Predictive Analysis of Remote Interviews using Computer Vision

R. Poornima[1], S. Kaja Mohiddin[2], R.V.S. Yaswanth Kumar[3], P. Deekshith[4]
[1,2,3,4] *Department of Computer Science and Engineering,*
Raghu Engineering College, Visakhapatnam, India

***Abstract*—In this modern era, analyzing human behavior through videos is not a complex task, we have many advancements in computer vision and natural language processing (NLP) through which we can accomplish the desired task. This study presents a web based multimodal application that performs human behavior analysis based on multiple characteristics such as gaze tracking, emotion recognition, and sentimental speech analysis. The application uses OpenCV, Media Pipe, Deep Face, and NLP based sentimental analysis to process facial expressions, gaze direction in the eyes, and spoken content. Gaze estimation is performed using facial landmark detection, providing insights into the attentiveness and focus of a user. Emotion recognition is carried out using a deep learning facial analysis library which provides pre trained models for emotion detection, categorizing emotions such as happiness, sadness, anger, disgust, fear and surprise. Simultaneously, speech analysis identifies tone, sentiment, and linguistic patterns including the usage of filler words and pauses, to assess the amount of engagement and the overall confidence levels. The extracted behavioral features are processed and visualized through an interactive web interface, enabling real time feedback for applications in education, human computer interaction and psychological assessments.**

The proposed framework enhances traditional video analysis by integrating multiple behavioral indicators, providing a holistic comprehension of human engagement and emotional state. By synergistically combining visual and auditory cues, the system offers a robust evaluation of user behavior, which can be applied in remote learning environments, interview assessments, and human centric AI applications. The real time processing capabilities of the application guarantee seamless user interaction while preserving accuracy and reliability. Experimental results demonstrate the system's efficacy in identifying behavioural patterns, rendering it a valuable resource for researchers and practitioners in cognitive science, affective computing, and auto mated interaction analysis. Future enhancements may encompass advanced deep learning architectures and expanded datasets to augment the accuracy of gaze estimation and sentiment classification, thereby further advancing the field of multi modal behavior recognition.

***Index Terms*—Emotion recognition, gaze detection, speech analysis, deep learning, Flask application, behavioral analysis, machine learning, natural language processing, video processing, engagement metrics.**

I. INTRODUCTION

1.1 Background and Motivation

Understanding human behavior plays an important role in many of the domains, such as education, healthcare, scrutiny and human computer interaction. A deeper analysis of emotions, gaze speech and lot more communicative cues shows a person's cognitive processes and emotional state, making it suitable for real world applications. Traditionally, most of the behavioral analysis assessment is relied on manual observations and it is time consuming, inconsistent and biased. Advances in artificial intelligence (AI) and deep learning have enhanced automated, video based behavioral analysis in a major way to assess user loyalty, emotional conditions and attention in real time. Technologies such as facial recognition, gaze tracking, and speech analysis allow researchers to extract meaningful insights from video data with high accuracy. This project leverages these advancements to develop a multimodal behavioral analysis system capable of evaluating engagement, confidence, nervousness, and emotional state. By integrating OpenCV, Media Pipe, Deep Face, and speech recognition techniques, this system offers an objective, scalable, and comprehensive approach to behavioral analysis.

1.2 Importance of Behavioral Analysis in Various Applications Behavioral analysis plays a vital role in several fields:

a) Education: Teachers can keep an eye on students gaze patterns to detect disengage d students or those who are struggling with the lecture, allowing customized interventions to improve learning curve of each student.
b) Healthcare: Emotion detection and speech analysis can help in identifying and monitoring mental health dis orders like anxiety, depression, and autism spectrum disorders.
c) Surveillance and Security: Facial expressions and gaze analysis can enhance security systems by identifying suspicious behaviors and preventing potential threats.
d) Marketing and User Experience: Businesses can ana lyze customer responses to advertisements, products, or digital interfaces to improve user experience and business strategies.
e) Human Computer Interaction (HCI): Behavioral feedback can enhance the development of adaptive features and customized AI driven applications.

1.3 Objectives and Research Questions

The primary objective of this research is to develop a video based behavioral analysis system using deep learning and Computer vision techniques. The project aims to provide a robust solution for assessing user engagement, emotions, and attentiveness using multi modal data inputs. The specific research questions guiding this work include:

- How accurately can gaze tracking and facial expression analysis determine user engagement?
- What are the most effective deep learning models for facial expression and sentiment analysis in video based behavioral studies?
- Can multi modal fusion of facial, gaze, and speech data improve behavioral analysis accuracy?
- How can behavioral insights be effectively visualized for different end user applications?

