EMOTION RECOGNITION AND SENTIMENT ANALYSIS FOR

RELATIONSHIP IMPROVEMENT.

Hewa Gamage Chanika Dilini

IT20651442

B.Sc. (Hons) Degree in Information Technology Specialized in

Information Technology.

Department of Software Engineering and Faculty of Computing.

Sri Lanka Institute of Information Technology

Sri Lanka

April 2024

EMOTION RECOGNITION AND SENTIMENT ANALYSIS FOR

RELATIONSHIP IMPROVEMENT.

Hewa Gamage Chanika Dilini

IT20651442

B.Sc. (Hons) Degree in Information Technology Specialized in

Information Technology.

Department of Software Engineering and Faculty of Computing.

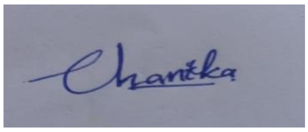
Sri Lanka Institute of Information Technology

Sri Lanka

April 2024

DECLARATION

We declare that this is our own work, and this proposal does not incorporate without acknowledgement any material previously submitted for a degree or diploma in any other university or Institute of higher learning and to the best of our knowledge and belief it does not contain any material previously published or written by another person except where the acknowledgement is made in the text.

Group Member Name	**Student ID**	**Signature**
H.G.C. Dilini	IT20651442	

Signature of the Supervisor: **Date:** 2024/04/05

ABSTRACT

Humans can show many different expressions when they communicate, and these expressions can have various meanings. Emotions play a key role in mental well-being, and many people face different mental health issues. This study explores the creation of a mobile app designed to improve users' emotional health using advanced technology. The app includes features for recognizing emotions from audio, detecting emotions from video, analyzing sentiments, and creating a virtual reality (VR) character to build relationships.

The app uses smart AI technology to analyze users' voices and facial expressions in real-time to identify their emotions. For recognizing emotions from audio, we use a deep learning model. For detecting emotions from video, we use the ResNet50 model. Sentiment analysis is done using a Random Forest Classifier to understand users' feelings better.

Based on this analysis, the app gives personalized recommendations, such as relaxation mini-games and contact information for counselors if someone is feeling very depressed. Additionally, the app includes an AI-powered VR character created using the RASA framework. This virtual character acts as a companion, offering support and guidance to users, aiming to build a strong emotional connection. The goal of this study is to use technology to promote emotional well-being. By combining advanced models for audio and video emotion detection, effective sentiment analysis, and an interactive VR character, this app aims to be a helpful tool for improving mental health and emotional wellness.

Keywords: ***emotions, sentiment analysis, VR environment, image processing, audio recognition ResNet50, Random Forest Classifier, Unity, Convai***

ACKNOWLEDGEMENT

First and foremost, I want to express my heartfelt gratitude to my supervisor, Dr. Dilshan De Silva. His constant guidance and support were invaluable to me throughout my undergraduate research. I would also like to extend my sincere thanks to Mr. Samitha Vidhanaarachchi, my co-supervisor. He was always willing to help and provide advice whenever I needed it. His support was crucial in overcoming the various challenges that arose during the research. His insights and encouragement helped me stay focused and motivated, leading to the successful completion of this project.

This study is a blend of technology and mental health, so I required the expertise of both technology experts and mental health professionals. I am deeply appreciative of Ms. Shalindi Pandithakoralage from the Sliit Help Desk. Her immense support in understanding people's feelings and suggesting remedies was essential to this project. Her guidance was incredibly helpful and significantly contributed to the project's success.

Finally, I want to express my gratitude to my teammates, family, friends, and everyone who directly or indirectly supported me throughout this project. Their encouragement and assistance were vital in keeping me motivated and on track. Thank you to all for your unwavering support and belief in me.

TABLE CONTENT

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS

Abbreviation	Description
CNN	Convolutional Neural Network
VGG	Visual Geometry Group
FER	Facial Expression Recognition
CK+	Cohn-Kanade
AI	Artificial Intelligence
WBS	Work Breakdown Structure
WHO	World Health Organization
GDPR	General Data Protection Regulation
AWS	Amazon Web Services
ReLU	Rectified Linear Unit
API	Application Programming Interface
Expo	A framework for building universal native apps

RD	Real-time Emotion Detection
VCD	Video Call Data Collection
ER	Emotion Recognition
DLF	Deep Learning Framework
IP	Image Processing
CPAD	Cross-Platform App Development
MER	Multi-Emotion Recognition
PGS	Personalized Goal Setting
IVCP	Integration with Video Call Platforms
PTR	Progress Tracking and Reporting
RD	Research and Development
ML	Machine Learning
TU	Target Audience
CR	Commercialization Strategy

TCR	Test Case Result
TDS	Test Data Set
GDPR	General Data Protection Regulation
H5	Hierarchical Data Format version 5

1. INTRODUCTION

1.1 General Introduction.

Understanding emotions is essential for understanding human behavior and interactions. As digital communication grows, the ability to read emotions from nonverbal cues is becoming more important. Recent improvements in image processing, machine learning, and deep learning have made it possible to identify emotions from facial expressions, helping to improve relationships and overall experiences [1].

Facial expressions are a primary way humans show emotions. Research has shown that certain muscle movements in the face are linked to specific emotions, which has helped create automated systems to recognize these emotions. Image processing plays a crucial role in this field [4]. It involves gathering, preparing, and analyzing image data to find key features. Advanced computer vision methods, supported by deep learning, are effective in identifying facial features and emotions. Tools like OpenCV have made these technologies accessible to more researchers, allowing them to develop complex facial recognition and tracking systems.

One major challenge in detecting emotions is making models that can see small changes in facial expressions. Deep learning, especially convolutional neural networks (CNNs), has shown great promise in this area [2]. These models are trained on large sets of facial expressions and can find patterns related to emotions such as happiness, sadness, anger, and surprise. Pre-trained CNN architectures like VGG, ResNet, and Inception have been fine-tuned for emotion detection and show promising results [5].

Key datasets like CK+ (Cohn-Kanade) and FER2013 (Facial Expression Recognition 2013) have been essential in advancing this field [3]. These datasets have labeled images that show different emotions, which are important for training and testing emotion detection models. Transfer learning, which involves using pre-trained models for specific tasks, has become popular because of the limited availability of labeled data. Additionally, researchers are exploring methods that combine visual cues with audio and text to improve the accuracy of emotion recognition, understanding that human emotions are complex.

The use of emotion recognition through image processing extends beyond personal interactions. It shows promise in areas like healthcare, marketing, and entertainment. In healthcare, watching a patient's emotional state can help find mental health issues early. In marketing, understanding consumer emotions can guide advertising strategies and product design. In entertainment, emotion recognition can be used to tailor content to viewers' emotional responses.

Despite its potential, emotion recognition technology raises ethical concerns. Privacy issues related to collecting and analyzing personal data, especially facial images, are significant. It is important to balance technological advancements with respect for individual privacy and rights, ensuring the responsible use of emotion recognition systems. Image-based emotion recognition for mood analysis is a rapidly growing field with a lot of potential [6]. Combining computer vision, deep learning, and psychological insights has led to models that can understand emotions from facial expressions. Although ethical considerations remain, the potential to improve relationships and deepen our understanding of human emotions through technology is exciting. Our project aims to more accurately identify users' emotions and provide appropriate advice, marking a pioneering effort in Sri Lanka and offering personalized guidance based on emotion detection.

1.2 An overview on emotion identification.

Emotion identification is a key research area that combines psychology, computer science, and neuroscience to recognize and understand human emotions from signals like facial expressions, voice tones, and body language. In terms of historical development research on emotions began long ago, with Charles Darwin's 1872 work on how emotional expressions have evolved, laying the foundation for modern studies.

Paul Ekman identified six basic emotions those are happiness, sadness, anger, fear, disgust, and surprise. His research showed that people can recognize these emotions with over 70% accuracy across different cultures [7]. James Russell's Circumflex Model describes (Fig 1.2.1) emotions based on arousal (energy level) and valence (pleasantness). This model captures 85% of the differences in how people experience emotions [8].

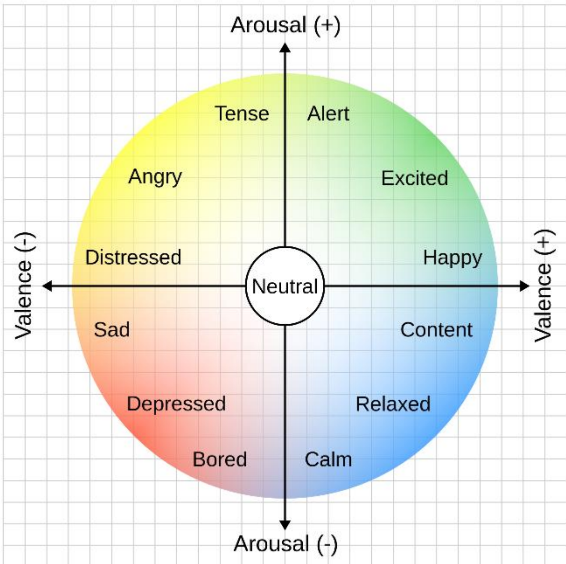


Figure 1. 2.1 : James Russell's Circumflex Model

Techniques like facial detection and feature extraction are vital. Open CV, a popular tool, has more than 47 thousand people of user community and estimated number of downloads exceeding 18 million. Deep learning models, especially Convolutional Neural Networks (CNNs), have transformed emotion identification. For example, the ResNet architecture, which has over 100 layers, achieved a top-5 error rate of 3.57% in a big image recognition challenge, proving its strength in recognizing features [9]. Combining visual, audio, and text data can improve the accuracy of emotion recognition [10]. Research shows that using multiple types of data together can improve accuracy by about 10-15% compared to using just one type.

Emotion identification helps in diagnosing and monitoring mental health conditions. Automated emotion detection can identify depression with about 85% accuracy, helping in early intervention [11]. Depression rates increasing around in the World. Depression affects about 1 in 15 adults in any given year, and 1 in 6 people will experience depression at some time in their life. An Our World in Data study estimates about 3.4% (2-6% when including the margin of error) of the global population has depression. This is about 264 million people worldwide. According to WHO estimates, the ten countries with the highest prevalence of depression shown in table 1.2.1.

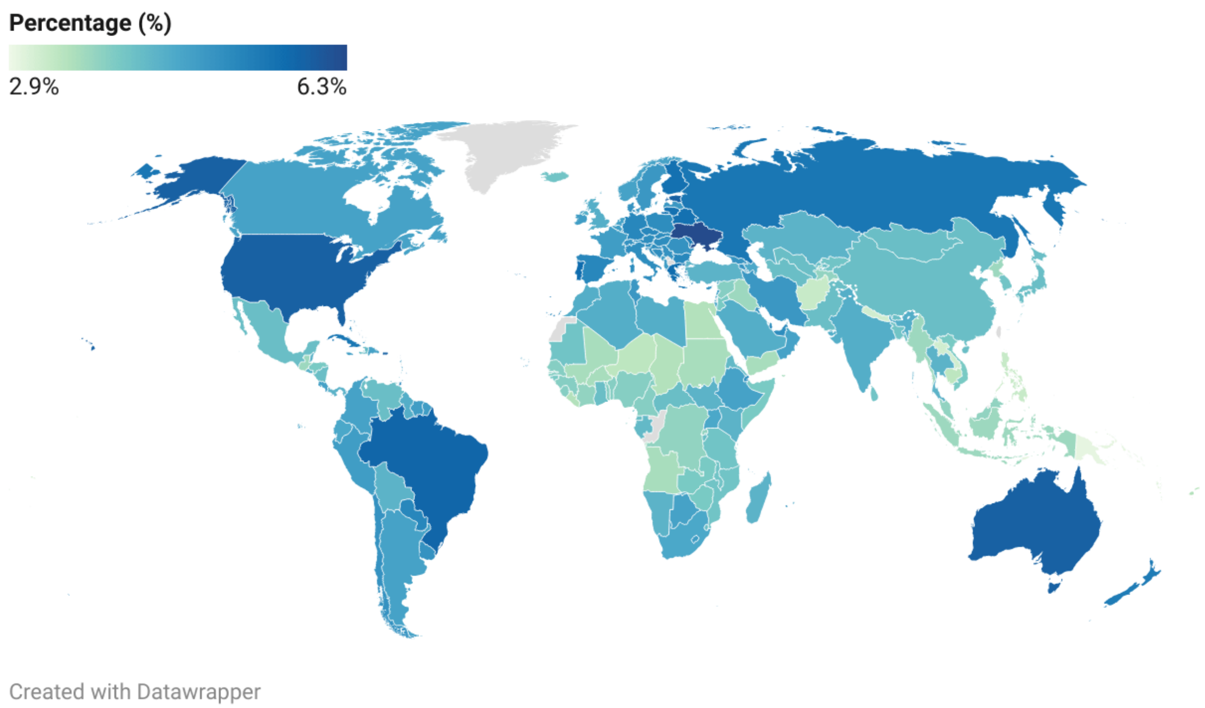


Figure 1. 2.2 : Depression rates by country 2024

Table 1.2.1: Ten countries with the highest prevalence of depression.

