidate's knowledge base into account when making its judgment. The candidates are chosen for interviews based on their video recordings using the current AI-based approach as the globe approaches the technological era. . Despite the fact that the pandemic has passed, 82% of recruiters still do online interviews. However, a number of candidates have trouble during online interviews. For example, it might be challenging to remember social cues like smiles, nods, head tilts, and others when we are already under pressure. Practice, practice, and more practice is the only way out. This research uses an artificial intelligence (AI)-based mock interview evaluator IJTOR, which is a dynamic interview evaluator, to solve the issue.

The system uses speech frequencies and knowledge to determine confidence, and it accepts input in the form of a facial expression to check emotion. This module is divided into three pieces. Section I serves as an introduction, outlining the requirements of the interviewer, the field of study, the kinds of technologies that are employed, and the ways in which our research will benefit the candidates. The system's flow is also explained in Section In Section II, which is devoted to literature study, we looked at the current system and the various technologies that it employs. We were able to identify the conflict in the currently suggested system thanks to the research.. The entire system's architecture, methods, and types of algorithms utilized in the peer-reviewed research study are covered in section III, along with diagrams and a flow chart. Section IV compares and contrasts the current and suggested systems. The analysis made it easier to identify the gaps in the current system. It demonstrates that the current systems are not dynamic and that the system generates the output after a drawn-out procedure. The output of the entire system is displayed in Section V.

II. LITERATURE SURVEY

This research is about creating a special software for practicing job interviews. It gives users new interview questions and provides feedback using artificial intelligence. They made a video platform where people can practice interviews. This helps companies pick the right candidates and helps job seekers get better at interviews. The platform looks at how people talk, their body language, personality traits, and overall performance in the interview. This software, called MIP, focuses a lot on analyzing interviews in Chinese. It uses what people say, how they say it, and how they act to figure out how well they did in the interview. Based on nonverbal clues, a machine-learning technique for identifying and analyzing changes in interviewee behavior and personality characteristics [10]. The emotions, eye movements, grins, and head movements of the candidate have all been studied. A method can help people prepare for interviews when they are unsure about various sorts of questions and how to answer them. The interviewee can prepare for the technical aptitude exam according to their difficulty level, which also aids in preparation for face-to- face interviews. Emoticons are an essential part of who we are and how we behave has been found using EEG to track emotions. For the classification and identification of emotion, the arousal and valence model has been used. thermostat, are one type of device that use IoT technologies. The model includes arousal and valence dimensions and distributes the emotions in a two- dimensional space. They employed machine learning approaches to train classifiers and categorize emotional states such as arousal and positive and negative valence. We've talked a lot about measuring how happy customers are and figuring out which features are good for understanding their feelings. To figure out how people are feeling, we're using a famous collection of written or spoken words that show different emotions. And for checking how satisfied customers are, we're listening to what they honestly say about the service they got during phone calls. For posterior analyses, the author used recordings available in the call center [11]. The three "standard" databases that the authors take into consideration for the classification of speech emotion are the Ryerson Audio-Visual Database of Emotional Speech and Song (RAVDESS), the Interactive Emotional Dyadic Motion Capture (IEMOCAP), and the Emotional Database (EMODB). The authors created a virtual interview training system in order to address the issues that students were having during the interview process. The system focuses on three main parts: three virtual agents with distinct types of personalities.

