Automated Video Interviews Using Machine Learning

Yash Agarwall[1]
Dept. of Information Science and Engineering
BMS College of Engineering
Bangalore , Karnataka,India

Shubham Pamecha[2]
Dept. of Information Science and Engineering
BMS College of Engineering
Bangalore , Karnataka,India

Shreyansh Bansal[3]
Dept. of Information Science and Engineering
BMS College of Engineering
Bangalore , Karnataka,India

Asst. Prof. Rashmi K.B[4]
Dept. of Information Science and Engineering
BMS College of Engineering
Bangalore , Karnataka,India

***Abstract*—**
After the pandemic, Every sector has experienced the impact of change. corona. Many people are unemployed and soon we will see how the world needs more job seekers. Since there is no such process where you can conduct interviews and screening, you must be ready to engage in the recruitment process with IT solutions.
Now we have voice recorders or skype phone alerts that are not yet undergone feature analysis. So, overcoming all these shortcomings and introducing a new era of the interview process is the main goal of the project.
With artificial intelligence systems, we can predict human skills and recommended types of work, as AI will be introduced in today's industry, namely recruitment agencies. By utilizing a range of machine learning algorithms, including the convolutional neural network, Naive Bayes, Random Forest, and SVM, significant progress can be made.. Employee sentiment, voice to text, face to face company verification process will be implemented in this project.
In the future, the system can access user logs that can analyze the effectiveness and efficiency of users. Also, employees are able to get benefit from this system by analyzing their emotions and personality, we can see such a system in future interviews.

Keywords— Naïve bayes, random forest, SVM, convolutional neural network (CNN), TensorFlow, emotion, company verification,resume parsing,image processing,

I. INTRODUCTION

Interviewing is a long process these days. Due to the pandemic, employees are too busy coming to work to interview. AI systems can help make things easier in many ways.
In the traditional method of interviewing candidates for specific jobs, the human resources unit of the organization requests employment based on the candidate's resume.
This personnel department Analyze a candidate's skills manually through their resume to see if they are a good fit for the jobresources unit of the organization requests employment based on the candidate's resume.
This personnel department Analyze a candidate's skills manually through their resume to see if they are a good fit for the job. They conduct interviews and the committee has the important responsibility of evaluating suitable candidates for the job. Here not only the skills but also the personality of the candidate is checked, as every recruited candidate must have the right attitude and discipline towards the work.
However, the conventional recruitment approach lacks technological integration. To enable future reference, personal audio-visual recordings can be stored. The primary objective of this application is to develop a system that utilizes artificial intelligence to identify emotions in facial expressions and vocal cues, as well as analyze resumes. Consequently, our decision was to construct an end-to-end interview system that employs AI. This system not only categorizes candidates' resumes but also assesses their distinctive personality traits by conducting asynchronous video interviews.
Work completed before this time is associated with many machine learning algorithms and can be very time-consuming and affect system performance. CNNs have been proven to perform well in image processing tasks. Capable of automatically processing images and inferring first impressions from cameras, this project is implemented with a semi-supervised DL model including CNN to develop an AI-based automated interviewer that can automatically identify candidates.

II. LITERATURE SURVEY

A. Overview of Convolution Neural Network(CNN)

Convolutional Neural Network (CNN) is a deep learning algorithm that can recognize and classify objects. It is a multilayer neural network designed to process visual input and perform very useful tasks such as image recognition, segmentation, and object recognition

Two key parts of a CNN:

- A complexity technique that shows certain analytical properties of the input.
- A fully connected layer that uses the definition function of convolution layers to predict the best definition.

The CNN architecture is inspired by the function and organization of the visual cortex and is designed to mimic the connectivity patterns of neurons. In a CNN, neurons are divided into three-dimensional structures. These sets of neurons analyze the features of the image, or each group of neurons specializes in recognizing a part of the image. The CNN then uses the layer-based predictions to generate the final output. It provides a vector of probability scores indicating that a particular feature belongs to a particular class.

