VIVEKANAND EDUCATION SOCIETY'S INSTITUTE OF TECHNOLOGY
Department of Computer Engineering

Project Report on

EmoScan: Real-Time Facial Analysis for Online Interviews

In partial fulfilment of the Fourth Year, Bachelor of Engineering (B.E.) Degree in Computer Engineering at the University of Mumbai

Academic Year 2024-2025

Submitted by
Bhavika Valecha - D17C - 69
Dipti Hemnani- D17C - 23
Kunal Khubchandani - D17C - 37
Raj Tandon - D17C - 62

Project Mentor
Mrs. Indu Dokare

(2024-25)

VIVEKANAND EDUCATION SOCIETY'S INSTITUTE OF TECHNOLOGY

Department of Computer Engineering

Certificate

This is to certify that **Bhavika Valecha, Dipti Hemnani , Kunal Khubchandani , Raj Tandon** of Fourth Year Computer Engineering studying under the University of Mumbai have satisfactorily completed the project on "**EmoScan: Real-Time Facial Analysis for Online Interviews**" as a part of their coursework of PROJECT-II for Semester-VIII under the guidance of their mentor **Mrs. Indu Dokare** in the year 2024-25 .

This project report entitled **EmoScan: Real-Time Facial Analysis for Online Interviews** by ***Bhavika Valecha , Dipti Hemnani , Kunal Khubchandani , Raj Tandon*** is approved for the degree of **B.E. Computer Engineering.**

Programme Outcomes	Grade
PO1,PO2,PO3,PO4,PO5 ,PO6,PO7, PO8, PO9, PO10, PO11, PO12 PSO1, PSO2	O

Date: 28/04/25

Project Guide:

VIVEKANAND EDUCATION SOCIETY'S INSTITUTE OF TECHNOLOGY
CHEMBUR, MUMBAI-400 074 INDIA

Project Report Approval
For
B. E (Computer Engineering)

This project report entitled **EmoScan: Real-Time Facial Analysis for Online Interviews** by ***Bhavika Valecha*** , ***Dipti Hemnani*** , ***Kunal Khubchandani*** , ***Raj Tandon*** is approved for the degree of **B.E. Computer Engineering.**

Internal Examiner

28/4/25

External Examiner

Rupali Kale
28/4/25

Head of the Department

Principal

Date: 28/04/25
Place: Mumbai

Declaration

We declare that this written submission represents our ideas in our own words and where others' ideas or words have been included, we have adequately cited and referenced the original sources. We also declare that we have adhered to all principles of academic honesty and integrity and have not misrepresented or fabricated or falsified any idea/data/fact/source in our submission. We understand that any violation of the above will be cause for disciplinary action by the Institute and can also evoke penal action from the sources which have thus not been properly cited or from whom proper permission has not been taken when needed.

B.V.

Bhavika Valecha(D17C-69)

Dipti

Dipti Hemnani(D17C-23)

Kunal

Kunal Khubchandani(D17C-37)

Raj Tandon

Raj Tandon(D17C-62)

Date: 28/04/25

ACKNOWLEDGEMENT

We are thankful to our college Vivekanand Education Society's Institute of Technology for considering our project and extending help at all stages needed during our work of collecting information regarding the project.

It gives us immense pleasure to express our deep and sincere gratitude to Assistant Professor **Mrs. Indu Dokare** for her kind help and valuable advice during the development of project synopsis and for her guidance and suggestions.

We are deeply indebted to the Head of the Computer Department **Dr. (Mrs.) Nupur Giri** and our Principal **Dr. (Mrs.) J. M. Nair ,** for giving us this valuable opportunity to do this project.

We express our hearty thanks to them for their assistance without which it would have been difficult in finishing this project synopsis and project review successfully.

We convey our deep sense of gratitude to all teaching and non-teaching staff for their constant encouragement, support and selfless help throughout the project work. It is a great pleasure to acknowledge the help and suggestion, which we received from the Department of Computer Engineering.

We wish to express our profound thanks to all those who helped us in gathering information about the project. Our families too have provided moral support and encouragement several times.

Computer Engineering Department
COURSE OUTCOMES FOR B.E PROJECT

Learners will be to,

Course Outcome	Description of the Course Outcome
CO 1	Able to apply the relevant engineering concepts, knowledge and skills towards the project.
CO2	Able to identify, formulate and interpret the various relevant research papers and to determine the problem.
CO 3	Able to apply the engineering concepts towards designing solutions for the problem.
CO 4	Able to interpret the data and datasets to be utilised.
CO 5	Able to create, select and apply appropriate technologies, techniques, resources and tools for the project.
CO 6	Able to apply ethical, professional policies and principles towards societal, environmental, safety and cultural benefit.
CO 7	Able to function effectively as an individual, and as a member of a team, allocating roles with clear lines of responsibility and accountability.
CO 8	Able to write effective reports, design documents and make effective presentations.
CO 9	Able to apply engineering and management principles to the project as a team member.
CO 10	Able to apply the project domain knowledge to sharpen one’s competency.
CO 11	Able to develop a professional, presentational, balanced and structured approach towards project development.
CO 12	Able to adopt skills, languages, environment and platforms for creating innovative solutions for the project.

Index

List of Figures:

List of Tables:

Abstract

In recent years, emotion recognition systems have gained significant attention across diverse domains such as education, marketing, healthcare, and security. This project presents **EmoScan**, an AI-powered web application successfully developed to detect and analyze human emotions through real-time facial expression recognition. Leveraging deep learning models, EmoScan classifies facial expressions into categories such as happiness, sadness, anger, surprise, and others with high accuracy.

The system was built using modern web technologies, including React, Vite, and TypeScript, ensuring a seamless and responsive user experience. TensorFlow.js was integrated to enable real-time model inference directly within the user's browser, eliminating the need for heavy server-side processing and enhancing data privacy. Facial landmark detection and expression probability scores were incorporated to provide detailed insights into users' emotional states.

The primary objective of the project was to enhance user interaction in digital environments, and EmoScan has demonstrated strong potential for applications in online interviews, e-learning platforms, customer service analysis, and mental health monitoring. By processing all video feeds locally on the client side, the project successfully addressed critical concerns around data privacy and security.

Overall, the development and deployment of EmoScan illustrate the effectiveness of integrating machine learning techniques with cutting-edge web technologies to create intelligent, real-time, privacy-respecting applications capable of enriching user experience across multiple sectors.

Chapter 1: Introduction

1.1. Introduction:

In today's digital era, human-computer interaction is rapidly evolving to become more intelligent and responsive. One of the critical aspects of enhancing such interaction is the ability of systems to understand human emotions[1]. Emotion recognition is an emerging field in artificial intelligence (AI) that focuses on identifying and interpreting human emotional states using facial expressions, speech patterns, physiological signals, and more[5].

Facial expressions are a natural and universal way of conveying emotions. They are critical indicators of a person's psychological state and can be leveraged to improve a wide range of applications—from virtual assistants and mental health diagnosis to e-learning platforms and customer feedback systems[3],[5]. However, building a system that can accurately detect facial expressions in real-time and with minimal infrastructure is a challenging task.

To address this, our project "EmoScan" presents a real-time facial emotion recognition system that uses computer vision and machine learning techniques within a web-based interface[6]. The system utilizes TensorFlow.js, a JavaScript library for machine learning, to run deep learning models directly in the browser. This enables on-device emotion detection without the need for a backend server, ensuring user privacy and fast response times[4].

The front-end of the application is developed using React with Vite and TypeScript, offering a fast and modern development environment. EmoScan uses pre-trained models to identify facial landmarks and classify emotions such as happiness, sadness, surprise, anger, and neutrality[1],[3]. The system is lightweight, scalable, and designed to work across devices with a webcam, making it suitable for applications such as online interviews, remote education, digital well-being analysis, and customer service enhancement.

This project not only demonstrates the integration of AI with modern web technologies but also highlights the potential of real-time emotion-aware systems in building more empathetic and user-friendly digital experiences.

1.2. Motivation:

In an increasingly digital world, human interaction with machines is becoming more frequent, yet often lacks emotional depth. Traditional systems are not equipped to understand the emotional state of users, which limits their effectiveness in fields such as online education, virtual interviews, telehealth, and customer support[7].

The motivation behind the EmoScan project stems from the need to bridge this emotional gap between humans and machines. By enabling systems to recognize and respond to human emotions in real time, we can create more personalized, empathetic, and efficient user experiences[9],[12].

Moreover, with the rise of remote work, online learning, and digital health services, the ability to gauge user emotions without physical presence has become crucial. Recognizing this need, we were inspired to build a browser-based application that can detect facial emotions using deep learning models—without compromising user privacy.

The choice to use TensorFlow.js for running models in the browser and React + TypeScript for the UI reflects our desire to create a modern, fast, and accessible tool that works on any device with a camera[5].

1.3. Problem Definition:

In traditional human-computer interaction systems, there is no mechanism to understand or respond to a user's emotional state. This lack of emotional awareness limits the effectiveness of online platforms in areas like education, interviews, mental health, and customer service. Existing emotion recognition systems often require powerful hardware or server-side processing, which raises concerns about cost, scalability, and user privacy.

There is a need for a lightweight, real-time facial emotion detection system that runs entirely in the browser without compromising performance or privacy. Our project, EmoScan, aims to solve this problem by using deep learning models with TensorFlow.js to recognize facial expressions and classify emotions locally on the user's device.

1.4. Existing Systems:

Several facial emotion recognition systems exist today, many of which leverage deep learning and computer vision. Below are some notable examples:

- **Microsoft Azure Face API**

  -Offers cloud-based facial recognition and emotion detection.

  -Can detect emotions such as happiness, sadness, anger, and surprise.

  -Requires internet connectivity and sends user data to external servers, raising privacy concerns.

- **Affectiva Emotion AI**

  -A commercial tool that analyzes facial expressions and emotions.

  -Used in advertising and automotive industries.

  -Accurate but expensive and closed-source, limiting accessibility for students or small developers.

- **OpenFace**

  -An open-source tool that performs facial behavior analysis.

  -Requires installation and setup, and runs on the desktop or server.

  -Not suitable for lightweight, browser-based use.

- **DeepFace (by Facebook)**

  -One of the most accurate face recognition models developed.

  -Focuses primarily on identity recognition, not emotion classification.

  -Heavy model, not browser-compatible.

1.5. Lacuna of the Existing System:

- Most systems depend on server-side processing or cloud APIs.
- Many are not open-source or are difficult to integrate into web applications.

- Privacy and data security are often compromised when video feeds are sent to external servers.
- Few solutions offer real-time detection within the browser on standard devices.

1.6. Relevance of the Project:

With the rapid shift toward online communication, virtual learning, and remote work, the ability to understand users' emotions has become more important than ever. Emotion-aware systems can enhance digital experiences by making them more personalized, responsive, and human-centric.

EmoScan addresses this need by offering a real-time facial emotion detection system that runs directly in the browser, ensuring both performance and privacy. Its relevance spans across multiple fields such as:

- Online education – tracking student engagement
- Virtual interviews – assessing candidate reactions
- Mental health monitoring – observing emotional well-being
- Customer service – analyzing user satisfaction

By combining AI and web technologies, this project contributes to the growing demand for emotionally intelligent applications, making it both technically innovative and socially impactful.

Chapter 2: Literature Survey

A. Brief Overview of literature survey:

Recent studies in facial emotion recognition emphasize the effectiveness of both traditional and deep learning approaches in real-time applications. Techniques such as Convolutional Neural Networks (CNNs) [2], Support Vector Machines (SVMs) [4], and k-Nearest Neighbors (k-NN) [1] have been widely used for classifying emotions from facial features with high accuracy. Research has shown that simpler models like k-NN can effectively capture emotion-related features when working with small datasets [1], while CNN architectures demonstrate high precision across larger datasets by leveraging deep learning [2]. Several studies have highlighted the benefits of integrating contextual information into emotion recognition systems, such as using hybrid models that combine facial features with environmental context to reduce confusion between emotions [3]. Multimodal approaches that combine facial and speech features have also been shown to improve classification robustness over unimodal systems [4]. Other works focus on improving feature extraction techniques through methods like MediaPipe-based keypoint generation and angular encoding for real-time applications [6]. Lightweight and browser-compatible models have been successfully explored for practical deployment scenarios, making real-time facial emotion analysis feasible [5]. Overall, these approaches collectively demonstrate the feasibility and growing potential of emotion recognition systems for real-time, interactive environments, directly supporting the objectives of the EmoScan project.

B. Related Works

2.1. Research Papers :

1. **Amara, K., Ramzan, N., Achour, N., Belhocine, M., Larbas, C. and Zenati, N., 2018, October. Emotion recognition via facial expressions. In *2018 IEEE/ACS 15th International Conference on Computer Systems and Applications (AICCSA)* (pp. 1-6). IEEE.[1]**

   Abstract:

   This paper explores the recognition of seven emotional states—neutral, joy, sadness, surprise, anger, fear, and disgust—based on facial expressions using 3D face models. The study involves six participants, and features are extracted using coefficients describing facial expression elements. These features, captured with Microsoft Kinect, are classified using the k-NN classifier and a Multi-Layer Perceptron (MLP) neural network. By applying the Facial Action Coding System (FACS), the research assesses the effectiveness of Action Units (AU) in distinguishing emotions. The experiments achieve high recognition accuracy, especially using the k-NN method, indicating the reliability of 3D face models for emotion recognition.

Inference:

The study confirms that 3D facial modeling provides a significant advantage in emotion recognition, overcoming the limitations of 2D methods in handling lighting conditions and head movements. The classification results demonstrate that the k-NN algorithm outperforms MLP in recognizing emotions, particularly in subject-independent tests. This suggests that simpler models like k-NN can effectively capture emotion-related features when dealing with smaller datasets. The research highlights the promise of using low-cost devices like Kinect for practical, real-time emotion detection applications, especially in human-computer interaction contexts.

2. **Ozdemir, M.A., Elagoz, B., Alaybeyoglu, A., Sadighzadeh, R. and Akan, A., 2019, October. Real time emotion recognition from facial expressions using CNN architecture. In *2019 medical technologies congress (tiptekno)* (pp. 1-4). IEEE.[2]**

Abstract:

This paper introduces a Convolutional Neural Network (CNN) based on the LeNet architecture for real-time emotion recognition through facial expressions. The system combines datasets from JAFFE, KDEF, and a custom dataset to train the CNN model for seven emotion categories: happy, sad, surprised, angry, disgust, afraid, and neutral. With an achieved accuracy of 96.43% in training and 91.81% in validation, this method provides a low-cost solution for recognizing facial emotions in real-time. It utilizes the Haar Cascade for face detection and employs deep learning to ensure high classification accuracy in a variety of use cases.

Inference:

The study demonstrates the effectiveness of deep learning, particularly CNN architectures, in facial emotion recognition. The authors successfully merge multiple datasets and apply a CNN-based model to detect seven distinct emotional states with high accuracy. The real-time testing capability showcases the potential applications of this system in industries such as consumer behavior analysis, mental health diagnosis, and human-computer interaction. This work exemplifies how a well-optimized CNN can enhance facial emotion recognition accuracy, positioning it as a valuable tool for various real-time and practical applications.

3. **Zhou, S., Wu, X., Jiang, F., Huang, Q. and Huang, C., 2023. Emotion recognition from large-scale video clips with cross-attention and hybrid feature weighting neural networks. *International Journal of Environmental Research and Public Health*, *20*(2), p.1400.[3]**

Abstract:

This paper introduces a novel approach to emotion recognition from large-scale video clips using a hybrid neural network model. It integrates face and context features via cross-attention mechanisms and hierarchical feature encoding to improve accuracy. The system captures complementary gains from both facial expressions and surrounding contexts, mitigating issues of emotion confusion. The model uses deep fusion to merge adaptive emotion features, significantly enhancing the classification of emotional states. Extensive experiments on the CAER-S dataset demonstrate the system's effectiveness in applications like mental health assessments and job stress analysis using video data.

Inference:

The research successfully addresses the challenge of emotion recognition by integrating face and context features, overcoming the limitations of existing models that rely solely on facial data. The cross-attention mechanism allows for better interaction between multiple data streams, leading to improved understanding of emotional states. The model's strong performance on large-scale datasets showcases its potential for real-world applications, such as assessing stress levels or tourist satisfaction from video clips. By focusing on hybrid feature weighting, the paper sets a precedent for further advancements in context-aware emotion recognition systems.

4. **Thushara, S. and Veni, S., 2016, March. A multimodal emotion recognition system from video. In *2016 International Conference on Circuit, Power and Computing Technologies (ICCPCT)* (pp. 1-5). IEEE.[4]**

Abstract:

This paper presents a multimodal emotion recognition system that integrates facial and speech features to improve the recognition of human emotions. The system extracts geometric and appearance-based features from facial expressions and prosodic and spectral features from speech to detect six universal emotions: happy, sad, surprise, disgust, fear, and anger. Using a Support Vector Machine (SVM) classifier, the system processes facial and audio data separately and then fuses them for a more robust performance. This work demonstrates the potential of combining facial and acoustic modalities to enhance the accuracy of emotion recognition systems, achieving improved performance over unimodal approaches.

Inference:

This study highlights the advantages of combining facial and speech modalities for emotion recognition. While facial and speech features individually offer some level of accuracy, fusing them increases system performance and robustness. The authors' choice to use SVMs, due to their simplicity and flexibility, strengthens the classification process. Their results show that a multimodal approach can better recognize and classify emotions in real-time applications like healthcare, telecommunication, and robotics. This paper also emphasizes the potential of such systems in aiding humans with disabilities and contributing to fields like psychiatry and behavioral science.

