VIVA-TECH IJRI
ANNUAL RESEARCH JOURNAL
VIVA-TECH INTERNATIONAL JOURNAL FOR RESEARCH AND INNOVATION
ANNUAL RESEARCH JOURNAL
ISSN(ONLINE): 2581-7280

Smart Interviews Using AI

Aditi More[1], Samiksha Mobarkar[2], Siddhita Salunke[3], Reshma Chaudhari[4]
[1](Computer Engineering Department, VIVA Institute of Technology, India)
[2](Computer Engineering Department, VIVA Institute of Technology, India)
[3](Computer Engineering Department, VIVA Institute of Technology, India)
[4](Computer Engineering Department, VIVA Institute of Technology, India)

Abstract : *With the advent in technology, a lot of our common things have become smart. But our interviewing system still seems to be stuck at same point. If one has low marks it's fine but a person with a bad personality cannot be hired even if they are satisfying in other aspects, as they do more harm than good. This is the reason online interviews or chatbots are not preferred as many are of the opinion that although every other detail can be thoroughly checked, there is no way they can correctly have a grasp of the interviewee's personality.Because of COVID-19 pandemic all interviews are taken online but their concerns have increased regarding the aforementioned point. Taking this into consideration, we have built an interview system that analyzes the personality traits of the candidates with the help of facial and speech emotion recognition whose resumes have been approved. For facial emotion recognition we have used CNN model and for speech emotion recognition we have used Google API.*

Keywords - *Artificial Intelligence, Convolutional Neural Network, Deep Neural Network, Facial Emotion Recognition, personality, Speech emotion recognition*

I. INTRODUCTION

"Interview" means a one-to-one or one-to-many conversation between an interviewee and an interviewer or a panel. The panel or interviewer will ask questions to which the interviewee will reply, providing information. This is what, it generally means. The way these interviews are taken has also evolved, from face-to-face and in person to videoconferencing to AI chatbots, keeping up with the passing eras.

But somehow, today's interview system has stagnated at some point. Even with the advancement, many still prefer the old way of interviewing the potential candidates. Maybe the reason is that, while other aspects can be checked online, you can't have a grasp on the candidate's personality, which can be considered the most important. Our current interview system is a far cry from intelligent systems found to be implemented across the world. Not only is it time consuming but also needs a lot of efforts and attention from our side. Such a system cannot be said to be efficient considering any of the above-mentioned factors. To analyze the personality of a candidate a lot of parameters are considered such as facial expression, speech, emotion recognition, sentiment analysis, handwriting or text analysis etc. Our project is specifically built to analyze the personality traits of candidates whose resume are passed. This process involves and considers a lot of factors.

Out of all of these, we have decided to use facial and speech emotion recognition and analysis to build our project. Previous Work

II. PREVIOUS WORK

The Yu-Sheng Su, et.al [1], proposed a real-time image and video processor enabled with an artificial intelligence (AI) agent that can predict a job candidate's behavioral competencies according to his or her facial expressions. This is accomplished using a real-time video-recorded interview with a histogram of oriented gradients and support vector machine (HOG-SVM) plus convolutional neural network (CNN) recognition. Different from the classical view of recognizing emotional states, this prototype system was developed to automatically decode a job candidate's behaviors by their micro-expressions based on the behavioral ecology view of facial displays (BECV) in the context of employment interviews using a real-time video recorded interview. Alin Dragos Bogdan Moldoveanu, et.al [2], proposed a VR Job Interview Simulator which has the