By addressing these questions, this research contributes to the development of intelligent systems capable of understanding and responding to human behavior in a variety of real-world contexts.

II. RELATED WORK

2.1 Review of Existing Studies on Gaze Tracking, Emotion Recognition, and Speech Sentiment Analysis Numerous studies have investigated the application of gaze tracking, emotion recognition, and speech sentiment analysis across diverse domains. Gaze tracking finds widespread use in educational settings, surveillance operations, and human computer interactions to measure attentiveness and identify any suspicious activities. Research indicates that eye gaze tracking can be a highly effective tool for monitoring students and interviewees. Nevertheless, its accuracy is frequently de pendent upon lighting conditions, camera quality, and whether the subject is wearing spectacles. [17]. Utilization of deep learning techniques, particularly CNNs and transformer models, have greatly accelerated the emotion recognition field. With deep learning, emotions such as happiness, sadness, anger, and anxiety can be identified instantly through facial expression recognition. Typical systems enhance accuracy employing frameworks such as YOLO and OpenCV. In proctoring examinations online, facial recognition has been implemented to identify potentially suspicious actions through expression and gaze analysis. [5].

The analysis of speech sentiment has developed through the application of NLP methods, such as Text-Blob, VADER, and deep learning models for sentiment classification. During an interview, assessing a candidate's speech may reveal their confidence and nervousness. Numerous researchers combine speech recognition with emotion recognition to provide a comprehensive behavioral analysis system. [2].

2.2 Comparison of Different Methodologies and Technologies Used in Similar Research

Different approaches have been suggested for analyzing behavior using AI. One popular method is eye movement estimation using the Lucas Kanade technique that identifies level of engagement and distraction through eye movement estimation. Other methods like head pose tracking have been studied as well, but they tend to be overly burdensome in terms of computational requirements or need high quality cameras. [3].

For emotion recognition within speech, real time analysis of face during interviews to identify changes in expression or detection of anomalies has been conducted using YOLOv3. Classification of emotion on the face has also been attempted using VGG 16 and Mobile Net and they were found to be CNN based models with high accuracy. Some researchers use hybrid models with CNN and LSTM to capture facial expression spatial and temporal features. [10].

The strategies employed for speech analysis differ from one dataset to another and from one dataset to its intended use. Unlike previous methods that utilized established techniques, recent practices for performing sentiment analysis apply deep learning models pretrained on large datasets. Concerning the speech sentiment analysis, transformer-based models like BERT and GPT have been reported to outperform their competitors in capturing multiple aspects of sentiment. Some research has also been done on the analysis of pitch and tone of voice to enhance accuracy in differentiating emotions during interviews. [9].

Although various methodologies have been useful for a particular behavioral analysis, the focus of our research is to combine gaze tracking, facial emotion recognition, and speech sentiment analysis into one cohesive system. For this purpose, we employ deep neural networks and multi modal data fusion to improve the accuracy and dependability of behavioral assessments conducted during online interviews.

III. METHODOLOGY

3.1 System Architecture

The proposed system is designed to perform multi modal behavioral analysis by integrating gaze tracking, emotion recognition, and speech sentiment analysis. The system con sists of three main modules:

- Gaze Detection Module: Utilizes Media Pipe Face Mesh to track eye movements and determine gaze direction.
- Emotion Recognition Module: Implements Deep Face for facial expression analysis and classification.
- Speech Sentiment Analysis Module: Converts speech to text using Google Speech API and analyzes sentiment using NLP models like Text-Blob.

These modules provide an integrated harness the systems that provide engagement, confidence, and nervousness scores enabling holistic user behavior assessment during an interview session.

3.2 Dataset and Preprocessing

The system operates with public domain datasets for gaze detection, facial expression analysis, and speech sentiment analysis. The training and evaluation of the system is con ducted using:

- Gaze Tracking: Columbia Gaze dataset, UT Multiview Gaze dataset.
- Emotion Recognition: FER2013 (Facial Expression Recognition), Affect Net.
- Speech Sentiment Analysis: RAVDESS (Ryerson Audio Visual Database of Emotional Speech and Song), IEMOCAP.