B. Overview of Emotion and Emotion Recognition

The accuracy of recognizing emotions varies from person to person. Technology-assisted emotion recognition is currently in its infancy. The most common way to recognize emotions is through facial expressions in text, speech, and audio and video clips. In today's context, Murray and Arnott attempted to distinguish emotion from speech changes and intelligibility by looking at simulated emotions in speech and composite facial expressions (Table 1).

	Anger	Happiness	Sadness	Fear	Disgust
Speech rate	Slightly faster	Faster or slower	Slightly slower	Much faster	Very much slower
Pitch average	Very much higher	Much higher	Slightly slower	Very much higher	Very much lower
Pitch range	Much wider	Much wider	Slightly narrower	Much wider	Slightly wider
Intensity	Higher	Higher	Lower	Normal	Lower
Voice quality	Breathy, chest tone	Breathy, blaring	Resonant	Irregular voicing	Grumbled chest tone
Pitch changes	Abrupt, on stressed syllables	Smooth, upward inflections	Downward inflections	Normal	Wide downward terminal inflections
Articulation	Tense	Normal	Slurring	Precise	Normal

Human emotions play an important role in relationships. Programmed emotional recognition has been a valid research topic since its early days. So we are seeing progress in this field. Emotions are reflected through speech, hands, body gestures and looks. Therefore, the extraction and understanding of emotions is crucial for cooperation between human and machine response.

C. Overview of Company Verification

The skill set of a particular candidate is derived from the analysis of each resume.like Naukari.com has several profiles. A DNN is applied to develop and apply a model to these profiles.
Using the generated model, you can quickly determine which jobs are suitable for applicants with specific skill sets. Keras is an effective library to help develop deep learning models that are Easy to use. The DNN algorithm is applied to a total of 25 classes (job reference roles) and 10,000 rows to help organizations determine whether an applicant is a good fit for a particular job.

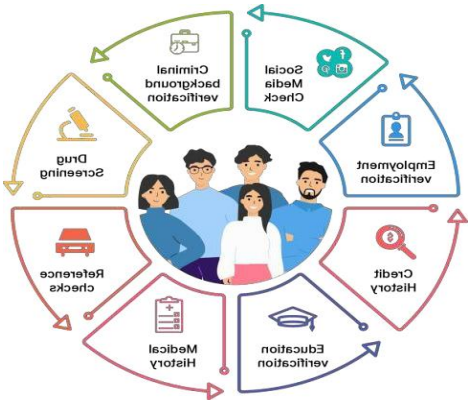

Figure 1: Steps in Company Verification

D. Overvieww of Dataset-Ryerson Audio-Visual Database ofEmotional Speech and Song(RAVDESS)

The Ryerson Auditory-Visual Database of Emotional Speech and Song (RAVDESS) was chosen as the first dataset for the model. There are several other datasets such as the Indian dataset, but RAVDESS is the most accessible, widely distributed and up-to-date sentiment dataset. It contains speech and audio-visual data of 12 men and 12 women expressing eight characteristic emotions in English.
Each sentence stimulus is a long seven-syllable sentence. Each facial expression is created with natural emotional intensity and strong emotional intensity, and an additional neutral expression is added. The age of this actor is between 21 and 33 years. The gender distribution is even with 12 males and 12 females and ethnicity is predominantly Caucasian. The main language of this actor is English with a neutral North American accent. Humans have a low accuracy rate of 67% in this dataset, which shows that classifying emotions is not that easy even for humans, and we aim to beat this accuracy rate in understanding emotions.

III. PREVIOUS WORKS

Authors Lei chen, Ru zhao, and Chee Wee Leong, in a research paper on automated video interview judgment against a large collection collected online, AVI with an AI decision agent to predict the communication skills of job seekers, introduce the development and evaluation of this platform. . The personality characteristics of this finding suggest that while the AVI-AI can effectively predict some traits, it may have limitations in predicting other traits, assessing conscientiousness and extraversion in that context. be This highlights the complexity of doing the work.