III. METHODOLOGY

Our system will evaluate the student's performance in the interview. The parameters that will be considered by the system at the time of evaluating the score are the facial emotion of the candidate, confidence based on speech, and knowledge. Our system is divided into 5 phases, Face recognition (Authentication), Data Separation, Facial expression recognition, Confidence recognition based on speech score and Knowledge Base.
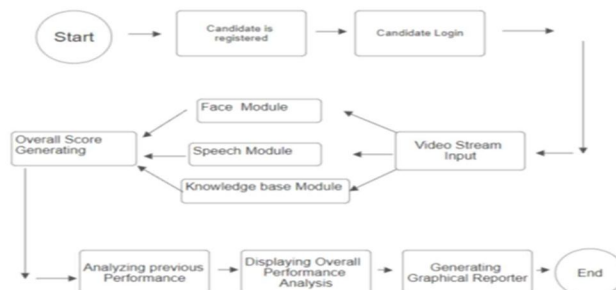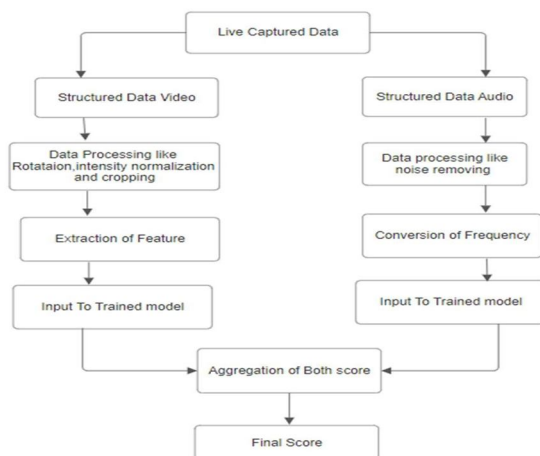

Fig.1 Flow Diagram

Fig.2 Proposed Block Diagram

A. Face Identification

1) *Image Acquisition:* The system will require the candidate's photo, which he uploaded at the time of registration. Based on that photo, the system will recognize that person, and if the person is confirmed, then it will continue.

2) *Image pre-processing:* First received image will get converted into BGR to grayscale format. Then it will crop inthe shape of 50*50 pixels. And then it is used for recognition.

3) *Recognition:* The system uses a Har-cascade classifier to classify the different faces. For face recognition, the system uses the predefined python library Face Recognition. In that case, the live photo of the candidate will be converted into face encodings and compared with the encoding of the photothat the candidate uploaded at the registration comparison system using compare face function of the Face Recognitionlibrary.

B. Data Segregation

The system will receive live data it will get separated in video audio format for future use for getting live video data using the OpenCV library and for getting live audio data system using the library.

C. Recognition of Facial Expression

1) *Deep Learning Model Training*

a) Dataset: a dataset is pre-divided into a training and testing folder which contains 28821 images in the training and 7066 images in testing. It has 7 different categorical values i.e., happy, sad, angry, disgust, anger, neutral, and fear image has a dimension of 48*48.

b) Training the model: for training, we have used CNN algorithm because we have to perform operations on an image and it takes a lot of time and resources. CNN uses parameter-sharing techniques and dimensionality reduction so it reduces the required time and computation.
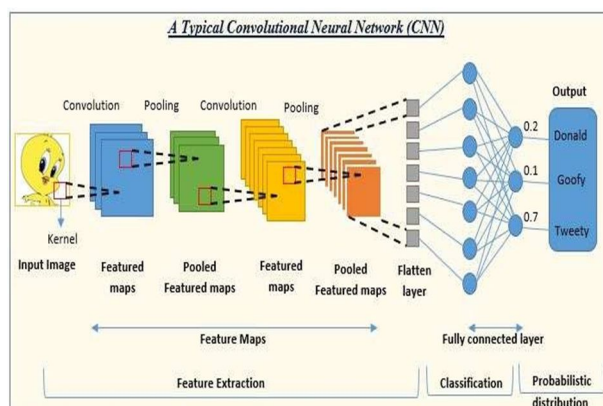

Fig 3. Working of CNN Algorithm

Fit= (W−F+2P)/S+1

$$W_{out} = \frac{W - F}{S} + 1$$

Pre-processing the data: to obtain more precision, the system divides the live video into frames and pre-processes each frame as it is received. The image will first be cropped to a 48*48 ratio. After then, it will be transformed into a grayscale picture. After then, the intensity of the model will normalize based on the requirement as stated by the ng. Feature extraction: the nose is extracted using four convolutional layers following the pre-processing features, such as the location of the lips and eyes. Prediction: the model generated the output emotion after running the retrieved feature through it. The score will rise or fall in accordance with the anticipated feeling.