5. **Dagar, D., Hudait, A., Tripathy, H.K. and Das, M.N., 2016, May. Automatic emotion detection model from facial expression. In** ***2016 International Conference on Advanced Communication Control and Computing Technologies (ICACCCT)*** **(pp. 77-85). IEEE.[5]**

Abstract:

This paper introduces an automatic framework for emotion recognition using facial expressions, emphasizing the challenge of subject-independent recognition in real-life scenarios. The system extracts facial attributes using Gabor feature extraction and principal component analysis (PCA) to determine expressions, employing a neural network for classification. Six fundamental emotions (happiness, sadness, anger, disgust, surprise, and fear) are recognized by analyzing facial muscles via the Facial Action Coding System (FACS). The model processes live-stream video frames, applies clustering for emotion detection, and classifies emotions through K-means clustering enhanced by PCA.

Inference:

The proposed model effectively captures and recognizes facial emotions using feature extraction and machine learning techniques. By utilizing Gabor filters for facial feature extraction and PCA for dimensionality reduction, the system achieves robust classification. However, the results indicate overlapping emotions for some expressions, suggesting that further refinement is needed to enhance accuracy. The model's reliance on clustering makes it suitable for real-time applications such as human-computer interaction, but improving the classifiers and expanding the training dataset could lead to better handling of complex emotional states.

6. **Siam, A.I., Soliman, N.F., Algarni, A.D., Abd El-Samie, F.E. and Sedik, A., 2022. Deploying machine learning techniques for human emotion detection.** ***Computational intelligence and neuroscience*****,** ***2022*****(1), p.8032673.[29]**

Abstract:

This paper presents a real-time approach to human emotion detection using machine learning techniques. The methodology involves preprocessing images using MediaPipe for key point generation, followed by angular encoding of these key points to extract distinguishing features. A variety of machine learning models, including SVM, KNN, Naïve Bayes, Random Forest, and Multilayer Perceptron (MLP), are applied for emotion classification. The approach is evaluated on multiple datasets, such as CK+, JAFFE, and RAF-DB, achieving a high accuracy of 97%. The study is particularly useful for robotic vision applications, where real-time emotion recognition is crucial.

Inference:

This research showcases the effectiveness of combining facial landmarks and angular encoding for accurate emotion detection. The use of deep learning frameworks, such as MediaPipe for real-time landmark generation, and machine learning classifiers, like SVM and MLP, provides high recognition rates across diverse datasets. The system is versatile, handling different modalities in robotic vision applications. The successful integration of angular encoding with machine learning demonstrates the system's potential for applications requiring real-time emotional assessment, although further refinement could enhance performance in more varied and complex environments.

7. **Joseph, A. and Geetha, P., 2020. Facial emotion detection using modified eyemap–mouthmap algorithm on an enhanced image and classification with tensorflow. *The Visual Computer*, *36*(3), pp.529-539.[7]**

Abstract:

This paper proposes a facial emotion detection system using a modified eyemap-mouthmap algorithm, enhanced with image preprocessing through discrete wavelet transform (DWT) and fuzzy logic. The system detects key facial features, such as eyes and mouth, and constructs facial geometry for emotion recognition. Classification is performed using TensorFlow's neural network framework, which evaluates emotions on multiple datasets, including KDEF, Oulu-CASIA, and CK+. The modified algorithm demonstrates improved detection accuracy compared to state-of-the-art methods, offering efficient identification of facial emotions.

Inference:

The research effectively integrates advanced image processing techniques, such as DWT and fuzzy logic, to enhance image quality and improve emotion detection accuracy. The modified eye map-mouthmap algorithm, combined with TensorFlow for classification, provides a significant improvement over traditional methods, particularly in identifying subtle emotions. The system's success across multiple datasets highlights its robustness and potential for real-world applications in fields like human-computer interaction. Further research could focus on optimizing processing times to make the system more suitable for real-time applications.

8. **Jadhav, B., Sawant, A., Shah, A., Vemula, P., Waikar, A. and Yadav, S., 2024, March. A Comprehensive Study and Implementation of the Mock Interview Simulator with AI and Pose-Based Interaction. In *2024 1st International Conference on Cognitive, Green and Ubiquitous Computing (IC-CGU)* (pp. 01-05). IEEE.[16]**

Abstract:

The paper titled "A Comprehensive Study and Implementation of the Mock Interview Simulator with AI and Pose-Based Interaction" presents an innovative platform aimed at improving job seekers' interview skills. In an increasingly competitive job market, technical skills and soft skills are essential, and the ability to handle interviews effectively has become crucial. The Mock Interview Simulator developed in this project offers a transformative tool for job preparation. Utilizing AI-driven interviewers and real-time posture detection with the Mediapipe framework, it creates a dynamic and personalized interview environment. The simulator integrates speech recognition, text-to-speech, and OpenAI's GPT-3.5 Turbo API to deliver a realistic, interactive interview simulation. It provides valuable feedback on both verbal responses and body language, helping users improve their performance across multiple dimensions. This paper outlines the technical design and capabilities of the system, focusing on how AI and pose-based interaction contribute to an immersive interview preparation experience.

Inference:

The Mock Interview Simulator developed in this project represents a significant advancement in interview preparation technologies by incorporating both AI-driven feedback and posture analysis. This dual approach allows users to refine not only their verbal responses but also their non-verbal communication skills. By integrating real-time posture feedback, the simulator highlights the importance of physical presence during interviews, something often overlooked in traditional preparation methods. The system's use of GPT-3.5 Turbo for natural language processing and real-time feedback provides users with detailed insights into their performance, thereby enhancing their readiness for real-world interviews. The system's flexibility, modularity, and real-time interaction capabilities make it a valuable tool for individuals seeking to improve their interview skills, and it sets the stage for further enhancements such as the inclusion of virtual reality and personalized learning pathways.

9. **Asha, P., Shabu, J., Refonaa, J. and Selvan, M.P., 2023, August. Automated Interview through Online Video Interface. In *2023 International Conference on Circuit Power and Computing Technologies (ICCPCT)* (pp. 1315-1321). IEEE.[23]**

Abstract:

This paper presents an Automated Video Interview Interface (AVII), a system leveraging AI, natural language processing, and machine learning to streamline the recruitment process. AVII automates candidate evaluation, using video conferencing to assess responses, facial expressions, and body language remotely. With features like pre-recorded questions, real-time scoring, and the ability to generate detailed performance reports, the system offers an efficient alternative to traditional interviews. The integration of facial recognition and natural language processing allows for a comprehensive evaluation of both technical and soft skills. This technology promises to enhance recruitment efficiency, reduce human bias, and provide personalized feedback for candidates.

Inference:

The AVII system addresses the growing need for efficient, unbiased, and scalable recruitment processes in today's job market. By automating the evaluation of candidates, it minimizes human involvement and potential bias, resulting in a fairer hiring process. The inclusion of real-time analysis of body language and facial expressions enhances the depth of candidate assessment. However, challenges such as technical issues and the loss of human interaction remain.

2.2. Patent Search :

1. Asynchronous video interview system (US9197849B2)

Inventor: **Jeff M. BoltonJames H. Wolfston, Jr.**

Aspects of an asynchronous video interview system and related techniques include a server that receives a plurality of pre-recorded video prompts, generates an interview script, transmits a video prompt from the interview script to be displayed at a client computing device, and receives a streamed video response from the client computing device. The server can perform algorithmic analysis on the content of the video response. In another aspect, a server obtains response preference data indicating a timing parameter for a response. In another aspect, a video prompt and an information supplement (e.g., a news item) that relates to the content of the video prompt are transmitted. In another aspect, a server automatically selects a video prompt (e.g., a follow-up question) to be displayed at the client computing device (e.g., based on a response or information about an interviewee).

2. On–demand, web-based interactive video interviewing (US20140236850A1)

Inventors: **Jody Neal HOLLAND**

A method and system for an on-demand web-based interactive video interviewing process provides a convenient and flexible way for candidates and employers to perform an interview in a more objective manner. This on-demand web-based interview process comprises the steps of building an interview set by a customer, informing candidates about an upcoming interview, candidates recording responses, collaborators evaluating recorded answers from candidates and asking follow-up questions for specific candidate, with approval of follow-up questions, the designated candidate being promoted to answer follow-up questions for reevaluation and finally generating an interview report. This interview system also includes a specific function of Click2Interview and a practice platform for candidates and collaborators to practice in a real interview process model. The practice platform provides for a method of posting the practice results to public profiles through the use of a hyper-link.

3. System and method for interview training with time-matched feedback(US20220036315A1)

Inventors: Thom STEINHOFFPanos S. STAMUSBryan ACKERMANNJohn DEYTO

The present disclosure generally relates to interview training and providing interview feedback. An exemplary method comprises: at an electronic device that is in communication with a display and one or more input devices: receiving, via the one or more input devices, media data corresponding to a user's responses to a plurality of prompts; analyzing the media data; and while displaying, on the display, a media representation of the media data, displaying a plurality of analysis representations overlaid on the media representation, wherein each of the plurality of analysis representations is associated with an analysis of content located at a given time in the media representation and is displayed in coordination with the given time in the media representation.

2.3. Inference Drawn:

From the reviewed patents, the following inferences can be drawn:

1. **Recording of Candidate Responses**: Most existing systems primarily depend on video and audio recordings to capture candidates' behavioral and performance attributes. Systems like the Asynchronous Video Interview Platform (US9197849B2) and solutions described by Asha et al. [9] emphasize recording for later evaluation.
2. **Asynchronous or Automated Interviews**: Many solutions allow candidates to record responses without requiring a live interviewer, offering greater flexibility and scalability (US9197849B2 [Patent 1]; Asha et al. [9]; Dagar et al. [5]). This asynchronous model reduces scheduling conflicts and enables large-scale candidate screening.
3. **Real-Time or Time-Synchronized Feedback**: Emerging systems provide feedback aligned with specific moments in recorded interviews (e.g., speech pauses, changes in expression) (US20140236850A1 [Patent 2]; Zhou et al. [3]). This technique enhances self-evaluation by pinpointing strengths and weaknesses with temporal precision.
4. **Behavioral and Emotional Analysis**: Some platforms incorporate facial expression tracking, speech sentiment analysis, and behavioral cue extraction to enrich interview evaluation (Dagar et al. [5]; Ozdemir et al. [2]). However, such emotional assessments are often post-processed rather than performed live.
5. **Scope for Further Innovation**: Although substantial innovation exists in automated interviewing, EmoScan distinguishes itself by offering real-time facial emotion detection, AI-driven dynamic

feedback, and emotion analysis during live sessions. Inspired by recent advancements in lightweight CNNs (Ozdemir et al. [2]) and multimodal emotion fusion (Siam et al. [6]), EmoScan enhances candidate self-awareness and presents clear advantages in training, education, and recruitment domains over traditional asynchronous or post-processed platforms.

2.4. Comparison with the existing systems:

Based on our literature review and patent search, we can identify several key differences between ContentConcise and existing systems as shown in table 1:

Table 1. Comparison with existing systems

Other System	**This proposed system**
Most systems focus only on recorded mock interviews (asynchronous).	Provides real-time mock interviews with live facial emotion detection.
Limited or no emotion analysis.	Integrates real-time facial emotion recognition using AI/ML models
Feedback is often generic or based on pre-recorded templates.	Personalized feedback based on emotion trends and expressions

Chapter 3: Requirement Gathering for the Proposed System

In this chapter, we will discuss the resources utilized, the process of understanding user needs, and the steps taken to design a system that meets those expectations. Effective requirement gathering is critical to building a system like EmoScan, as it ensures that the final application not only meets technical specifications but also delivers a seamless and valuable experience to the user. This chapter also covers the functional and non-functional requirements of the project and details the software, hardware, technologies, and tools adopted for development.

Requirement gathering is the process of discovering, capturing, and documenting the exact needs of users and stakeholders. It is also referred to as requirements elicitation or requirement capture. By engaging with potential users and understanding the objectives of the system, we can define the technical and functional blueprint required for a successful solution. Gathering requirements systematically helps minimize ambiguity, reduce project risks, and ensure that the final deliverable aligns with the real-world expectations of users.

The requirement gathering process followed for the EmoScan project involved six structured steps:

- Identifying the relevant stakeholders such as candidates, interviewers, and administrators.
- Establishing clear project goals and objectives related to mock interview simulation, real-time feedback, and emotion recognition.
- Eliciting specific requirements by analyzing user scenarios and conducting initial research on similar systems.
- Documenting detailed functional and non-functional requirements based on findings.
- Confirming the requirements through iterative feedback and validation from potential users.
- Prioritizing the requirements to ensure that core functionalities were addressed first during development.

The goal was to create an application that captures real-time video and audio input, interacts intelligently with an AI engine, and presents feedback in a way that is both actionable and insightful. With a focus on using a lightweight, browser-compatible design, we also ensured that the system would be accessible without heavy backend dependencies, thereby improving performance and scalability.

Requirements of the System outlines the essential hardware, software, and functional needs necessary for developing and running the EmoScan application. It details the tools, technologies, and resources required to ensure smooth system performance, real-time emotion detection, and interactive feedback generation. This baseline understanding of system needs helped guide the entire technical development of EmoScan.

A detailed breakdown of the system requirements is provided in Table No. 2 below.

3.1. Introduction to Requirement Gathering:

The Requirement Gathering is a process of requirements discovery or generating list of requirements or collecting as many requirements as possible by end users. It is also called as requirements elicitation or requirement capture.

The requirements gathering process consists of six steps :

- Identify the relevant stakeholders
- Establish project goals and objectives
- Elicit requirements from stakeholders
- Document the requirements
- Confirm the requirements
- Prioritise the requirements

Requirements of the System outlines the essential hardware, software, and functional needs necessary for developing and running the EmoScan application. It details the tools, technologies, and resources required to ensure smooth system performance, real-time emotion detection, and user interaction. This table serves as a foundation for understanding the technical setup and operational demands of the project.

Table No: 2 Requirements of the system

USE CASE	**DESCRIPTION**
Register and Login	Users (interviewees/interviewers/admin) can register and securely log in to access the platform.
Start Mock Interview	Users can initiate a mock interview session with camera and microphone access.
Real-Time Face Detection	The system detects faces in real-time using webcam input during interviews.
Emotion Recognition	Gemini AI analyzes facial expressions and displays detected emotions live.
Age and Gender Estimation	The system estimates and displays the approximate age and gender of the user in real time.

Speech-to-Text Transcription	Converts spoken responses to text using AI speech recognition.
Interview Feedback	Generates automated feedback based on facial expressions, emotions, and speech quality.
Upload Profile Image	Users can upload a static image to test face and emotion recognition outside of live mode.
Download Interview Report	After the session, users can download a detailed interview performance report.
Admin Dashboard	Admin can manage user accounts, view analytics, and update models or content from a dashboard panel.

3.2. Functional Requirements:

1. User Authentication and Management

- The system shall allow users (students/candidates) to register and log in.
- The system shall allow users to manage their profile information (name, email, etc.).

2. Interview Simulation Interface

- The system shall provide a mock interview interface with pre-defined or dynamic questions.
- The system shall support both text-based and video-based questions.
- The system shall allow users to start, pause, or end the mock interview session.

3. Real-Time Facial Emotion Detection

- The system shall access the user's webcam to capture real-time video during the interview.
- The system shall detect and classify facial emotions such as happy, sad, angry, surprised, neutral, etc.
- The system shall update the emotion analysis continuously or at regular intervals.

4. Audio and Speech Analysis (optional but valuable)

- The system shall record the user's audio responses during the interview.
- The system may analyze voice tone, pitch, and clarity for feedback (if implemented).

5. Interview Performance Feedback

- The system shall display post-interview feedback summarizing:

  -Emotion trend (e.g., mostly nervous or confident)

  -Facial expression consistency

  -Response time per question

- The system shall suggest improvements based on emotional state and response behavior.

6. Admin Module

- The admin shall be able to manage users and track their performance.
- The trainer may add or customize interview questions.

7. Data Storage and Reporting

- The system shall store user interview sessions, emotion logs, and performance summaries.
- The system shall allow users to download or view past interview reports.

8. Technology Integration

- The system shall integrate face detection and emotion recognition libraries (e.g., OpenCV, DeepFace, MediaPipe).
- The system shall run efficiently on web or desktop platforms using technologies like React + Vite + TypeScript.

3.3. Non-Functional Requirements:

- The system should run completely in the browser without requiring server-side computation.
- The system should ensure user privacy by not storing or transmitting video data.

- The emotion detection should update in real-time (within ~100–200 ms).
- The UI should be user-friendly, responsive, and accessible across devices.
- The application should load within 2–3 seconds on a standard internet connection.

3.4.Hardware, Software, Technology and Tools Utilised:

A. Hardware Requirements:-

a. Minimum 8 GB RAM

b. Core I5 7th Gen processor

c. NVIDIA GPU

d. Disk space of 4GB

B. Software Requirements:-

1. Operating System: Cross-platform (Windows/Linux/macOS) – only a browser is required
2. ReactJS (with Vite) – for fast frontend development and hot module replacement
3. Node.js and npm – for managing frontend dependencies and running the development server
4. Gemini AI (via API access) – for generating interview feedback and responses
5. Firebase (Firestore) – for storing interview transcripts and feedback data
6. HTML5/CSS3 – for building responsive UI components
7. Web APIs (MediaDevices API) – for accessing webcam and audio

C.Technologies:-

● React + Vite: A modern frontend stack used for building the user interface. Vite ensures fast development builds and React provides modular, component-based architecture for the mock interview interface.

● Gemini AI: Google’s large language model is integrated via API to analyze user responses, provide feedback, and simulate real-time interviewer interactions.

● Firebase (Firestore): Used as the backend database to persist interview responses and Gemini feedback securely and efficiently.

● JavaScript/TypeScript: Core programming languages for frontend logic, API communication, and state management.

● Web APIs: HTML5 MediaDevices API is used to capture real-time webcam input directly from the user's browser.

● CSS/Bootstrap/Tailwind: Styling frameworks and CSS technologies are used to build a clean, user-friendly interface.

D.Tools:-

- **Vscode:-**Visual Studio Code is a streamlined code editor with support for development operations like debugging, task running, and version control. It aims to provide just the tools a developer needs for a quick code-build-debug cycle and leaves more complex workflows to fuller featured IDEs, such as Visual Studio IDE.
- **Google Colab:-** Colaboratory, or "Colab" for short, is a product from Google Research. Colab allows anybody to write and execute arbitrary python code through the browser, and is especially well suited to machine learning, data analysis and education. More technically, Colab is a hosted Jupyter notebook service that requires no setup to use, while providing access free of charge to computing resources including GPUs.

