



KIET
GROUP OF INSTITUTIONS
Connecting Life with Learning



A
Project Report
on
AI Interviewer
submitted as partial fulfillment for the award of
BACHELOR OF TECHNOLOGY
DEGREE

SESSION 2023-34
in
Computer Science and Engineering

By
Kushagra Srivastava (2000290100085)

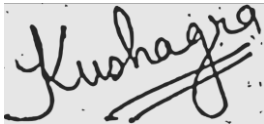
Under the supervision of
Sanjiv Sharma
KIET Group of Institutions, Ghaziabad

Affiliated to
Dr. A.P.J. Abdul Kalam Technical University, Lucknow
(Formerly UPTU)
June 2024

DECLARATION

We hereby declare that this submission is our own work and that, to the best of our knowledge and belief, it contains no material previously published or written by another person nor material which to a substantial extent has been accepted for the award of any other degree or diploma of the university or other institute of higher learning, except where due acknowledgment has been made in the text.

Signature:

A handwritten signature in black ink, appearing to read 'Kushagra', with a stylized flourish underneath.

Name: Kushagra Srivastava

Roll No.: 2000290100085

Date: 03-06-2024

CERTIFICATE

This is to certify that Project Report entitled AI Interviewer which is submitted by Kushagra Srivastava in partial fulfillment of the requirement for the award of degree B. Tech. in Department of Computer Science & Engineering of Dr. A.P.J. Abdul Kalam Technical University, Lucknow is a record of the candidates own work carried out by them under my supervision. The matter embodied in this report is original and has not been submitted for the award of any other degree.

.

Sanjiv Sharma

Addn Head of Department

Dr. Vineet Sharma

(HoD-Computer Science & Engineering)

Date: 3rd June 2024

ACKNOWLEDGEMENT

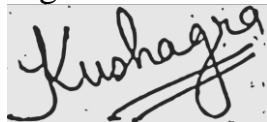
It gives us a great sense of pleasure to present the report on the B. Tech Project undertaken during B. Tech. Final Year. We owe a special debt of gratitude to Mr. Sanjiv Sharma, Department of Computer Science & Engineering, KIET, Ghaziabad, for his constant support and guidance throughout the course of our work. His sincerity, thoroughness and perseverance have been a constant source of inspiration for us. It is only his cognizant efforts that our endeavors have seen the light of the day.

We also take the opportunity to acknowledge the contribution of Dr. Vineet Sharma, Head of the Department of Computer Science & Engineering, KIET, Ghaziabad, for his full support and assistance during the development of the project. We also do not want to miss the opportunity to acknowledge the contribution of all the faculty members of the department for their kind assistance and cooperation during the development of our project.

We also do not like to miss the opportunity to acknowledge the contribution of all faculty members, especially faculty/industry person/any person, of the department for their kind assistance and cooperation during the development of our project. Finally, we acknowledge our friends for their contribution to the completion of the project.

Date: 3rd June 2024

Signature:

A handwritten signature in black ink, appearing to read 'Kushagra', with a stylized flourish underneath.

Name : Kushagra Srivastava

Roll No.: 2000290100085

ABSTRACT

Skill Development and Training is one of the industries which has been on the rise in the last few years. Training of individuals to gather hard skills and soft skills has been a major part of skill development, which targets at refining the qualities of an individual to be more industry relevant and get a strong base for a good career into their respective industries. Interview preparation and skill training has been a major part of the rise of EdTech industry in recent years. This is reflected in budding product-based startups which highly focus on delivering online solutions for candidates, which can be utilized to give resources and roadmaps tailored according to the candidate's profile targeting the industrial requirements.

Though there has been major focus on the industrial training, this training is highly concentrated on the hard skills requirements for the jobs and not the soft skills which are required for qualifying the recruitment processes. This leads to improper delivery of the candidate profile before the HR Manager leading to disqualification. This pays companies with blank positions due to less skilled candidates and candidates with no or less paying offers than they deserve with the level of their hard skills. This gap between industrial relevance and candidate preparation can be bridged by promoting soft skills training and mock interview preparations. 'AI INTERVIEWER' is one such project which aims at building a virtual interviewer which can conduct interview of the candidate and provide personalized feedback to the candidates for improvement.

TABLE OF CONTENTS	Page No.
DECLARATION.....	ii
CERTIFICATE.....	iii
ACKNOWLEDGEMENTS.....	iv
ABSTRACT.....	v
LIST OF FIGURES.....	viii
LIST OF TABLES.....	ix
LIST OF ABBREVIATIONS.....	x
 CHAPTER 1 (INTRODUCTION).....	 1
1.1. Introduction.....	1
1.2. Project Description.....	3
 CHAPTER 2 (LITERATURE REVIEW).....	 5
2.1. Literature Review from papers.....	5
2.2. Online Communities and References.....	6
 CHAPTER 3 (PROPOSED METHODOLOGY)	 7
3.1. Proposed Functionality and Technologies	7
3.2. Technological Background and Implementation.....	10
3.2.1. Realistic Online Interview Simulation with Camera Support.....	10
3.2.2. Enhanced Self-Awareness through Facial Expression Recognition.....	11
3.2.3. Sharper Focus Detection with Eye Tracking.....	16
3.2.4. Speech Delivery Analysis for Confident Communication.....	17

3.2.5. Personalized Feedback and Performance Tracking.....	18
CHAPTER 4 (RESULTS AND DISCUSSION)	19
CHAPTER 5 (CONCLUSIONS AND FUTURE SCOPE).....	21
5.1. Conclusion.....	21
5.2. Future Scope.....	21
REFERENCES.....	23
APPENDIX.....	25

LIST OF FIGURES

Figure No.	Description	Page No.
1.1	Sample Dataset for FER 2013	12
1.2	Plot for Accuracy and Loss for Custom Models (4)	13-14
1.3	Plot for Accuracy and Loss for Transfer Learning Models	15

LIST OF TABLES

Table. No.	Description	Page No.
3.2.2.1	Table with custom model architectures & performances	13
3.2.2.2	Table with performances of model made from transfer learning.	14

LIST OF ABBREVIATIONS

HR	Human Resources
ML	Machine Learning
DL	Deep Learning
AI	Artificial Intelligence
ANN	Artificial Neural Networks
CNN	Convolutional Neural Networks
VGG	Visual Geometric Group
FER	Facial Emotion Recognition
STT	Speech To Text
WPM	Words Per Minute

CHAPTER 1

INTRODUCTION

1.1 INTRODUCTION

The recruitment landscape, once static and rigid, is witnessing a transformative wave. While human interaction has long been the anchor of talent assessment, its limitations – subjectivity, unconscious bias, and resource constraints – are becoming increasingly apparent. Enter AI Interviewer, a revolution in its making, poised to redefine both interview preparation and the way we hire.

This innovative platform leverages the power of deep learning and computer vision to create hyper-realistic virtual interview environments. Imagine a candidate seamlessly interacting with a programmed interviewer, receiving real-time feedback on their responses, and undergoing an analysis of not just what they say, but also how they say it. Deep learning models sift through their words, extracting key themes and gauging sentiment, while computer vision deciphers the subtle language of facial expressions and body language. This comprehensive assessment paints a holistic picture of the candidate's communication skills, confidence, and overall demeanor, offering recruiters invaluable insights beyond the resume.

The benefits, however, transcend mere interview simulation. For candidates, AI Interviewer becomes a personalized training ground. They can practice under pressure, receive unbiased feedback on their strengths and weaknesses, and hone their communication skills in a stress-free environment. No longer constrained by geographical barriers or scheduling conflicts, they can access this platform at their own pace, gaining the confidence and polish to excel in real interviews.

For organizations, AI Interviewer translates into efficiency and transparency. By pre-screening candidates and identifying the most promising individuals, it streamlines the hiring process,

eliminates time-consuming initial interviews, and allows recruiters to focus their resources on in-depth assessments with the most qualified candidates. This not only saves time and cost but also fosters a fairer, more objective talent acquisition framework. Unconscious biases are minimized, diversity is enhanced, and ultimately, the right individuals are selected for the right roles.