Another article, "Big 5 Personality Review Trait Prediction" Using TensorFlow Author Manisha Nilugonda[1] and Dr. Karanam Madavi[2] study the recognition of Big 5 personality traits with facial expressions. It proposes a non-surgical device that can accurately determine these personality traits based on facial features alone. This device has potential applications in professional development, health assistance, and diagnosis of mental health problems and physical illnesses that manifest as personality changes. In this paper, we provide a brief review on Big 5 feature detection using facial expressions and discuss various methods used for this purpose. Finally, we provide a comparison of different approaches to identifying the Big 5 personality traits. This article does not refer to specific models that exhibit these features detection using facial expressions and discuss various methods used for this purpose. Finally, we provide a comparison of different approaches to identifying the Big 5 personality traits. This article does not refer to specific models that exhibit these features.

On the other hand, the article "Interpretation of CNN Models for Regression of Apparent Personality Traits" by Carls Ventura and David Masip focuses on the interpretability of deep learning models for inference of apparent personality traits. The researchers validated the advanced models, specifically ignoring audio signals and analyzing video content. They found that facial information plays an important role in predicting personality traits. Based on this finding, we trained a model using only face regions to achieve improved accuracy.To gain insight into trait prediction, researchers used recursive visualization tools. The results showed that specific areas of the face affect trait prediction. The face parts detector is automatically created from the last blended layer The presence and activation of these units in the network is related to the prediction of personality traits. They found a significant relationship between activation of specific units and action units.

The paper "Learning short-term and long-term facial behaviors for personality recognition" by Dong Mingrui describes the recognition of personality traits using short-term and long-term modeling of facial behaviors. A new framework for empirical results highlights the effectiveness of examining long-term facial behavioral modeling and combining it with automatically predicted facial features to improve personality recognition accuracy.

Ashmita Verma and Prakhar Deep's paper, "Cognitive Profiling for Job Recruitment", presents a clustering-based system studied by Keshgra Aman to assess the personality and cognitive abilities of job seekers for recruitment purposes. I'm here. Their research examines the relationship between personality, cognitive ability, and age, revealing underlying patterns and tendencies. The results are consistent with previous studies, but more extensive research is needed to confirm and extend these findings.

IV. PROPOSED SOLUTION

We are developing a character recognition system using asynchronous video analysis. I am using the TensorFlow library to create this system. Create a model to classify resumes using machine learning algorithms such as Naive Bayes, Support Vector Machines (SVM) and Random Forest. And the same algorithm is used to analyze songs that are converted into text.
CNN algorithms and facial features capture a user's face with a camera and perform facial feature extraction to generate results with facial emotions such as happy faces, sad faces, perfect poses and smiles. A convolutional neural network is used for face classification. Interviewer tone analysis uses supervised learning algorithms such as Random Forest, Naive Bayes and SVM models.

The developed system first obtains a registration of candidates willing to be interviewed, then takes a few photos of the candidate to verify their identity. After registration, candidates can register in the system by uploading their details such as name, background, university, etc. Candidates must upload a resume and a voice recording, which is sent to interviewers to verify the job suitability of the candidate generated by the system through a trained dataset. The candidate's emotions (e.g., sad or happy) are then recognized by trained algorithms as neutral, surprised, or angry.
and based on that a chart with the emotion count is being generated and all this information is further sended to the interviewer.