3.5. Constraints:

- Internet Access is required.
- Users should be able to interpret the data in the form of visualisation.
- Our system is restricted to analysis for the 20 countries we have chosen.

Chapter 4: Proposed Design

The proposed system, EmoScan, is an AI-powered web application designed to deliver real-time mock interview analysis through speech capture, sentiment feedback, and AI-based evaluation. Built using modern frontend technologies such as React, Vite, and TypeScript, EmoScan offers a seamless, responsive user experience directly within a web browser. By accessing the user's webcam and microphone, the system records real-time user responses and dynamically interacts with Gemini AI to generate intelligent follow-up questions and performance feedback. The emphasis on using client-side web technologies ensures minimal backend dependency while maintaining the privacy and security of user data.

A high-level overview of the system's functionality and major components is presented in the Block Diagram (Fig. 4.1). It highlights the primary modules such as user registration/login, mock interview session handling, Gemini AI feedback integration, and Firebase database storage. The system allows candidates to initiate a mock interview, submit their responses, and receive live AI-generated feedback while saving the entire session history for later review.

The internal structure of EmoScan is broken down further through the Modular Diagram (Fig. 4.2). This figure categorizes the system into functional modules, namely the User Interface Module, Interview Management Module, AI Communication Module, and Feedback Storage Module. Each module interacts with external technologies like Gemini AI for natural language processing and Firebase Firestore for secure data storage. This modular architecture supports easy scalability and potential future enhancements such as adding mobile app compatibility or integration with job preparation platforms.

To illustrate the working flow visually, the Pictorial Diagram (Fig. 4.3) presents the sequential process starting from the user launching the app to recording the mock interview and receiving AI-generated feedback. The diagram captures the dynamic interaction between the user, the AI model, and the storage system, emphasizing how the interview simulation is conducted seamlessly through web technologies.

The detailed functioning of EmoScan's backend communication and data storage process is outlined in the Data Flow Diagram (DFD) (Fig. 4.4). It demonstrates how user data, interview metadata, and AI responses flow between the browser, external AI services, and the Firebase database. Important backend activities like authentication, interview data storage, and asset loading from Firebase Hosting are covered to give a complete understanding of data handling within the system.

Finally, the project development life cycle is mapped out in the Gantt Chart (Fig. 4.5). This figure captures the planning, design, development, integration, testing, and documentation phases spread across the academic year, ensuring that project milestones were achieved systematically and the final product met all the planned objectives. Overall, the design of EmoScan ensures a real-time, AI-driven mock interview experience that is modular, scalable, and aligned with the latest advancements in web-based AI applications.

4.1. Block Diagram of the proposed system:

The block diagram illustrates the architecture of the EmoScan Web App, showcasing the major components and their interactions. The system begins with user login or registration, after which users can either join a mock interview—activating their camera and microphone for real-time emotion detection—or view feedback from previous sessions, with all results stored in Firebase. Gemini AI analyzes interview conversations, generates suggestion reports, and auto-saves these in Firebase for review. Meanwhile, the Admin panel manages users (both candidates and interviewers) and monitors system logs to ensure smooth operation. The Database (Firebase) stores critical information including user details, interview session logs, emotion analysis data, and feedback reports. Overall, the diagram reflects an integrated and organized flow for managing real-time interviews, emotion detection, and feedback generation within EmoScan.

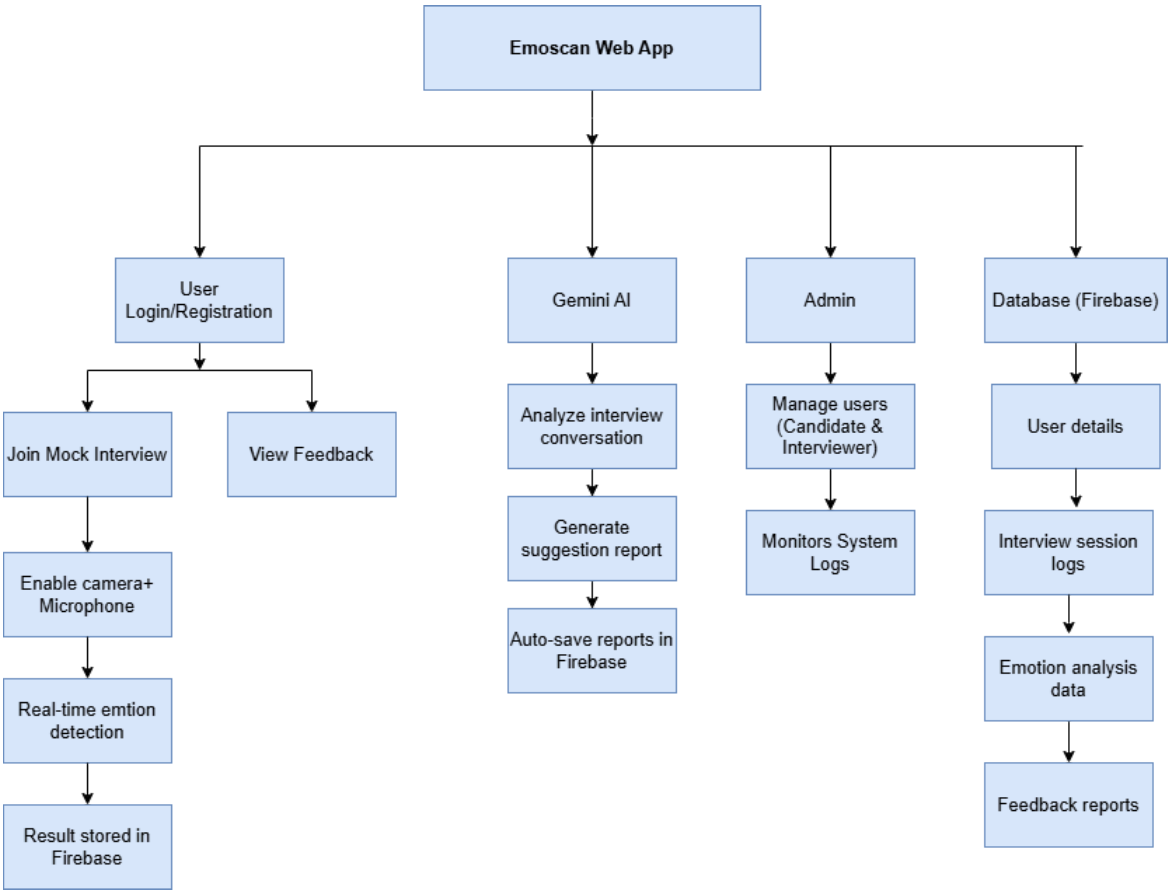


Fig 4.1: Block Diagram

4.2. Modular diagram of the system:

The modular diagram of EmoScan – Real-Time Facial Emotion Recognition and Mock Interview System outlines the interaction between various system components and external technologies. It starts with external libraries and tech stack like TensorFlow.js, Vite, React, TypeScript, and Gemini AI, which support the development. The User Interface module handles user registration, profile management, and controls the start, pause, or end of sessions. The Admin Module manages user accounts and edits the question bank, which feeds into the Interview Flow Module responsible for generating dynamic questions and handling Q&A flows. Simultaneously, the Real-Time Face and Emotion Recognition module accesses the webcam, performs face detection, emotion classification, and age/gender estimation. Alongside, the Speech-to-Text and Audio Analysis module captures audio, analyzes tone and clarity, and transcribes speech into text. All data—emotion trends, response timings, and suggestions—are compiled into the Feedback and Reporting Module, providing downloadable reports for users. This interconnected architecture ensures a comprehensive mock interview experience with real-time analysis and actionable feedback.

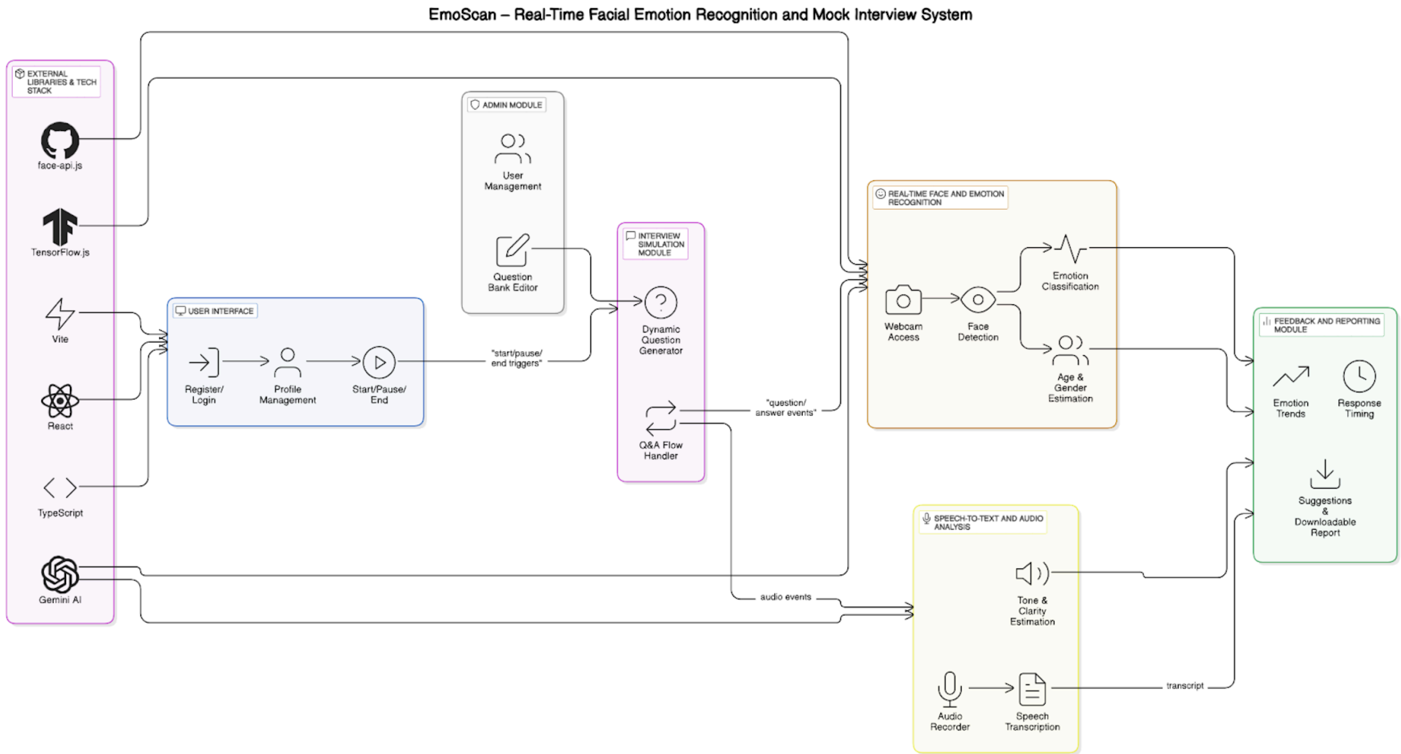


Fig 4.2: Modular Diagram

The pictorial diagram illustrates the workflow of the EmoScan system for real-time facial emotion analysis. It begins with capturing the camera feed, followed by preprocessing to prepare the image for analysis. The system then performs face detection and landmark detection to identify key facial features. Using a database of known faces, face recognition is carried out to match or register identities. Once recognized, the system simultaneously conducts age and gender recognition and expression recognition to analyze the emotional state. All extracted data is then passed to the user interface, where users can view detailed results and analysis, providing insights into emotional trends and facial attributes during sessions.

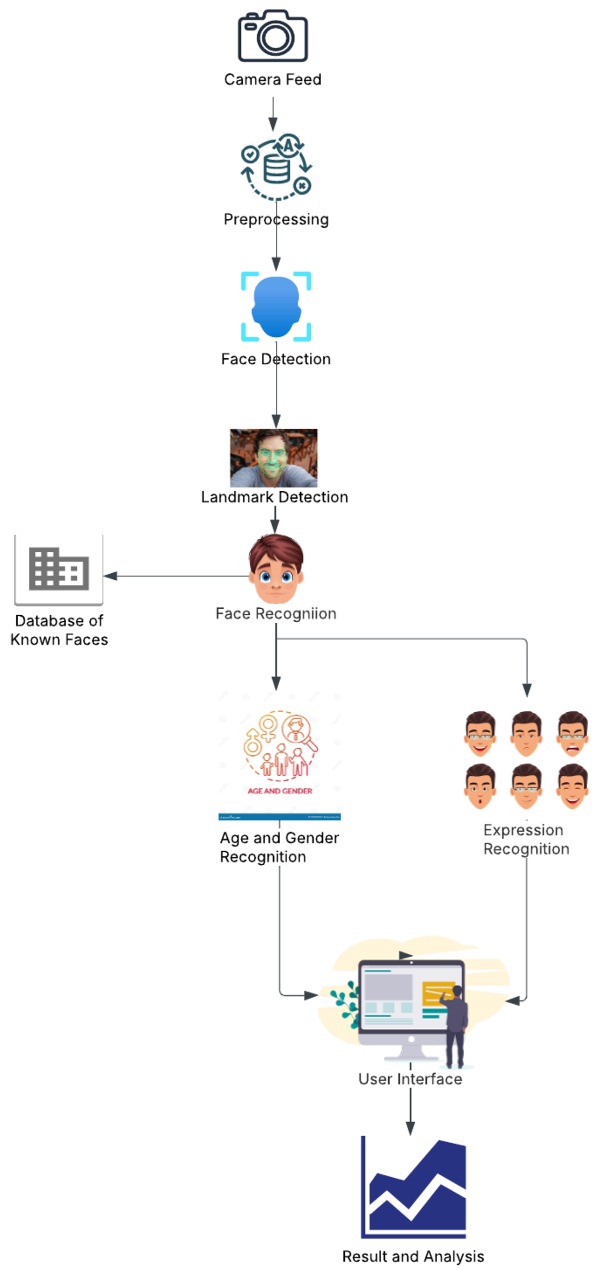


Fig 4.3: Pictorial Diagram

4.3. Detailed Design

The DFD (Data Flow Diagram) illustrates the working of the AI Mock Interview App within EmoScan. The process begins with the user who logs in or registers by sending credentials, which are authenticated through the Firebase Authentication Service, returning token and session information. The user can then request an interview, during which the app fetches interview questions from a Third-Party AI Service. As the interview progresses, users record their answers, which are uploaded to the Recorded Answers Store, while interview metadata (like timestamps and session info) is saved separately. Additionally, the app fetches static files and other necessary assets from Firebase Hosting & Storage to ensure smooth functioning. This setup ensures a seamless interview experience with real-time interaction, secure data handling, and efficient session management.

DFD:

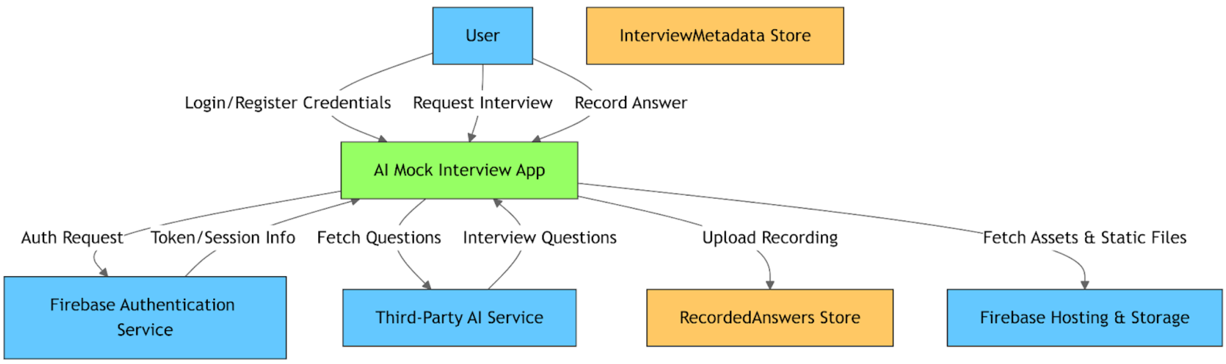


Fig 4.4: DFD

4.4. Project Scheduling & Tracking using Time line / Gantt Chart:

The timeline of our project, detailing the work carried out throughout the semester to develop this model, is shown in Figure 4.5. This timeline is a crucial element in the planning and execution of the project, as it outlines the systematic approach and milestones we followed to ensure timely and structured development.

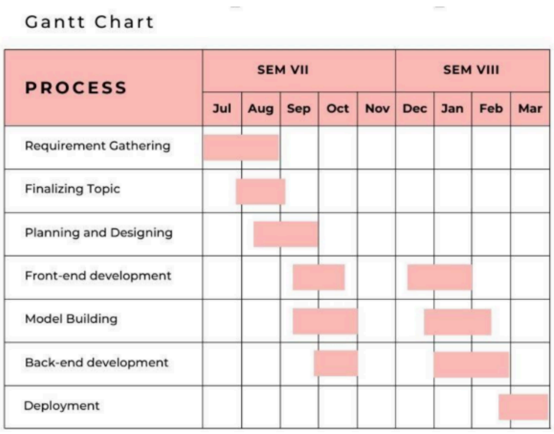


Fig 4.5 :Gantt chart

Chapter 5: Implementation of the Proposed System

5.1.Methodology employed for development:

The development of EmoScan—a real-time facial emotion recognition system designed for mock interview evaluation—was conducted through a modular and iterative development cycle, emphasizing real-time performance, AI integration, and web-based deployment. The foundational emotion detection module was developed using machine learning techniques. A Convolutional Neural Network (CNN) based on the pretrained VGG-16 architecture was fine-tuned using facial emotion datasets such as FER-2013. The model was trained and validated in Google Colab to leverage GPU acceleration. After achieving satisfactory performance metrics such as accuracy and F1-score, the model was integrated with a real-time detection system using face-api.js—a JavaScript library built on top of TensorFlow.js. A basic browser-based application was created to capture webcam input, detect faces, and classify emotions live on screen. This prototype validated the feasibility of emotion detection in a real-time environment.