But the impact of AI Interviewer goes beyond talent acquisition. It holds the potential to revolutionize skill development by offering personalized training simulations. Imagine a platform that pinpoints individual skill gaps, prescribes targeted exercises, and tracks progress in real-time. This fosters a continuous learning environment where individuals can upskill and adapt to evolving industry demands, contributing to a more agile and skilled workforce.

The future of skill development and talent acquisition lies at the intersection of human insight and AI's analytical prowess. AI Interviewer provides a platform for this powerful synergy, where technology amplifies our understanding of candidates' skills and potential. This empowers both individuals and organizations, propelling a more efficient, equitable, and skill-centric talent ecosystem.

1.2 PROJECT DESCRIPTION

The project aims at delivering a web application which can provide candidates (users) with two main functionalities: Rehearsing online interviews with camera monitoring and speech-based feedback. These two functionalities are provided on a web application to reduce the time required to access the services and make the user experience better. Web Application has been developed with streamlit v1.35.0. to ease the integration of machine learning and deep learning services in the platform.

The first functionality is implemented by using streamlit's components for video recording which receives inputs from camera (integrated camera or additional cam) and provides video and audio feed to be used as inputs for processing functionalities. This video feed, which comes up as images are processed to find the visual features like facial emotion and eye focus of the user. Facial Emotion Recognition is implemented by using deep learning model based on FER2013 dataset, which was further tuned with interview recording of college students to make it ready for use for the sample audience in the development audience. The other service utilized to provide this functionality is based on eye tracking techniques. Eye tracking techniques cover a wide range of algorithms ranging from tracking of eye to tracking iris only to find the angle at which the user is seeing before the screen. Eye tracking for this use case doesn't necessarily involve very precise angles but simple check over the candidate if he/she is focusing on the screen or involving in unfair means/showing under confident reflexes before the interviewer. These two main models are currently used in the project, which also opens scope for additional features but being restricted due to near real time processing and feedback requirements. This feedback is stored and marked for each interview for each user.

The second functionality focuses on the speech delivered. Candidate's speech delivery affects the overall interview experience due to factors such as speed of speaking, stammering or the repetitive pronunciation of same lines. This module takes in audio input and converts it into text for further processing. Pretrained model's API is used for this purpose to understand a wide range of pronunciation and speech qualities. The count check functionality checks the words spoken per minute and provides feedback on the number of words spoken and the

preferred word count. This module has other features which check the occurrences of stammered words in between sentences. This is reflected to the user to avoid pronouncing the word or sound multiple times in an interview.

Thus, this project will be able to serve the purpose of skill training and interview preparation.

CHAPTER 2

LITERATURE REVIEW

2.1 LITERATURE REVIEW FROM PAPERS

The paper "Gradient-based learning applied to document recognition" by LeCun et al. (1998)[1] stands as a landmark in the history of deep learning, specifically for convolutional neural networks (CNNs). This work introduced a novel architecture inspired by the visual processing hierarchy of the brain, utilizing stacked convolutional layers to extract spatial features from image data. The authors demonstrated the effectiveness of CNNs in handwritten digit recognition, achieving superior performance compared to traditional methods.

Yu et al. (2011) were among the first to explore deep learning for FER, proposing a Convolutional Neural Network (CNN) architecture that significantly outperformed traditional methods based on handcrafted features like Ekman and Friesen's (1978) Facial Action Coding System (FACS). This work paved the way for further research into CNN-based architectures for FER, demonstrating the potential of deep learning to capture the subtle nuances of facial expressions.

Building upon this foundation, AlexNet, introduced by Krizhevsky et al. (2012)[2], marked a significant breakthrough in the field. This deep CNN architecture, with its multiple convolutional and pooling layers, achieved remarkable results on the ImageNet Large Scale Visual Recognition Challenge (ILSVRC), significantly surpassing previous state-of-the-art methods. This success sparked widespread interest in CNNs and deep learning, paving the way for further advancements in computer vision tasks.

Following AlexNet's success, VGGNet, proposed by Simonyan & Zisserman (2014)[3], pushed the boundaries of performance even further. This deeper network, consisting of numerous convolutional layers with small filter sizes, demonstrated superior accuracy on the ILSVRC, showcasing the potential of even deeper architectures for image classification.

Huang et al. (2017)[11] proposed DenseNet, a novel architecture characterized by dense connectivity between layers. DenseNet connects each layer to all preceding layers, promoting efficient feature reuse and alleviating the vanishing gradient problem. This dense connectivity led to significant performance improvements in ImageNet classification compared to state-of-the-art models like VGG and ResNet, while simultaneously requiring fewer parameters.

Ding and Tao (2019) leveraged VGGFace (Parkhi et al., 2015)[6], a pre-trained model for facial recognition, as a feature extractor for FER, achieving state-of-the-art performance. This work highlights the effectiveness of transferring knowledge from related domains to improve FER accuracy.

2.2 ONLINE COMMUNITIES AND REFERENCES

There are various websites like ‘paperswithcode.com’ which provide state-of-the-art papers and the models proposed in them, which have shown the best accuracy over the datasets. Each task has been marked separately with papers, their code, and models. Thus, the custom models can be separately compared to the latest research and models presented for a specific dataset or task.

CHAPTER 3

PROPOSED METHODOLOGY

3.1. PROPOSED FUNCTIONALITY AND TECHNOLOGIES

This project aims to create a web application that empowers candidates to excel in online interviews. Here's a detailed breakdown of the goals and the technologies that make them possible:

Goal 1: Realistic Online Interview Simulation with Camera Support

Functionality: Users can record mock interviews using their webcam (integrated or external). This allows them to practice in a setting that closely mimics a real online interview.

Technology:

Webcam Access: Streamlit provides built-in functionalities to access a user's webcam device. With user permission, the application can capture video and audio data during the interview recording. WebRTC is used for this specific purpose in the project.

Goal 2: Enhanced Self-Awareness through Facial Expression Recognition

Functionality: Analyze facial expressions during the interview recording using a pre-trained Deep Learning model. This provides users with feedback on their emotional state during the interview.

Technology:

Deep Learning (DL): A subfield of Machine Learning (ML) that uses artificial neural networks (CNNs – Convolutional Neural Networks) to learn complex patterns (facial features) from data. In this case, a pre-trained DL model trained on the FER2013 dataset (a collection of facial expressions) will be used.

FER2013 Dataset: A publicly available dataset containing labeled images of various facial expressions (happy, sad, angry).

Model Tuning: To improve the model's accuracy for interview scenarios, the pre-trained model (EfficientNet) will be further fine-tuned with a dataset of actual interview recordings from college students. This helps the model recognize interview-specific expressions better.

Goal 3: Sharper Focus Detection with Eye Tracking

Functionality: Implement basic eye tracking to assess if the user is maintaining eye contact with the "virtual interviewer" (the screen). This helps identify potential distractions or nervousness.

Technology:

Eye Tracking: Techniques used to analyze eye movements. Here, a basic version will be implemented, focusing on detecting if the user's gaze is directed towards the screen. High-precision eye tracking is not crucial for this application.

Goal 4: Speech Delivery Analysis for Confident Communication

Functionality: Analyze the user's speech during the interview recording. Provide feedback on factors such as speaking rate (words per minute) and stammering occurrences. This helps users work towards clear and confident communication.

Technology:

Speech-to-Text (STT) API: A pre-trained API (Application Programming Interface) that converts spoken audio into text format. This allows the application to analyze the user's speech content.

Words Per Minute (WPM) Count: A metric that measures the rate at which a person speaks. The application will calculate the WPM of the user's speech and provide feedback on how it compares to a desired range.

Stammer Detection: The application will analyze the speech audio for instances of stammering (repeated sounds or hesitations). Users will receive feedback on the frequency of stammering to help them improve fluency.