Rank	Country	Depression Rate (%)
1	Ukraine	6.3%
2	United States	5.9%
3	Australia	5.9%
4	Estonia	5.9%
5	Brazil	5.8%
6	Greece	5.7%
7	Portugal	5.7%
8	Belarus	5.6%
9	Finland	5.6%
10	Lithuania	5.6%

These figure 1.2.2 highlight the widespread issue of depression across various countries, each facing unique challenges and cultural factors affecting mental health. For instance, Ukraine has the highest rate at 6.3%, likely influenced by ongoing socio-political stress. In high-income countries like the United States and Australia, the prevalence is also significant, emphasizing the universal nature of this mental health issue despite economic stability.

Emotion recognition technology can improve virtual assistants. Customer satisfaction with AI assistants increases by 20% when the assistant can recognize and respond to user emotions. Understanding consumer emotions helps in making better ads. Studies show that ads aimed at emotional responses have a 23% higher engagement rate than neutral ads. Emotion recognition can personalize user experiences in games and media. Adaptive gaming environments that respond to player emotions can increase user engagement.

The development of emotion recognition technology raises ethical concerns, especially about privacy and consent. The European Union's General Data Protection Regulation (GDPR) has strict rules about personal data, including emotional data, to protect privacy rights. Ensuring models are trained on varied datasets is crucial. Current datasets like CK+ and FER2013 have thousands of images (Fig 1.2.3) , but more diverse data is needed for better generalization [3]. Detecting subtle emotions is difficult. Models often mix up similar expressions, reducing accuracy by up to 15% in real-world applications.

Figure 1. 2.3 : CK+ and FER2013 dataset sample

Efficient algorithms for real-time emotion detection are essential. Advances in hardware like GPUs have cut down processing times, making real-time applications possible. Finally, emotion identification is a rapidly growing field with many uses and important ethical considerations. It offers exciting possibilities for better human-computer interactions and understanding human emotions. This project aims to accurately identify users' emotions and give suitable advice, marking a pioneering effort in Sri Lanka and offering personalized guidance based on emotion detection.

1.3 Research Gap

While there has been significant progress in emotion recognition, particularly through video and facial expression analysis, existing systems fall short in several key areas. The primary gap lies in the integration of emotion recognition with applications aimed at improving relationships, providing real-time feedback and remedies, and assessing the progression of emotional states. This gap is particularly evident in the context of video call feed data.

Research “A” utilizes bilinear pooling B-CNN and fusion feature F-CNN to identify facial expressions in video streams, achieving efficient and accurate emotion recognition [12]. The system employs Convolutional Neural Networks (CNNs) to process images through layers of convolutions, pooling, and fully connected layers, concluding with a SoftMax function to classify emotions. Despite its accuracy in identifying human emotions, this system does not provide post-recognition activities such as relationship improvement strategies or effective management techniques. Users cannot manage their emotions or improve their relationships based on the identified feelings.

Research “B” paper presents a system that uses CNNs to classify facial emotions from grayscale images and real-time videos. It employs feature extraction through convolution and pooling layers and classification using a SoftMax layer [14]. To mitigate overfitting, techniques like dropout, cluster standardization, and L2 regularization are used. The model, tested on the FER2013 dataset, shows good accuracy in predicting emotions. However, similar to Research A, this study is limited to emotion identification and does not provide strategies for relationship improvement or emotional management post-identification.

Research “C” study analyzes emotions from social networking sites and photo-sharing websites using a specialized CNN and pretrained models [16]. It categorizes emotions into five types affection, happiness, cruelty, fear, and sadness. Although this approach offers insights into general emotions from images, it lacks a mobile or web application to provide solutions such as relationship goal trackers or status updates on users' relationships after identifying emotions.

Research “D” research focuses on emotion recognition using multi-modal approaches, combining facial expressions, voice intonations, and physiological signals. The system integrates data from these various sources using deep learning models to improve the accuracy of emotion recognition. While this multi-modal approach provides a more comprehensive understanding of emotional states, it does not address how to use this information to improve relationships or provide real-time remedies for emotional management. Additionally, the system's complexity and the need for multiple data inputs limit its practicality for everyday use and integration into mobile or web applications [15].

None of the studies provide practical applications for improving relationships based on the identified emotions. The existing systems do not offer real-time feedback or remedies to help users manage and improve their emotional well-being.

But proposed system utilizing advanced image processing and deep learning techniques, particularly ResNet50, to accurately identify emotions from video call feed data. Integrating real-time feedback and providing practical remedies to help users manage their emotions. This feature aims to help users understand their emotional responses and take steps to improve their interactions and relationships. Implementing a feature to track the progression of user emotions over time, offering insights into emotional trends and patterns. This can help users and therapists understand long-term emotional changes. Developing user-friendly mobile and web applications that combine emotion recognition with tools for emotion management. This integration aims to make the technology accessible and useful in everyday situations, enhancing its practical application.

Table1.3.1: Comparison between existing system

Research	Emotion Recognition	Display remedies to control after identifying the emotion.	Progression level of user emotion.	Mobile application, Web application.
Research A	✓	✗	✗	✗
Research B	✗	✗	✗	✗
Research C	✓	✗	✗	✗
Research D	✓	✗	✓	✗
Proposed System	✓	✓	✓	✓

2. RESEARCH PROBLEM

Understanding emotions during live video calls is a significant challenge in affective computing. The primary issue is accurately identifying emotions from real-time video feeds, which are dynamic and constantly changing. Unlike static images, live videos involve continuous facial expressions that need instant analysis. Variations in lighting, facial movements, and camera angles make it difficult for emotion recognition systems to work effectively outside controlled lab conditions.

Most current systems focus primarily on visual data like facial expressions, often ignoring important auditory cues such as speech tone, pitch, and the words spoken. Combining these elements could greatly improve the accuracy and robustness of emotion recognition. Emotions also vary significantly across different cultures, and an effective system must understand these differences to interpret emotions correctly for a diverse range of users. This cultural variability poses a complex challenge for developing universally effective emotion recognition systems.

Another major concern is privacy. Continuous monitoring of facial expressions and behaviors during live video calls raises significant privacy issues. Ensuring user privacy while accurately detecting emotions is a critical challenge. Users may feel uncomfortable or manipulated knowing they are being closely watched by a computer system, which necessitates developing systems that are both effective and respectful of privacy.

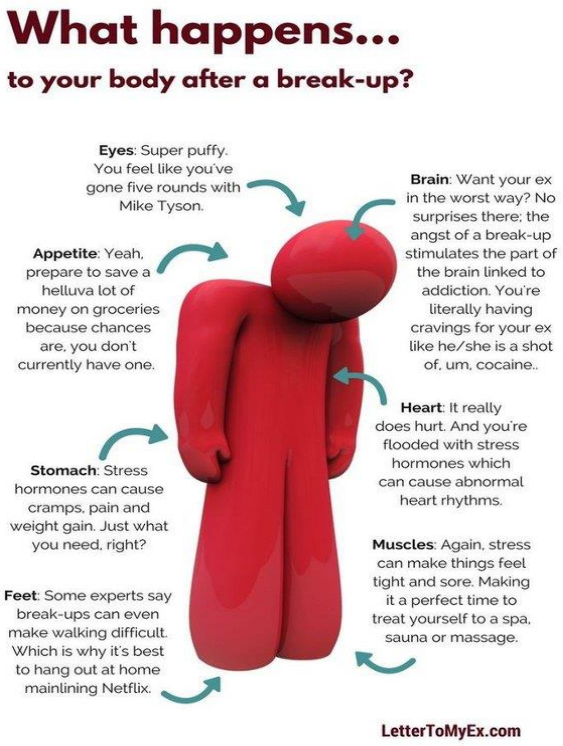


Figure 2.1: Survey of after breakup what happened to your body

Moreover, emotions are not static, they change over time. Systems must be capable of tracking these changes to provide a comprehensive understanding of the user's emotional state. This involves capturing micro-expressions, body movements, and behaviors like blinking rates and gestures. Integrating visual, auditory, and textual data is essential for a complete understanding of emotions. For instance, combining data from facial expressions, speech, and textual analysis can provide a more holistic view of the user's emotional state.

In summary, the research problem is about creating advanced, culturally sensitive emotion recognition systems that work in real-time during live video calls while addressing privacy concerns. This research aims to bridge the gap between lab-based accuracy and real-world application, helping people connect better through technology. This can significantly improve relationships by providing insights into emotional states, enabling better empathy and understanding. The goal is to develop a system that can accurately identify emotions during live video calls, which has significant implications for fields such as mental health, customer service, and human-computer interaction. The system will use sophisticated techniques to analyze and understand the emotions conveyed through facial expressions, speech, and textual content, providing real-time feedback and insights.

3. RESEARCH OBJECTIVES

3.1 Main Objectives

The main objective of this component is to create a mobile application that can accurately and quickly recognize emotions using data from video calls. The aim is for the app to understand how users feel during conversations. This will help improve communication in relationships by providing personalized advice and suggestions for improvement based on the detected emotions. The app will use advanced image processing techniques to analyze facial expressions and other visual cues in real-time, making it a useful tool for better emotional understanding and interaction.

3.2 Specific Objectives

There are the specific objectives that need to be fulfilled in order to achieve the overall objective described above.

Real-Time Emotion Recognition Model Development

Develop an efficient real-time emotion recognition model using image processing techniques. This involves selecting suitable deep learning models for detecting facial features and recognizing expressions. Specifically, a Convolutional Neural Network (CNN) will be trained to accurately classify emotions, ensuring quick and precise identification of various emotional states during mobile video calls.

Multimodal Integration:

Explore the integration of multiple sensory cues, including visual and auditory data, to improve the accuracy of emotion recognition. Develop a system that combines these cues to provide a comprehensive understanding of users' emotions during mobile video calls. This approach aims to enhance the system's ability to detect and interpret emotions more accurately.

User Interface and Integration Using Computer Vision (OpenCV):

Design a user-friendly interface that displays real-time emotion recognition results during video calls. Utilize computer vision techniques, such as OpenCV, to process and analyze visual data. Integrate the emotion recognition module into a broader framework that offers personalized suggestions for improving communication based on the identified emotions. This will help users better understand and respond to each other's emotional states.