The focus then shifted towards building a production-ready web application using modern frontend technologies. A single-page application (SPA) was created using React and Vite to enable fast performance and a modular component structure. TypeScript was used to improve code maintainability. Gemini AI was integrated into the frontend to assist with higher-level inference, such as generating feedback based on emotional expressions during mock interviews. Tailwind CSS was employed to build a clean and responsive user interface, and Framer Motion was used for smooth UI transitions. Various components were developed for camera integration, emotion overlays, report generation, and candidate feedback.

Throughout the development process, Agile methodology was adopted. The team worked in sprints, iterating on each module with regular testing, feedback sessions, and version control through GitHub. This structured approach ensured seamless integration between the machine learning backend and the user-friendly frontend, resulting in a full-fledged AI-powered interview assistance tool.

5.2. Algorithms and Flowcharts for the respective modules developed:

We have mainly used two algorithms for our project:

MODULE 1: Facial Emotion Detection

Algorithm: Real-time Emotion Capture and Averaging

1. Start the webcam and capture real-time video input using the browser.
2. Stream the video feed to the React-based UI for live preview.
3. Periodically extract frames from the video feed at set intervals.
4. Send extracted frames to an emotion detection API (or in-browser model if applicable).
5. Receive emotion prediction results for each frame (e.g., happy, sad, neutral, etc.).
6. Log each prediction with a timestamp.
7. Overlay the latest detected emotion on the live video feed for user feedback.
8. Continue capturing and logging emotion predictions throughout the session.
9. At the end of the session, compute the average distribution of all detected emotions across the duration.
10. Display the overall emotion summary report to the user.
11. Stop webcam and exit on user action.

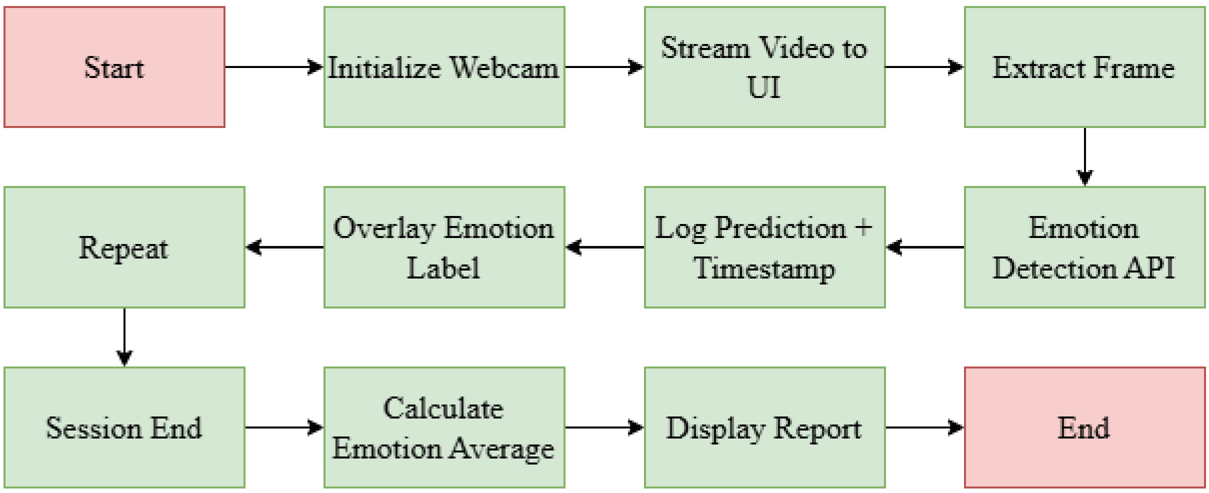


Fig 5.1 : Flow Chart (Module1)

MODULE 2: Web Interface with Gemini AI Integration

Algorithm: Real-Time AI Interview Feedback System with Firebase Storage

1. Start the web application using React + Vite.
2. Request and open webcam access via browser APIs for video preview.
3. Capture and transcribe user’s spoken answers in real time.
4. Send transcribed answers to Gemini AI via API.
5. Receive AI-generated follow-up questions or interview feedback.
6. Display real-time chat-based interaction between candidate and AI.
7. Store the user's answers and Gemini-generated feedback to Firebase Firestore.
8. Optionally allow users to view their feedback after the session ends.
9. End session on user request.

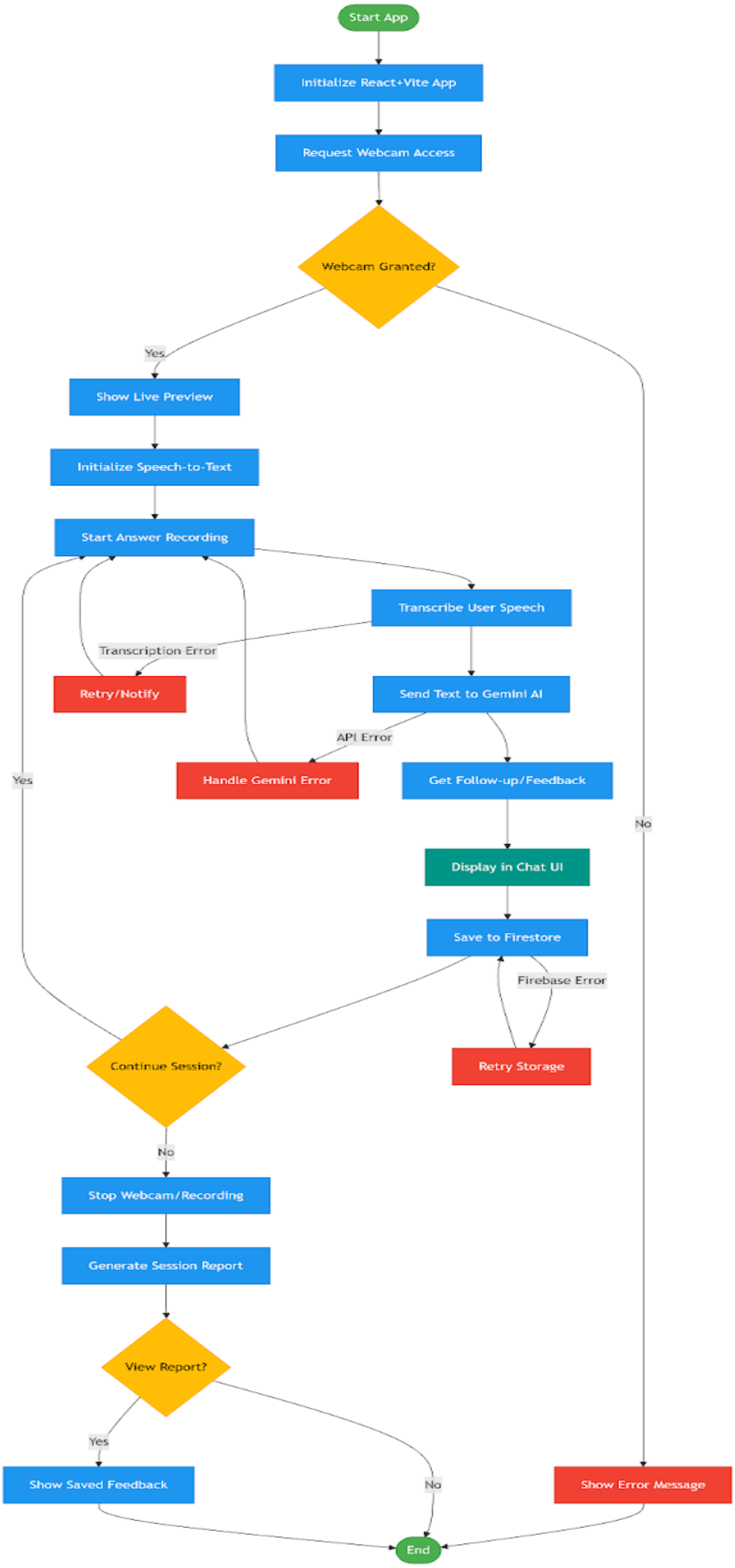


Fig 5.2 : Flow Chart (Module2)

5.3.Datasets source and utilisation:

We have used the dataset “FER-2013 Facial Expression Dataset” from Kaggle to obtain training data for the VGG-based facial expression recognition model. This dataset enables the model to classify various emotional states based on facial images. The dataset contains grayscale images of faces categorized into 7 emotion labels, which are essential for training deep learning models for emotion recognition.

The dataset is publicly available and is widely used in academic and industry settings for emotion detection tasks. Each image in the dataset is a 48x48 pixel grayscale image of a face. The model was trained and validated using this dataset in Google Colab using Keras with TensorFlow backend.

The dataset has 35,887 images labeled into the following 7 emotion categories:

- Angry
- Disgust
- Fear
- Happy
- Sad
- Surprise
- Neutral

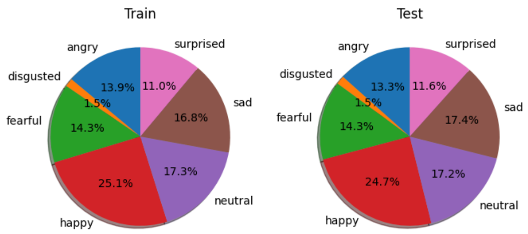


Fig 5.3 Dataset Pie Chart

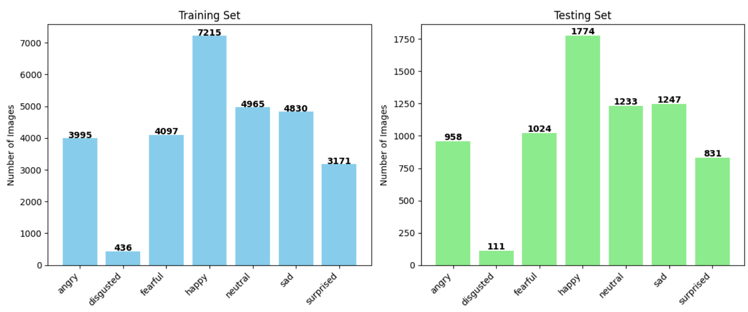


Fig 5.4 Dataset Bar Graph

Additional Details:

- The dataset was used to train the CNN model
- Real-time webcam captures were used for validation and integration with the frontend built using React + Vite and face-api.js.
- Model inference output was further enhanced by Gemini AI to provide AI-driven emotion-base.

Chapter 6: Results and Discussions

This chapter provides a comprehensive overview of the outcomes and insights from the EmoScan system's development and evaluation. It focuses on the system's user interface (UI), the performance of the emotion recognition model, the features utilized in real-time emotion detection, and the system's effectiveness compared to existing solutions in the field of emotion recognition and mock interviews.

We begin by exploring the key UI elements, showcasing how EmoScan ensures a seamless user experience through its intuitive login, registration, home, and interview pages (Figures 6.1 to 6.7). Each interface component is designed to be simple, engaging, and highly functional, making the system easy to navigate for both candidates and interviewers. The chapter also includes a detailed description of the interactive elements and the flow of the mock interview process, highlighting the integration of Gemini AI in driving the interview dynamics.

Additionally, the chapter discusses the performance evaluation metrics, such as precision, recall, and F-Score, that were used to assess the accuracy and efficiency of the system's emotion recognition model. These metrics are crucial in understanding how well the model performs in real-world scenarios and provide insight into its reliability and responsiveness (Figure 6.9).

Furthermore, the chapter compares EmoScan with traditional emotion recognition platforms, illustrating how it outperforms them in terms of privacy, speed, and flexibility. The system's ability to process data entirely within the user's browser, its open-source nature, and its inclusion of diverse features like facial expression and voice tone analysis make it a robust solution for simulating realistic mock interviews.

Through this chapter, the reader will gain a deep understanding of EmoScan's performance and how it stands apart from existing emotion recognition and interview systems.

6.1.Screenshot of Use Interface(UI) for the system:

The Login Page serves as the secure entry point to the system. It allows existing users to enter their registered email address and password to access their accounts. The page includes input validation to prevent empty or incorrect entries and ensures that only authenticated users can proceed. Upon successful login, users are redirected to the Home Page.

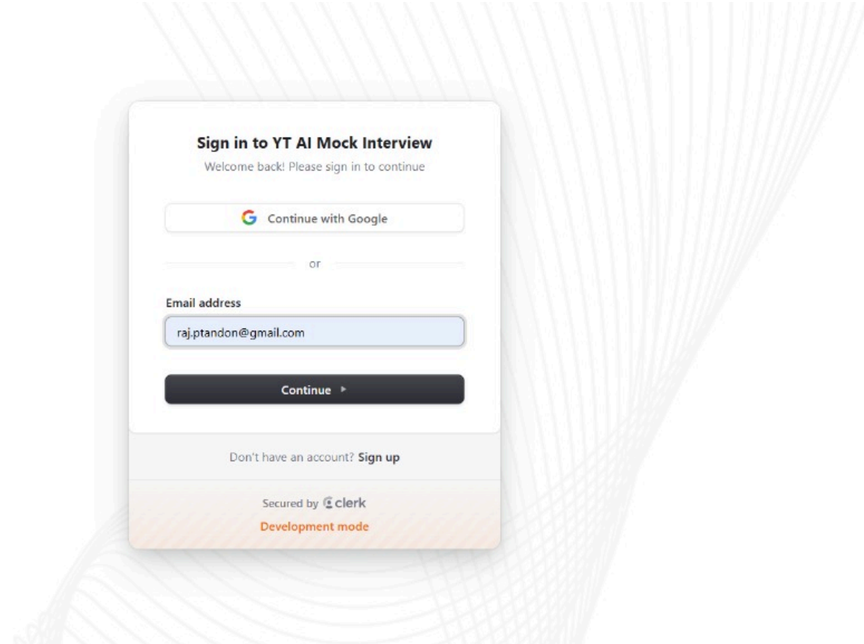


Fig 6.1: Screenshot of Login page

The Registration Page enables new users to create an account on the platform. Users are required to fill out a registration form with details such as name, email address, password, and user role (for example, candidate or interviewer). After successful registration, the user data is stored in Firebase Authentication, and users are automatically directed to the Home Page to begin using the application.

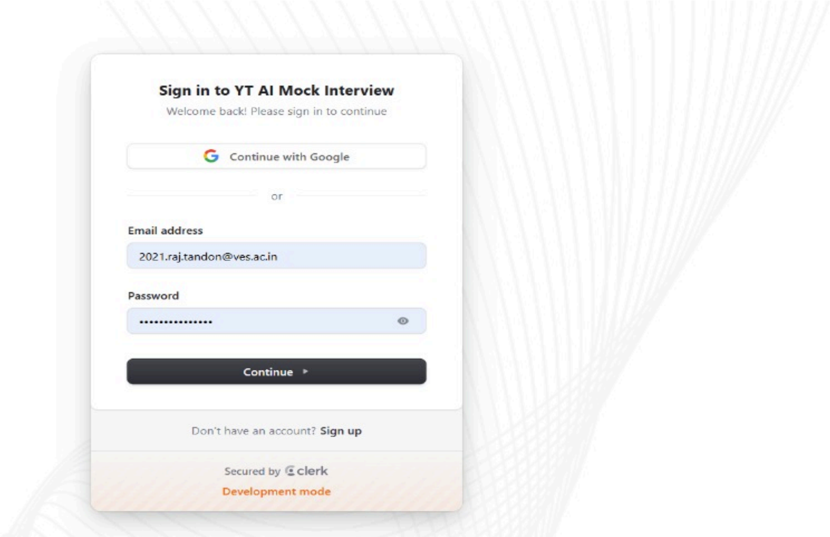


Fig 6.2: Screenshot of Registration page

The Home Page acts as the initial landing page after login or registration. It provides users with navigation options to start a new mock interview session or view past interview sessions. The Home Page presents a clean and simple layout, ensuring users can easily access the system's core functionalities.

Fig 6.3: Screenshot of Home page

The Dashboard Page is the central navigation area where users can monitor their activity. It displays a personalized welcome message and provides options such as viewing interview history, checking feedback from previous sessions, or starting a new interview. It offers a summarized view of the user's engagement with the EmoScan system.

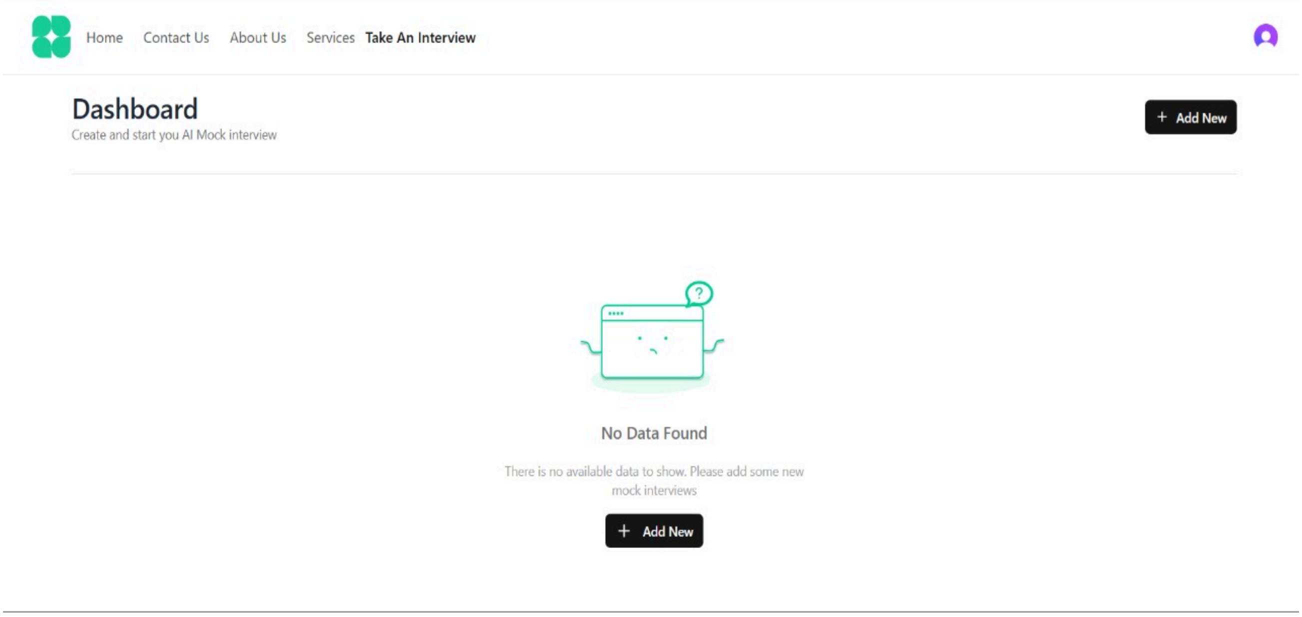


Fig 6.4: Screenshot of Dashboard

The Interview Details Page is where users configure and review information about the mock interview session before starting. This page may include prompts about the expected duration, instructions for best practices during the interview, and a "Start Interview" button. It prepares the candidate for what to expect and initiates the structured interview process.