purpose of helping software engineers increase their job interview performances by practicing their hard and soft skills. The VR-Job application includes a series of innovative technologies, such as virtual reality, chatbots, and measurement of the electrodermal activity (EDA), facial recognition or emotion analysis. All three types of immersion- sensory immersion, mental immersion and emotional immersion have been tried to accomplish. Computer vision and machine learning are used together to achieve certain tasks, such as facial detection, semantic analysis or emotion recognition. Electrodermal activity (EDA) is used to track the changes in the conductance of the skin. It is related to emotional and cognitive processes of the human body. Another parameter which can be useful for identifying stress and nervousness is the heart rate. All these can be tracked by using various devices, including skin monitor sensors, fitness bracelets or even smartwatches. Hung-Yue Suen, et.al [3], proposed an asynchronous video interview (AVI) platform with an artificial intelligence (AI) decision agent based on a TensorFlow convolutional neural network (CNN), called AVI-AI, that can be used to partially displace human raters' work in the initial stage of employment screening and to successfully predict a job candidate's communication skills and personality traits is developed. The experimental results show that AVI-AI can predict not only a candidate's interpersonal communication skills but also his or her openness, agreeableness, and neuroticism, as perceived by experienced human resource professionals. The interrater reliability values were all acceptable to support the ground truth assumption. Sarthak Katakwar, et.al [4], proposed a system which uses Convolutional neural network (CNN). The personalized details of some sample candidates and their features released are used to train the APR model, which uses specific CNN, built using the Python engine and TensorFlow deep learning. Before uploading images to the neural network model, things are made standard by adjusting the feature value range to [0, 1]. The extracted elements are then combined with other elements and presented in the extraction layer for final classification. Hung-Yue Suen, et.al [5], proposed an end-to-end AI interviewing system developed using asynchronous video interview (AVI) processing and a TensorFlow AI engine to perform automatic personality recognition (APR) based on the features extracted from the AVIs and the true personality scores from the facial expressions and self-reported questionnaires of 120 real job applicants. The experimental results show that the AI-based interview agent can successfully recognize the "big five" traits of an interviewee. The AI-based interview agent can supplement or replace existing self-reported personality assessment methods that job applicants may distort to achieve socially desirable effects. Dong Hoon Shin,et.al [6], proposed the detection of user emotions using multi-block deep learning in a self-management interview application. Unlike the basic structure for learning about whole-face images, the multi-block deep learning method helps the user learn after sampling the core facial areas (eyes, nose, mouth, etc.), which are important factors for emotion analysis from face detection. Through the multi-block process, sampling is carried out using multiple AdaBoost learning. For optimal block image screening and verification, similarity measurement is also performed during this process. A performance evaluation of the proposed model compares the proposed system with AlexNet, which has mainly been used for facial recognition in the past. As comparison items, the recognition rate and extraction time of the specific area are compared. The extraction time of the specific area decreased by 2.61%, and the recognition rate increased by 3.75%. Eduard frant, et.al [7], proposed an architecture which is an adaptation of an image processing CNN, programmed in Python using Keras model-level library and Tensor Flow backend. The theoretical background that lays the foundation of the classification of emotions based on voice parameters is briefly presented. According to the obtained results, the model achieves the mean accuracy of 71.33% for six emotions (happiness, fear, sadness, disgust, anger, surprise), which is comparable with performances reported in scientific literature. A person's speech can be altered by various changes in the autonomic nervous system and affective technologies can process this information to recognize emotion. As an example, speech produced in a state of fear, anger, or joy becomes loud and fast, with a higher and wider range in pitch, whereas emotions such as sadness or tiredness generate slow and low-pitched speech. Some emotions have been found to be more easily computationally identified, such as anger or approval. Inshirah Idris,et.al [8], proposed a system which was developed for investigating the detection of speech emotion using different sets of voice quality, prosodic and hybrid features. There are a total of five datasets of emotion features experimented in this work which are: Two from voice quality features, One set from prosodic features and Two from hybrid features. The experimental data used from Berlin Emotional Database Multi-Layer Perceptron; Neural Network are used for classification. Results show that hybrid features gave better overall recognition rates compared to voice quality and prosodic features alone. The best overall detection of hybrid features is 75.51% while prosodic and voice quality features are 64.67% and 59.63% respectively. Pavol Harár, Radim el.at [9], has developed a method for Speech Emotion Recognition (SER) using Deep Neural Network (DNN) architecture with convolutional, pooling and fully connected layers. They used 3 class sets (angry, neutral, sad) of German Corpus containing 271 labeled recordings with a total length of 783 seconds. Audio files were split into 20ms segments without overlap. Voice Activity Detection (VAD) algorithm used to eliminate blank segments and divided all data into training (80%)validation (10%) and testing (10%) sets. DNN is optimized using Stochastic Gradient Descent. As input authors used raw data without and