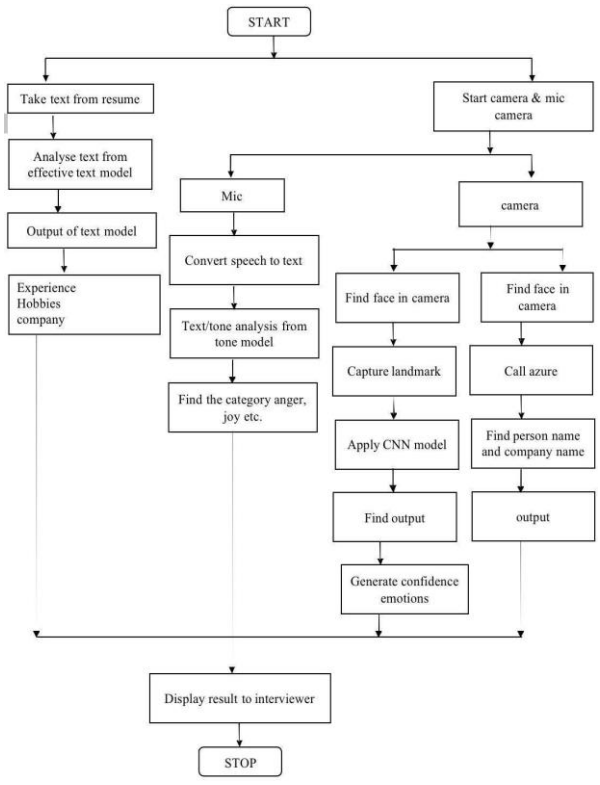

Figure 2: System Architecture Diagram

A. RESULTS AND DISCUSSIONS

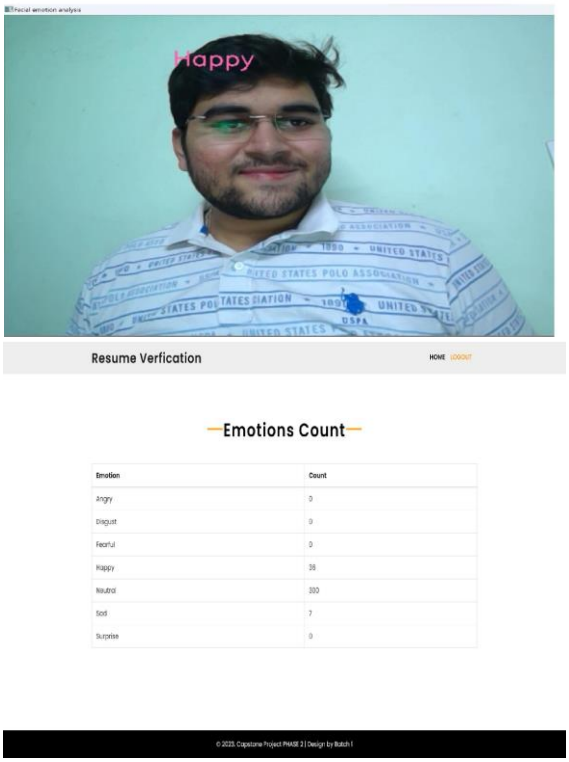

Figure 3: Screenshot of the emotion detected and emotion count chartof the candidate .

Figure 3 shows how candidate video interviews are filmed based on facial expression emotion recognition. A sentiment score chart will be developed in the future and can be integrated with a set of questions that candidates must answer to further reinforce this theme. And all this is recorded in the system along with happy, sad, neutral, angry, surprised, etc. reactions of the volunteer when answering the question.

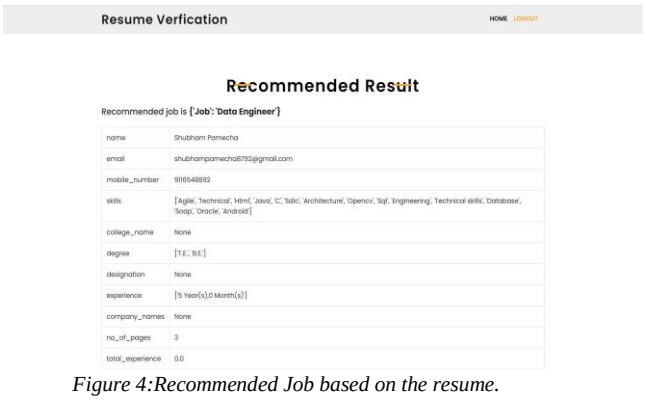

Figure 4:Recommended Job based on the resume.