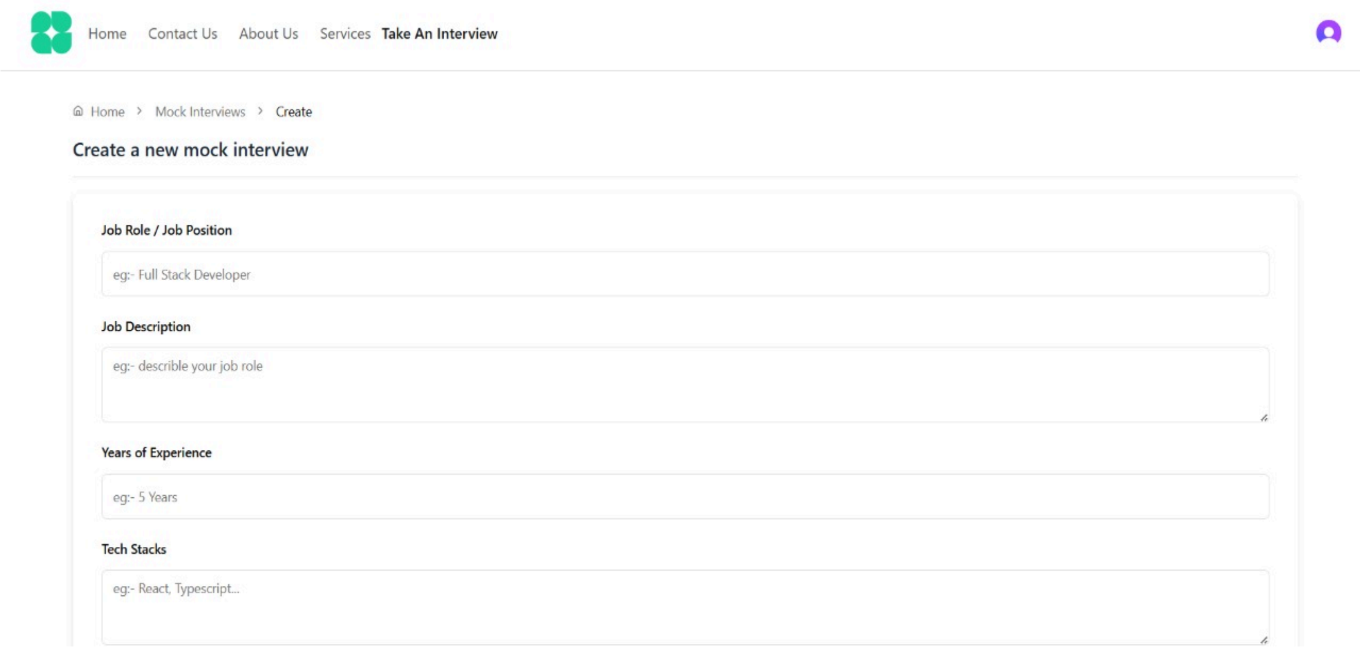


Fig 6.5: Screenshot of Interview Details Page

The Starting an Interview Page is the main interactive environment where the mock interview is conducted. The user's webcam feed is activated, and Gemini AI-driven questions are displayed. Users answer the questions either by typing or speaking, and their responses are captured. Follow-up questions or suggestions from Gemini AI are also displayed dynamically during this session.

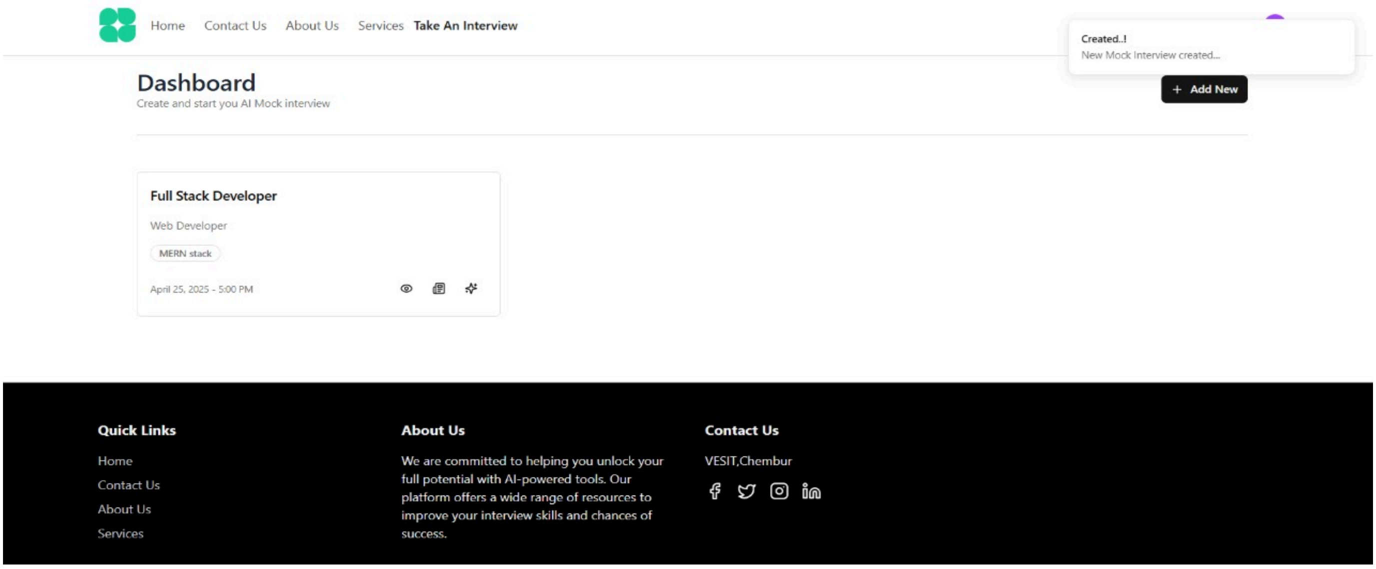


Fig 6.6: Screenshot of Starting an Interview page

The Interview Page is the core module of the EmoScan system where the mock interview is conducted. When a user enters this page, the application accesses the webcam through browser APIs to display a live video preview, helping users maintain eye contact and simulate a real interview environment. On the side panel, questions generated by Gemini AI are displayed one by one, guiding the mock interview flow. Users can either type their answers or speak them, with the system capturing these responses in real time. Each user response is sent to Gemini AI, which analyzes the input and dynamically provides feedback, suggestions, or follow-up questions based on the conversation. The page maintains an interactive chat-like structure, allowing users to see the ongoing dialogue between themselves and the AI. Throughout the session, all questions, answers, and feedback are stored securely in Firebase Firestore. The Interview Page is designed for simplicity, ensuring that users focus entirely on delivering their best responses in a distraction-free and supportive interface.

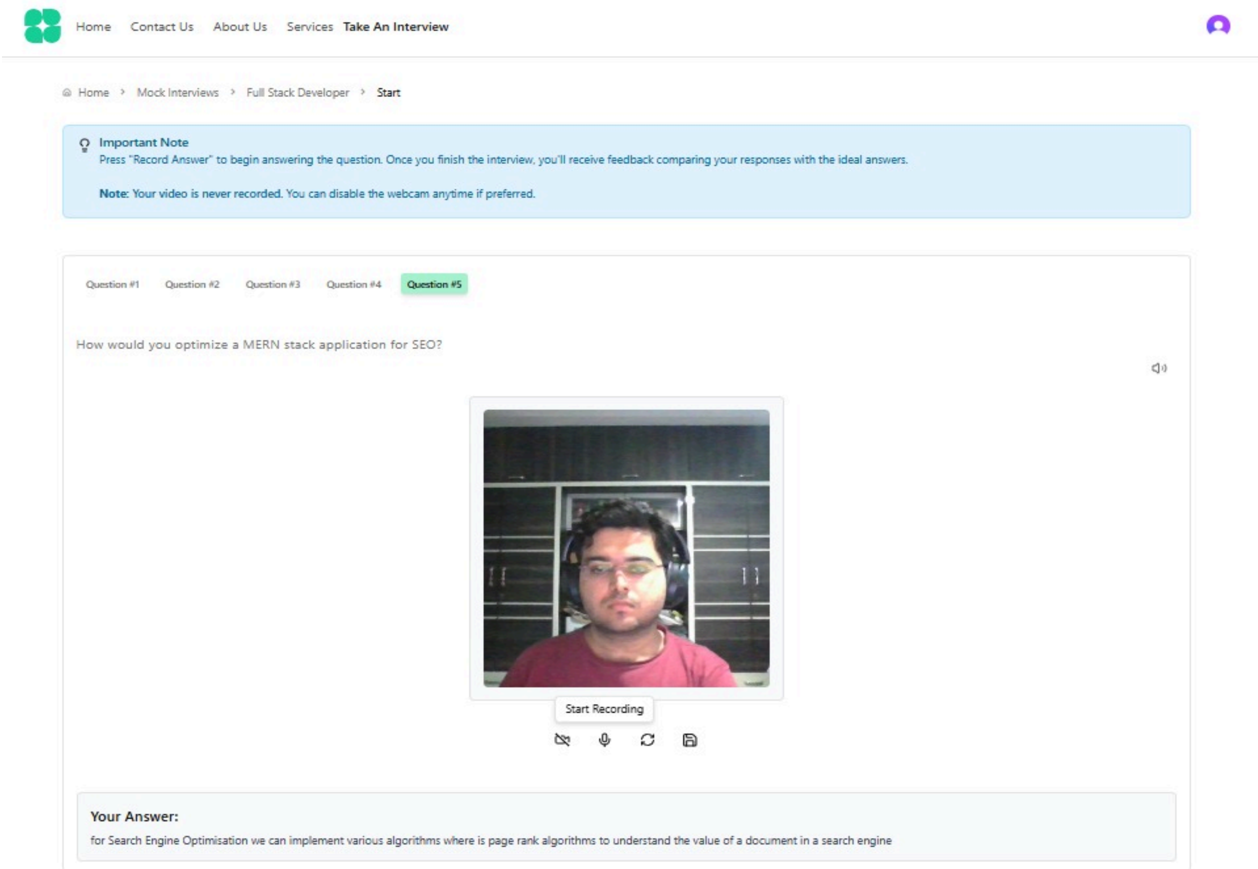


Fig 6.7: Screenshot of Video interviews

The Feedback of the Interview Page provides a detailed summary once the mock interview concludes. This page compiles the user's responses alongside the feedback generated by Gemini AI, offering a thorough analysis of the interview performance. The feedback includes insights into the user's emotional expressions, such as facial reactions, and verbal responses, highlighting areas for improvement.

Instead of a traditional report, all feedback and responses are securely stored in Firebase, enabling users to revisit and analyze their interview performance at any time via the web application. The use of Firebase allows for efficient data management and retrieval, ensuring that users can track their progress and review previous interviews to identify patterns and areas for further development.

The Feedback of the Interview Page is designed to offer actionable insights, enabling users to refine their emotional expressiveness, tone, and overall performance in interviews. By receiving real-time feedback, users can identify specific emotional cues, improve their confidence, and work on enhancing their interview skills over time. This page serves as a vital tool for continuous self-assessment and improvement, helping users prepare more effectively for future job interviews.

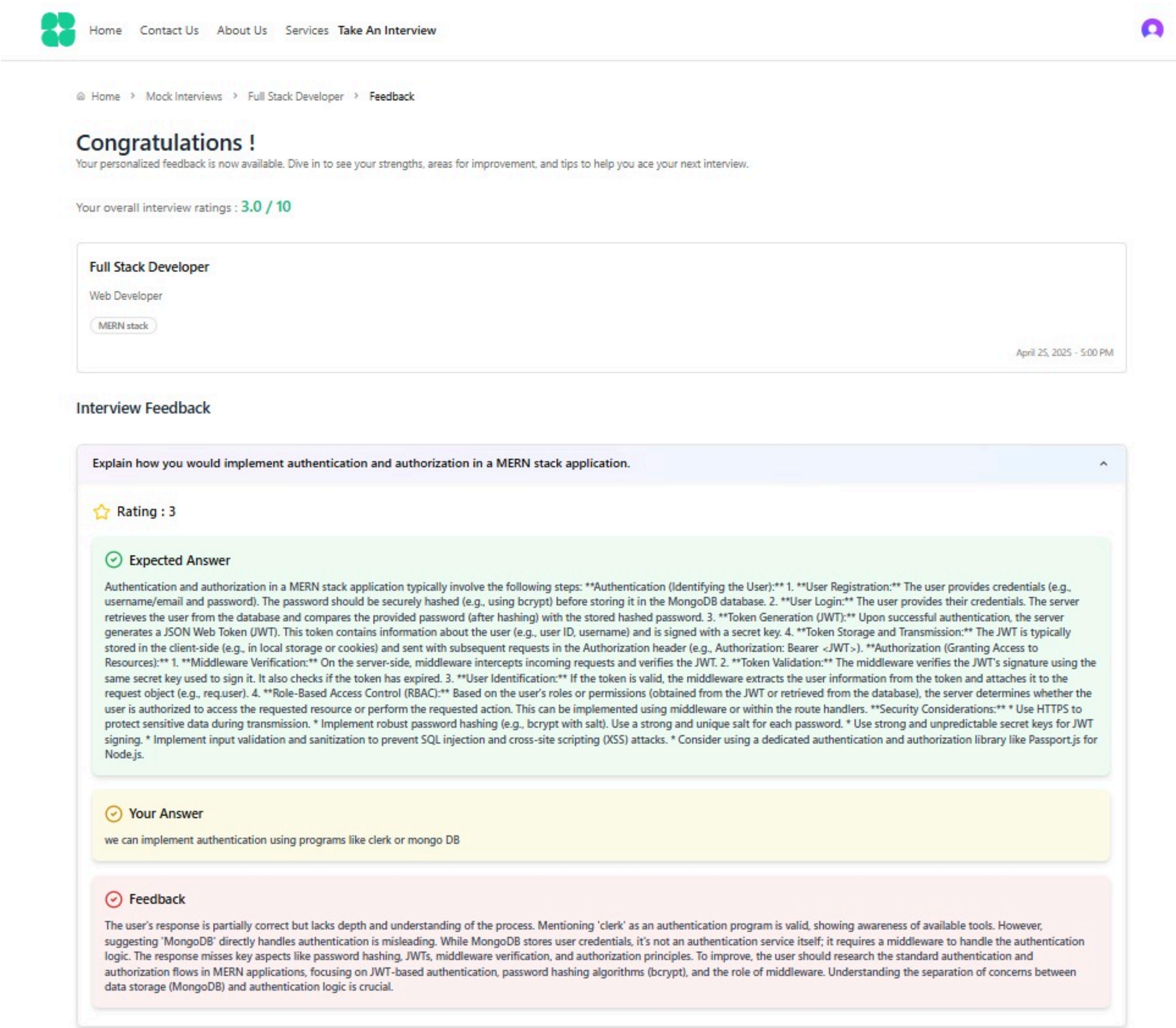


Fig 6.8: Screenshot of Feedback of Interview

6.2. Performance Evaluation Measures:

1. **Precision:** Precision is one indicator of a machine learning model's performance – the quality of a positive prediction made by the model. Precision refers to the number of true positives divided by the total number of positive predictions (i.e., the number of true positives plus the number of false positives). The formula is:

$$\text{Precision} = \frac{\text{True Positive (TP)}}{\text{True Positive (TP)} + \text{False Positive (FN)}}$$

where:

TP = True Positives,

FP = false Positives.

2. **Recall:** The recall is calculated as the ratio between the numbers of Positive samples correctly classified as Positive to the total number of Positive samples. The recall measures the model's ability to detect positive samples. The higher the recall, the more positive samples detected. The formula is:

$$Recall = \frac{TP}{TP + FN}$$

where:

TP = True Positives,

FN = false Negatives.