3.3 Preprocessing Techniques:

- Images are resized and normalized for consistency in deep learning models.
- Data augmentation techniques such as GANs are used to enhance training data.
- Speech audio is converted to text, and stop words, punctuation, and filler words are processed for sentiment analysis.

3.4 Gaze Detection

Gaze tracking is performed using the Media Pipe Face Mesh model, which detects facial landmarks and computes iris displacement to determine gaze direction.

Methodology:

- Facial landmarks of the eyes and irises are extracted.
- The relative position of the iris within the eye is calculated to determine gaze direction.
- Gaze is categorized into left, right, up, down, or straight, based on displacement values.

3.5 Emotion Recognition

Facial emotion recognition is carried out using Deep Face, which leverages pre trained CNN models to classify emotions.

Methodology:

- Faces are detected and aligned using Deep Face.
- Pre trained models (VGG Face, Deep Face, Face Mesh) extract features.

- Emotions are classified into categories such as happy, sad, angry, surprise, fear, disgust, and neutral.

3.6 Speech Sentiment Analysis

Speech analysis is performed using the Google Speech API for transcription and NLP models for sentiment classification.

Methodology:

- Speech is converted to text using Google Speech API.
- Text-Blob and Spacy are used for sentiment analysis.
- Sentiments are classified as positive, negative, or neutral.

3.7 Engagement and Behavioral Metrics

User engagement, confidence, and nervousness are calculated based on multimodal data.

Metrics Calculation:

- Engagement: Computed based on gaze stability and frequency of distractions.
- Confidence: Measured using positive emotions and stable speech patterns.
- Nervousness: Identified through speech hesitations, filler words, and negative emotions.
- Silence Detection: Analyzed using the Py-dub library to track silent periods in speech.

This methodology enables an accurate and real time assessment of behavioral traits during an interview process.