Candidates are required to upload a resume which is sent to our system for analysis. After scanning resumes and using trained data, our developed system suggests suitable job profiles for candidates, making it easy and time-saving for HR professionals. Show candidates, the system we propose allows HR professionals to quickly determine which job is a good fit for a candidate based on a candidate's resume.

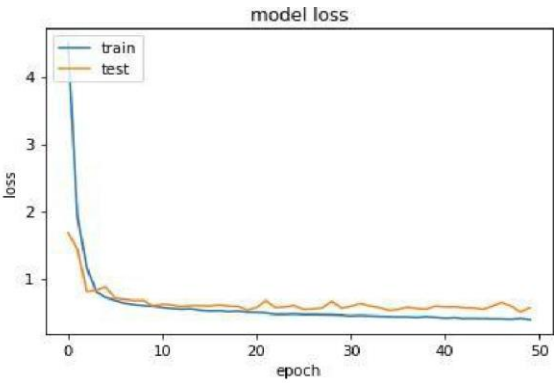

Figure 5: Model loss

This project successfully implemented resume parsing and job recommendation functions. The resume parsing task involved extracting relevant information from resumes using a machine learning CNN model. Based on the parsed data, the system provided accurate job recommendations from a pool of 25 jobs.

In addition, the project includes speech recognition capabilities that identify the gender of the speaker and classify emotions such as sadness and anger based on their tone.

The initial accuracy of the model was 0.5500, which significantly improved to 0.9150 after training the data. Likewise, the initial information loss was 1.2252, but after training and testing it was reduced to 0.1619.

As a result, the project successfully performed resume parsing, job recommendation, and speech recognition tasks. The implemented machine learning model showed promising results in accurately parsing resumes, providing job recommendations and classifying emotions based on speech tone.

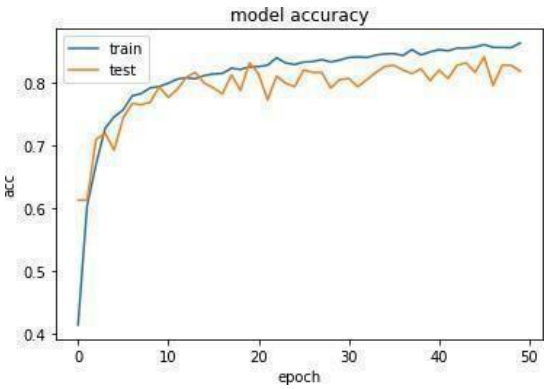

Figure 6: Model accuracy

In the field of video emotion recognition, this project successfully identified various emotions such as happiness, sadness, disgust and anger. This feature is useful for automated job recommendation purposes. The system had an initial accuracy of 0.2560, which improved significantly with subsequent training and reached 0.9476. Similarly, the initial data loss started at 1.8090 and decreased to 0.1471 after rigorous testing and training.

Using this model, the system effectively recommends vacancies and helps companies with recruitment projects. Accurately identifying emotions in videos strengthens the overall system and enables more personalized and relevant career recommendations.
Educational index:
During the training and testing process, we split the dataset 10/90. The model achieved significant accuracy after 50 epochs, with a corresponding loss of validity of 54% and a validity accuracy of 82%. The classification report (Figure 7) and confusion matrix (Figure 8) provided insight into the accuracy of each emotion category and highlighted potential areas for future model refinement.