Goal 5: Personalized Feedback and Performance Tracking

Functionality: After each interview recording, the application will provide users with a comprehensive feedback report. This report will detail facial expressions, eye focus, speech analysis results (WPM & Stammering), and any additional features implemented in the future. Users can access past interview recordings and feedback for reference and progress tracking.

3.2. TECHNOLOGICAL BACKGROUND AND IMPLEMENTATION

3.2.1 Goal 1: Realistic Online Interview Simulation with Camera Support

Streamlit excels at crafting user interfaces, even though it doesn't handle complex video processing itself. In this application, Streamlit serves as the user's primary touchpoint for controlling the interview recording process.

1. **User Interaction:** Streamlit provides intuitive elements like buttons for users to initiate and stop interview recordings, mimicking the flow of a real interview.
2. **Live Preview Window:** The user interface can be enhanced with a live video preview window. By leveraging external libraries like OpenCV, users can see their webcam feed and adjust their positioning before recording.

WebRTC: WebRTC establishes a secure, real-time connection between the user's web browser and the application server. This technology facilitates the capture and transmission of video data during interview simulations.

1. **Live Video Capture:** WebRTC captures the user's video feed directly from the webcam in real-time, mimicking the experience of a live online interview.
2. **Potential Performance Benefits:** WebRTC's peer-to-peer nature can potentially reduce latency compared to traditional server-based video calls, leading to a more realistic and responsive simulation experience.

The backend acts as the processing hub, receiving the real-time video stream from WebRTC. Depending on the architecture, it might utilize additional libraries to manage and potentially manipulate the video data. This powerful backend serves as the future launchpad for machine learning models. Once integrated, these models can analyze the captured video, providing valuable feedback on the user's interview performance, including aspects like facial expressions and eye focus.

3.2.2. Goal 2: Enhanced Self-Awareness through Facial Expression Recognition

EfficientNet-based Facial Expression Recognition: The application utilizes a pre-trained facial expression recognition model built on the EfficientNet architecture. This model is specifically chosen for its efficiency in processing video frames captured through WebRTC.

Key functionalities:

1. **Model Selection:** The EfficientNet architecture offers a good balance between accuracy and computational efficiency, making it suitable for real-time processing within the application.
2. **Real-time Frame Processing:** Within the Streamlit WebRTC callback function (triggered by incoming video frames), individual frames are extracted from the video stream.
3. **Emotion Recognition:** Each extracted frame is fed into the EfficientNet model, which generates a score for each of the seven pre-defined emotions (e.g., happy, sad, angry).

Real-time Emotion Feedback: Based on the emotion scores provided by the model, the application offers real-time feedback to the user:

1. **Dominant Emotion:** The application identifies the emotion with the highest score as the user's dominant emotion during that specific frame.
2. **Emotional Balance:** The application can calculate an "emotional balance" metric by analyzing the distribution of scores across all emotions. This metric helps the user understand if they are exhibiting a healthy range of emotions or leaning too heavily towards a particular emotion (e.g., excessive nervousness).

Dataset: FER-2013 Facial Expressions Dataset

The data consists of 48x48 pixel grayscale images of faces. The faces have been automatically registered so that the face is centered and occupies about the same amount of space in each image. The task is to categorize each face based on the emotion shown in the facial expression

in to one of seven categories (0=Angry, 1=Disgust, 2=Fear, 3=Happy, 4=Sad, 5=Surprise, 6=Neutral). Data distribution is as follows:

1. Train Data: 28,709 Images
2. Test Data: 3589 Images

Sample Dataset: (Random Images from the dataset)



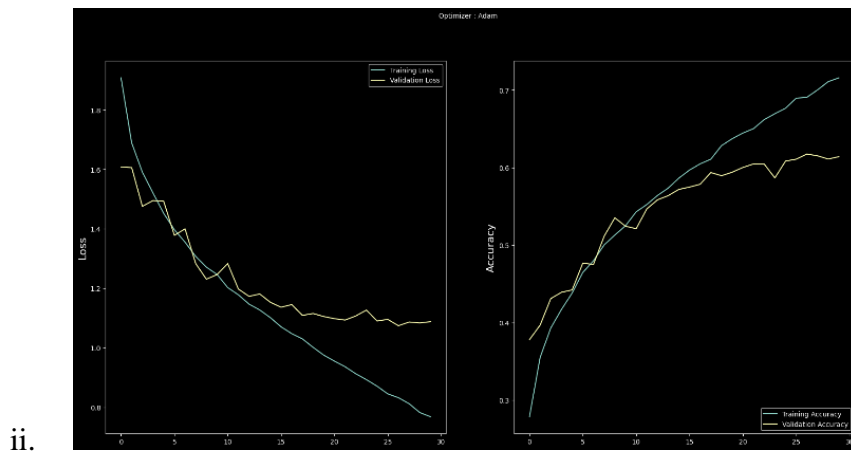
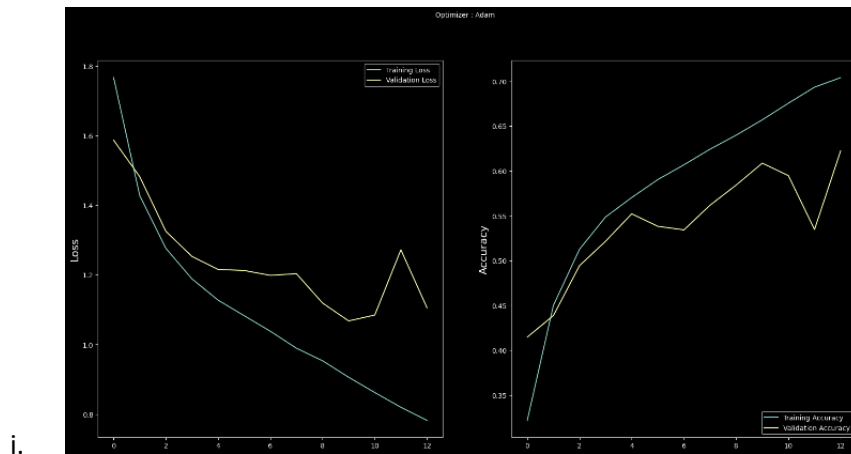
Model Training (Refer to Appendix for full research paper):

The training process for the facial emotion recognition (FER) CNN model leverages several techniques to optimize performance and prevent overfitting. Early stopping halts training when validation loss stagnates, while model checkpointing preserves the best performing model based on validation accuracy. Adam optimization with learning rate adjustments is employed for efficient training. Categorical cross-entropy loss measures model performance, while accuracy serves as the primary evaluation metric.

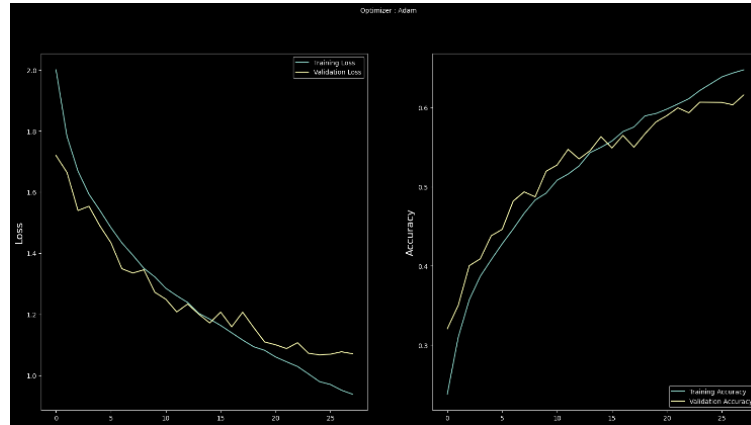
Custom Models	Training Accuracy (%)	Validation Accuracy (%)	Number & Type of layers
Model 1	70.40	62.23	4 Conv (64, 128, 512, 512), 2 FC (256, 512)
Model 2	71.57	61.41	3 Conv (64,128,256), 2 FC (256, 512)
Model 3	64.75	61.59	4 Conv (64, 128, 256, 512), 2 FC (256, 512)
Model 4	66.50	60.34	5 Conv (64, 128, 256, 512, 1024, 512) 2 FC (256, 512)