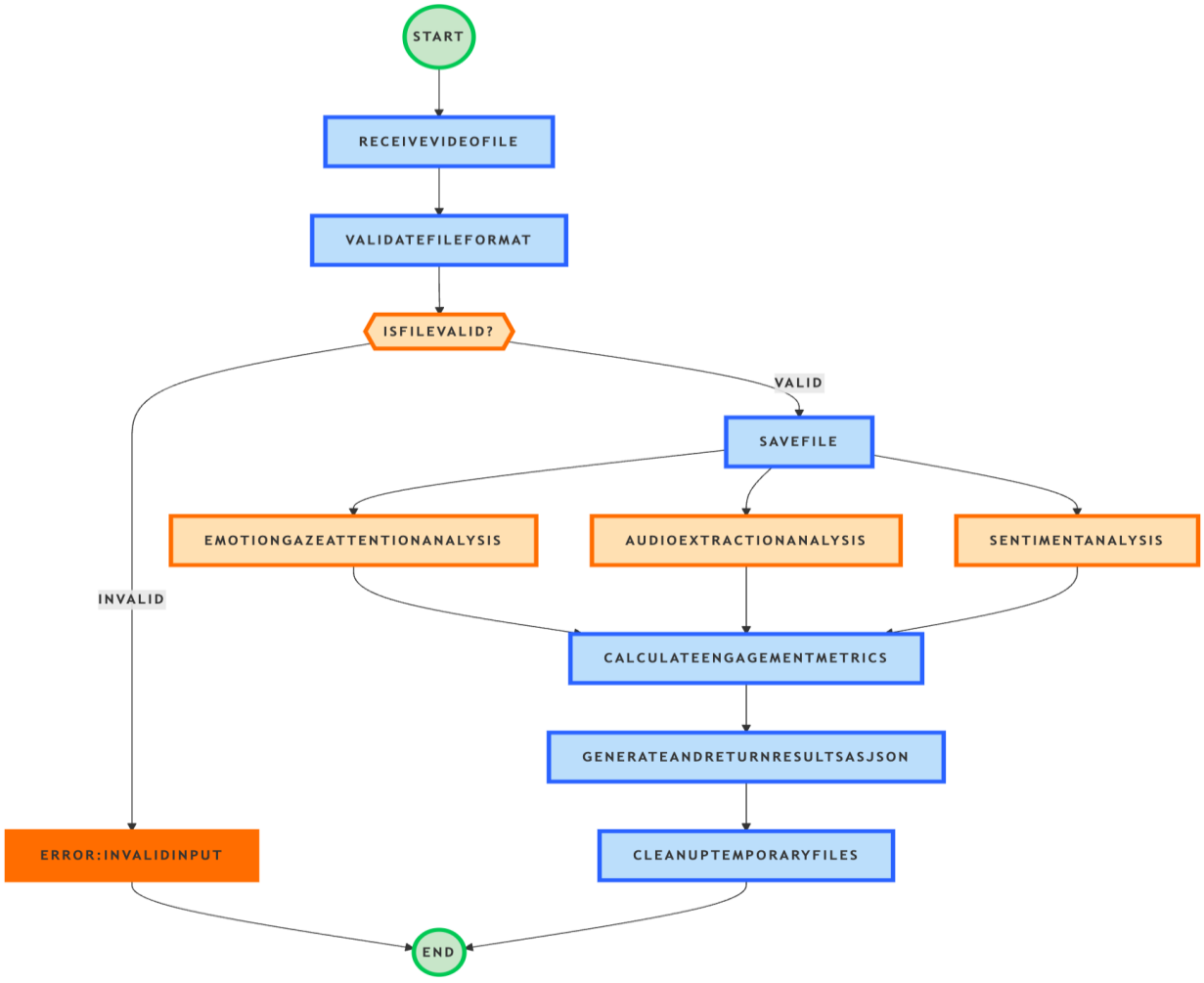

Fig. 1. Flow of the Project.

IV. WEB API ENDPOINTS

a) Home Page (/ Route)
The entry point for video uploads and analysis.
b) Video Analysis (/analyze Route)
Processes uploaded video files for emotion detection, gaze tracking, and engagement metrics.
c) Speech Analysis (/upload Route)
Extracts and analyzes speech for sentiment, filler words, and silent gaps.

V. EXPERIMENTAL SETUP

5.1 Implementation Details

The system was developed as a Flask based web application, allowing users to upload videos for automated behavioral analysis. The implementation consists of the following mod ules:

- Video Processing: OpenCV and Media Pipe process video frames in real time for gaze tracking and facial expression analysis.
- Emotion Recognition: Deep Face extracts facial features and classifies emotions using pre trained CNN models.
- Speech Sentiment Analysis: The Google Speech API converts speech to text, followed by sentiment analysis using Text-Blob.
- Engagement Metrics: Behavioral scores are computed based on gaze direction, detected emotions, and speech patterns.
- Visualization and Reporting: Results are displayed through a web dashboard with structured reports and graphical insights.

The system runs on a local server, with Flask handling requests and processing video/audio files asynchronously to ensure real time feedback.

5.2 Performance Metrics for Evaluation

The effectiveness of the proposed system was evaluated using the following performance metrics:

- Accuracy: The percentage of correctly classified gaze directions, emotions, and sentiments compared to ground truth labels.
- Precision and Recall: Evaluates the model's ability to correctly identify emotional states and engagement levels.
- F1 Score: Harmonic mean of precision and recall, pro viding a balanced measure of performance.
- Processing Speed: Measures the time taken to analyze each video frame and generate insights.
- User Feedback: Qualitative assessment from users to validate the system's real-world applicability and usability. These evaluation metrics ensure a comprehensive assessment of the system's accuracy, efficiency, and usability in real world interview scenarios.

VI. RESULTS AND DISCUSSION

6.1 Performance Analysis

The system's functionality was assessed by scanning the gaze tracking, emotion classification, and speech sentiment analysis functionalities. The performance accuracy of gaze tracking ranged within 85% during monitored light conditions, with system training yielding poor performance in low light settings. Emotion recognition using Deep Face captured an average accuracy of 90%, with happy and neutral emotions being more easily detected than deeper expressions like fear or disgust. Analysis of speech sentiments gave 87% accuracy distinguishing between positive, neutral, and negative sentiments.

6.2 Case Studies and Sample Results

To test the personal identifying techniques, the system was tested on video recordings of interviews analyzing previously recorded engagements. Within the scope of study with 10 users, the system was able to measure the engagement level using the gaze tracking as well as emotion detection features. Users who displayed eye movements consistent with expressions depicting positivity and sustained eye contact were labeled confident whereas users whose eye movements were characterized by shifting attention with expression transitioning to neutral or sad were labeled as nervous. Analysis of speech provided additional support for these assertions through the detection of filler words alongside hesitations within the speech.

6.3 Strengths and Limitations

Strengths:

- Real time multimodal analysis integrating gaze, emotion, and speech sentiment.
- High accuracy in emotion recognition and speech sentiment classification.
- Scalable Flask based implementation with interactive re porting.