D. Speech score-based Recognition of Confidence

This module focuses on speech recognition and categorization based on 7 emotions; this module consists of sub-modules:

1) *Data preprocessing:* This module focuses on preprocessing the data before it fits into the prediction model. Preprocessing consists of different steps,

- Speech Segment: We use a library called 'Speech Segment'to load the speech audio
- Normalization: We make sure that every piece of sound ismade louder by 5 decibels. Then, we turn this sound into a list of numbers that show how loud it is at different times. This step is really important for getting the sound ready forthe next steps of processing.

2) *Model Training:* For training, every sample was labeled with specific emotional names, and then samples were module using into one flattened array of different classes named X hen labels were named Y s Y. Finally input variables were fed into a training model for model selection,used LSTM: long short-term memory. Here models were trained using different dense layers of neural networks.

E. For Recognition of Speech

The system used developed models with the help of a deep learning algorithm for development firstly system trained model with 8 emotions for which it used the dataset having audio files of 7 different categorical emotions (happy, sad, Neutral, disgust, angry, fearful, surprised total 2800 files = 2 actor's x 200 phrases x 7 emotions per actor.

F. Knowledge-driven

The received audio will be first converted into text which will be extracted for evaluation. The concept of speech-to text conversion has been used here.

Further keyword extraction is done from the sentence. The First syntax will get checked from the received sentence and then semantic correctness will get checked. Owing to this correctness of the answer will be checked based on thekeywords present in our dataset. It will now get matched withthe given answer and based on the keyword's presence the score will be generated. In addition to this syntax and semantics scores will also be generated. Once all the scores are generated results will be declared based on the average score. There will be a threshold of specific points set by the interviewer. Above that threshold all, the students will get selected.

G. E- Module for Scoring

The scoring system of our system is divided into 3 parts:facial emotion score, confidence score and knowledge base.

1) Emotion score has a weightage of 20% which is in therange of 1 to 7.
2) Speech has a weightage of 20% which is divided into positive or negative.
3) Confidence with a score of 10%.
4) Knowledge has a score of 50% which is again divided intosubtypes which are given below.

- Semantic Correctness has a weightage of 10%.
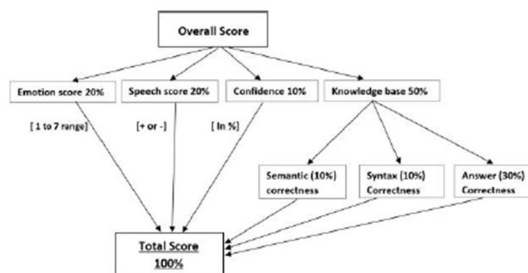- Syntax Correctness also has 10%.
- Answer Correctness is 30%

Fig.4 Scoring Module

IV.   RESULT

To evaluate the model performance, we used three types of parameters, which are emotion, knowledge base, and confidence. Facial expression is very important for judging the emotion of humans. Also, face recognition is essential for the system because it supplies security by not giving unauthorized candidates for interviews.
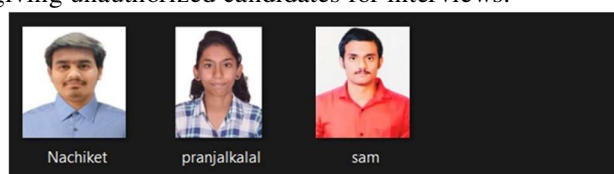

Fig.5 Databases of candidate


```
emotions = ["angry", "disgust", "fear", "happy", "neutral", "sad", "surprise"]

# Initialize the webcam
cap = cv2.VideoCapture(0)
```

Fig.6 7-Emotion Predication

After that, the model predicts the value of the emotion of thephoto given in fig. 5. And next it converts the value of emotion to the lab of the emotion. There is one happy labelled image that our model predicted as neutral.