Table 3.2.2.1. Table with custom model architectures and performances

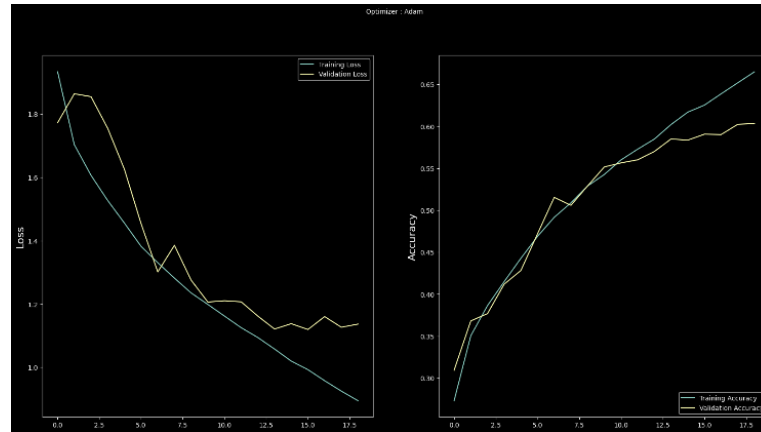
Plots for respective models:



iii.



iv.

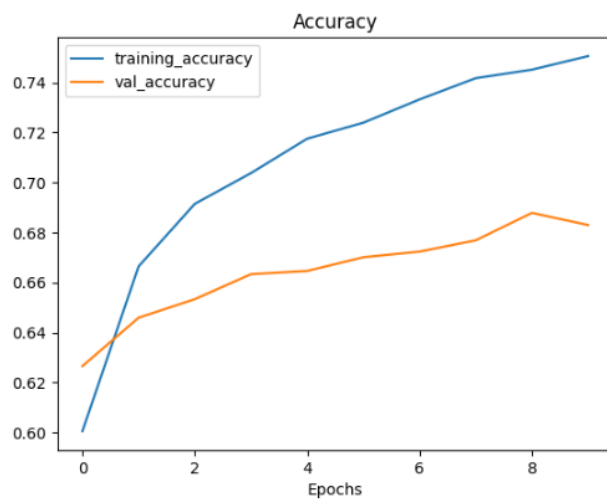
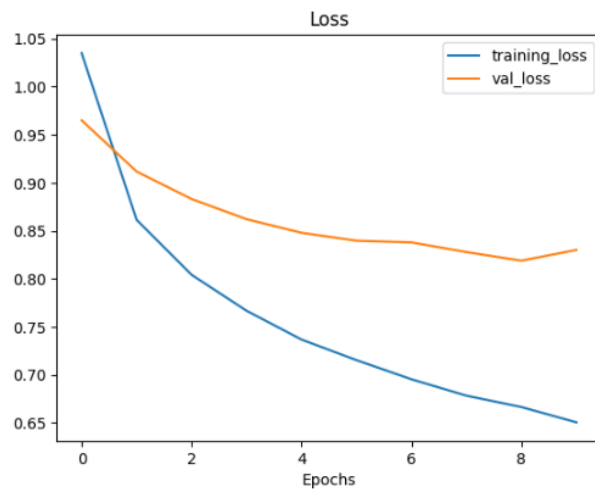


Transfer Learning Benefits: An EfficientNet model, leveraging pre-trained weights, achieves significantly higher validation accuracy (68.30%) compared to the best custom model (62.23%). This highlights the advantage of transfer learning in FER tasks. Pre-trained models capture general visual features, reducing complexity and accelerating development of high-performing FER systems.

Training Accuracy	0.6506
Training Loss	0.7506
Validation Accuracy	0.8301
Validation Loss	0.6830

Table 3.2.2.2. Table with performances of model made from transfer learning.

Plots (Accuracy and Loss):



3.2.3. Goal 3: Sharper Focus Detection with Eye Tracking

Custom Deep Learning Model for Eye Tracking: The application utilizes a deep learning model specifically trained to categorize eye focus during an interview scenario. This model development process involves:

1. Focus Category Classification: A dataset is created by the developer, consisting of images or video frames manually categorized into three distinct classes:
 - a. Center of Screen: Represents images where the user's eyes are fixated on the central area of the screen, mimicking eye contact with a virtual interviewer.
 - b. Border of Screen: Represents images where the user's eyes are directed towards the edges of the screen, potentially indicating divided attention.
 - c. Out of Screen: Represents images where the user's eyes are looking away from the screen entirely, suggesting a lack of focus on the interview.
2. Simpler Deep Learning Architecture: Due to the less complex nature of this task compared to facial expression recognition, a simpler deep learning architecture can be employed. This reduces processing time and computational resources needed for real-time performance within the application. Convolutional Neural Networks (CNNs) are well-suited for this purpose due to their effectiveness in image recognition tasks.

3.2.4. Goal 4: Speech Delivery Analysis for Confident Communication

By integrating the STT API with WebRTC, the application can analyze the user's speech delivery in real-time, offering immediate feedback during the mock interview.

1. **API Selection:** Choose a robust and accurate STT API that can handle a variety of accents and speech patterns. Consider factors like pricing, supported languages, and real-time processing capabilities.
2. **Stream Integration:** Integrate the STT API calls within the Streamlit WebRTC callback function responsible for processing incoming audio frames. This ensures real-time conversion of speech to text as the user speaks during the interview recording.

Custom Functions for Speech Analysis

The application leverages custom Python functions built on top of the STT API output to analyze the user's speech delivery. Here's a breakdown of these functionalities:

1. **Words Per Minute (WPM) Calculation:** The STT API provides the transcribed text from the user's speech. A custom function can be developed to calculate the WPM by dividing the total number of words in the transcribed text by the elapsed interview recording time (in minutes). This WPM value can then be compared to a desired range for effective communication and feedback provided to the user.
2. **Stammer Detection:** Another custom function can be implemented to identify stammering occurrences within the transcribed text. This function might analyze for repetitive words, phrases, or prolonged sounds indicative of stammering. The frequency of stammering can be calculated and presented as feedback to the user, allowing them to work towards smoother speech delivery.

3.2.5. Goal 5: Personalized Feedback and Performance Tracking

While not directly involved in complex data processing, Streamlit plays a vital role in this goal by creating the user interface for feedback reports and past recording access. Users can interact with the application to view their performance history and detailed feedback reports. Additionally, Streamlit facilitates data storage by managing the association between feedback data and corresponding user recordings.

Feedback Report Generation:

The backend server processes data from various sources to generate a comprehensive feedback report for each interview recording. This data includes:

1. Facial Expression Recognition: Results from the deep learning model, identifying the user's facial expressions throughout the recording.
2. Eye Tracking Analysis: Data from the eye tracking module, indicating whether the user-maintained focus on the screen during the interview.
3. Speech Analysis Results: Speech analysis modules provide insights on Words Per Minute (WPM) and instances of stammering detected in the audio recording.

Feedback Report Presentation: The generated feedback report is presented to the user through the Streamlit interface. This report might include:

1. Visualizations: Charts or graphs representing the user's facial expressions over time, potentially highlighting patterns or areas for improvement.
2. Eye Focus Metrics: A summary of eye focus data, indicating the percentage of time spent looking at the screen compared to looking away.
3. Speech Delivery Statistics: The report might showcase the user's average WPM and the number of occurrences of stammering identified during the interview recording.

CHAPTER 4

RESULTS AND DISCUSSION

This project developed a web application to enhance interview preparation for candidates. Streamlit (v1.35.0) facilitates a user-friendly interface and integrates machine learning models. The application offers a realistic online interview experience through webcam recording. It analyzes facial expressions using a deep learning model, further fine-tuned with college student recordings, to provide feedback throughout the interview. Eye tracking techniques assess user focus during the recording. Speech analysis calculates WPM and identifies occurrences of stammering to help users refine fluency. Finally, users receive comprehensive feedback reports with visualizations and actionable pointers for improvement. Additionally, they can access past recordings and reports to track progress. This combination of functionalities empowers users to practice and receive feedback in a realistic online interview setting, preparing them for success.