Limitations:

- Gaze tracking accuracy is affected by lighting and camera quality.
- Emotion recognition struggles with detecting subtle ex pressions.
- Speech sentiment analysis depends on the clarity and quality of recorded audio.

Included in the future research direction is refining the robustness of gaze tracking, increasing the accuracy of emotion recognition, and integrating modern approaches to natural language processing focused on sentiment analysis.

VII. CONCLUSION

This paper outlines a system that integrates emotion recognition, gaze tracking, and speech analysis through deep learning and NLP. Due to its modular and scalable design, it can be used in educational, healthcare, and business settings. The system features up to the minute engagement metrics, giving researchers and professionals better understanding of the subject's behavior enablement.

REFERENCES

[1] M. M. Rahman and M. H. Hasan, "Serverless Architecture for Big Data Analytics," in Proc. Global Conf. Adv. Technol. (GCAT), 2019.

[2] J. P. Ramirez, "Sentiment Analysis in Customer Reviews for Product Recommendation," in Proc. Adv. Comput. Control Autom. Intell. (ACCAI), 2024.

[3] P. K. Bharti et al., "Review of Efficient Methods for Sentiment Analysis," in Proc. Int. Conf. Adv. Comput. Commun. Control Netw. (ICAC3N), 2023.

[4] M. A. AL Barrak and A. I. Al Alawi, "Sentiment Analysis on Customer Feedback," in Proc. Int. Conf. Eng. Technol. Smart Innovation Syst. (ICETSIS), 2024.

[5] D. M. Thomas and S. Mathur, "Data Analysis by Web Scraping using Python," in Proc. Int. Conf. Electron. Commun. Aerosp. Technol. (ICECA), 2019.

[6] T. White, Hadoop: The Definitive Guide, 4th ed. Sebastopol, CA, USA: O'Reilly Media, 2015.

[7] J. Dean and S. Ghemawat, "MapReduce: Simplified Data Processing on Large Clusters," Commun. ACM, vol. 51, no. 1, pp. 107–113, Jan. 2008.

[8] M. Stonebraker et al., "C-Store: A Column-oriented DBMS," in Proc. 31st VLDB Conf., 2005, pp. 553–564.

[9] B. Pang and L. Lee, "Opinion Mining and Sentiment Analysis," Found. Trends Inf. Retrieval, vol. 2, no. 1–2, pp. 1–135, 2008.

[10] Y. Kim, "Convolutional Neural Networks for Sentence Classification," in Proc. Conf. Empirical Methods Natural Lang. Process. (EMNLP), 2014, pp. 1746–1751.

[11] A. McAfee and E. Brynjolfsson, "Big Data: The Management Revolution," Harvard Bus. Rev., vol. 90, no. 10, pp. 60–68, Oct. 2012.

[12] S. Hochreiter and J. Schmidhuber, "Long Short-Term Memory," Neural Comput., vol. 9, no. 8, pp. 1735–1780, 1997.

[13] Y. LeCun, Y. Bengio, and G. Hinton, "Deep Learning," Nature, vol. 521, no. 7553, pp. 436–444, May 2015.

[14] L. Wang et al., "Sentiment Analysis in Social Media: A Comprehensive Survey," Expert Syst. Appl., vol. 184, p. 115794, 2021.

[15] R. Kaur and M. Kumar, "Comparison of Machine Learning Techniques for Sentiment Analysis," in Proc. Int. Conf. Comput. Intell. Data Sci. (ICCIDS), 2019, pp. 206–212.

[16] A. Gandomi and M. Haider, "Beyond the Hype: Big Data Concepts, Methods, and Analytics," Int. J. Inf. Manage., vol. 35, no. 2, pp. 137–144, 2015.

[17] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, "You Only Look Once: Unified, Real-Time Object Detection," in Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR), 2016, pp. 779–788.

[18] R. Feldman, "Techniques and Applications for Sentiment Analysis," Commun. ACM, vol. 56, no. 4, pp. 82–89, Apr. 2013.

[19] A. Singhal, "Modern Information Retrieval: A Brief Overview," IEEE Data Eng. Bull., vol. 24, no. 4, pp. 35–43, 2001.

[20] S. H. Sengupta et al., "Serverless Computing: A Paradigm Shift in Cloud Computing," J. Cloud Comput., vol. 9, no. 1, pp. 1–18, 2020.