SMART INTERVIEWS USING AI

[1]Aditi S. More, [2]Samiksha S. Mobarkar, [3]Siddhita S. Salunke, [4]Reshma R. Chaudhari
Department of Computer Engineering
VIVA Institute Of Technology
Mumbai, India
[1]moreaditi10@gmail.com, [2]mobarkarsamiksha@gmail.com, [3]siddhitasalunke9311@gmail.com, [4]reshmachaudhari@viva-technology.org

Abstract— With the advent in technology, a lot of our common things have become smart. This has not only elevated the quality of our living standards but also increased our comfort and safety levels. Things like smart homes, smart traffic management, smart offices, smart surveillance etc. has been made readily available to the general public across the world. Many more features like this are readily evolving with our technological advance and will keep on going like this in the future. Now all interviews are taken online and not only marks but also other aspects are considered, the most important being personality. Taking this into consideration, we have built an interview system that analyzes the personality traits of the candidates whose resumes have been approved. Using the ideas mentioned in various research papers and combining it with suitable technology, this project has been built.

Index Terms— interview, personality, smart, technology, face emotion recognition, 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.

II. PREVIOUS WORKS

Yu-Sheng Su, et.al [1], aims to develop 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.

Alin Dragos Bogdan Moldoveanu, et.al [2], the VR Job Interview Simulator has the purpose of helping software engineers increase their job interview performances by practicing their hard and soft skills. 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.

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 candidates communication skills and personality traits.

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.

Hung-Yue Suen, et.al [5], proposed an end-to-end AI interviewing system which is 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 job applicants.

Dong Hoon Shin, et.al [6], this paper proposes detection of user emotions using multi-block deep learning in a self-management interview application. 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.

Eduard frant, et.al [7], the architecture 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.

Inshirah Idris, et.al [8], in this paper a system is 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.

Pavol Harár, Radim el.at [9], in this paper the author 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. 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.

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.

III. LIMITATIONS IN THE PREVIOUS WORKS

The current interview system is still lagging despite the advent in technology, where we still conduct it, in a time-consuming manner and use up too much effort. Many of the existing systems still take the user's information manually before interview after which the candidate is provided with a unique id to go with. Also the 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.

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.

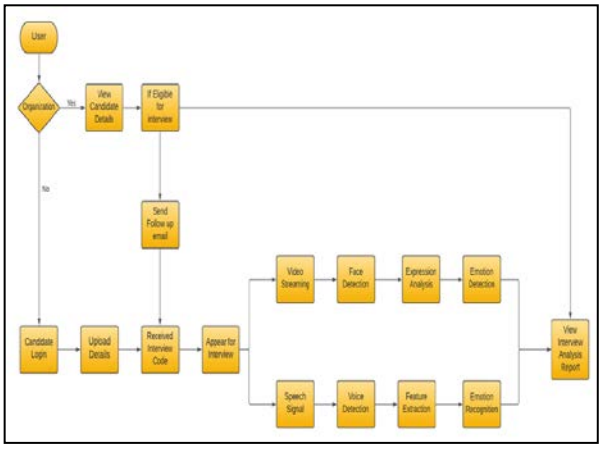

Fig.1: Block diagram

In Fig. 1 shows the block diagram of the proposed system. The block diagram has two deferent section. One is for facial emotion detection and other is for speech emotion recognition. The output of the two model will be checked.

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, the dataset contains two columns, "pixels" and "emotion". A single digit ranging from 0 to 6 (inclusive) corresponding to the emotion present in the image was contained in the "emotion" column. 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.

V. ALGORITHM

The algorithm for the proposed system is as follows:

There are actually two parts in this algorithm.

A] Speech Recognition -

Step 1: Import all necessary packages.

Step 2: Specify the path of the dataset.

Step 3: Loop through each file in the dataset folder and map labels for that particular file

Step 4: Loop through each file in the dataset folder and extract Acoustic features such as MFCC, STFT, Contrast, Mel Spectrum, Chroma and Tonnetz from the audio clips of the dataset.