Key Functionalities of the resultant application:

1. **Realistic Online Interview Simulation:** This functionality replicates a real online interview experience. Users can record themselves using their webcam, allowing them to practice their interview skills in a setting that closely resembles an actual interview. This immersive environment helps users become more comfortable and confident before their real interview.
2. **Facial Expression Recognition:** This feature leverages a deep learning model trained on a vast dataset of facial expressions. The model analyzes the user's facial expressions throughout the interview recording and provides feedback. This feedback helps users identify and manage their emotional state during practice sessions, allowing them to project a more professional and confident demeanor in real interviews.
3. **Eye Focus Detection:** Eye tracking technology plays a crucial role in assessing a user's focus during the interview recording. This functionality analyzes where the user is looking on the screen. Maintaining eye contact with the simulated interviewer

(represented by the screen) is an indicator of attentiveness and engagement. This feedback helps users develop the necessary focus and avoid behaviors that might suggest nervousness or a lack of interest.

4. **Speech Delivery Analysis:** The application analyzes the audio component of the interview recording to provide valuable feedback on speech delivery. It calculates the user's Words Per Minute (WPM) to assess speaking pace and clarity. Additionally, it detects occurrences of stammering or repetitive speech patterns. By addressing these aspects, users can refine their speech fluency and project confidence during interviews.
5. **Personalized Feedback & Performance Tracking:** Following each interview recording, the application generates a comprehensive feedback report that incorporates insights from facial expression recognition, eye tracking, and speech analysis. Users can access all their past recordings and corresponding feedback reports. This allows them to track progress over time, identify areas for improvement, and tailor their preparation strategies for future interviews.

CHAPTER 5

CONCLUSION AND FUTURE SCOPE

5.1. Conclusion

This project successfully addressed the critical need for soft skills development alongside technical expertise in today's job market. The "AI Interviewer" web application empowers candidates by providing a realistic online interview simulation experience. Through webcam recording and machine learning analysis, users receive personalized feedback on their facial expressions, eye focus, and speech delivery. These features translate into practical guidance on managing emotions, projecting confidence, and refining communication skills. By offering access to past recordings and progress tracking, "AI Interviewer" allows users to tailor their preparation strategies and confidently showcase their full range of abilities during real interviews. This comprehensive solution positions "AI Interviewer" as a valuable tool within the EdTech industry, fostering success for both individual users and the evolving landscape of skill development.

5.2 Future Scope

Future Scope of this project can explore following functionalities, each of which can serve the purpose of the application:

1. **Advanced Emotion Recognition:** Incorporating more sophisticated deep learning models could provide a deeper understanding of a user's emotional state during the interview. This could include recognizing specific emotions like nervousness, excitement, or boredom, allowing for more targeted feedback.
2. **Body Language Analysis:** Integrating body language analysis could provide valuable insights into a user's nonverbal communication. This could involve analyzing posture, gestures, and hand movements to identify areas for improvement.
3. **Mock Interview Functionality:** The application could be expanded to include a mock interview feature where users can practice with a virtual interviewer. This could

involve pre-programmed responses or utilize AI techniques to create a more dynamic and interactive experience.

4. Interviewer Feedback Integration: The ability to incorporate feedback from real interviewers could further enhance the application's effectiveness. Users could upload past interview feedback or connect with human coaches for personalized insights.
5. Mobile Application Development: Developing a mobile application would allow users to access the platform and practice interview skills on the go, offering greater flexibility and convenience.

REFERENCES

- [1] Y. Lecun, L. Bottou, Y. Bengio, and P. Haffner, "Gradient-based learning applied to document recognition," in Proceedings of the IEEE, vol. 86, no. 11, pp. 2278-2324, Nov. 1998, doi: 10.1109/5.726791.
- [2] Krizhevsky, Alex & Sutskever, Ilya & Hinton, Geoffrey. (2012). ImageNet Classification with Deep Convolutional Neural Networks. Neural Information Processing Systems. 25. 10.1145/3065386..
- [3] Simonyan, Karen and Andrew Zisserman. "Very Deep Convolutional Networks for Large-Scale Image Recognition." CoRR abs/1409.1556 (2014).
- [4] Bojarski, Mariusz et al. "End to End Learning for Self-Driving Cars." ArXiv abs/1604.07316 (2016): n. pag.
- [5] Ekman, P., & Friesen, W. V. (1978). Facial Action Coding System (FACS) [Database record]. APA PsycTests.
- [6] Parkhi, Omkar & Vedaldi, Andrea & Zisserman, Andrew. (2015). Deep Face Recognition. 1. 41.1-41.12. 10.5244/C.29.41.
- [7] Yosinski, Jason & Clune, Jeff & Bengio, Y. & Lipson, Hod. (2014). How transferable are featured in deep neural networks? 3320-3328.
- [8] A. S. Razavian, H. Azizpour, J. Sullivan and S. Carlsson, "CNN Features Off-the-Shelf: An Astounding Baseline for Recognition," 2014 IEEE Conference on Computer Vision and Pattern Recognition Workshops, Columbus, OH, USA, 2014, pp. 512-519, doi: 10.1109/CVPRW.2014.131.
- [9] <https://paperswithcode.com/task/facial-expression-recognition>
- [10] Ekman, P., & Friesen, W. V. (1978). Facial Action Coding System (FACS) [Database record]. APA PsycTests.

[11] Huang, Gao & Liu, Zhuang & van der Maaten, Laurens & Weinberger, Kilian. (2017).
Densely Connected Convolutional Networks. 10.1109/CVPR.2017.243.

APPENDIX

Comparative Analysis of Deep Learning Models for Facial Emotional Recognition

Kushagra Srivastava
Computer Science & Engineering
KIET Group of Institutions
Ghaziabad
kushagrathisside@gmail.com

Sanjeev Sharma
Computer Science & Engineering
KIET Group of Institutions
Ghaziabad
martin.mmmec@gmail.com

Abstract—Emotional Intelligence and Facial Emotion Recognition are some of the fields that come under high focus when computer vision is finding its way to develop one of the smartest AI solutions. Implementation of custom models often requires elaborate and well-defined processes at all stages of model development i.e. dataset preparation, model training, tuning, and optimization to run them effectively in applications. Thus, the use of pre-trained models comes into play, which can significantly reduce the training/tuning cost and time requirements. Thus, analysis of publicly available pre-trained models contributes significantly to their use for building efficient and faster AI solutions.

Keywords— *Sentiment Analysis, Deep Learning, Pretrained Models, Transfer Learning*

I. INTRODUCTION

Machine Learning has found its way into many applications starting from basic regression-based problems like price prediction or classification problems which include multiple classes as per the requirement of specific use cases. Deep Learning comes into discussion when machine learning steps into the domain of complex problems that require multiple variables that can't be exactly defined in a specific order. Neural Networks within Machine Learning are studied under this subset referred to as Deep Learning. Neural Networks are classified into many types based on the type of architecture and functions. Each type itself has different applications, most of which require some additional techniques for better results. Hence, a comparative study of the existing pre-trained models makes it less time-consuming to build application-ready models. For a better understanding of the above-mentioned terms in the context of the use case of sentiment analysis, some terminologies are provided below.

A. Machine Learning and Deep Learning

Machine learning (ML) is a subfield of artificial intelligence (AI) that enables computers to learn without being explicitly programmed. It involves algorithms that learn from data and improve their performance over time. Deep learning (DL) is a subset of ML that uses artificial neural networks with multiple hidden layers to learn complex patterns from data. DL has achieved significant success in various applications, including computer vision, natural language processing, and speech recognition.