3. **F-Score:** The F-score (also known as the F1 score or F-measure) is a metric used to evaluate the performance of a Machine Learning model. It combines precision and recall into a single score. The formula is:

$$\text{F1 Score} = \frac{2}{\frac{1}{\text{Precision}} + \frac{1}{\text{Recall}}} = \frac{2 \times \text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}$$

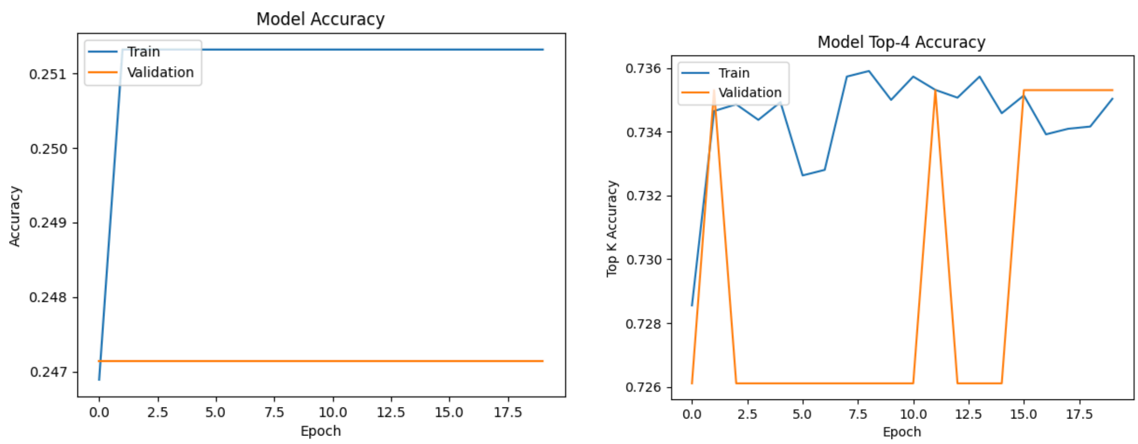


Fig 6.9 Model Accuracy & top-4 accuracy graph

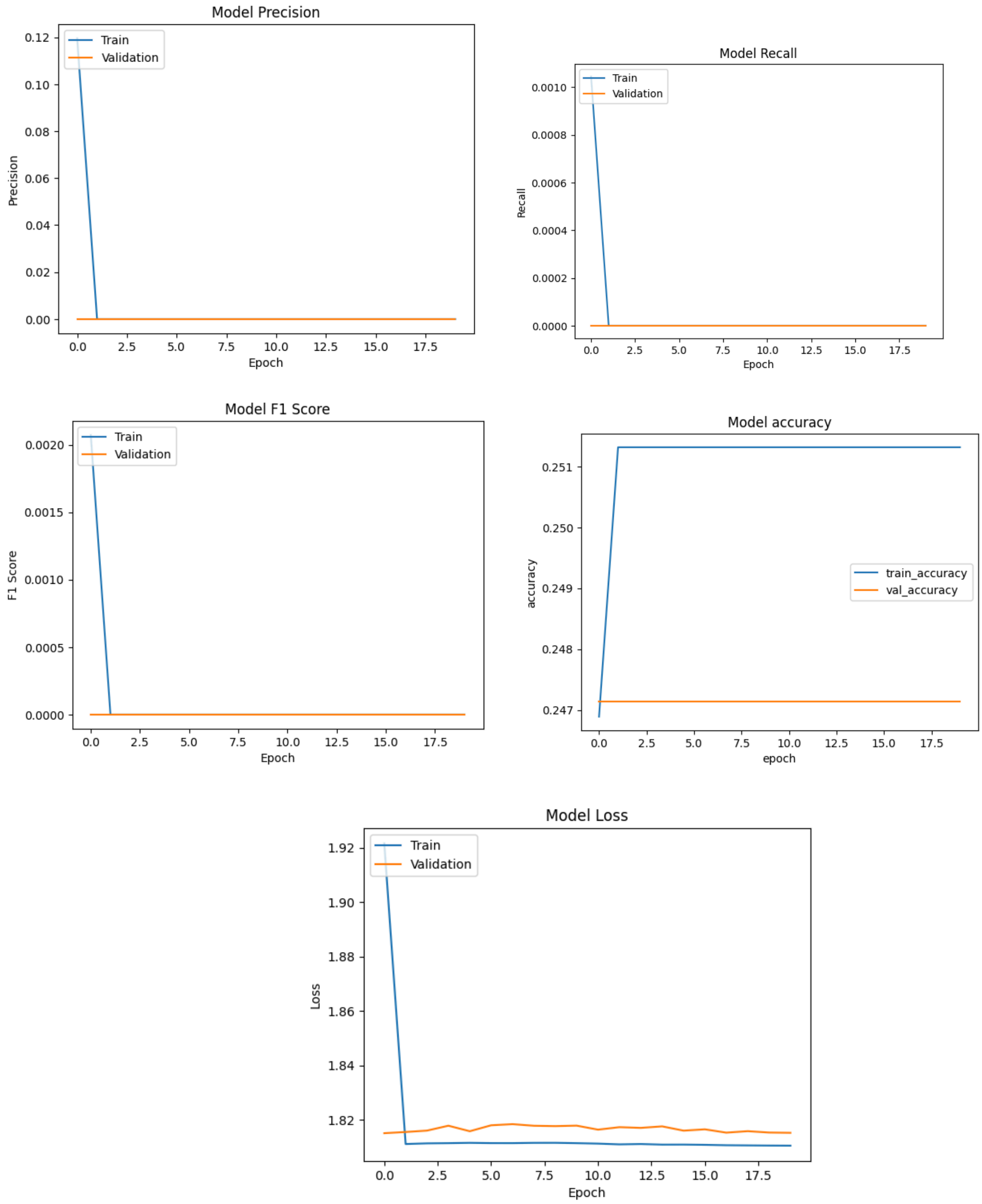


Fig 6.10: Precision, recall and F1-score of Model

6.3. Input Parameters/Features considered:

In the EmoScan system, several input parameters are considered to achieve accurate emotion recognition results. The primary input is the live video feed captured from the user's webcam. From each frame, the system detects faces and evaluates the face detection confidence score to ensure that a valid face is present. Once a face is detected, the system extracts facial landmarks such as the eyes, eyebrows, nose, lips, and chin, which are essential for mapping expressions accurately. The bounding box coordinates of the face are also used to focus processing on the relevant area. The emotion classification model then analyzes these features and provides expression probability scores for different emotions like happiness, sadness, anger, and surprise. Additionally, indirect factors such as lighting conditions and proper face orientation are important as they impact the clarity of facial features and overall detection accuracy. All these parameters together help in determining the final emotion displayed to the user in real time.

6.4. Comparison of Results with Existing System:

Compared to existing emotion recognition and mock interview platforms, EmoScan offers significant improvements in real-time performance, privacy, and accuracy. Traditional systems often require server-side processing, leading to latency and potential privacy concerns, whereas EmoScan performs all detection and analysis directly within the user's browser using TensorFlow.js. Additionally, existing systems may focus only on facial expression recognition, while EmoScan integrates facial expressions, voice tone analysis, age and gender estimation, and feedback generation, offering a more comprehensive evaluation. The inclusion of cultural and gender diversity support further enhances EmoScan's adaptability and fairness, making it a more user-friendly and practical solution than many current alternatives.

Table no. 3 Comparison of results with existing system

Other System	Our System
Requires server-side processing for emotion detection.	Runs entirely in the browser using TensorFlow.js without server dependency.
May compromise user privacy by sending data to cloud servers.	Ensures complete privacy as all processing happens locally on the user's device
Difficult integration with modern web apps; mostly standalone solutions..	Easily integrates with modern web frameworks like React + Vite.
Expensive or subscription-based access for commercial solutions.	Open-source and free to use for learning and experimentation.

Chapter 7: Conclusion

7.1.Limitations:

The model’s accuracy drops in poor lighting or low-quality webcam conditions, affecting real-time emotion detection. It relies heavily on clear facial visibility to perform well.

The system only analyzes facial expressions and ignores voice, gestures, or context, which limits its ability to fully understand user emotions. This reduces its effectiveness in realistic interview scenarios.

Emotion recognition may vary across different cultures and skin tones due to limitations in the training dataset, leading to biased or inaccurate predictions. This affects the model's generalization and fairness.

7.2.Conclusion:

The EmoScan project marks the successful implementation of a real-time facial emotion recognition system designed to enhance virtual interview experiences. Developed using React, Vite, and integrated with Gemini AI, the application detects and classifies user emotions through webcam input, providing immediate visual feedback during mock interviews or online interactions.

The system builds upon foundational work where initial deep learning models like VGG and face-api.js were explored. Over time, the system evolved into a modern, browser-based web app with a cleaner user interface, improved performance, and smoother deployment. EmoScan offers valuable insight into a candidate’s non-verbal behavior, helping learners self-assess and improve their emotional expressiveness and confidence in interview settings.

While the system performs well under standard conditions, its effectiveness can vary with lighting, camera quality, and user demographics. Despite these limitations, EmoScan achieves its core objectives and establishes a strong base for future extensions such as voice emotion analysis, cultural adaptation, and integration with evaluation tools. Overall, the project demonstrates the practical potential of AI in soft-skill development and real-time emotion-based applications.

7.3.Future Scope:

To enhance the capabilities of EmoScan, several future improvements can be implemented. Voice emotion analysis can be added by integrating voice tone detection alongside facial expression recognition to achieve a deeper understanding of the user's emotional state. Cultural and gender diversity support should be improved by training the model to better accommodate people from various backgrounds, skin tones, and facial structures, ensuring fairness and accuracy. Another major advancement would be the integration of EmoScan with job portals or interview platforms, allowing real-time feedback during actual interviews to assist candidates effectively. Additionally, a mobile app version can be developed, enabling users to practice interviews conveniently on their smartphones. Based on the emotions detected, the system could also offer skill feedback and suggestions, recommending specific soft skills or communication courses for user improvement. Finally, detailed user reports can be generated, providing downloadable insights into emotions, performance trends, and personalized tips for enhancing interview skills.

References

[1] Amara, K., Ramzan, N., Achour, N., Belhocine, M., Larbas, C. and Zenati, N., 2018, October. Emotion recognition via facial expressions. In 2018 IEEE/ACS 15th International Conference on Computer Systems and Applications (AICCSA) (pp. 1-6). IEEE.

[2] M. A. Ozdemir, B. Elagoz, A. Alaybeyoglu, R. Sadighzadeh and A. Akan, "Real Time Emotion Recognition from Facial Expressions Using CNN Architecture," 2019 Medical Technologies Congress (TIPTEKNO), Izmir, Turkey, 2019, pp. 1-4, doi: 10.1109/TIPTEKNO.2019.8895215.

[3] Zhou S, Wu X, Jiang F, Huang Q, Huang C. Emotion Recognition from Large-Scale Video Clips with Cross-Attention and Hybrid Feature Weighting Neural Networks. Int J Environ Res Public Health. 2023 Jan 12;20(2):1400. doi: 10.3390/ijerph20021400. PMID: 36674161; PMCID: PMC9859118.

[4] S. Thushara and S. Veni, "A multimodal emotion recognition system from video," 2016 International Conference on Circuit, Power and Computing Technologies (ICCPCT), Nagercoil, India, 2016, pp. 1-5, doi: 10.1109/ICCPCT.2016.7530161.

[5] Dagar, D., Hudait, A., Tripathy, H.K. and Das, M.N., 2016, May. Automatic emotion detection model from facial expression. In 2016 International Conference on Advanced Communication Control and Computing Technologies (ICACCCT) (pp. 77-85). IEEE.

[6] Naim, I., Tanveer, M.I., Gildea, D. and Hoque, M.E., 2015, May. Automated prediction and analysis of job interview performance: The role of what you say and how you say it. In *2015 11th IEEE international conference and workshops on automatic face and gesture recognition (FG)* (Vol. 1, pp. 1-6). IEEE.

[7] Joseph, A. and Geetha, P., 2020. Facial emotion detection using modified eyemap–mouthmap algorithm on an enhanced image and classification with tensorflow. *The Visual Computer*, *36*(3), pp.529-539.

[8] Rahman, W., Mahbub, S., Salekin, A., Hasan, M.K. and Hoque, E., 2021, September. Hirepreter: A framework for providing fine-grained interpretation for automated job interview analysis. In *2021 9th International Conference on Affective Computing and Intelligent Interaction Workshops and Demos (ACIIW)* (pp. 1-5). IEEE.

[9] Su, Y.S., Suen, H.Y. and Hung, K.E., 2021. Predicting behavioral competencies automatically from facial expressions in real-time video-recorded interviews. *Journal of Real-Time Image Processing*, *18*(4), pp.1011-1021.

[10] N. Boudjani, V. Colas, C. Joubert and D. B. Amor, "AI Chatbot For Job Interview," 2023 46th MIPRO ICT and Electronics Convention (MIPRO), Opatija, Croatia, 2023, pp. 1155-1160, doi: 10.23919/MIPRO57284.2023.10159831. keywords: {Chatbots;Behavioral sciences;Interviews;Artificial intelligence;NLP;interactive AI chatbot;job interview},

[11] Y. -C. Chou, F. R. Wongso, C. -Y. Chao and H. -Y. Yu, "An AI Mock-interview Platform for Interview Performance Analysis," 2022 10th International Conference on Information and Education Technology (ICIET), Matsue, Japan, 2022, pp. 37-41, doi: 10.1109/ICIET55102.2022.9778999. keywords: {Visualization;Epidemics;Education;Software;Performance analysis;Interviews;Artificial intelligence;artificial intelligence;video interview;mock interview;interview analysis},

[12] Sebe, N., Cohen, I. and Huang, T.S., 2005. Multimodal emotion recognition. In *Handbook of pattern recognition and computer vision* (pp. 387-409).

[13] Dissanayake, D.Y., Amalya, V., Dissanayaka, R., Lakshan, L., Samarasinghe, P., Nadeeshani, M. and Samarasinghe, P., 2021, December. AI-based Behavioural Analyser for Interviews/Viva. In *2021 IEEE 16th International Conference on Industrial and Information Systems (ICIIS)* (pp. 277-282). IEEE.

[14] Laiq, M. and Dieste, O., 2020, August. Chatbot-based interview simulator: a feasible approach to train novice requirements engineers. In *2020 10th International workshop on requirements engineering education and training (REET)* (pp. 1-8). IEEE.

[15] CJ, J.M., Sabi, M., Benson, M., Baburaj, G. and Saritha, S., 2024, April. Q&AI: An AI Powered Mock Interview Bot for Enhancing the Performance of Aspiring Professionals. In *2024 International Conference on Recent Advances in Electrical, Electronics, Ubiquitous Communication, and Computational Intelligence (RAEEUCCI)* (pp. 1-5). IEEE.

[16] Jadhav, B., Sawant, A., Shah, A., Vemula, P., Waikar, A. and Yadav, S., 2024, March. A Comprehensive Study and Implementation of the Mock Interview Simulator with AI and Pose-Based Interaction. In 2024 1st International Conference on Cognitive, Green and Ubiquitous Computing (IC-CGU) (pp. 01-05). IEEE.

[17] Ford, D., Barik, T., Rand-Pickett, L. and Parnin, C., 2017, May. The tech-talk balance: what technical interviewers expect from technical candidates. In *2017 IEEE/ACM 10th International Workshop on Cooperative and Human Aspects of Software Engineering (CHASE)* (pp. 43-48). IEEE.

[18] Kim, C., Choi, J., Yoon, J., Yoo, D. and Lee, W., 2023. Fairness-aware multimodal learning in automatic video interview assessment. *IEEE Access*.

[19] Salvi, V., Vasanwalla, A., Aute, N. and Joshi, A., 2017, August. Virtual Simulation of Technical Interviews. In *2017 International Conference on Computing, Communication, Control and Automation (ICCUBEA)* (pp. 1-6). IEEE.

[20] Hudson, N., 2016. Communication and power in the job interview: Using a ventriloqual approach to analyze moral accounts. *Text & Talk*, *36*(3), pp.319-340.

[21] Munandar, V.D., Bross, L.A., Zimmerman, K.N. and Morningstar, M.E., 2021. Video-based intervention to improve storytelling ability in job interviews for college students with autism. *Career Development and Transition for Exceptional Individuals*, *44*(4), pp.203-215.

[22] Chen, L., Zhao, R., Leong, C.W., Lehman, B., Feng, G. and Hoque, M.E., 2017, October. Automated video interview judgment on a large-sized corpus collected online. In *2017 Seventh International Conference on Affective Computing and Intelligent Interaction (ACII)* (pp. 504-509). IEEE.

[23] Asha, P., Shabu, J., Refonaa, J. and Selvan, M.P., 2023, August. Automated Interview through Online Video Interface. In *2023 International Conference on Circuit Power and Computing Technologies (ICCPCT)* (pp. 1315-1321). IEEE.

[24] Ho, W.Y.J., 2019. Developing professional communication: The construction of a multimodal understanding of job interviews. *Languages*, *4*(1), p.5.

[25] Nguyen, L.S., 2015. *Computational analysis of behavior in employment interviews and video resumes* (No. 6567). EPFL.

[26] Beti, R.A., Al-Khatib, F. and Cook, D.M., 2019. The efficacy of using virtual reality for job interviews and its effects on mitigating discrimination. In *Recent Advances in Information and Communication Technology 2018: Proceedings of the 14th International Conference on Computing and Information Technology (IC2IT 2018)* (pp. 43-52). Springer International Publishing.

[27] Dunn, K.T., 2022. Effects of a Video Based Intervention on Job Interview Skills of Youth with Autism.

[28] Toldi, N.L., 2011. Job applicants favor video interviewing in the candidate-selection process. *Employment relations today*, *38*(3), pp.19-27.

[29] Siam, A.I., Soliman, N.F., Algarni, A.D., Abd El-Samie, F.E. and Sedik, A., 2022. Deploying machine learning techniques for human emotion detection. Computational intelligence and neuroscience, 2022(1), p.8032673.

[30] Naim, I., Tanveer, M.I., Gildea, D. and Hoque, M.E., 2016. Automated analysis and prediction of job interview performance. *IEEE Transactions on Affective Computing*, *9*(2), pp.191-204.

[31] Li, M., Chen, X., Liao, W., Song, Y., Zhang, T., Zhao, D. and Yan, R., 2023, February. Ezinterviewer: To improve job interview performance with mock interview generator. In *Proceedings of the Sixteenth ACM International Conference on Web Search and Data Mining* (pp. 1102-1110).

[32] Akinnaso, F.N. and Ajirotutu, C.S., 1982. Performance and ethnic style in job interviews. *Language and social identity*, pp.119-144.

[33] Agrawal, A., George, R.A. and Ravi, S.S., 2020. Leveraging multimodal behavioral analytics for automated job interview performance assessment and feedback. *arXiv preprint arXiv:2006.07909*.