feature selection. The trained model achieved overall test accuracy of 96.97% on whole-file classification. The model works well, the model does not have any pre-given context, and still it gives better results. According to the author, providing context can improve the efficiency to the next level. This model can be used to find out whether the person speaking is angry or happy. Hao Hu, et.al [10], the GMM supervector based SVM is applied to this field with spectral features. A GMM is trained for each emotional utterance, and the corresponding GMM supervector is used as the input feature for SVM. Experimental results on an emotional speech database demonstrate that the GMM supervector based SVM outperforms standard GMM on speech emotion Recognition. Since the gender-dependent emotion recognition system is preferred, they analyzed the confusion between different emotions in condition of separate-gender subject System proposes solution via Confusion matrix of GMM supervector based SVM for female subject and male subject individually constituting 5 emotions anger; fear; happiness, neutral, sadness. Tatjana Liogienė et.al [11], proposed a technique sequential forward selection for multistage emotion recognition.SFS technique is a greedy search algorithm with a relatively low demographic load. It removes the dataset of features by maximizing the efficiency of the feature dataset. By sequentially extending feature subset the efficiency is maximized. This SFS technique is a greedy search that gives a sub-optimal solution as not all the possible feature subsets are analysed. The proposed scheme gives higher classification accuracy than single stage classification scheme by 0.5-4.3 %. A.Revathy et.al [12], proposed adequacy of Hidden Markov Model tool compartment (HTK) for perceiving speech, speaker and emotion from the emotional speeches utilizing Mel frequency cepstral coefficients (MFCC) as a component. HTK preparing apparatuses are utilized for assessing the boundaries for portraying the HMMs for the speeches and their related records during recognition stage, obscure expressions are deciphered utilizing the recognition instruments of HTK, HCompV is used for computing overall mean and variance and generating prototype HMM. The presentation of speech and speaker recognition frameworks is viewed as great and is somewhat low for emotion recognition. This is presumably because of the utilization of same arrangement of speech of same arrangement of speakers in various emotions. The performance of the system is very much degraded for the noisy test speech without adaptive RLS filtering. Versatile RLS filtering is ended up being a decent strategy, since it diminishes the commotion without changing the speech frequencies. R. M. A. H. Manewa et.al [13], proposed system that detects facial emotions using deep learning and convolutional neural network, that is able to recognize important parts of the face, and maximum difficult dataset FER2013 is used to measure the performance of the proposed method. This unique work stands itself precise from existing works being as a comprehensive work that specializes in figuring out facial expressions, detecting feelings by way of analyzing the facial features and based totally on human emotions notifying the user approximately any security risks. The Deep Learning Technique while combining CNN-based Machine Learning with the help of Keras framework and TensorFlow backend concept is an efficient and accurate approach for detecting emotions by using facial expression in which the output can be determined in a completely quick duration. Finally, it is concluded that the Deep Learning technique with CNNs works better without the usage of any more training statistics or GPU. Noel Jaymon et.al [14], Facial expressions depict emotions and produce information on the personalities and thoughts of people. Tensorflow framework, Keras library and the Xception Architecture of CNN are used to train the model on the Fer2013 dataset. The model detected all 7 emotions on an image provided by the user but during the Real Time Detection of emotion the model lacked robustness. It successfully detected 6 out of 7 emotions. It gave the accuracy of 34%. Denis Rangulov et.al [15], proposed system that detects facial emotions using deep learning and convolutional neural network, that is able to recognize important parts of the face, and maximum difficult dataset FER2013 is used to measure the performance of the proposed method. This unique work stands itself precise from existing works being as a comprehensive work that specializes in figuring out facial expressions, detecting feelings by way of analyzing the facial features and based totally on human emotions notifying the user approximately any security risks. The Deep Learning Technique while combining CNN-based Machine Learning with the help of Keras framework and TensorFlow backend concept is an efficient and accurate approach for detecting emotions by using facial expression in which the output can be determined in a completely quick duration. Finally, it is concluded that the Deep Learning technique with CNNs works better without the usage of any more training statistics or GPU.