B. Computer Vision and CNNs

Within the expansive domain of deep learning, computer vision (CV) emerges as a subfield enabling computers to perceive and interpret the visual world akin to humans. It tackles intricate tasks like object detection and image classification, drawing heavily on techniques like convolutional neural networks (CNNs). These specialized

architectures excel at processing grid-like data like images, extracting valuable spatial features through their unique layered structure. Their ability to learn hierarchical representations of visual information has revolutionized CV, paving the way for groundbreaking applications like self-driving cars and medical image analysis.

C. Facial Emotion Recognition (FER)

Further exploration leads us to sentiment analysis (SA), a branch of natural language processing (NLP) dedicated to automatically deciphering the emotional undertones woven into text. By classifying opinions, feelings, and attitudes as positive, negative, or neutral, SA offers invaluable insights across various domains, from social media analysis to market research.

Diving deeper still, we encounter facial emotion recognition (FER), a subfield of CV focused on deciphering human emotions from facial expressions. By analyzing features like eyebrows, eyes, and mouth position, FER aims to infer emotional states like happiness or sadness. Potential applications span human-computer interaction and psychological research, offering a deeper understanding of human emotional expression through facial cues.

D. Transfer Learning

Unlocking the full potential of deep learning hinges on the powerful technique of transfer learning (TL). This approach leverages knowledge gained from a pre-trained model on a new, related task. Imagine pre-trained models like VGG or ResNet, initially trained on massive datasets for image classification, being fine-tuned for CV tasks like FER. These models have already learned general-purpose features like edge detection or object recognition, acting as a strong foundation for the target task. Fine-tuning the final layers on the specific facial expression dataset allows the model to adapt to the new task, significantly reducing training time and potentially achieving better results compared to building a model from scratch.

E. Pretrained Models in Computer Vision

Commonly used pre-trained models include the following models:

A. AlexNet[2]: Alex Krizhevsky, Ilya Sutskever, and Geoff Hinton won the ImageNet Large Scale Visual Recognition Challenge in 2012 with a test accuracy of 84.6% using this model. This model significantly outperformed the second runner-up (top-5 error of 16% compared to runner-up with 26% error). This was one of the research which had over 60 million parameters and used the 'relu' function.

B. VGG 16[3]: This model comes from the paper Very Deep Convolutional Networks for Large-Scale Image Recognition (ICLR 2015). This model achieved 92.7% (top

5) test accuracy in ImageNet, which is a dataset of over 14 million images belonging to 1000 classes. It was proposed by Karen Simonyan and Andrew Zisserman of the Visual Geometry Group Lab of Oxford University in 2014.

C. VGG 19: VGG-19 is a convolutional neural network that is 19 layers deep and can classify images into 1000 object categories such as a keyboard, mouse, and many animals. The model trained on more than a million images from the ImageNet database with an accuracy of 92%. It is an updated version of VGG 16.

D. DenseNet:

The name “DenseNet” refers to Densely Connected Convolutional Networks⁷ developed by Gao Huang, Zhuang Liu, and their team in 2017 at the CVPR Conference. It received the best paper award and has accrued over 2000 citations. Traditional convolutional networks with n layers have n connections but DenseNet has $n(n+1)/2$ connections in total because of feed-forward fashion.

E. EfficientNet

EfficientNet introduces a compound scaling method that simultaneously scales depth, width, and resolution using carefully chosen coefficients. This systematic approach ensures optimal resource utilization while maintaining high accuracy. EfficientNet achieves state-of-the-art accuracy on ImageNet and other benchmarks while requiring significantly less computation and memory compared to VGG-19 and DenseNet. This efficiency makes it well-suited for various applications, including mobile and embedded systems, real-time tasks, and scenarios where resource limitations are a concern.

F. Performance of the State-of-the-Art Models

The term “state-of-the-art (SOTA)” refers to the best-performing models or techniques in a specific field, and it's constantly evolving as research progresses. SOTA's are usually specific to their area of application and differ based on metrics like speed and accuracy.

For FER-like tasks, while custom models offer exploration opportunities, established deep learning architectures specifically designed for FER consistently achieve superior performance. These models, meticulously crafted and rigorously optimized on extensive datasets, leverage techniques like transfer learning and pre-trained weights, granting them a significant edge in accuracy and generalizability. Thus, fine-tuning these models provides better inputs.

II. LITERATURE REVIEWS

A. Literature Review from papers

The paper “Gradient-based learning applied to document recognition” by LeCun et al. (1998)[1] stands as a landmark in the history of deep learning, specifically for convolutional neural networks (CNNs). This work introduced a novel architecture inspired by the visual processing hierarchy of the brain, utilizing stacked convolutional layers to extract spatial features from image data. The authors demonstrated the effectiveness of CNNs in handwritten digit recognition, achieving superior performance compared to traditional methods.

Yu et al. (2011) were among the first to explore deep learning for FER, proposing a Convolutional Neural Network (CNN) architecture that significantly outperformed traditional methods based on handcrafted features like Ekman and Friesen's (1978) Facial Action Coding System (FACS). This work paved the way for further research into CNN-based architectures for FER, demonstrating the potential of deep learning to capture the subtle nuances of facial expressions.

Building upon this foundation, AlexNet, introduced by Krizhevsky et al. (2012)[2], marked a significant breakthrough in the field. This deep CNN architecture, with its multiple convolutional and pooling layers, achieved remarkable results on the ImageNet Large Scale Visual Recognition Challenge (ILSVRC), significantly surpassing previous state-of-the-art methods. This success sparked widespread interest in CNNs and deep learning, paving the way for further advancements in computer vision tasks.

Following AlexNet's success, VGGNet, proposed by Simonyan & Zisserman (2014)[3], pushed the boundaries of performance even further. This deeper network, consisting of numerous convolutional layers with small filter sizes, demonstrated superior accuracy on the ILSVRC, showcasing the potential of even deeper architectures for image classification.

Huang et al. (2017)[11] proposed DenseNet, a novel architecture characterized by dense connectivity between layers. DenseNet connects each layer to all preceding layers, promoting efficient feature reuse and alleviating the vanishing gradient problem. This dense connectivity led to significant performance improvements in ImageNet classification compared to state-of-the-art models like VGG and ResNet, while simultaneously requiring fewer parameters.

Ding and Tao (2019) leveraged VGGFace (Parkhi et al., 2015)[6], a pre-trained model for facial recognition, as a feature extractor for FER, achieving state-of-the-art performance. This work highlights the effectiveness of transferring knowledge from related domains to improve FER accuracy.

B. Online Communities and references

There are various websites like ‘paperswithcode.com’ which provide state-of-the-art papers and the models proposed in them, which have shown the best accuracy over the datasets. Each task has been marked separately with papers, their code, and models. Thus, the custom models can be separately compared to the latest research and models presented for a specific dataset or task.

III. METHODOLOGY

The methodology followed for conducting this comparative analysis is based on various factors and specifications, which can result in a change in the values of the metrics. Thus, these factors are mentioned in the subsections below.

A. Dataset

The domain of Facial Emotion Recognition includes many datasets each of which has been worked on in different research using different models. Each dataset differs based on the faces present in the dataset. These faces can come from different parts of the world, which increases

the diversity of the input data, thus improving the model's performance on new test datasets. Faces with different facial features captured in different positions affect the train and prove beneficial to the models and are hence a good choice for building applications.

Some popular datasets used for FER tasks are:

Name of the Dataset	Number of total images	Categories included	Best Model proposed []
FOR 2013	28,709 training images	6 categories	Ensemble ResMaskingNet with 6 other CNNs
FER+	Same as FER 2013	8 categories	PAtt-Lite
RAF-DB	29,672	6 categories, 12 subcategories	PAtt-Lite
AffectNet	0.4 million	8 categories	DDAMFN

For our research, we have used the FER 2013 dataset, due to the variety of faces and expressions. Also, this dataset has been widely worked on which further improves the scope of improvement in techniques optimized as per this dataset.