	precision	recall	f1-score	support
neutral	0.91	0.88	0.90	59
calm	0.81	0.70	0.75	43
happy	0.69	0.90	0.78	68
sad	0.88	0.74	0.81	78
angry	0.88	0.85	0.86	85
fearful	0.93	0.81	0.86	78
disgust	0.72	0.84	0.78	62
surprised	0.84	0.89	0.86	53
accuracy			0.83	526
macro avg	0.83	0.82	0.82	526
weighted avg	0.84	0.83	0.83	526

Figure 7: Classification Report

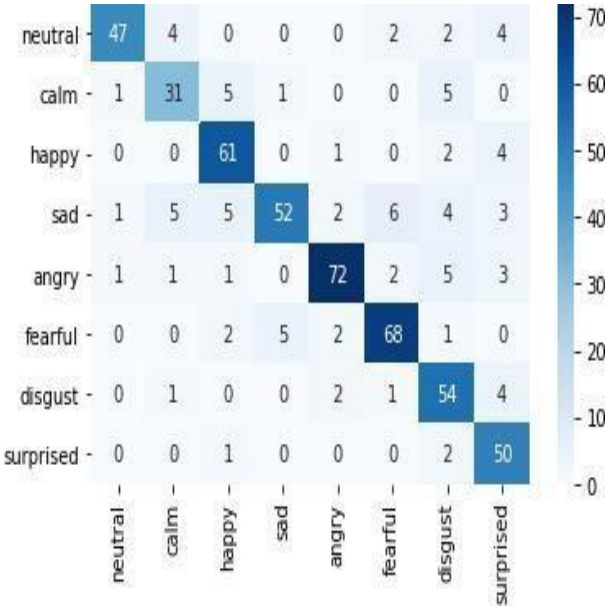

Figure 7: Classification Report

V. CONCLUSION

The proposed system places emphasis on character computation and the development of an AVI (audio-visual interface) that integrates semi-supervised deep learning models. This model effectively and automatically identifies job candidates during interviews, addressing the limitations of conventional personality estimation methods. The AVI's automatic personality recognition (APR) approach achieves an impressive accuracy rate exceeding 90%, surpassing previous studies on non-verbal communication. By leveraging multimodal features learned by deep neural networks, the system enhances its prediction performance of personality traits. Advanced APR systems have the potential to complement or replace self-reported personality assessments, reducing bias resulting from social desirability among job seekers.

Future work in this field aims to integrate visual methodologies and prosodic features to enhance understanding of interviewees' personalities. It is important to note that this study relied on a specific group of professionals as participants, which may restrict the generalizability of the findings, necessitating future research involving more diverse participant populations. Additionally, to enhance the system's accuracy, we will explore different machine learning algorithms, incorporate additional facial features for improved face analysis, and compare the results with alternative algorithms to determine the most effective approach. In summary, future work will concentrate on enhancing the system through these proposed modifications.

REFERENCES

[1] C. Wang, Q. Wang, K. Ren, N. Cao, and W. Lou,"Toward secure and dependable storage services in cloud computing", IEEE Transactions on Services Computing, vol. 5, no. 2, pp. 220–232, 2012

[2] A New Approach for Automatic Face Emotion Recognition and Classification Based on Deep Networks. (n.d.). A New Approach for Automatic Face Emotion Recognition and Classification Based on Deep Networks | IEEE Conference Publication |

[3] A. Verma, P. Deep, K. Aman, V. Khemchandan i, S. Chandra and G. Sharma, "Cognitive Profiling fo rJobRecruitments:AClusteringBased Approach," 20 21 International Conference on Computational Perfo rmance Evaluation (ComPE), Shillong, India, 2021, pp.604608, doi: 10.1109/ComPE53109.2021.9751962.

[4] D. Mingrui and Z. Dong, "Learning Short-TermandLongTerm Facial Behaviors for Personality Trai ts Recognition," 2022 IEEE 2nd International Conference on Software Engineering and Artificial Intelligence (SEA I),Xiamen, China, 2022, pp. 41-45, doi: 10.1109/SEAI55746.2022.9832040.