III. OVERCOMING LIMITATIONS

The current interview system which is in picture is still having many loopholes. Many of the existing systems still take the user's information manually before interview using different forms then try to evaluate the entered data after which they select the candidate for the interview. They interviews are asynchronous which is a biggest drawback because the interviewer will not know the pressure of handling live situations. Also the existing systems cannot recognize an interviewee's personality in a diverse participant population. As it is said, no matter how much a person is hardworking, has excellent grades and looks, is almost an ideal employee but doesn't have a good personality then there is no use of hiring them. They do more harm than good to the

company. Human recruiters will quite often be one-sided somehow or another. Whether or not they realize it, a few enrollment specialists might settle on recruiting choices in light of orientation, nationality, age, looks, etc. An AI focuses on significant factors like competitors' character, abilities, experience and capabilities.

One of the fundamental difficulties for HR spotters is to recognize the best ability out of the numerous applications they get every day. AIs can assist with taking out these manual errands as they are modified to get greatest proficiency as far as time, expenses and quality.

The limitations of the system can be overcome by our proposed system. We are going to display questions on screen and the candidate has to start recording his answers and during live streaming the emotions of the candidate will be extracted. In our method we are going to consider both facial emotions as well as speech emotions. This approach has advantages for both candidate and interviewer. It lets the employer as well as candidates to review their interview reports when convenient. The employer can select them whereas candidates get time to reflect on their answers and there is no fear of dealing with prejudices.

IV. PROPOSED SYSTEM

The proposed system is a website. It serves as a helping aid for conducting interviews with more ease and efficiency. It helps to eliminate candidate's fear of dealing with prejudice during interview. The candidate has to enter his details and those details are viewed on admin side. If the candidate qualifies the criteria the mail regarding the interview is sent on his registered email id. The selected candidate has to appear for the video interview followed by the speech interview. The reports of video interview and speech interview are generated at the end and they are displayed to the candidate as well as to the admin.