Goal Tracking and Emotional Growth

Incorporate a goal-tracking feature into the app to help users enhance their relationship skills. Provide real-time emotional analysis to offer insights into progress and personalized advice. This feature will promote self-awareness and positive changes, fostering emotional growth and improving relationships. Users will be able to track their emotional development over time, setting and achieving personal goals for better communication and relationship management.

4. METHODOLOGY

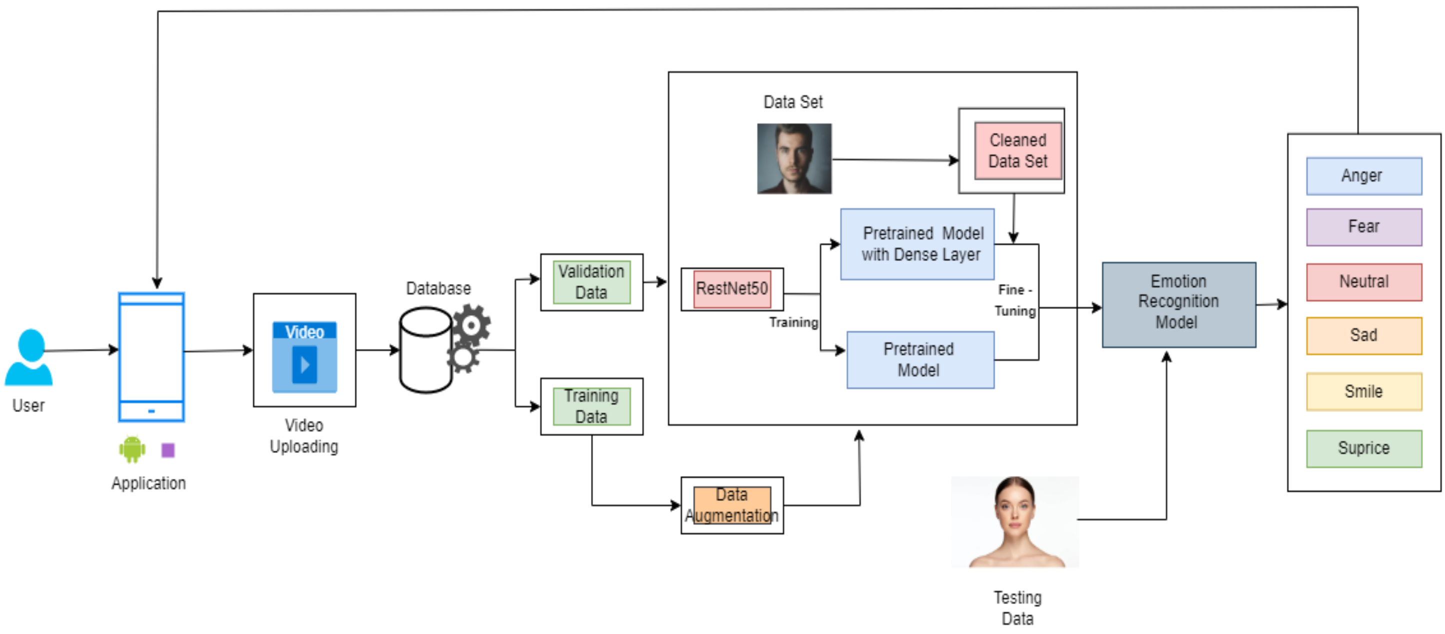


Figure 4.1: Overview of system diagram

The primary goal of this component is to develop an accurate, real-time emotion recognition system for mobile video calls. The process begins with data collection, where registered users upload video call feeds to an AWS back-end server. These videos are processed into individual frames, yielding a total of 1,818 images, which are split into 70% for training and 30% for testing. Six emotion classes are identified, Anger, Fear, Neutral, Sad, Smile, and Surprise.

Model development involves utilizing the ResNet50 CNN architecture, selected for its superior performance in image processing tasks. Transfer learning is employed using a pre-trained ResNet50 model from TensorFlow's Keras API, with additional custom layers added for emotion classification. The model is compiled with the Adam optimizer and sparse categorical cross-entropy loss function.

Training the model involves feeding the training dataset into the model for ten epochs, with accuracy and loss monitored for both training and validation datasets to ensure effective learning. Post-training evaluation includes generating a confusion matrix to analyze classification performance.

The trained model is then deployed on the AWS server and integrated into a mobile application using OpenCV for real-time video frame analysis. The app captures live video calls, processes frames in real-time, and predicts users' emotions, displaying the results through an intuitive user interface. Additionally, the system incorporates AI-enhanced audio-based sentiment analysis for a multimodal approach.

4.1 Requirement Gathering

The data collection phase for the emotion recognition component was comprehensive, drawing on multiple rich sources to ensure the robustness and accuracy of the system. The primary aim was to gather diverse and representative data to train and validate the emotion recognition models effectively.

Figure 4.2: Data Gathering

We began by leveraging data from previous research groups(figure 4.2). These groups had already conducted extensive studies on emotion analysis, providing a foundational dataset rich with annotated emotional expressions. This historical data was invaluable as it contained various controlled and spontaneous emotional expressions, serving as a critical resource for understanding patterns in emotional responses.

Additionally, we collaborated with professional counselors to obtain emotion-related data from their sessions. Counselors interact with clients in a range of emotional states, from joy and excitement to distress and sadness. This data was particularly useful for capturing nuanced emotional expressions in therapeutic settings, where emotions can be both intense and subtle. The involvement of counselors ensured that the data included professionally validated emotional labels, enhancing the quality and reliability of our dataset.

To further diversify our data, we collected video call feeds from campus students (Figure 4.3). This dataset was crucial for several reasons. Firstly, it provided real-world data from everyday interactions, ensuring that our model could generalize well to typical use cases. Campus students engaged in a variety of conversations, from casual chats to more emotionally charged discussions, offering a broad spectrum of emotional expressions. This real-life data was essential for training our models to recognize emotions in naturalistic settings, where factors like varying lighting conditions, different camera angles, and diverse facial movements come into play.

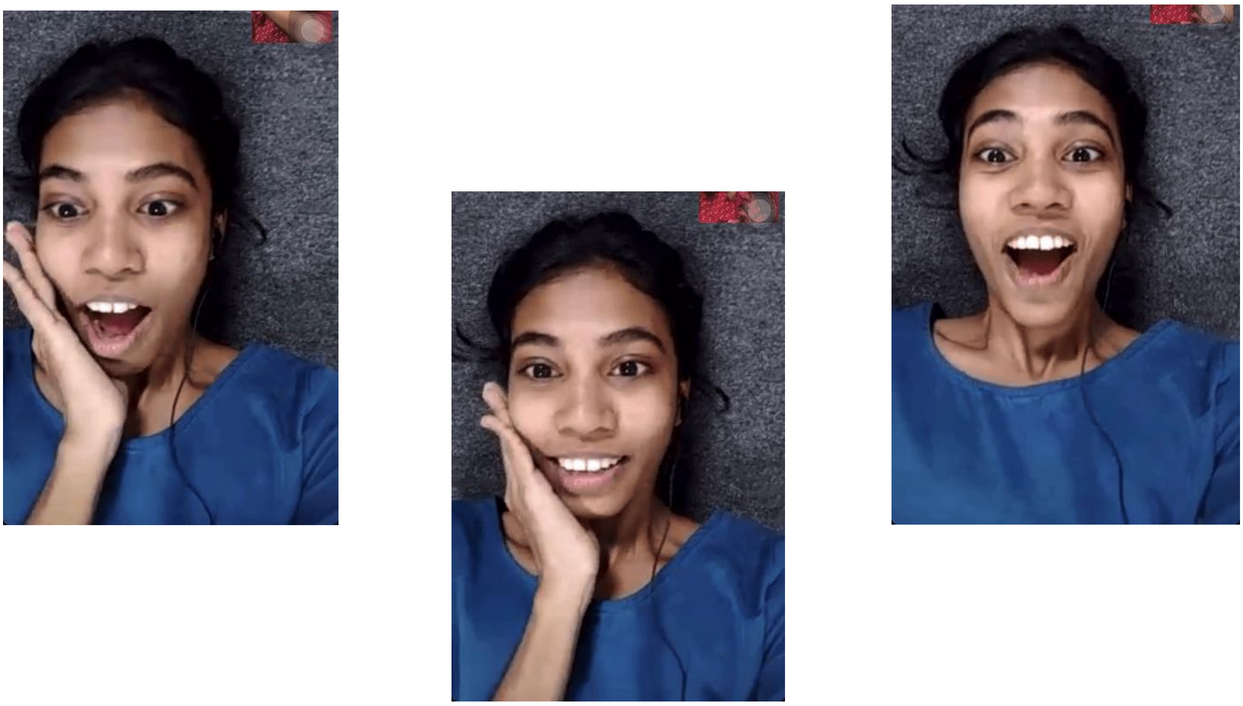


Figure 4.3: Video call feed data from campus student

The process of collecting video call feeds involved obtaining informed consent from all participants, ensuring ethical standards were maintained. Videos were then uploaded to google drive, where they were securely stored and processed. To prepare the data for model training, extracted individual frames from these videos, effectively breaking down the video feed into a series of images. This step was crucial for creating a robust dataset suitable for training deep learning models.

In total, collected 1,818 images representing six key emotion classes: Anger, Fear, Neutral, Sad, Smile, and Surprise. These images were split into two subsets, with 70% allocated for training the models and 30% reserved for testing and validation. This split ensured that the models were trained on a diverse set of data while also being rigorously tested on unseen data to evaluate their performance. By integrating data from previous research, counselor-provided sessions, and real-world student interactions, we ensured that our dataset was both comprehensive and representative. This multi-source approach allowed us to build a robust emotion recognition system capable of accurately identifying emotions in real-time during mobile video calls.

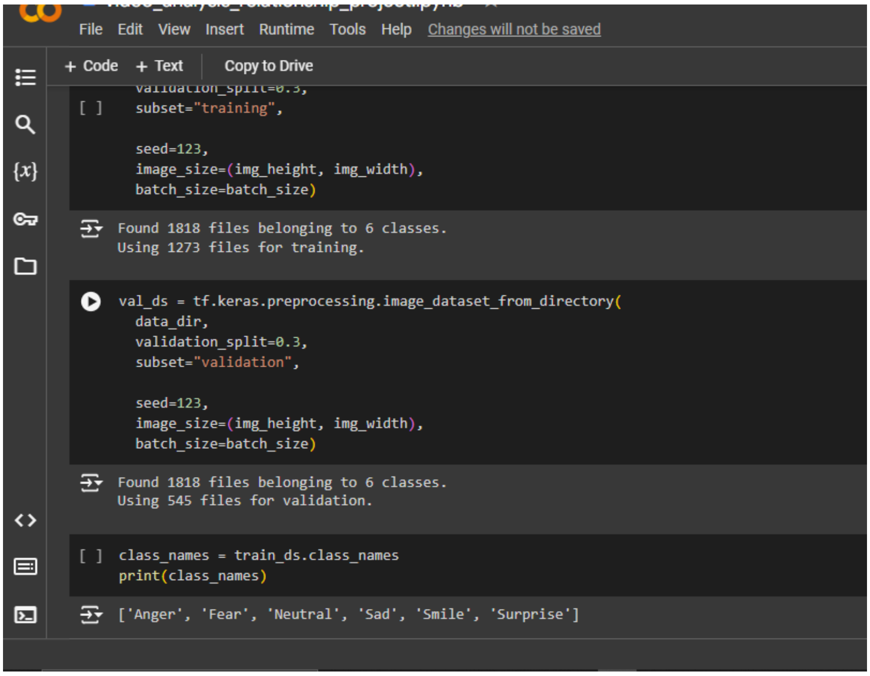
4.2 Feasibility Study

The feasibility study for the emotion recognition component evaluated technical, operational, and economic aspects. Technically, leveraging AWS for data processing and TensorFlow with Keras for model development ensured robust infrastructure and advanced capabilities. Operationally, collaboration with counselors and collecting real-world data from students proved viable, providing diverse emotional expressions for training. Economically, using existing cloud resources and open-source frameworks minimized costs. The project is feasible, given the solid technical foundation, available data sources, and cost-effective implementation strategy.