Step 5: Concatenate Features data frame with associated emotion data frame.

Step 6: Split the data into training, testing and validation sets.

Step 7: Create a Deep Neural Network for classification.

Step 8: Fit the model to training data and validation data.

Step 9: Make predictions about emotion detected using test data.

B] Facial Emotion Recognition –

Step 1: Import all necessary packages.

Step 2: Specify the path of the dataset.

Step 3: Loop through each file in the dataset folder and map labels for that particular file.

Step 4: Loop through each file in the dataset folder and perform expression analysis by referring to the video clips of the dataset.

Step 5: Concatenate Features data frame with associated emotion data frame.

Step 6: Split the data into training, testing and validation sets.

Step 7: Create a Deep Neural Network for classification.

Step 8: Fit the model to training data and validation data.

Step 9: Make predictions about emotion detected using test data.

VI. RESULTS

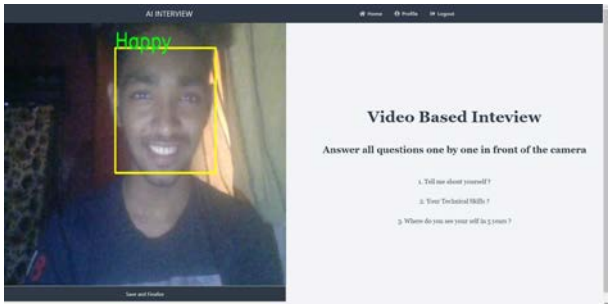

Fig. 2 : Video Interview

Fig. 2 shows the output of the model in GUI. This is how the video interviews will take place and on that basis the facial expression emotion recognition and analysis will be concluded.

Fig. 3 : Speech Interview

Fig. 3 shows the output of the model in GUI. This is how the speech interviews will take place and on that basis the speech emotion recognition and analysis will be concluded.

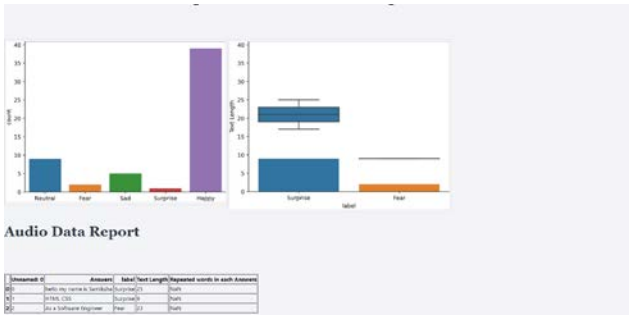

Fig. 4 : Report For Video and Speech Interview

Fig. 4 shows the report which has been generated on the basis of the various factors concluded by using facial expre- ssion, speech emotion recognition and text analysis.

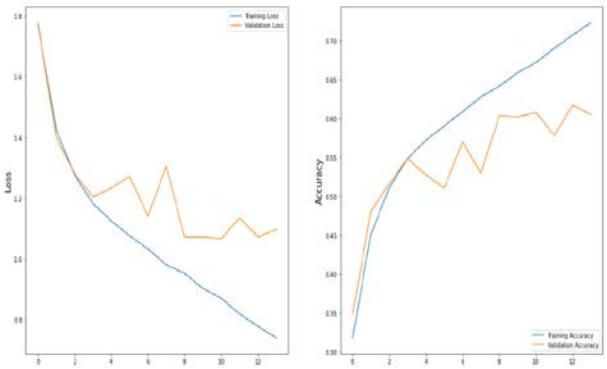

Fig. 5 : Accuracy on Test Data

Fig. 5 is the result obtained after testing the accuracy of the data. The graph on the left shows training and validation loss while the graph on the right shows training and validation accuracy.

VII. CONCLUSION AND FUTURE SCOPE

The proposed system will classify the resumes that it has received of all the 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.

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.

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.