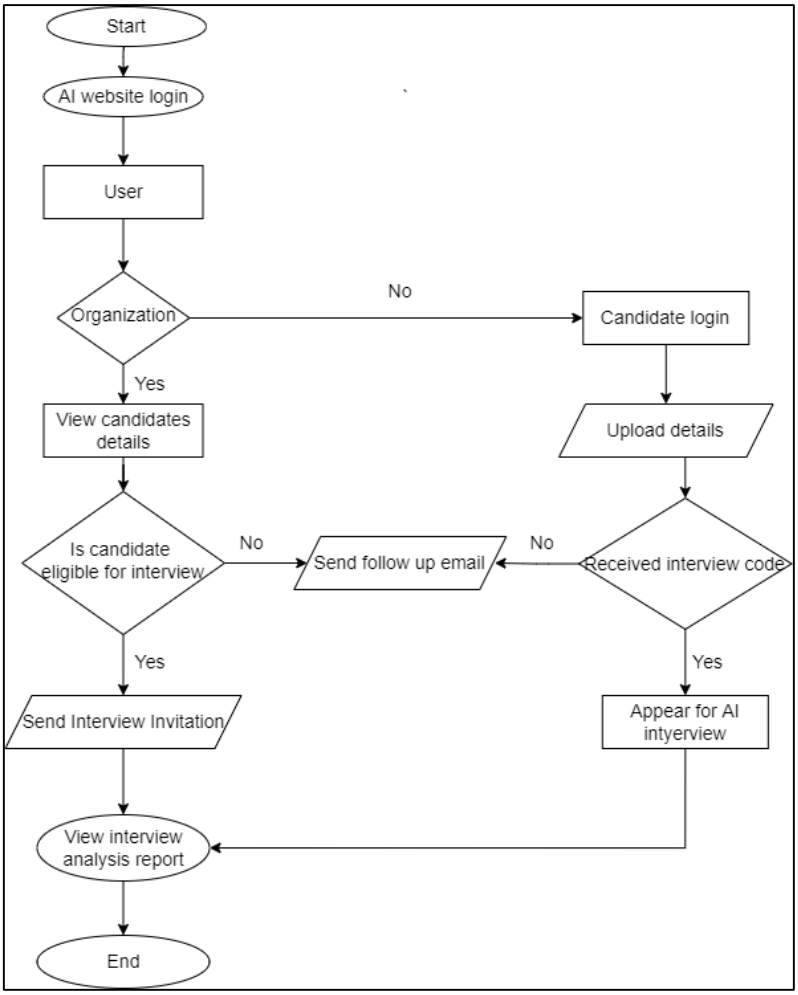

Fig.1: System flow diagram

The above fig.1 shows the system flow diagram of the proposed system. The flow starts from when the candidate receives the interview code for AI interview. On the basis of facial emotion detection and speech

emotion recognition the analysis report for the interview is generated. For facial emotion detection we have trained the model by using Kaggle dataset. The data consists images of faces of 48x48 pixel grayscale. These images have been registered so that the face covers large parts in the center of the image and fits in same amount of space in each image. Dataset of the moods have been categorized like the following (0=Angry, 1=Disgust, 2=Fear, 3=Happy, 4=Sad, 5=Surprise, 6=Neutral).

In this project, two columns namely "pixels" and "emotion" are considered in dataset. In emotion column a single digit ranging from 0 to 6 (inclusive) corresponding to the emotion present in the image is present. The second column contains a string for each image. These strings are space-separated pixel values in row order. The dataset contains only the "pixels" column and the task was to categorize the emotion column. The training set consists of approximately 28,709 examples. CNN/Conv-Net or Convolutional Neural Network is an algorithm of Deep learning. In this, the algorithm is fed with an input image so that it can assign learnable biases and weights. It also tries to find importance of the various aspects in the provided image. Each characteristic can be differentiated from one another using these networks. As compared to other algorithms (classification) the pre-processing needed in CNN is much lower.

For speech emotion recognition we need to train the model with sound extract various values and parameters which depicts and are associated different sentiments such features like MFCC, STFT, Contrast, Mel Spectrum, Chroma and Tonnetz are extracted from the audio clips of the dataset. Using deep learning models for Sentiment analysis or emotion analysis with natural language processing using Python package NLTK and another model DNN (Deep Neural Networks) for Audio Feature Extraction Python package Librosa.

For speech to text we have used speech to text library The Web Speech API has a principle regulator interface for this - SpeechRecognition - in addition to various intently related points of interaction for addressing language structure, results, and so on. The speech is perceived using microphone and the detected words are passed as string.

SpeechRecognition is the controller interface used for the speech recognition service. In Chrome it is known as webkitSpeechRecognition. SpeechRecognition handles the SpeechRecognitionEvent sent from the recognition service. SpeechRecognitionEvent.results returns a SpeechRecognitionResultList object representing all the speech recognition results for the current session.

V. CONCLUSION

The proposed system will conduct the interviews of the eligible candidates and then analyze the personality traits of candidates through video interviews. To analyze the personality of a candidate a lot of parameters are considered such as facial expression, speech, emotion recognition, sentiment analysis, handwriting or text analysis in this system we are using facial and speech emotion recognition as well as text emotion recognition.

Thus with the help of deep analysis using AI the assessment of an interview is beyond human undertaking and it prevents intended or unconscious biases that often prevent a fair recruitment process. It also makes the recruitment process more accurate and provides best results to organizations in less time.