4.3 Preprocessing and data augmentation

For the emotion recognition system, data preprocessing and augmentation are essential to improve model performance(figure 4.3.1). Initially, video call feeds were decomposed into individual frames, producing a dataset of 1,818 images representing six emotion classes: Anger, Fear, Neutral, Sad, Smile, and Surprise. Each image was resized to 224x224 pixels to match the input requirements of the ResNet50 model. Data augmentation techniques, such as random rotations, flips, and brightness adjustments, were applied using TensorFlow's image preprocessing functions to enhance variability and prevent overfitting. These techniques ensured the model could generalize well across different lighting conditions, facial orientations, and expressions, improving its robustness and accuracy during real-time emotion recognition tasks. This comprehensive preprocessing pipeline was crucial for training an effective and reliable emotion recognition model.

Figure 4.3.1: Data preprocessing code snippet



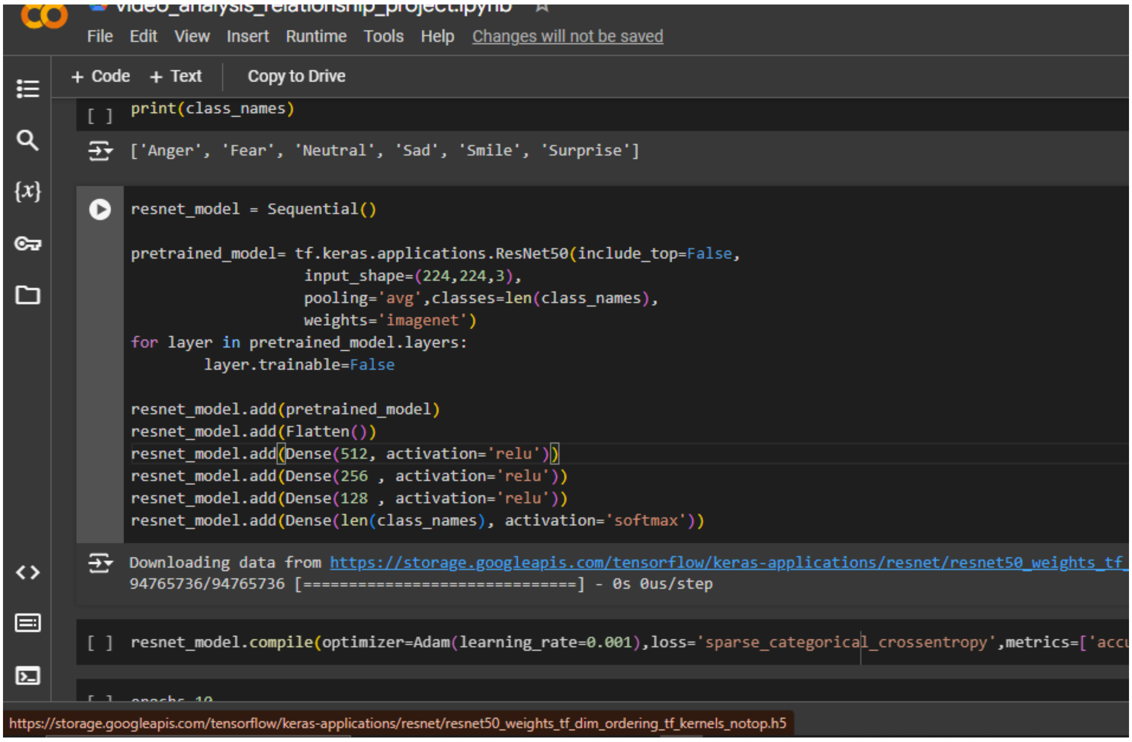
4.4 Identification and classification of emotion.

Identifying and classifying emotions using machine learning, particularly deep learning models, is a structured process involving several key steps, data collection, preprocessing, model selection and training, and evaluation. The comprehensive breakdown of each stage in this complex but fascinating endeavor.

Data Collection and Preparation, the initial step involves gathering a substantial dataset of images, each annotated with the corresponding emotional label. For instance, you might have images categorized into emotions such as Anger, Fear, Neutral, Sad, Smile, and Surprise. The quality and size of the dataset are crucial, as they directly impact the model's performance. A typical dataset might contain thousands of images to ensure diversity and robustness. Once collected, these images need to be resized to a uniform dimension, such as 224x224 pixels, to match the input size required by most convolutional neural networks (CNNs). Additionally, data augmentation techniques are applied to the dataset to increase its variability and enhance the model’s ability to generalize. Augmentation techniques can include random rotations, horizontal and vertical flips, and brightness adjustments, creating variations of the images that the model can learn from. Data preprocessing is a critical step where the raw data is transformed into a suitable format for the model. This step often includes normalizing the pixel values to a range of 0 to 1 or -1 to 1, which helps in speeding up the training process and improving the convergence of the model. Splitting the dataset into training and validation sets is also a crucial part of this phase. Typically, 70% of the data is used for training, and the remaining 30% is used for validation to assess the model's performance during training.

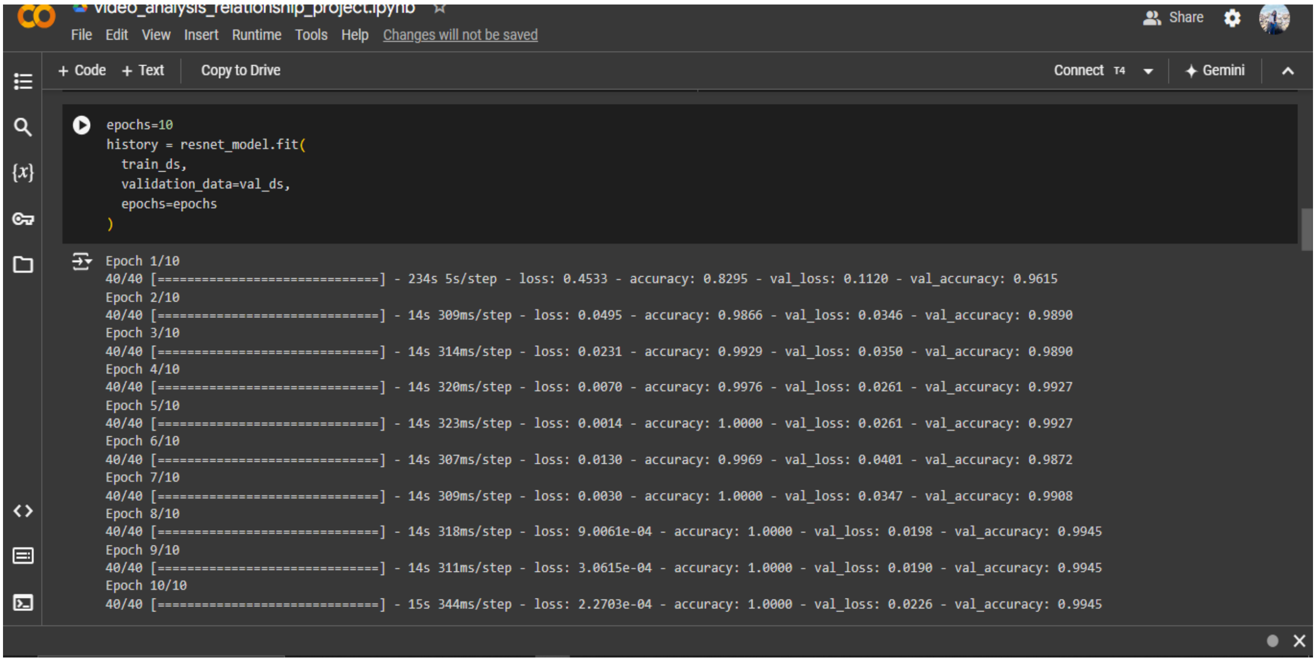
For emotion classification, a pre-trained CNN such as ResNet50 is often employed. ResNet50, known for its deep architecture and ability to handle the vanishing gradient problem, serves as an excellent feature extractor. The process involves using the ResNet50 model pre-trained on a large dataset like ImageNet, excluding its top layer, and adding custom layers to adapt it for emotion classification. Flatten Layer converts the 3D output of the convolutional layers into a 1D feature vector. Dense Layers fully connected layers with ReLU activation functions(figure 4.4.1), introducing non-linearity to the model. These layers might have units in decreasing order, such as 512, 256, and 128. Output Layer, a dense layer with a SoftMax activation function those outputs probabilities for each emotion class. Once the architecture is defined, the model is compiled using an optimizer like Adam, which adapts the learning rate during training for efficient learning.

Figure 4.4.1: Dense layers with ReLU activation code snippet



The loss function typically used is sparse categorical cross-entropy, ideal for multi-class classification problems. Training involves feeding the model with the training dataset and adjusting the weights through backpropagation over several epochs, often around 10 to 20, depending on the dataset size and complexity(figure4.4.2).

Figure 4.4.2:Training of the resNet model



During training, the model's performance is monitored using metrics like accuracy and loss on both training and validation datasets. This helps in identifying overfitting or underfitting issues. If the validation performance stagnates or worsens while the training performance improves, it might indicate overfitting. After training, the model's performance is evaluated on a separate test set or through cross-validation. The evaluation includes generating confusion matrices, which provide a detailed breakdown of the model's predictions versus the actual labels, normalized to percentage values. This helps in identifying which emotions the model predicts accurately and which ones it confuses. Additionally, plotting accuracy and loss curves over epochs gives insights into the model's learning process. These plots can highlight whether the model has learned effectively or if further tuning is required.

Once the model is trained and evaluated, it can be deployed for real-time emotion recognition. In practical applications, such as a mobile app for emotion recognition during video calls, the model is integrated into a system where it processes live video frames, predicts emotions, and displays them in real-time. For deployment, the model is often saved in a standard format like TensorFlow's Saved Model or Keras's H5. The deployment environment, such as an AWS server, hosts the model and the application, ensuring that the system can handle real-time data efficiently.

The identification and classification of emotions using deep learning involve a systematic approach from data collection and preprocessing to model training, evaluation, and deployment. Each step is crucial in ensuring the model's accuracy and reliability in recognizing emotions. By leveraging advanced CNN architectures like ResNet50 and employing robust data augmentation and preprocessing techniques, the developed model can effectively classify emotions in real-time applications, enhancing human-computer interactions and providing valuable insights into emotional dynamics. This process not only demonstrates the power of deep learning in understanding human emotions but also showcases its potential in various practical applications.