Fig.7 Emotion Detection


```
1  # [happy, neutral, contempt, sad, angry, disgust, fear]
2  # here emotion taken as inndex of above list
3  emotional_score=[5,4,3,4,0,0,0,1,2,1,1,2,2,1,1,3,2,1,2]
4  a=[]
5  b=[]
6  temp=[6,5,4,3]
7  for i in emotional_score:
8      if i in temp:
9          a.append(i)
10     else:
11         b.append(i)
12  ans=(abs(len(a)-len(b))/20)*100
13  # here 0 means negative and 1 means positive
14  speech_score=[0,0,0,0,1,1,1,1,0,0,1,0,0,1,1,1,0,0,1,1]
15  confidence=87
16  knowledge_base=[76,76,76*3]
17  print("overall_score",((ans+(sum(speech_score)//2)*10+confi
```

```
overall score 70.875
```

Fig.8 Score Evaluation Algorithm

As the score plays a very crucial part in our system so it should be handled very carefully so we have built this algorithm to calculate the overall score shown in fig.7.
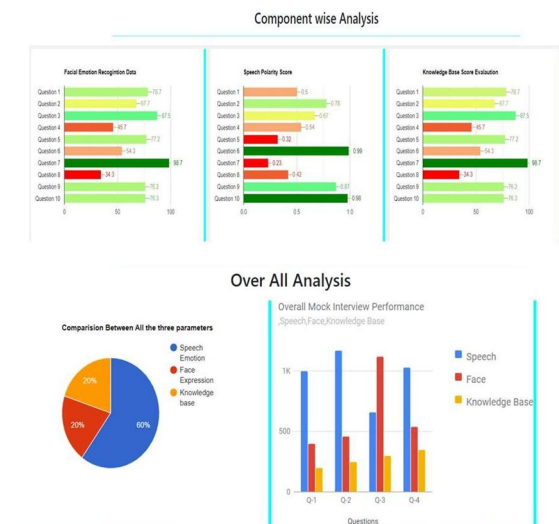

Fig.9 Over All Analysis Result

V. ADVANTAGES

As an AI Mock Interview Evaluator, the "EmoConfident Interviewer" provides a number of important benefits. It guarantees constant, unbiased feedback and data-driven insights, assisting applicants in identifying their areas of strength and growth. It effectively manages a large number of applicants and offers tailored comments to ease interview nervousness. It is accessible around-the-clock. The AI provides a scalable solution that lowers training costs and saves HR time by evaluating soft skills and simulating numerous scenarios. Advanced analytics facilitate trend analysis and performance monitoring, and it is a flexible tool to improve interview preparation and assessment due to its industry adaptability and integration capabilities.

VI. CONCLUSION

This study presents a system for applicants who, due to shyness or lack of confidence, do not perform well in interviews. The dynamic mock-interview system that is suggested in this work is. The algorithm uses the live video as input and evaluates the candidate's performance based on three factors: confidence, emotion, and knowledge. The system determines the final score based on the interview performance of the candidate. Kind criteria like a candidate's degree of confidence are absent from the current system. We employed one more metric, confidence, for the curated result. Additionally, a thorough examination of the interview candidates' qualities can be obtained through our system. Speech frequencies can be used for the confidence check in the system, and facial expressions can be used to verify emotion. CNN was utilized for facial recognition.

REFERENCES

[1] Chou, Yi-Chi, Felicia R. Wongso, Chun-Yen Chao, and Han-Yen Yu. "An AI Mock-interview Platform for Interview Performance Analysis." In 2022 10th International Conference on Information and Education Technology (ICIET), pp. 37-41. IEEE, 2022.

[2] Dissanayake, Dulmini Yashodha, Venuri Amalya, Raveen Dissanayaka, Lahiru Lakshan, Pradeepa Samarasinghe, Madhuka Nadeeshani, and Prasad Samarasinghe. "AI-based Behavioral Analyzer for Interviews/Viva." In 2021 IEEE 16th International Conference on Industrial and Information Systems (ICIIS), pp. 277-282. IEEE, 2021.