While in this model we have used facial and speech emotion recognition as the parameters for personality analysis, in future we can add more factors. Increasing the accuracy of the model is also the focus.

Acknowledgement

We would like to express a deep sense of gratitude towards our mentor Prof. Reshma Chaudhari, Department of Computer Engineering for her constant encouragement and valuable suggestions. The work that we have been able to present is possible because of timely guidance and support.

REFERENCES

[1] Y. Sue, H. Suen and K. Hung “Predicting behavioral competencies automatically from facial expressions in real-time video-recorded interviews”, *Journal of Real-Time Image Processing, vol.18,* 2021, pp. 1011–1021.

[2] I. Stanica, M. Dascula, C. Bodea and A. Moldoveam “VR Job Interview Simulator: Where Virtual Reality Meets Artificial Intelligence For Education”, *2018 Zooming Innovation in Consumer Technologies Conference (ZINC),* 2018.

[3] 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.

[4] S. Katakwar, O. Mahamuni, N. Inamdar and S. Sadanand “Emotion and Personality Analysis in Recorded Video Interview Using TensorFlow*”, International Research Journal of Engineering and Technology (IRJET), vol.08, no.04,* 2021.

[5] H. Suen, K. Hung and C. Lin “TensorFlow-based Automatic Personality Recognition Used in Asynchronous Video Interviews”, *IEEE Access, vol.07,* 2019.

[6] 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*.

[7] E. Frant, I. Ispas, V. Dragomir, M. Dascalu, E. Zoltan and I. Stoica “Voice Based Emotion Recognition with Convolutional Neural Networks”, *Romanian Journal Of Information Science And Technology, vol.20, no.03,* 2017, pp. 222–240.

[8] M. Salam and I. Idris "Emotion Detection with Hybrid Voice Quality and Prosodic Features using Neural Network", 2014 *4th World Congress on Information and Communication Technologies (WICT 2014),* 2014.
[9] R. Burger and M. Dutta "Speech Emotion Recognition with Deep Learning", *4th International Conference on Signal Processing and Integrated Networks (SPIN),* 2017.
[10] H. Hu, M. Xu and W. Wu, "GMM Supervector Based SVM with Spectral Features for Speech Emotion Recognition", Acoustics, Speech and Signal Processing, ICASSP 2007, *IEEE International Conference on vol.04,* 2007.
[11] T. Liogienė and G. Tamulevičius, "SFS feature selection technique for multistage emotion recognition," *2015 IEEE 3rd Workshop on Advances in Information, Electronic and Electrical Engineering (AIEEE),* 2015, pp. 1-4, doi: 10.1109/AIEEE.2015.7367299.
[12] A. Revathy, P. Shanmugapriya and V. Mohan, "Performance comparison of speaker and emotion recognition," *2015 3rd International Conference on Signal Processing, Communication and Networking (ICSCN),* 2015, pp. 1-6, doi: 10.1109/ICSCN.2015.7219844.
[13] R. M. A. H. Manewa and B. Mayurathan, "Emotion Recognition and Discrimination of Facial Expressions using Convolutional Neural Networks.," *2020 IEEE 8th R10 Humanitarian Technology Conference (R10-HTC),* 2020, pp. 1-6, doi: 10.1109/R10-HTC49770.2020.9357008.
[14] N. Jaymon, S. Nagdeote, A. Yadav and R. Rodrigues, "Real Time Emotion Detection Using Deep Learning," *2021 International Conference on Advances in Electrical, Computing, Communication and Sustainable Technologies (ICAECT),* 2021, pp. 1-7, doi: 10.1109/ICAECT49130.2021.9392584.
[15] Rangulov, Denis & Fahim, Muhammad "Emotion Recognition on large video dataset based on Convolutional Feature Extractor and Recurrent Neural Network" *IEEE 4th International Conference on Image Processing. Applications and Systems (IPAS)*.