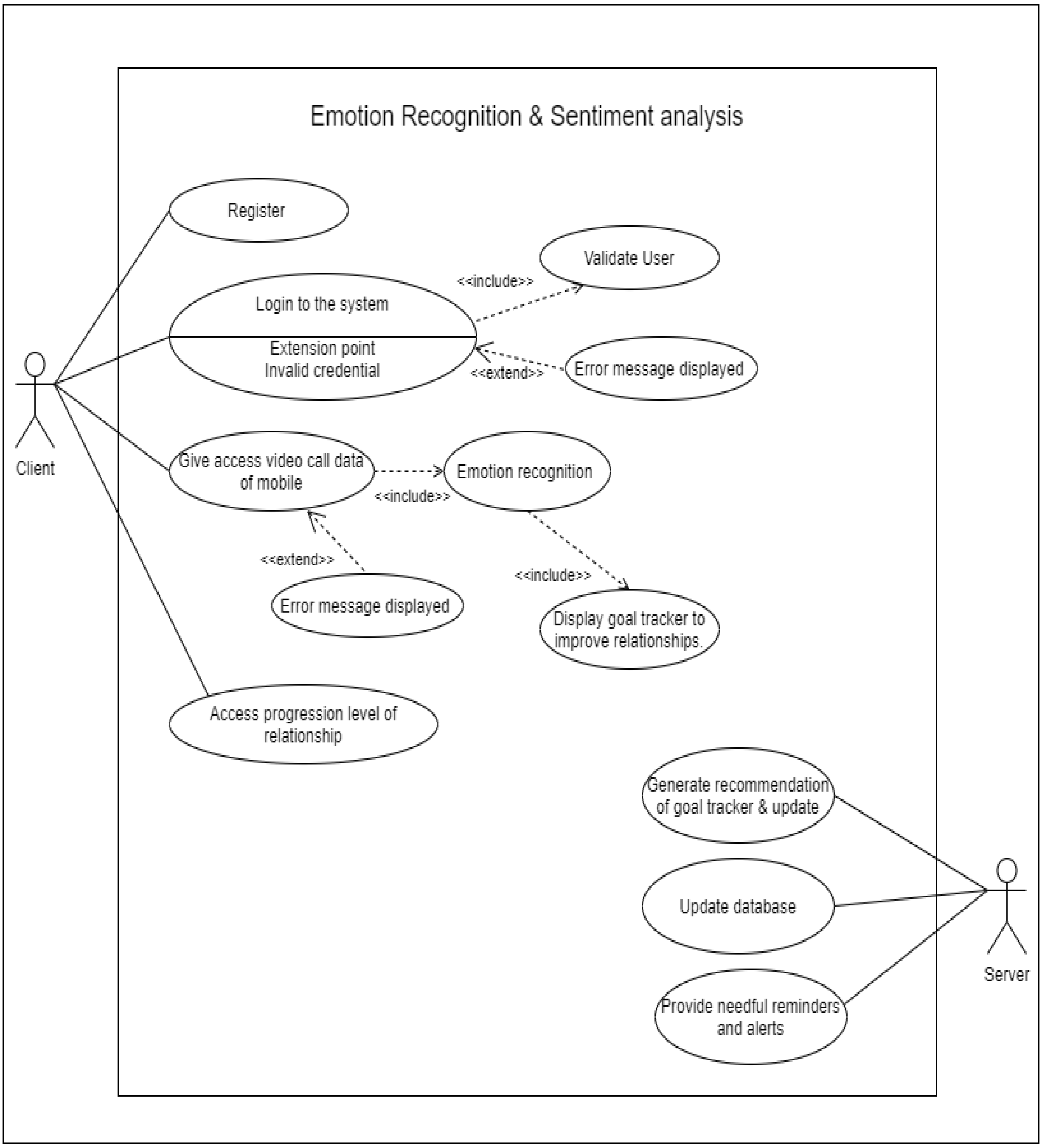
Table 4.4.1: Technologies, techniques, architectures, and algorithms used.

Technologies	TensorFlow, Keras, OpenCV, Amazon Web Services (AWS), Flask, Node.js , React Native, Expo
Techniques	Transfer Learning Image Processing Real-Time Emotion Recognition Multimodal Integration Data Augmentation Training and Validation Split (70% training, 30% validation)
Architectures	ResNet50 for image-based emotion recognition Convolutional Neural Networks (CNN) Pretrained model fine-tuning
Algorithms	Adam Optimizer (learning rate: 0.001) Sparse Categorical Cross-Entropy Loss Function Random Forest Classifier for emotional severity AI-enhanced audio-based sentiment recognition

4.5 Design Diagrams

System design diagram was created to identify the essential components and organize implementations related to developing the model. Fig 4.21 illustrates use case diagram helped to develop this research component.

Figure 4.1: Use case diagram



4.6 Functional Requirement

1. Real-time Emotion Detection: The system should accurately detect and analyze facial expressions and emotional cues in real-time during video calls.

2. Multi-Emotion Recognition: The component should be able to recognize and categorize a range of emotions, including happiness, sadness, anger, surprise, and more.

3. Personalized Goal Setting: Users should be able to set personalized relationship goals and milestones, with the system providing tailored recommendations based on emotional insights.

4. Integration with Video Call Platforms: The component needs seamless integration with popular video call platforms, allowing users to access emotional analysis during their calls.

5. Progress Tracking and Reporting: Users should be able to track their emotional progress over time, receiving regular reports and insights to gauge relationship improvement.

4.7 Non-Functional Requirement

1. User Privacy and Data Security: The system must ensure that all collected user data is securely stored, encrypted, and accessible only to authorized personnel, complying with privacy regulations.

2. Performance - The system is expected to demonstrate efficient performance, delivering rapid and precise results to users.

3. Availability - The application's accessibility should extend to users worldwide, accommodating various languages and ensuring uninterrupted availability as required.

4. Scalability and Concurrent Usage: The solution should be able to handle a growing number of simultaneous users without degradation in performance, ensuring a smooth user experience during peak usage times.

4.8 System Requirements

Software requirements delineate the functionalities and resources necessary for a software system to meet user needs effectively. They bridge user expectations and technical implementation, guiding development by defining features, performance, and interactions. By minimizing misunderstandings and aligning stakeholders, requirements ensure the creation of user-centered, reliable software solutions that fulfill business objectives.

- Video Call Data Collection: Obtain video call feed data from the Meta company.

- Emotion Recognition: Convolutional Neural Networks (CNN) for analyzing facial expressions and recognizing emotions.

- Deep Learning Framework: Keras, for a high-level deep learning framework, for training and deploying emotion recognition models.

- Image Processing: OpenCV for real-time face detection and feature extraction from video call data.

- Cross-Platform App Development: Develop a user-friendly application using React Native and Expo, ensuring compatibility with both Android and iOS.

- Backend Server: Use Flask to host and manage the deployed deep learning models for real-time emotion recognition.

- Integration: Establish a Node Server to connect the mobile and web applications with the Flask Server, facilitating real-time communication.

- Crowdsourcing: Implement a mechanism to engage users and gather feedback on the accuracy of emotion recognition results.

4.9 Commercialization aspects of the product

Commercialization Strategy for the Emotion Identification and Classification App

Target Audience

- Couples and Romantic Partners:
  The app can help couples understand each other's emotions better, fostering improved communication and relationship satisfaction.

- Families
  Family members can use the app to enhance emotional intelligence, leading to better family dynamics and conflict resolution.

- Individuals Seeking Relationship Improvement
  Individuals looking to improve their emotional skills and relationship dynamics can benefit from the insights provided by the app.

- Therapists and Counselors:
  Professionals can use the app as a tool to better understand their clients' emotional states, aiding in therapy and counseling sessions.

Market Space

- Age Limit:
  The app is designed to be user-friendly and accessible to users of all ages. It can be adapted for children, adolescents, adults, and the elderly, making it versatile across different age groups.

- Technological Proficiency:
  The app does not require advanced technological knowledge. Its intuitive design ensures that users with basic smartphone or computer skills can navigate and utilize the app effectively.

Commercialization Strategy

- Freemium Model
  - Basic Version: Offer a free version with essential features to attract a broad user base. This version can include basic emotion identification and simple reports.
  - Premium Version: Introduce a subscription-based model for advanced features such as detailed emotional analysis, personalized advice, integration with other apps, and additional resources for relationship improvement.

- Community Engagement
  - User Forums: Create an online community where users can share experiences, tips, and support each other.
  - Regular Updates: Engage with users through regular updates, including new features and improvements based on user feedback.
  - Social Media Presence: Maintain active social media profiles to interact with users, share success stories, and provide valuable content related to emotional intelligence and relationship advice.

- Marketing and Outreach:
  - Partnerships: Collaborate with relationship counselors, therapy centers, and educational institutions to promote the app.
  - Influencer Marketing: Partner with influencers in the relationship and mental health space to reach a wider audience.
  - Content Marketing: Develop a blog and video content that educates potential users about the benefits of the app and emotional intelligence.

- Customer Support and Engagement:
  - Help Center: Provide a comprehensive help center with FAQs, user guides, and video tutorials.

5. Implementation and Testing

5.1 Data preprocessing, augmentation & model implementation

Data preprocessing is crucial for preparing the dataset to be fed into the deep learning models. First, images are collected and resized to a standard size of 224x224 pixels. This resizing is necessary to ensure uniformity across the dataset, as models like VGGNet and ResNet require inputs of a fixed size. Next, pixel values are normalized to fall within the range of 0 to 1, which helps in stabilizing and speeding up the training process. Data augmentation techniques, such as random rotations, flips, zooms, and shifts, are applied to artificially expand the dataset, which helps in making the model more robust to variations in the input data. These augmentations are performed using libraries like TensorFlow's ImageDataGenerator, which allows for real-time augmentation during training.

Two versions of the VGGNet architecture were implemented to compare their performance on the emotion recognition task. The VGGNet19 model (figure 5.1.1) was trained and achieved 71.29% training accuracy and 58.01% validation accuracy. This architecture involves 19 layers, including convolutional layers with small receptive fields (3x3), followed by max-pooling layers, and fully connected layers at the end. Similarly, the VGGNet16 model(figure 5.1.2), which consists of 16 layers, was trained and achieved better results with 73.61% training accuracy and 69.21% validation accuracy. The key difference between the two models is the number of convolutional layers, which impacts the depth and learning capacity of the network. Both models utilized ReLU activation functions and dropout layers to prevent overfitting, with training conducted using the Adam optimizer and categorical cross-entropy loss function.