[3] Salvi, Vikash, Adnan Vasanwalla, Niriksha Aute, and Abhijit Joshi. "Virtual Simulation of Technical Interviews." In 2017 International Conference on Computing, Communication, Control and Automation (ICCUBEA), pp [4] Chou, Yi-Chi, and Han-Yen Yu. "Based on the application of AI technology in resume analysis and job recommendation." In 2020 IEEE International Conference on Computational Electromagnetics (ICCEM), pp. 291-296. IEEE, 2020.

[5] Priya, K., S. Mohamed Mansoor Roomi, P. Shanmugavadivu, M. G. Sethuraman, and P. Kalaivani. "An Automated System for the Assessment of Interview Performance through Audio & Emotion Cues." In 2019 5th International Conference on Advanced Computing& Communication Systems (ICACCS), pp. 1049-1054. IEEE, 2019.

[6] Moschona, Danai Styliani. "An affective service based on multi-modal emotion recognition, using eel enabled emotion tracking and speech emotion recognition." In 2020 IEEE International Conference on Consumer Electronics-Asia (ICCE-Asia), pp. 1-3. IEEE, 2020.

[7] Parra-Gallego, Luis Felipe, and Juan Rafael Orozco-Arroyave. "Classification of emotions and evaluation of customer satisfaction from speech in real world acoustic environments." Digital Signal Processing 120 (2022): 103286.

[8] Mandal, Rubi, Pranav Lohar, Dhiraj Patil, Apurva Patil, and Suvarna Wagh. "AI-Based mock interview evaluator: An emotion and confidence classifier model." In 2023 International Conference on Intelligent Systems for Communication, IoT and Security (ICISCoIS), pp. 521-526. IEEE, 2023.

[9] Yang, Jingwen, Zelin Chen, Guoxin Qiu, Xiangyu Li, Caixia Li, Kexin Yang, Zhuanggui Chen, Leyan Gao, and Shuo Lu. "Exploring the relationship between children's facial emotion processing characteristics and speech communication ability using deep learning on eye tracking and speech performance measures." Computer Speech& Language 76 (2022): 101389.

[10] Wang, Qian, Mou Wang, Yan Yang, and Xiaolei Zhang. "Multi-modalemotion recognition using EEG and speech signals." Computers in Biology and Medicine 149 (2022): 105907.Punit Gupta., Jasmeet Chhabra " IoT based Smart Home design using power and security management", vol. 12, pp.12-25, 2016.

[11] Ftoon Abu Shaqra, Rehab Duwairi, Mahmoud Al-Ayyoub, "Recognizing Emotion from Speech Based onAge and Gender Using Hierarchical Models" The 10th International Conference on Ambient Systems, Networks and Technologies (ANT) April 29 – May 2, 2019, Leuven, Belgium ELSEVIER

[12] Denae Ford, Titus Barik, Leslie Rand-Pickett, Chris Parnin "The Tech-Talk Balance: What TechnicalInterviewers Expect from TechnicalCandidates",2017.

[13] Dongdong Li, Jinlin Liu, Zhuo Yang, Linyu Sun, Zhe Wang," Speech emotion recognition using recurrentneural networks with directional self-attention",2021, IEEE

[14] Julie E. Sharp, "Work in Progress: Using Mock Telephone Interviews with Alumni to Teach Job Search Communication",2022, IEEEZigBeeand PLC," IEEE Trans. Consumer Electron, vol. 60, no. 2, pp. 198- 202, May 2014.

[15] D. Shin, K. Chung and R. Park "Detection of Emotion Using Multi- Block Deep Learning in a Self-Management Interview App", Applied Sciences, vol.09, no.22, 2019.

[16] I. Stanica, M. Dascula, C. Bodea and A. Moldoveam "VR Job InterviewSimulator: Where Virtual Reality Meets Artificial Intelligence For Education", 2018 Zooming Innovation in Consumer TechnologiesConference (ZINC), 2018.

[17] H. Suen, K. Hung and C. Lin "Intelligent video interview agent used to predict communication skill and perceived personality traits", Human- centric Computing and Information Sciences vol.10, no.03, 2020.

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Call : 08813907089   (24*7 Support on Whatsapp)