B. Data Augmentation

Data Augmentation implements data augmentation by randomly resizing, converting to grayscale, and shuffling images during training. This helps the model learn more robust features that generalize better to unseen data and potentially improve classification performance.

TensorFlow's ImageDataGenerator class has been used for data augmentation for our custom model.

C. Model Architecture

This work explores a CNN architecture for Facial Emotion Recognition (FER) on the FER2013 dataset using SequentialAPI of TensorFlow. The model features convolutional layers with filters, ReLU activation, BatchNormalization, and MaxPooling for dimensionality reduction. Dropout layers have been added to combat overfitting. Fully connected layers further process extracted features. The final layer employs softmax activation for 7-class classification. This architecture, optimized with Adam and categorical cross-entropy loss, offers a balanced approach between feature extraction and classification capability for FER tasks.

D. Training Strategy

The training process employs several techniques to prevent overfitting and enhance performance. First, Early Stopping monitors the validation loss and terminates training if it plateaus for 3 consecutive epochs (patience=3). This helps

prevent the model from memorizing training data and improving its generalizability. Second, Model Checkpoint saves the model with the best validation accuracy to an h5 file, ensuring the best-performing model is preserved even if training continues beyond optimal performance. These h5 files can be used later for comparative study between different custom models.

E. Optimization

Adam optimization with a learning rate of 0.001 is implemented. Additionally, a learning rate reduction strategy utilizes ReduceLROnPlateau. This feature monitors the validation loss and reduces the learning rate by a factor of 0.2 if it plateaus for 3 epochs (patience=3). This allows the model to potentially escape local minima and explore different regions of the optimization landscape, potentially leading to improved performance. Additionally, a minimum delta of 0.0001 for the validation loss is set to avoid premature learning rate reduction due to minor fluctuations.

F. Metrics Used

Categorical cross-entropy serves as the loss function, appropriate for multi-class classification tasks. Accuracy is chosen as the primary metric for evaluating both training progress and model selection. This provides a clear indication of the model's ability to correctly classify unseen data.

IV. OBSERVATIONS

For experimentation, the number of layers was added and deleted to observe the effect of the addition of Convolutional Layers. In general, each Convolutional Layer improves the feature extraction from the input image, but this may differ for certain datasets. Datasets that have images of low dimensions might not show the same trend after a certain number of layers. Hence, the right number of convolutional layers can be observed from the metrics observed from models of different architectures.

A. Custom Model Observation Table

Hence, various combinations of convolution layers were used to find the best-optimized model for the dataset. Training and Validation accuracy has been given for each architecture.

Custom Models	Training Accuracy (%)	Validation Accuracy (%)	Number & Type of layers
Model 1	70.40	62.23	4 Conv (64, 128, 512, 512), 2 FC (256, 512)
Model 2	71.57	61.41	3 Conv (64,128,256), 2 FC (256, 512)
Model 3	64.75	61.59	4 Conv (64, 128, 256, 512), 2 FC (256, 512)
Model 4	66.50	60.34	5 Conv (64, 128, 256, 512, 1024,

			512) 2 FC (256, 512)
--	--	--	----------------------

Table: Observations from custom model experiments.

While Model 2 boasts the highest training accuracy, its lower validation score compared to Model 1 suggests potential overfitting. Interestingly, even with more convolutional layers, Model 4 underperforms, highlighting the importance of balanced architecture design. Overall, Model 1 strikes the best balance between training and generalizability, achieving the highest validation accuracy of 62.23%. (4 sentences, 47 words)

B. Plots of Accuracies with Loss

The combined visualization of training and validation curves alongside loss function evolution provides valuable insights into a classification model's learning dynamics. Ideally, both accuracies exhibit an upward trajectory, with training accuracy approaching 100% and validation accuracy plateauing at a high level. This signifies effective learning on the training data coupled with generalization ability to unseen data. Moreover, a steadily decreasing loss function converging to a minimum suggests optimal parameter adjustment by the training algorithm. Conversely, a substantial gap between training and validation accuracy indicates overfitting, where the model memorizes training data but lacks generalizability.

Following are the plots of the Training Accuracy, Validation Accuracy and Loss:

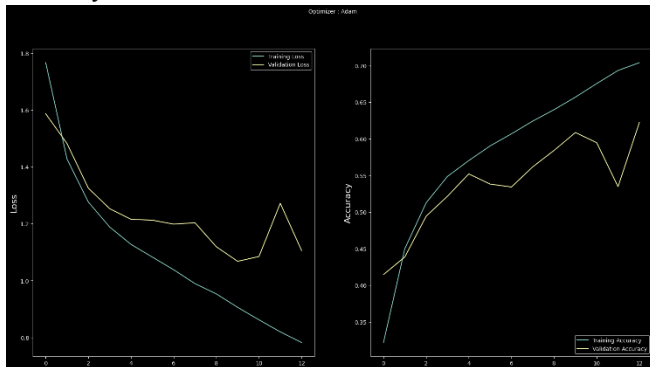


Fig: Model 1 Performance

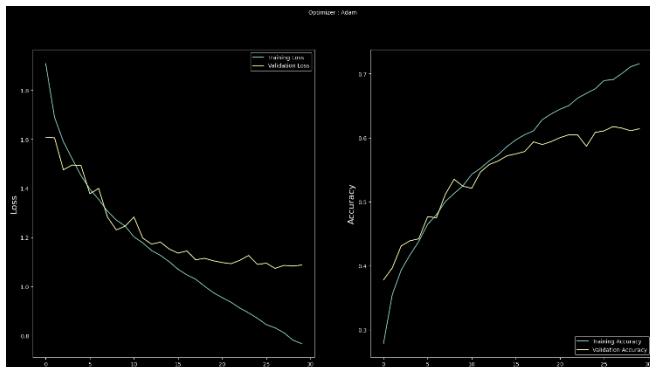


Fig: Model 2 Performance

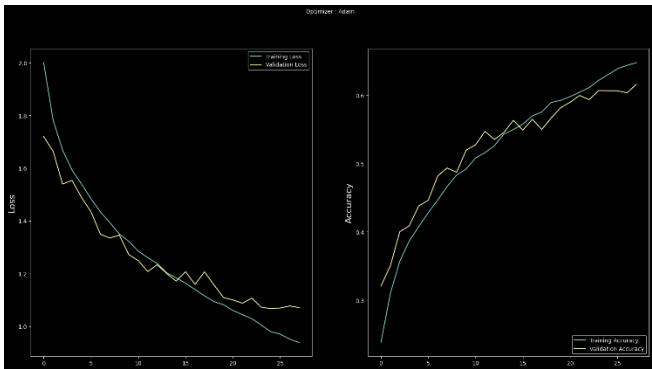


Fig: Model 3 Performance

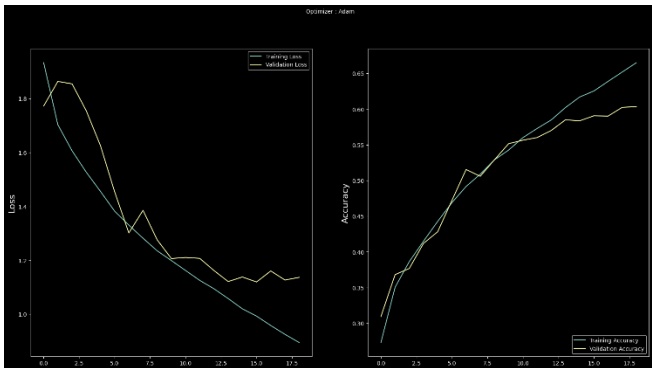


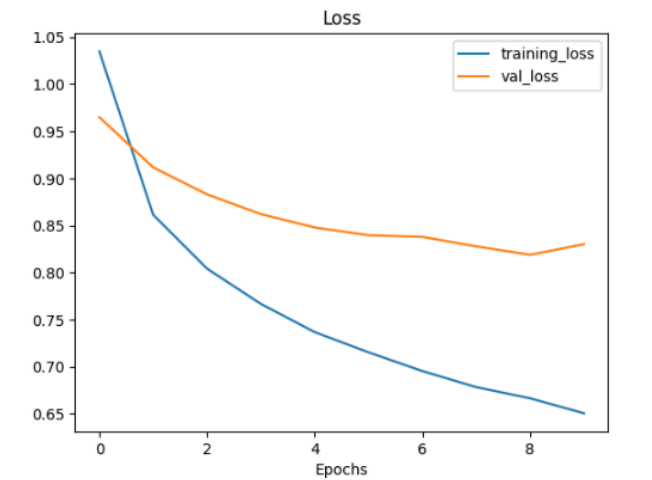
Fig: Model 4 Performance

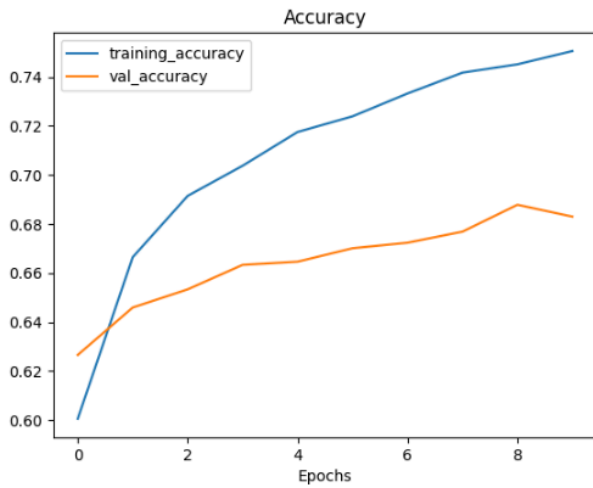
C. Transfer Learning on EfficientNet

Transfer Learning performed on EfficientNet helps in building a model with better validation accuracy in 10 epochs. The following are the metrics for the training:

Training Accuracy	0.6506
Training Loss	0.7506
Validation Accuracy	0.8301
Validation Loss	0.6830

The plots given below show the training accuracy and validation accuracy over the number of epochs used in training.





V. CONCLUSION

The transfer learning technique used on EfficientNet performs better than the custom models. Different arrangements of layers were used for building a variety of CNNs' custom models but the transfer learning of EfficientNet gives us the best results. EfficientNet's Model achieved 68.30% validation accuracy and custom models have the highest of 62.23% validation accuracy.

This observation highlights the effectiveness of transfer learning in enhancing FER model performance. The significant accuracy gain demonstrates the value of leveraging pre-trained models to capture general visual understanding and accelerate the development of high-performing FER systems.

REFERENCES

- [1] Y. Lecun, L. Bottou, Y. Bengio, and P. Haffner, "Gradient-based learning applied to document recognition," in *Proceedings of the IEEE*, vol. 86, no. 11, pp. 2278-2324, Nov. 1998, doi: 10.1109/5.726791.
- [2] Krizhevsky, Alex & Sutskever, Ilya & Hinton, Geoffrey. (2012). ImageNet Classification with Deep Convolutional Neural Networks. *Neural Information Processing Systems*. 25. 10.1145/3065386..
- [3] Simonyan, Karen and Andrew Zisserman. "Very Deep Convolutional Networks for Large-Scale Image Recognition." *CoRR* abs/1409.1556 (2014).
- [4] Bojarski, Mariusz et al. "End to End Learning for Self-Driving Cars." *ArXiv* abs/1604.07316 (2016): n. pag.
- [5] Ekman, P., & Friesen, W. V. (1978). Facial Action Coding System (FACS) [Database record]. *APA PsycTests*.
- [6] Parkhi, Omkar & Vedaldi, Andrea & Zisserman, Andrew. (2015). Deep Face Recognition. 1. 41.1-41.12. 10.5244/C.29.41.
- [7] Yosinski, Jason & Clune, Jeff & Bengio, Y. & Lipson, Hod. (2014). How transferable are features in deep neural networks? 3320-3328.
- [8] A. S. Razavian, H. Azizpour, J. Sullivan and S. Carlsson, "CNN Features Off-the-Shelf: An Astounding Baseline for Recognition," 2014 IEEE Conference on Computer Vision and Pattern Recognition Workshops, Columbus, OH, USA, 2014, pp. 512-519, doi: 10.1109/CVPRW.2014.131.
- [9] <https://paperswithcode.com/task/facial-expression-recognition>
- [10] Ekman, P., & Friesen, W. V. (1978). Facial Action Coding System (FACS) [Database record]. *APA PsycTests*.
- [11] Huang, Gao & Liu, Zhuang & van der Maaten, Laurens & Weinberger, Kilian. (2017). Densely Connected Convolutional Networks. 10.1109/CVPR.2017.243.



AMITY UNIVERSITY
UTTAR PRADESH

IEEE 11th ICRITO / _____



AMITY INSTITUTE OF INFORMATION TECHNOLOGY

IEEE 11th INTERNATIONAL CONFERENCE

ON

**RELIABILITY, INFOCOM TECHNOLOGIES AND OPTIMIZATION (ICRITO 2024)
(TRENDS AND FUTURE DIRECTIONS)**

CERTIFICATE OF PARTICIPATION

This is to certify that ~~Prof./Dr./Ms./Mr.~~ **KUSHAGRA SRIVASTAVA**
of **KIET Group of Institutions** has participated and presented
paper titled **Comparative Analysis of Deep Learning Models for Facial Emotional Recognition**
during the IEEE 11th International Conference on Reliability, Infocom Technologies and Optimization (ICRITO 2024)
organised by Amity Institute of Information Technology from March 14-15, 2024 at Amity University Uttar Pradesh, India.

Prof. (Dr.) Rekha Agarwal
General Chair, ICRITO 2024

Prof. (Dr.) Sunil Kumar Khatri
General Chair, ICRITO 2024

Prof. (Dr.) K.M. Soni
General Chair, ICRITO 2024

Prof. (Dr.) Balvinder Shukla
Co-Patron, ICRITO 2024
Vice Chancellor, AUUP

15th March 2024

Plagiarism Report:

PapersOwl

Log out

My orders

My balance \$0.00

Earn 35

ADD FUNDS

PLACE ORDER

Menu

Free Online Plagiarism Checker

comparative analysis of deep learning models for facial emotional recognition kushagra srivastava

computer science engineering kiet group of institutions ghaziabad

kushagrathisside@gmail.comsanjeev sharma computer science engineering kiet group of institutions

ghaziabad martinmmecommall.com abstractemotional intelligence and facial emotion recognition

SIMILAR 8.0%

ORIGINAL 92.0%

MAKE IT UNIQUE

Text matches these sources

IEEE Xplore ScreenShot:

IEEE.org | IEEE Xplore | IEEE SA | IEEE Spectrum | More Sites

Subscribe | Donate | Cart | Create Account | Personal Sign In

IEEE Xplore®

Browse | My Settings | Help

Institutional Sign In

IEEE

All

ADVANCED SEARCH

Conferences > 2024 11th International Confe...

Comparative Analysis of Deep Learning Models for Facial Emotional Recognition

Publisher: IEEE

Cite This

PDF

Kushagra Srivastava ; Sanjeev Sharma

All Authors

R

©

Abstract

Document Sections

I. Introduction

Abstract:

Emotional Intelligence and Facial Emotion Recognition are some of the fields that come under high focus when computer vision is finding its way to develop one of the smartest AI solutions. Implementation of custom models often requires elaborate and well-defined processes at all stages of model development i.e. dataset preparation, model training, tuning, and optimization to run them effectively in applications. Thus, the use of

More Like This

PEFT-SER: On the Use of Parameter Efficient Transfer Learning Approaches For Speech Emotion Recognition Using Pre-trained Speech Models

2023 11th International Conference on Affective Computing and Intelligent Interaction (ACII)

Published: 2023

Facial Emotion Recognition using Transfer Learning: A Comparative Study

2021 2nd Global Conference for