Figure 5.1.1:Training of the VGGNet19 model

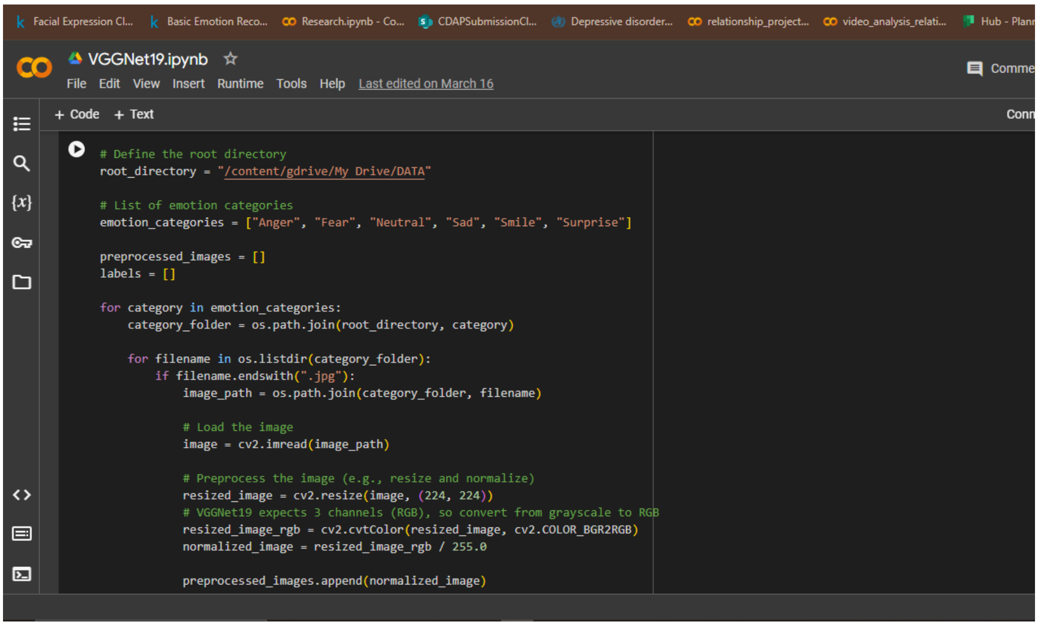
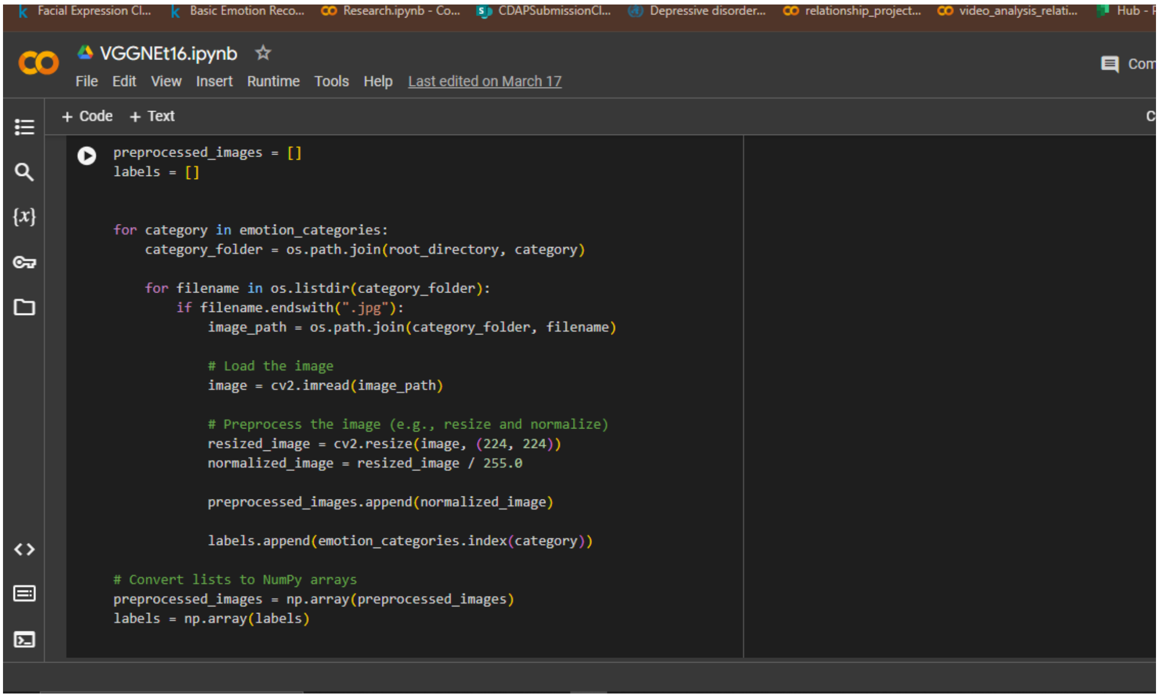


Figure 5.1.2: Training of the VGGNet16 model



The ResNet50 architecture, known for its residual learning framework, was implemented and outperformed the VGGNet models significantly. ResNet50 achieved a perfect training accuracy of 100% and a validation accuracy of 99.45%. This model consists of 50 layers, including convolutional layers, batch normalization, and identity shortcuts that help in mitigating the vanishing gradient problem. The architecture's strength lies in its ability to learn deeper representations without degrading performance, thanks to the residual connections that allow gradients to flow more effectively through the network. During training, the model also employed techniques like batch normalization and dropout to enhance generalization. The Adam optimizer and categorical cross-entropy loss were used for training, similar to the VGGNet models, but the addition of residual connections provided a significant performance boost, demonstrating the effectiveness of deeper architectures for emotion recognition tasks.

5.2 Test plan, Test strategy

Test planning serves as a foundational blueprint for ensuring software effectiveness. It outlines the necessary tasks, scope, and objectives to track project progress. A well-defined test strategy guides the testing process by specifying the steps and procedures required to ensure that all critical elements and functions are tested, considering the risks they may pose to users.

Steps and Procedures in Test Strategy:

- Define the Items to be Tested:
  Identify specific software components or features to be tested.

- Select Functions Based on Importance and Risk:
  Prioritize functions to test based on their significance and potential risk to users.

- Design Test Cases:
  Create detailed test cases according to use case descriptions to cover all identified scenarios.

- Execute Test Cases:
  Run the designed test cases on the software to check for proper functionality.

- Record Results:
  Document the outcomes of each test case execution, noting any deviations from expected behavior.

- Identify Bugs:
  Analyze test results to pinpoint any defects or issues within the software.

- Correct Bugs:
  Implement fixes for the identified bugs to ensure the software operates correctly.

- Repeat Testing:
  Retest the software after bug fixes to confirm that the issues have been resolved and the software meets expected results.

By following these steps, the test strategy ensures a systematic approach to software testing, improving overall quality and reliability.

5.3 Test Case Design

The following test cases were designed to ensure system reliability by testing all system functionalities.

Table 5.3.1: Test case 1

Test Case Id	01
Test Case	Verify video upload
Test Scenario	Verify whether the captured video is stored AWS cloud storage
Input	Captured users' small video
Expected Output	The video must be stored in the AWS cloud storage.
Actual Result	The video was stored in the AWS cloud storage.
Status (Pass/Fail)	Pass

Table 5.3.1: Test case 2

Test Case Id	02
Test Case	Identify emotion using CNN architecture.
Test Scenario	Testing images to identify emotion and select the best model.
Precondition	1273 labeled training & 545 testing images
Input	Test images
Expected Output	High accuracy.
Actual Result	High model accuracy with 99.45%
Status (Pass/Fail)	Pass

Table 5.3.1: Test case 3

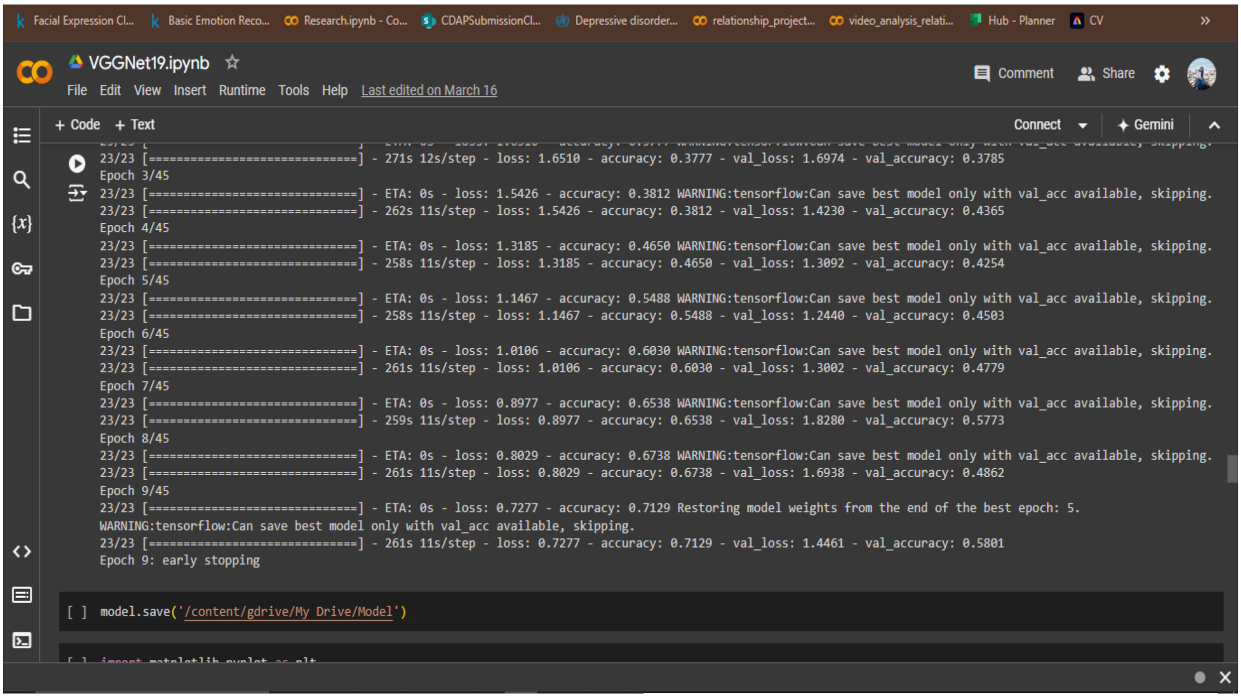
<table>
<tr><td>Test Case Id</td><td colspan="4">03</td></tr>
<tr><td>Description</td><td colspan="4">1818 images were tested to identify emotions and select the best architecture based on test accuracies.</td></tr>
<tr><td>Input</td><td colspan="4">1818 images.</td></tr>
<tr><td>Expected Output</td><td colspan="4">Expected 90% higher accuracy</td></tr>
<tr><td rowspan="4">Result</td><td>Architecture</td><td>Accuracy</td><td>Correct image count</td><td>Wrong image count</td></tr>
<tr><td>ResNet50</td><td>99.45%</td><td>1273</td><td>545</td></tr>
<tr><td>VGG16</td><td>69.21%</td><td>865</td><td>345</td></tr>
<tr><td>VGG19</td><td>58.01%</td><td>990</td><td>276</td></tr>
</table>

6. RESULTS AND DISCUSSION

6.1 Result

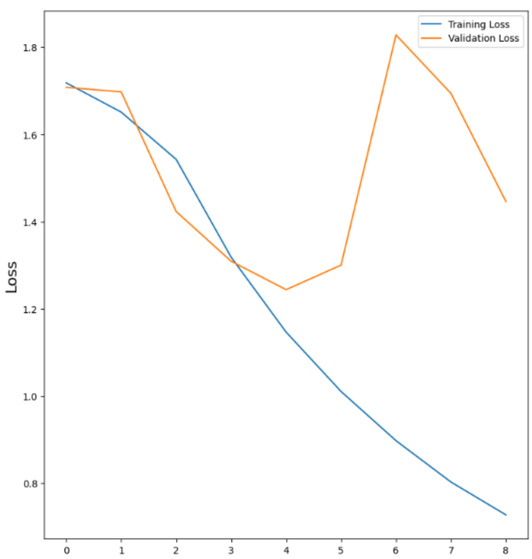
The performance of emotion recognition models based on various architectures revealed significant differences in their ability to accurately classify emotions. The VGGNet19 architecture, while showing promise, achieved a training accuracy of 71.29% and a validation accuracy of 58.01%. This discrepancy between training and validation accuracies indicates that the model might be overfitting, as it performs considerably better on the training data than on unseen data.