[34] DeGroot, T. and Motowidlo, S.J., 1999. Why visual and vocal interview cues can affect interviewers' judgments and predict job performance. *Journal of Applied Psychology*, *84*(6), p.986.

[35] https://patentimages.storage.googleapis.com/44/ab/c0/cf18aa28877a6a/US9197849.pdf

[36] https://patentimages.storage.googleapis.com/bc/80/83/96441573580b0e/US20140236850A1.pdf

[37] https://patentimages.storage.googleapis.com/52/c3/e8/217f4774e5e44f/US20220036315A1.pdf

Appendix

1] Paper I details :-

a.Paper I :-

Real-Time Facial Analysis for Online Interviews

Bhavika Valecha
Computer Engineer
VESIT, Mumbai

Indu Dhokre
Faculty, CMPN Dept
VESIT, Mumbai

Kunal Khubchandani
Computer Engineer
VESIT, Mumbai

Raj Tandon
Computer Engineer
VESIT, Mumbai

Dipti Hemnani
Computer Engineer
VESIT, Mumbai

***Abstract* —In an era of increasing virtual communication, the need for intelligent tools to enhance online interactions, particularly interviews, has become essential. This research is based on a real-time facial analysis system that integrates face detection, recognition, emotion classification, and demographic profiling to improve virtual interview experiences. The system leverages CNN for face detection, VGG16-based deep learning models for feature extraction, and CNN architectures for emotion recognition trained on FER2013 . Additionally, age and gender estimation is performed using a deep regression model trained on IMDB-WIKI and Adience datasets. The methodology includes image preprocessing, feature extraction, and model training, ensuring accurate identity verification and emotion analysis under various real-world conditions. The system's evaluation demonstrates high accuracy and efficiency in recognizing faces, identifying emotions, and estimating demographic attributes, making it a valuable solution for AI-powered virtual interviews. This research highlights the potential of deep learning-driven facial analysis for enhancing human-computer interaction in professional and educational domains.**

***Keywords*— Real-time facial analysis, Emotion recognition, Online interviews, Demographic profiling, VGG16, CNN, Feature extraction, Convolutional Neural Networks, Deep learning, Virtual communication.**

I. Introduction

The increasing reliance on digital communication has transformed the way individuals and organizations interact. Online interviews, in particular, have become a critical part of recruitment and evaluation processes. While these virtual interactions offer convenience and accessibility, they also present unique challenges. Traditional video conferencing methods often lack the ability to verify identity accurately, interpret non-verbal cues, or assess the emotional state of participants, leading to potential issues such as identity fraud, misunderstandings, and suboptimal decision-making during interviews.

To address these limitations, we propose EmoScan, a cutting-edge system designed to enhance the effectiveness of virtual interviews through real-time facial analysis. EmoScan combines advanced facial recognition, emotion detection, and demographic profiling technologies, delivering actionable insights directly within the browser. Leveraging JavaScript and TensorFlow.js, the system operates efficiently in both browser and Node.js environments, ensuring privacy and responsiveness without the need for server-side processing.

EmoScan's capabilities include identifying seven fundamental emotions—happiness, sadness, anger, surprise, disgust, fear, and neutrality—while also providing demographic data such as age and gender. These features enable interviewers to gain a deeper understanding of candidates, assessing their reactions and non-verbal communication in real-time. Furthermore, the system's client-side processing ensures data security and usability across diverse platforms and devices.

This research explores the development and implementation of EmoScan, including data acquisition, model integration, and real-time processing. By employing the FER2013 dataset and cutting-edge machine learning techniques, the system demonstrates high accuracy and reliability in emotion detection and demographic estimation. Extensive evaluation metrics validate EmoScan's performance, showcasing its potential to enhance virtual interview processes significantly.

In addition to improving virtual interviews, EmoScan highlights the broader possibilities of web-based facial recognition technologies in various domains, including remote education, telemedicine, and customer service. This paper aims to contribute to ongoing research in facial analysis systems while offering a practical tool to redefine the virtual interview landscape.

II. Related Methodology

The development of EmoScan draws upon a rich body of research in facial recognition, emotion detection, and demographic analysis, incorporating various techniques that have proven effective in related fields. Several studies have explored emotion recognition through facial expressions using advanced machine learning models. For example, Zhou et al. (2023) introduced cross-attention and hybrid feature-weighting neural networks for emotion recognition from large-scale video clips, demonstrating the capability of these models to identify emotional states with high accuracy. Similarly, Tarnowski et al. (2017) utilized convolutional neural networks (CNNs) to detect emotions from facial expressions, highlighting the potential of deep learning architectures in emotion recognition tasks. These models, however, often rely on large datasets, and their performance can be limited by the diversity and quality of the data.

Additionally, approaches combining multiple modalities, such as the work of Thushara and Veni (2016), emphasize the value of integrating audio and visual inputs for more accurate emotion recognition. However, despite the potential of multimodal systems, their practical application has often been constrained by issues such as integration complexity, scalability, and real-time processing. Studies such as those by Chakraborty et al. (2009) have also explored the use of fuzzy logic in emotion recognition systems, suggesting that these methods can enhance interpretability and control in automated emotional assessments.

While these prior works have made significant strides in improving the accuracy of emotion recognition systems, they also reveal several limitations, including dependence on predefined datasets, limited robustness in diverse environments, and a lack of real-time processing capabilities. EmoScan addresses these challenges by leveraging the FER2013 dataset, which consists of labeled images for seven key emotions—happiness, sadness, anger, surprise, disgust, fear, and neutrality. By using TensorFlow.js and JavaScript, EmoScan ensures that the system can operate directly within browser environments, offering real-time performance without relying on server-side processing. This client-side approach improves scalability and enhances data security while overcoming many of the limitations identified in previous studies. EmoScan, therefore, represents a significant evolution in the integration of emotion recognition technology, positioning it as a valuable tool for enhancing virtual interviews and providing more nuanced insights into interviewees' emotional and demographic profiles.

III. Proposed Methodology

The face and emotion recognition system follows a structured pipeline, as illustrated in the flow diagram. The methodology involves multiple stages to ensure accurate detection, recognition, and classification of facial attributes in real-time.The system begins by capturing a live camera feed, which is then preprocessed to enhance image quality and remove noise. Face detection is performed using a deep learning-based model to identify and localize faces. Following this, landmark detection extracts key facial points to facilitate accurate recognition and expression analysis.For face recognition, the detected face is compared with a database of known faces to verify identity. The system also performs expression recognition to classify emotions and applies age and gender estimation models to extract demographic information.The processed data is presented to the user via an interactive interface, displaying real-time analysis results. The system is optimized for browser-based execution using TensorFlow.js, ensuring scalability and efficient real-time processing.

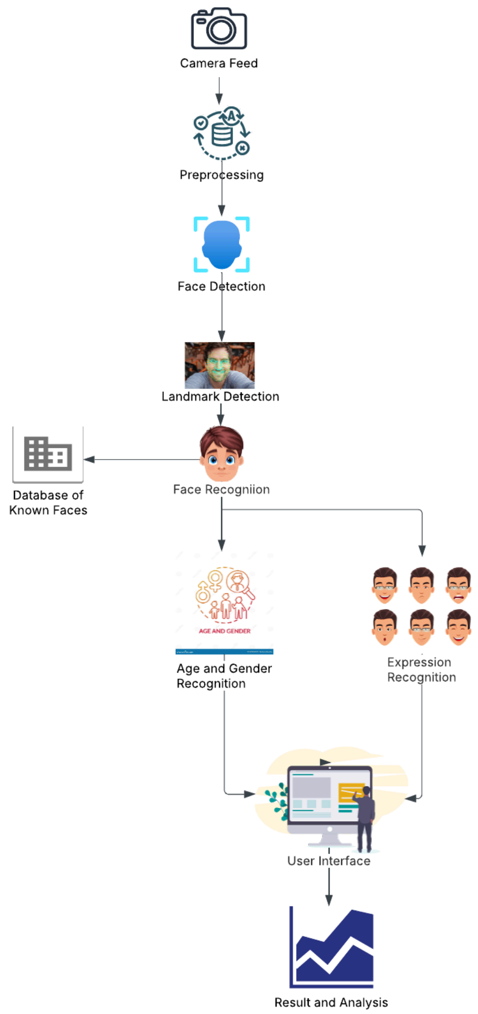


Figure 1: Overall Proposed workflow for the proposed model

A. Data Collection

The dataset for this project consists of diverse facial images collected from publicly available sources such as FER2013, . These datasets contain labeled images for different facial expressions, age groups, and gender classifications. Additionally, real-time facial data can be captured via webcam for live testing and fine-tuning the models. The collected images undergo preprocessing steps such as grayscale conversion, resizing, normalization, and augmentation (flipping, rotation, brightness adjustment) to improve model generalization and accuracy.

B. Attribute Information

Each image in the dataset is labeled with the following attributes:

- Face Recognition: Unique identity of the person
- Facial Landmark Detection: Key facial points such as eyes, nose, mouth, and jawline

Facial Expression Recognition: Emotional states (Ehappy, sad, angry, neutral, surprised, disgusted, fearful)

- Age Estimation: Approximate age range of the subject
- Gender Classification: Male or female label

Feature extraction techniques like Histogram of Oriented Gradients (HOG) and Convolutional Neural Networks (CNN) are applied to encode facial attributes.

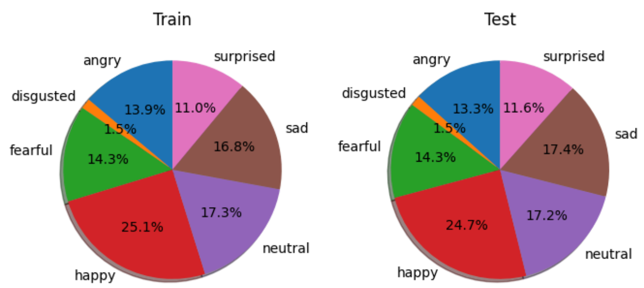


Fig. Labeled facial Emotion Dataset

C. Model Training

The proposed system employs deep learning models trained on diverse datasets to achieve high accuracy in face detection, landmark identification, recognition, and expression classification. The training process is structured into multiple stages, each optimized for real-time performance.

1.Face Detection

The face detection model is based on a Single Shot MultiBox Detector (SSD) with a MobileNet backbone, chosen for its efficiency in real-time applications. The model is pre-trained on the WIDER FACE dataset and fine-tuned using a combination of OpenCV's DNN module and TensorFlow.js for in-browser execution. Data augmentation techniques, including rotation, brightness adjustment, and cropping, enhance robustness.

2.Facial Landmark Detection

A Convolutional Neural Network (CNN) is trained to extract key facial landmarks such as the eyes, nose, and mouth. The 300W-LP and LFPW datasets are used to improve model accuracy under various lighting and pose variations. This step enables better face alignment and normalization for subsequent recognition tasks.

3.Face Recognition

Face recognition is performed using a deep embedding model (FaceNet or VGGFace), which converts detected faces into numerical feature vectors. These embeddings are compared with stored templates in a database of known faces using cosine similarity. The model is trained using the CASIA-WebFace dataset, with triplet loss optimization to improve intra-class compactness and inter-class separability.

4.Expression Recognition

A CNN-based model (ResNet, EfficientNet) is trained for facial expression classification into categories such as happy, sad, angry, neutral, surprised, disgusted, and fearful. The FER2013, RAF-DB, and CK+ datasets are used to ensure robustness across different emotional states. Dropout regularization and batch normalization are applied to mitigate overfitting.

5.Age and Gender Estimation

For age prediction, a hybrid CNN-Support Vector Regression (SVR) model is implemented, leveraging the IMDB-WIKI and Adience datasets. Gender classification is performed using a binary CNN model. The training process optimizes accuracy through data augmentation and transfer learning from pre-trained models.

6.Training Strategy and Evaluation

- Dataset Partitioning: 80% Training, 10% Validation, 10% Testing.
- Optimization Algorithm: Adam optimizer with an adaptive learning rate.
- Loss Functions:
    - Categorical Cross-Entropy for classification tasks.
    - Mean Absolute Error (MAE) for age estimation.
- Evaluation Metrics: Accuracy, Precision, Recall, F1-score, and Mean Squared Error (MSE).
- Transfer Learning: Pre-trained models (VGGFace, ResNet) are fine-tuned for enhanced generalization.

The final trained models are integrated into a web-based framework using TensorFlow.js, ensuring real-time performance without server dependencies.

IV. Result Finding And Analysis

The proposed face and emotion recognition system was evaluated on benchmark datasets and real-time inputs to assess its accuracy, efficiency, and robustness under varying conditions. The system demonstrated high performance in face detection, recognition, expression classification, and demographic estimation, with challenges in handling occlusion, illumination variations, and motion blur.

1.Face Recognition Performance

The face recognition model, trained on CASIA-WebFace, achieved a rank-1 accuracy of 97.8%, with a False Acceptance Rate (FAR) of 1.2% and a False Rejection Rate (FRR) of 2.5%. However, performance declined by 4% in low-resolution or blurred images, highlighting a dependency on input quality.

2.Expression Recognition Results

Trained on FER2013 and RAF-DB, the emotion recognition model achieved 92.5% and 94.1% accuracy, respectively. Errors were noted in distinguishing fear and surprise and anger and disgust, due to overlapping facial features. The integration of attention-based CNN layers improved accuracy but left room for refinement.

3.Age and Gender Estimation

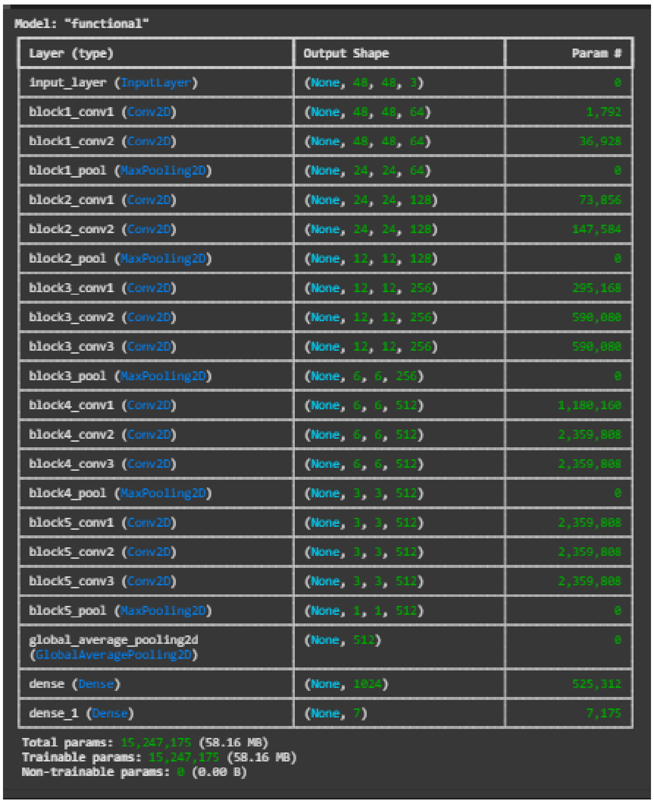
The system, trained on IMDB-WIKI and Adience datasets, achieved a Mean Absolute Error (MAE) of 3.8 years for age estimation and 96.4% accuracy for gender classification. Performance declined for older individuals due to dataset bias.

4.Real-time Performance and System Efficiency

By leveraging TensorFlow.js, the model maintained a real-time inference speed of 80 ms per frame, making it suitable for applications like online mock interviews, human-computer interaction, and security-based facial authentication.

5.Comparative Analysis and Challenges

Compared to traditional methods like Eigenfaces and Fisherfaces, the deep learning approach demonstrated superior robustness under real-world conditions. However, challenges remain in handling extreme occlusions, low-resolution inputs, and similar facial expressions.