Figure 6.1.1:Accuracy of the VGGNet19 model



Optimizer: ,

Figure 6.1.2: VGGNet19 model loss



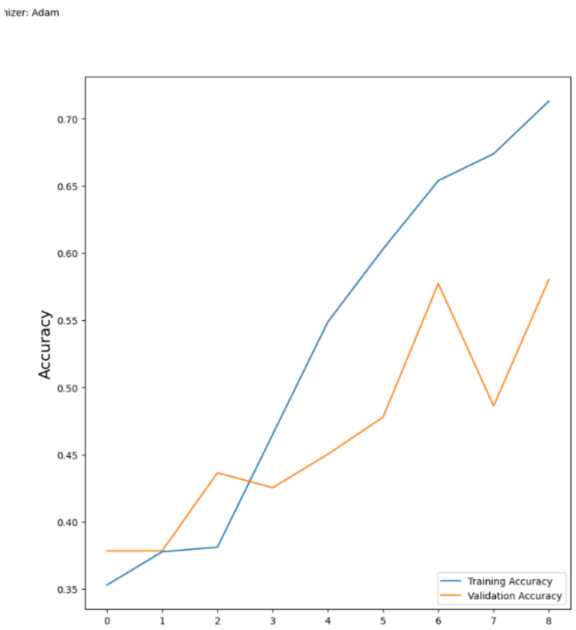
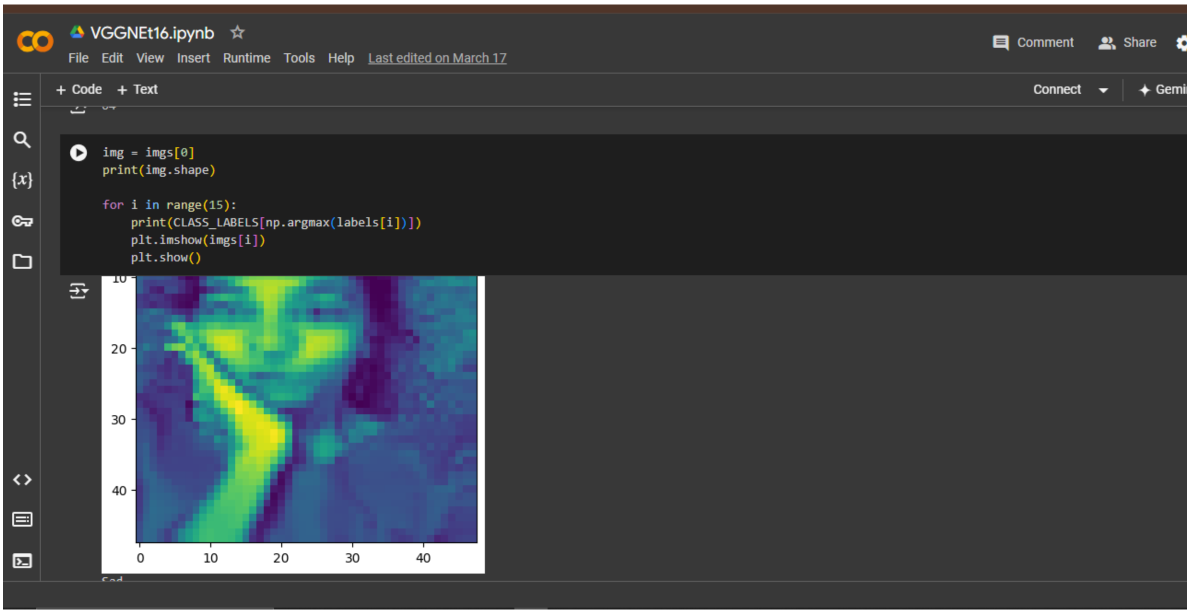


Figure 6.1.3: VGGNet19 model accuracy

Figure 6.1.4: VGGNet16 model performed




```
Accuracy: 0.6989247311827957

Classification Report:
              precision    recall  f1-score   support

           0       0.81      0.96      0.88        27
           1       1.00      0.08      0.15        12
           2       0.50      0.55      0.52        11
           3       0.73      0.97      0.83        31
           4       0.29      0.17      0.21        12

    accuracy                           0.70        93
   macro avg       0.67      0.55      0.52        93
weighted avg       0.70      0.70      0.64        93
```

The VGGNet16 model performed slightly better, with a training accuracy of 73.61% and a validation accuracy of 69.21%. The closer alignment between the training and validation accuracies suggests a more generalized model compared to VGGNet19. However, while these results indicate moderate success, there is still a significant gap to optimal performance.

In contrast, the ResNet50 model achieved outstanding results, with a perfect training accuracy of 100% and a near-perfect validation accuracy of 99.45%. These results highlight the robustness and efficacy of the ResNet50 architecture in emotion recognition tasks. The high validation accuracy indicates that the model generalizes exceptionally well to new data, minimizing both bias and variance.

Figure 6.1.5: ResNet50 model accuracy

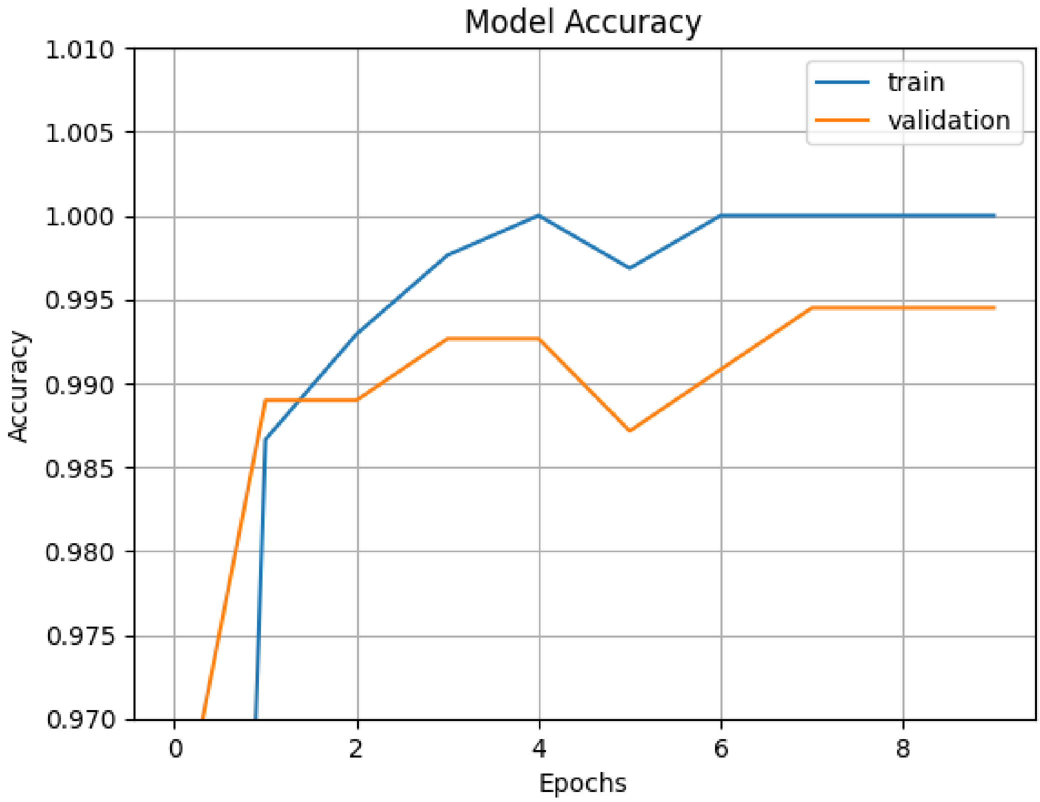


Figure 6.1.6: ResNet50 model loss

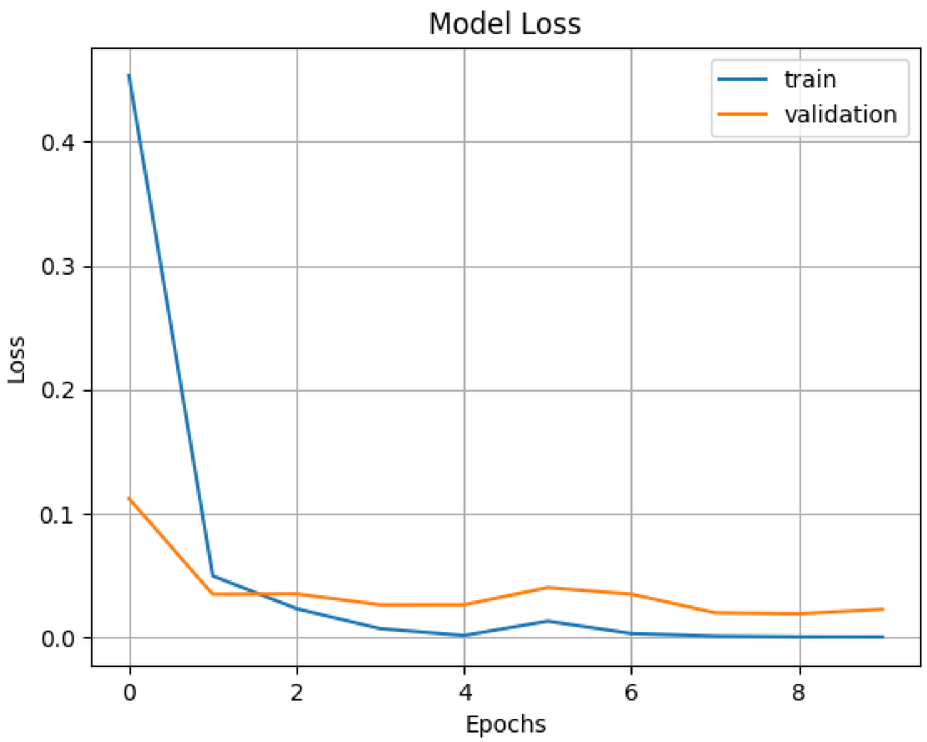
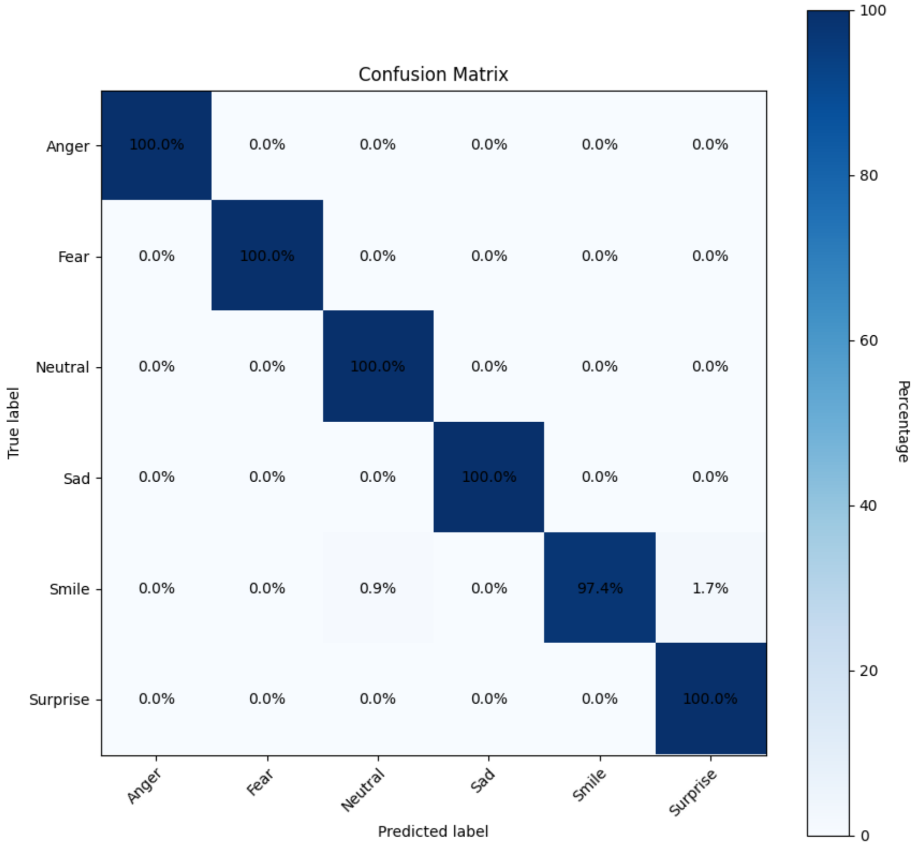


Figure 6.1.7: ResNet50 model conclusion matrix



6.2 Discussion

The comparative results from the different architectures underscore the importance of model selection in emotion recognition tasks. The VGGNet19 and VGGNet16 models, although reasonably effective, were outperformed by the ResNet50 model, which demonstrated superior learning capabilities and generalization.