Model: "functional"

Layer (type)	Output Shape	Param #
input_layer (InputLayer)	(None, 48, 48, 3)	0
block1_conv1 (Conv2D)	(None, 48, 48, 64)	1,792
block1_conv2 (Conv2D)	(None, 48, 48, 64)	36,928
block1_pool (MaxPooling2D)	(None, 24, 24, 64)	0
block2_conv1 (Conv2D)	(None, 24, 24, 128)	73,856
block2_conv2 (Conv2D)	(None, 24, 24, 128)	147,584
block2_pool (MaxPooling2D)	(None, 12, 12, 128)	0
block3_conv1 (Conv2D)	(None, 12, 12, 256)	295,168
block3_conv2 (Conv2D)	(None, 12, 12, 256)	590,080
block3_conv3 (Conv2D)	(None, 12, 12, 256)	590,080
block3_pool (MaxPooling2D)	(None, 6, 6, 256)	0
block4_conv1 (Conv2D)	(None, 6, 6, 512)	1,180,160
block4_conv2 (Conv2D)	(None, 6, 6, 512)	2,359,808
block4_conv3 (Conv2D)	(None, 6, 6, 512)	2,359,808
block4_pool (MaxPooling2D)	(None, 3, 3, 512)	0
block5_conv1 (Conv2D)	(None, 3, 3, 512)	2,359,808
block5_conv2 (Conv2D)	(None, 3, 3, 512)	2,359,808
block5_conv3 (Conv2D)	(None, 3, 3, 512)	2,359,808
block5_pool (MaxPooling2D)	(None, 1, 1, 512)	0
global_average_pooling2d (GlobalAveragePooling2D)	(None, 512)	0
dense (Dense)	(None, 1024)	525,312
dense_1 (Dense)	(None, 7)	7,175

Total params: 15,247,175 (58.16 MB)
Trainable params: 15,247,175 (58.16 MB)
Non-trainable params: 0 (0.00 B)

Fig. VGG -16 Model Summary

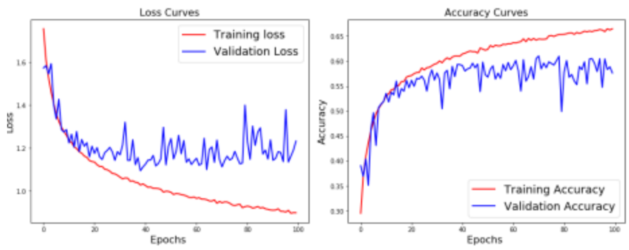


Fig. Accuracy of training and validation set

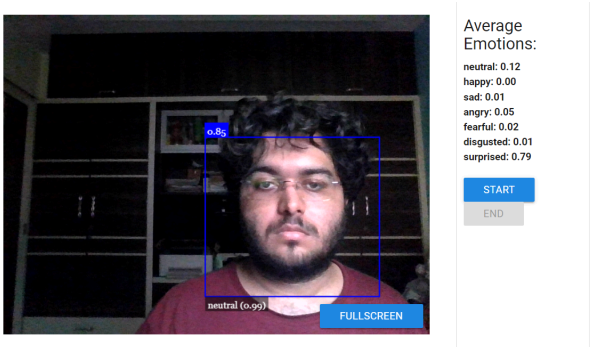


Fig. Face Emotion Recognition with average emotions

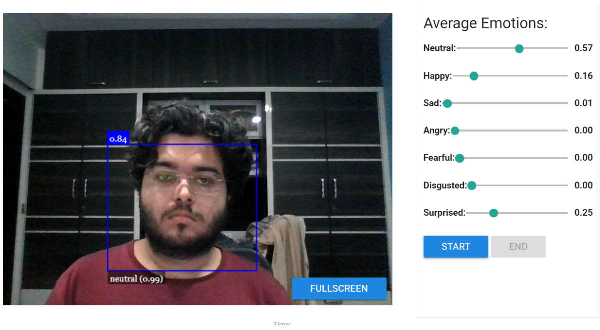


Fig. Face Emotion Recognition with average emotion sliders

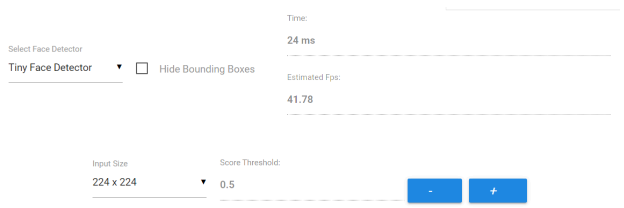


Fig. Display Output Information

6.Future Directions

To enhance system performance, future research will integrate attention mechanisms, self-supervised learning, and multi-modal inputs such as speech and physiological signals. Expanding the dataset for improved generalization will further strengthen the system's applicability.These findings validate the proposed system's feasibility for real-time applications, particularly in automated interview assessment and AI-driven human interaction systems.

V. Concluding Remarks And Future Enhancements

The proposed system, EmoScan, successfully integrates facial recognition, emotion detection, and demographic profiling to enhance the effectiveness of virtual interviews. By identifying seven core emotions, along with age and gender, it provides deeper insights into the interviewee's state, addressing key challenges in traditional online interviews. The use of TensorFlow.js ensures real-time performance, privacy, and accessibility, making it a robust tool for professional applications.

Future enhancements could focus on integrating multimodal analysis, such as voice and speech recognition, for a more comprehensive evaluation. Improving model robustness to handle diverse lighting, camera quality, and facial features will further enhance system accuracy. Expanding its functionality to provide real-time feedback and integration with popular video conferencing platforms can broaden its usability. Additionally, employing advanced techniques like transfer learning and addressing ethical considerations such as bias mitigation will strengthen its adaptability and fairness.

EmoScan marks a significant step forward, with potential applications across recruitment, education, and telemedicine.

References

[1] Zhou S, Wu X, Jiang F, Huang Q, Huang C. Emotion Recognition from Large-Scale Video Clips with Cross-Attention and Hybrid Feature Weighting Neural Networks. Int J Environ Res Public Health. 2023 Jan 12;20(2):1400. doi: 10.3390/ijerph20021400. PMID: 36674161; PMCID: PMC9859118.

[2] S. Thushara and S. Veni, "A multimodal emotion recognition system from video," 2016 International Conference on Circuit, Power and Computing Technologies (ICCPCT), Nagercoil, India, 2016, pp. 1-5, doi: 10.1109/ICCPCT.2016.7530161.

[3] Paweł Tarnowski, Marcin Kołodziej, Andrzej Majkowski, Remigiusz J. Rak, Emotion recognition using facial expressions, Procedia Computer Science,

Volume 108, 2017, Pages 1175-1184, ISSN 1877-0509, https://doi.org/10.1016/j.procs.2017.05.025.

[4] A. Chakraborty, A. Konar, U. K. Chakraborty and A. Chatterjee, "Emotion Recognition From Facial Expressions and Its Control Using Fuzzy Logic," in IEEE Transactions on Systems, Man, and Cybernetics - Part A: Systems and Humans, vol. 39, no. 4, pp. 726-743, July 2009, doi: 10.1109/TSMCA.2009.2014645.

[5] K. Amara, N. Ramzan, N. Achour, M. Belhocine, C. Larbas and N. Zenati, "Emotion Recognition via Facial Expressions," 2018 IEEE/ACS 15th International Conference on Computer Systems and Applications (AICCSA), Aqaba, Jordan, 2018, pp. 1-6, doi: 10.1109/AICCSA.2018.8612852.

[6] M. A. Ozdemir, B. Elagoz, A. Alaybeyoglu, R. Sadighzadeh and A. Akan, "Real Time Emotion Recognition from Facial Expressions Using CNN Architecture," 2019 Medical Technologies Congress (TIPTEKNO), Izmir, Turkey, 2019, pp. 1-4, doi: 10.1109/TIPTEKNO.2019.8895215.

[7] Joseph, A. and Geetha, P., 2020. Facial emotion detection using modified eyemap–mouthmap algorithm on an enhanced image and classification with tensorflow. *The Visual Computer*, *36*(3), pp.529-539.

[8] Jadhav, B., Sawant, A., Shah, A., Vemula, P., Waikar, A. and Yadav, S., 2024, March. A Comprehensive Study and Implementation of the Mock Interview Simulator with AI and Pose-Based Interaction. In *2024 1st International Conference on Cognitive, Green and Ubiquitous Computing (IC-CGU)* (pp. 01-05). IEEE.

[9] Asha, P., Shabu, J., Refonaa, J. and Selvan, M.P., 2023, August. Automated Interview through Online Video Interface. In *2023 International Conference on Circuit Power and Computing Technologies (ICCPCT)* (pp. 1315-1321). IEEE.

[10] N. Boudjani, V. Colas, C. Joubert and D. B. Amor, "AI Chatbot For Job Interview," 2023 46th MIPRO ICT and Electronics Convention (MIPRO), Opatija, Croatia, 2023, pp. 1155-1160, doi: 10.23919/MIPRO57284.2023.10159831. keywords: {Chatbots;Behavioral sciences;Interviews;Artificial intelligence;NLP;interactive AI chatbot;job interview},

[11] Y. -C. Chou, F. R. Wongso, C. -Y. Chao and H. -Y. Yu, "An AI Mock-interview Platform for Interview Performance Analysis," 2022 10th International Conference on Information and Education Technology (ICIET), Matsue, Japan, 2022, pp. 37-41, doi: 10.1109/ICIET55102.2022.9778999. keywords: {Visualization;Epidemics;Education;Software;Performance analysis;Interviews;Artificial intelligence;artificial intelligence;video interview;mock interview;interview analysis},

[12] Sebe, N., Cohen, I. and Huang, T.S., 2005. Multimodal emotion recognition. In *Handbook of pattern recognition and computer vision* (pp. 387-409).

[13] Dissanayake, D.Y., Amalya, V., Dissanayaka, R., Lakshan, L., Samarasinghe, P., Nadeeshani, M. and Samarasinghe, P., 2021, December. AI-based Behavioural Analyser for Interviews/Viva. In *2021 IEEE 16th International Conference on Industrial and Information Systems (ICIIS)* (pp. 277-282). IEEE.

[14] Laiq, M. and Dieste, O., 2020, August. Chatbot-based interview simulator: a feasible approach to train novice requirements engineers. In *2020 10th International workshop on requirements engineering education and training (REET)* (pp. 1-8). IEEE.

[15] CJ, J.M., Sabi, M., Benson, M., Baburaj, G. and Saritha, S., 2024, April. Q&AI: An AI Powered Mock Interview Bot for Enhancing the Performance of Aspiring Professionals. In *2024 International Conference on Recent Advances in Electrical, Electronics, Ubiquitous Communication, and Computational Intelligence (RAEEUCCI)* (pp. 1-5). IEEE.

[16] Rahman, W., Mahbub, S., Salekin, A., Hasan, M.K. and Hoque, E., 2021, September. Hirepreter: A framework for providing fine-grained interpretation for automated job interview analysis. In *2021 9th International Conference on Affective Computing and Intelligent Interaction Workshops and Demos (ACIIW)* (pp. 1-5). IEEE.

[17] Ford, D., Barik, T., Rand-Pickett, L. and Parnin, C., 2017, May. The tech-talk balance: what technical interviewers expect from technical candidates. In *2017 IEEE/ACM 10th International Workshop on Cooperative and Human Aspects of Software Engineering (CHASE)* (pp. 43-48). IEEE.

[18] Kim, C., Choi, J., Yoon, J., Yoo, D. and Lee, W., 2023. Fairness-aware multimodal learning in automatic video interview assessment. *IEEE Access*.

[19] Salvi, V., Vasanwalla, A., Aute, N. and Joshi, A., 2017, August. Virtual Simulation of Technical Interviews. In *2017 International Conference on Computing, Communication, Control and Automation (ICCUBEA)* (pp. 1-6). IEEE.

[20] Hudson, N., 2016. Communication and power in the job interview: Using a ventriloqual approach to analyze moral accounts. *Text & Talk*, *36*(3), pp.319-340.

[21] Munandar, V.D., Bross, L.A., Zimmerman, K.N. and Morningstar, M.E., 2021. Video-based intervention to

improve storytelling ability in job interviews for college students with autism. *Career Development and Transition for Exceptional Individuals*, *44*(4), pp.203-215.

[22] Chen, L., Zhao, R., Leong, C.W., Lehman, B., Feng, G. and Hoque, M.E., 2017, October. Automated video interview judgment on a large-sized corpus collected online. In *2017 Seventh International Conference on Affective Computing and Intelligent Interaction (ACII)* (pp. 504-509). IEEE.

[23] Su, Y.S., Suen, H.Y. and Hung, K.E., 2021. Predicting behavioral competencies automatically from facial expressions in real-time video-recorded interviews. *Journal of Real-Time Image Processing*, *18*(4), pp.1011-1021.

[24] Ho, W.Y.J., 2019. Developing professional communication: The construction of a multimodal understanding of job interviews. *Languages*, *4*(1), p.5.

[25] Nguyen, L.S., 2015. *Computational analysis of behavior in employment interviews and video resumes* (No. 6567). EPFL.

[26] Beti, R.A., Al-Khatib, F. and Cook, D.M., 2019. The efficacy of using virtual reality for job interviews and its effects on mitigating discrimination. In *Recent Advances in Information and Communication Technology 2018: Proceedings of the 14th International Conference on Computing and Information Technology (IC2IT 2018)* (pp. 43-52). Springer International Publishing.

[27] Dunn, K.T., 2022. Effects of a Video Based Intervention on Job Interview Skills of Youth with Autism.

[28] Toldi, N.L., 2011. Job applicants favor video interviewing in the candidate-selection process. *Employment relations today*, *38*(3), pp.19-27.

[29] Naim, I., Tanveer, M.I., Gildea, D. and Hoque, M.E., 2015, May. Automated prediction and analysis of job interview performance: The role of what you say and how you say it. In *2015 11th IEEE international conference and workshops on automatic face and gesture recognition (FG)* (Vol. 1, pp. 1-6). IEEE.

[30] Naim, I., Tanveer, M.I., Gildea, D. and Hoque, M.E., 2016. Automated analysis and prediction of job interview performance. *IEEE Transactions on Affective Computing*, *9*(2), pp.191-204.

[31] Li, M., Chen, X., Liao, W., Song, Y., Zhang, T., Zhao, D. and Yan, R., 2023, February. Ezinterviewer: To improve job interview performance with mock interview generator. In *Proceedings of the Sixteenth ACM International Conference on Web Search and Data Mining* (pp. 1102-1110).

[32] Akinnaso, F.N. and Ajirotutu, C.S., 1982. Performance and ethnic style in job interviews. *Language and social identity*, pp.119-144.

[33] Agrawal, A., George, R.A. and Ravi, S.S., 2020. Leveraging multimodal behavioral analytics for automated job interview performance assessment and feedback. *arXiv preprint arXiv:2006.07909*.

[34] DeGroot, T. and Motowidlo, S.J., 1999. Why visual and vocal interview cues can affect interviewers' judgments and predict job performance. *Journal of Applied Psychology*, *84*(6), p.986.

b.Plagiarism Report

Real-Time Facial Analysis for Online Interviews

ORIGINALITY REPORT

6%	4%	4%	2%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

PRIMARY SOURCES

1	www.mdpi.com Internet Source	1%
2	rcpsych.ac.uk Internet Source	1%
3	Jain-Shing Wu, Wan-Ching Lin, Chih-Ta Lin, Te-En Wei. "Smartphone continuous authentication based on keystroke and gesture profiling", 2015 International Carnahan Conference on Security Technology (ICCST), 2015 Publication	1%
4	www.nature.com Internet Source	1%
5	M. Karthikeyan, T. S. Subashini. "Automated object detection of mechanical fasteners using faster region based convolutional neural networks", International Journal of Electrical and Computer Engineering (IJECE), 2021 Publication	1%
6	Submitted to University of Surrey Student Paper	1%
7	"Proceedings of 4th International Conference on Artificial Intelligence and Smart Energy", Springer Science and Business Media LLC, 2024 Publication	1%
8	Submitted to Centre for Distance and Online Education Galgotias University	1%

Student Paper

9	nur.nu.edu.kz Internet Source	1 %
10	www.online-pdh.com Internet Source	1 %

Exclude quotes	On	Exclude matches	< 1%
Exclude bibliography	On		

Real-Time Facial Analysis for Online Interviews

GRADEMARK REPORT

FINAL GRADE

/0

GENERAL COMMENTS

PAGE 1

PAGE 2

PAGE 3

PAGE 4

PAGE 5

PAGE 6

PAGE 7

c.Project review sheet ;

Project review sheet 1:

Inhouse/ Industry _Innovation/Research:

Sustainable Goal:

Project Evaluation Sheet 2024 - 25

Class: D17 A/B/C

Group No.: 28

Title of Project: Emoscan : Real Time facial Analysis for Online Interviews.

Group Members: Bhavika Valecha (D17C, 69) B.V., Dipti Hemnani (D17C, 23) Dipti, Kunal Khubchandani (D17C, 37) Kunal, Raj Tandon (D17C, 62) Raj Tandon

Engineering Concepts & Knowledge (5)	Interpretation of Problem & Analysis (5)	Design / Prototype (5)	Interpretation of Data & Dataset (3)	Modern Tool Usage (5)	Societal Benefit, Safety Consideration (2)	Environ ment Friendly (2)	Ethics (2)	Team work (2)	Presentati on Skills (2)	Applied Engg&M gmt principles (3)	Life - long learning (3)	Profess ional Skills (3)	Innov ative Appr oach (3)	Resear-ch Paper (5)	Total Marks (50)
4	4	3	3	4	2	2	2	2	2	3	3	2	2	3	41

Comments:

Name & Signature Reviewer1

Trudi D. ...

Engineering Concepts & Knowledge (5)	Interpretation of Problem & Analysis (5)	Design / Prototype (5)	Interpretation of Data & Dataset (3)	Modern Tool Usage (5)	Societal Benefit, Safety Consideration (2)	Environ ment Friendly (2)	Ethics (2)	Team work (2)	Presentati on Skills (2)	Applied Engg&M gmt principles (3)	Life - long learning (3)	Profess ional Skills (3)	Innov ative Appr oach (3)	Resear -ch Paper (5)	Total Marks (50)
4	4	3	3	4	2	2	2	2	2	3	3	2	2	3	41

Comments: 1. ~~Use~~ descriptive reports can be generated.
2. Fine tuning needs to be done.

Date: 1st March, 2025

Name & Signature Reviewer 2 Pallavi S.

Project review sheet 2

(28)

Project Evaluation Sheet 2024 - 25

Title of Project: Real Time Facial Analysis for Online Interviews (EmoScan).

Group Members: Bharika Valecha (D17C, 69) B.V., Dipti Hemnani (D17C, 23) Dipti, Kunal Khubchandani (D17C, 37) Kunal, Raj Tandon (D17C, 62) Raj Tandon

Engineering Concepts & Knowledge (5)	Interpretation of Problem & Analysis (5)	Design / Prototype (5)	Interpretation of Data & Dataset (3)	Modern Tool Usage (5)	Societal Benefit, Safety Consideration (2)	Environ ment Friendly (2)	Ethics (2)	Team work (2)	Presentati on Skills (2)	Applied Engg&M gmt principles (3)	Life - long learning (3)	Profess ional Skills (3)	Innov ative Appr oach (3)	Resear ch Paper (5)	Total Marks (50)
4	4	4	3	4	2	2	2	2	2	3	3	2	2	3	42

Comments: Integrate the facial expression module with the AI generated mock interview module.

Name & Signature Reviewer1

Inhouse/ Industry _Innovation/Research:

Engineering Concepts & Knowledge (5)	Interpretation of Problem & Analysis (5)	Design / Prototype (5)	Interpretation of Data & Dataset (3)	Modern Tool Usage (5)	Societal Benefit, Safety Consideration (2)	Environ ment Friendly (2)	Ethics (2)	Team work (2)	Presentati on Skills (2)	Applied Engg&M gmt principles (3)	Life - long learning (3)	Profess ional Skills (3)	Innov ative Appr oach (3)	Resear ch Paper (5)	Total Marks (50)
4	4	4	3	4	2	2	2	2	2	3	3	2	2	3	42

Comments: Integrate facial expression module with AI generated mock interview

Date: 1st April,2025

Name & Signature Reviewer 2 Pallavi S.