The VGGNet19 model's lower validation accuracy compared to its training accuracy suggests overfitting. This issue is less pronounced in the VGGNet16 model, which shows a more balanced performance across training and validation sets. Overfitting occurs when a model learns not just the underlying patterns but also the noise in the training data, leading to poor performance on new, unseen data. Regularization techniques, such as dropout and weight decay, could potentially mitigate this issue in VGGNet models, although they were not sufficient in this case. The ResNet50 model's performance is a testament to the power of deeper networks with advanced architectures. ResNet50 uses residual learning, which helps in training very deep networks by addressing the vanishing gradient problem. This allows the model to learn more complex features without degradation in performance, which is evident in its near-perfect validation accuracy. The ability to maintain a high level of accuracy on unseen data is crucial for practical applications of emotion recognition, where the model will encounter a wide variety of faces and emotional expressions. The complexity of the ResNet50 architecture, with its 50 layers, allows it to capture intricate details in the data, leading to its superior performance.

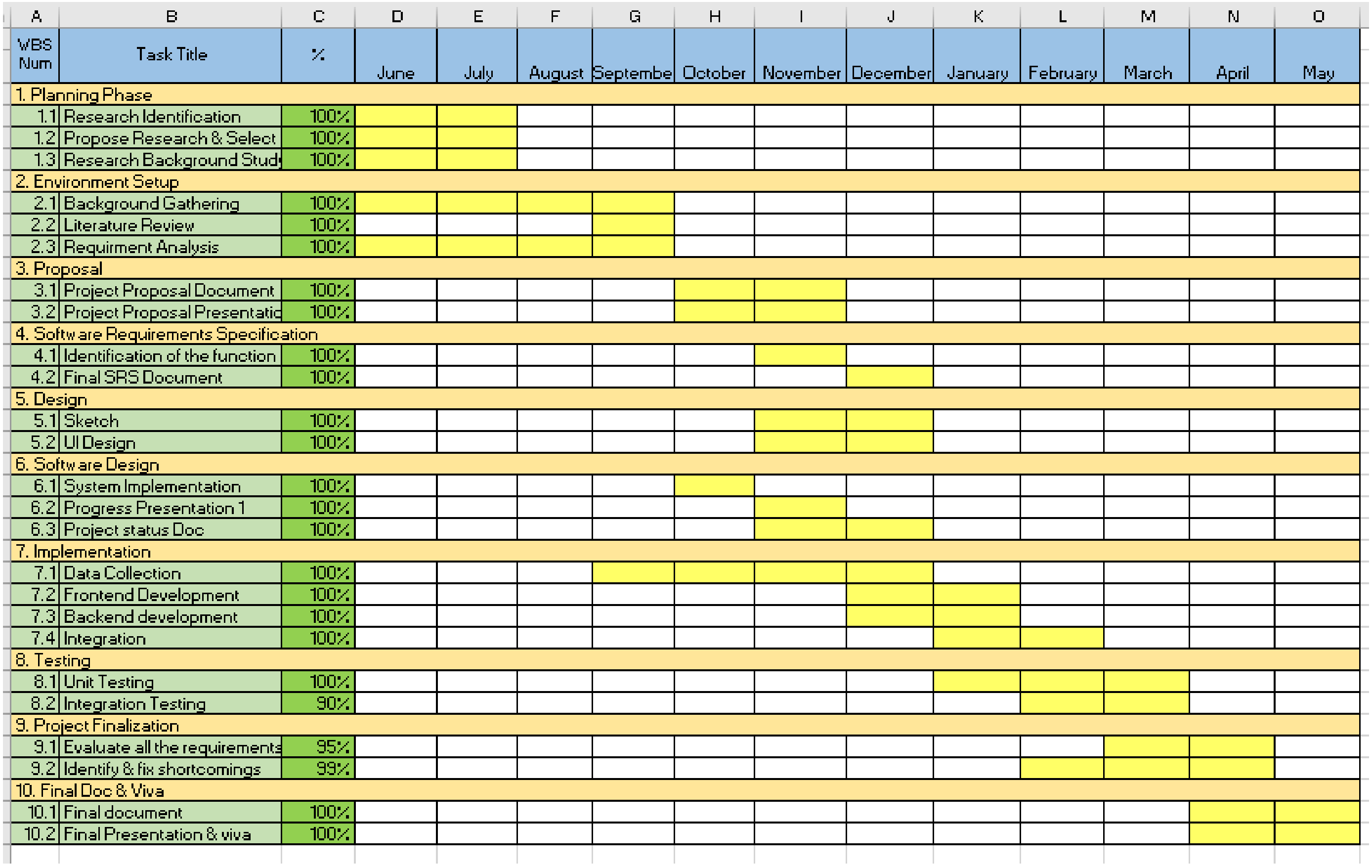
However, this complexity comes with increased computational requirements. Training such a deep network requires more powerful hardware and longer training times. Despite these requirements, the benefits of higher accuracy and better generalization make ResNet50 a preferable choice for emotion recognition tasks. The excellent performance of the ResNet50 model makes it highly suitable for real-world applications, such as automated emotion recognition in therapy sessions, relationship counseling, and mental health monitoring. Its ability to accurately identify emotions can enhance user interactions in these fields by providing precise feedback and analysis. For instance, therapists and counselors can rely on such a model to gain insights into their clients' emotional states, thereby tailoring their approaches more effectively.

While the ResNet50 model demonstrates impressive results, there is always room for improvement. Future work could explore ensemble methods, combining the strengths of multiple models to further enhance accuracy and robustness. Additionally, ongoing advancements in deep learning architectures, such as the introduction of even deeper networks or more efficient training algorithms, could provide further gains. Another area of potential improvement is the dataset itself. Emotion recognition models are highly dependent on the quality and diversity of the training data. Expanding the dataset to include a wider range of expressions, lighting conditions, and demographic variations could help improve the model's robustness and applicability across different contexts.

The results clearly show that the ResNet50 model outperforms VGGNet19 and VGGNet16 in emotion recognition tasks, with nearly perfect validation accuracy indicating strong generalization capabilities. This superior performance underscores the importance of selecting appropriate architectures for complex tasks like emotion recognition. Future efforts should focus on further improving model performance and expanding the dataset to ensure the model's applicability in diverse real-world scenarios.

7. GANT CHART

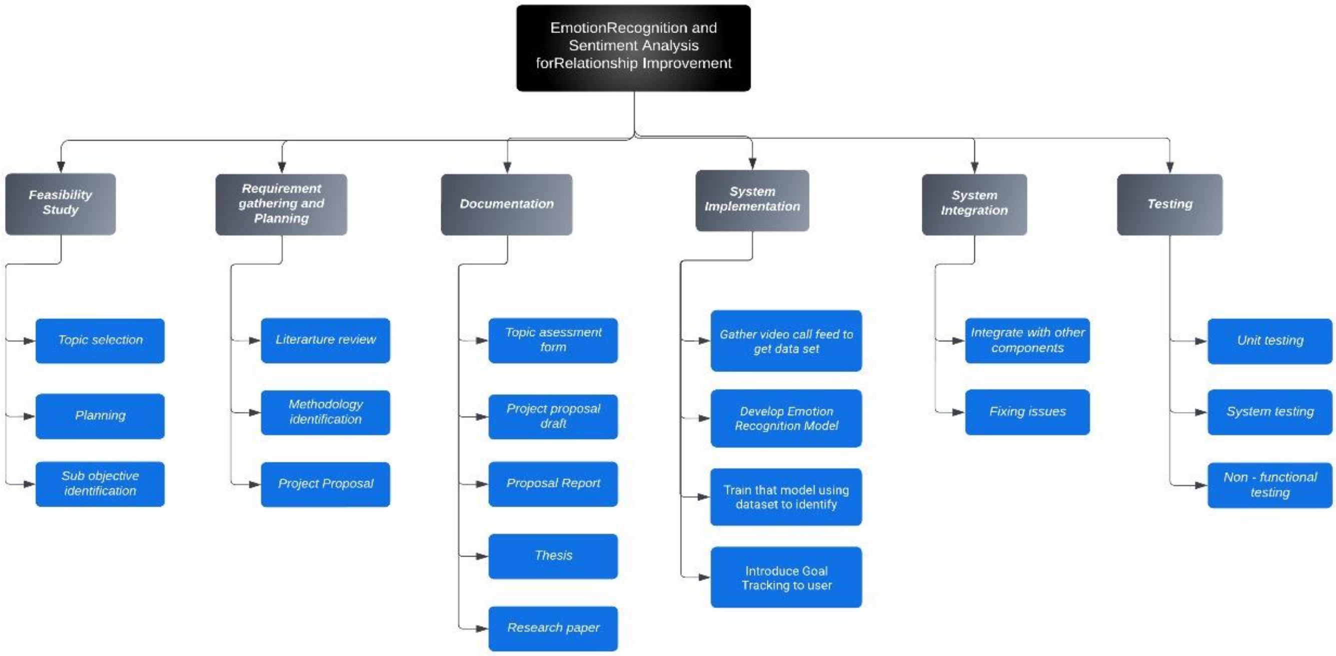
Figure 7.1: Gant Chart



A	B	C	D	E	F	G	H	I	J	K	L	M	N	O
VBS Num	Task Title	%	June	July	August	Septembe	October	November	December	January	February	March	April	May
1. Planning Phase														
1.1	Research Identification	100%												
1.2	Propose Research & Select	100%												
1.3	Research Background Stud	100%												
2. Environment Setup														
2.1	Background Gathering	100%												
2.2	Literature Review	100%												
2.3	Requirment Analysis	100%												
3. Proposal														
3.1	Project Proposal Document	100%												
3.2	Project Proposal Presentatio	100%												
4. Software Requirements Specification														
4.1	Identification of the function	100%												
4.2	Final SRS Document	100%												
5. Design														
5.1	Sketch	100%												
5.2	UI Design	100%												
6. Software Design														
6.1	System Implementation	100%												
6.2	Progress Presentation 1	100%												
6.3	Project status Doc	100%												
7. Implementation														
7.1	Data Collection	100%												
7.2	Frontend Development	100%												
7.3	Backend development	100%												
7.4	Integration	100%												
8. Testing														
8.1	Unit Testing	100%												
8.2	Integration Testing	90%												
9. Project Finalization														
9.1	Evaluate all the requirements	95%												
9.2	Identify & fix shortcomings	99%												
10. Final Doc & Viva														
10.1	Final document	100%												
10.2	Final Presentation & viva	100%												

8. WORK BREAK DOWN CHART

Figure 8.1: Work Break Down Chart



9. CONCLUSION

The development of emotion recognition using deep learning architectures, particularly focusing on models like VGGNet19, VGGNet16, and ResNet50, has yielded notable results and insights. This project highlights the performance disparities among these models, with ResNet50 significantly outperforming the others, achieving 100% training accuracy and 99.45% validation accuracy. This superior performance is attributed to ResNet50’s advanced architecture and residual learning, which effectively address issues like overfitting and the vanishing gradient problem. In contrast, VGGNet19 and VGGNet16, though robust, demonstrated overfitting and lower generalization capabilities, with training accuracies of 71.29% and 73.61% and validation accuracies of 58.01% and 69.21%, respectively.

The exceptional performance of ResNet50 underscores its potential for practical applications in fields such as therapy, relationship counseling, and mental health monitoring. This model’s ability to accurately classify emotions can significantly enhance user interactions, providing valuable insights for therapists and counselors to tailor their approaches more effectively. However, the complexity of ResNet50 also necessitates greater computational resources, highlighting the trade-off between model complexity and computational efficiency. For real-world deployment, optimizing this balance through techniques like model pruning, quantization, and hardware acceleration is crucial, especially in resource-constrained environments.

Future directions for enhancing emotion recognition systems include exploring ensemble methods, expanding datasets to improve diversity and robustness, and leveraging ongoing advancements in deep learning. Ensemble methods can combine the strengths of multiple models, leading to superior performance, while expanded datasets can enhance the model’s applicability across different contexts and populations. Additionally, incorporating advanced regularization techniques, such as dropout and weight decay, can help mitigate overfitting and improve generalization.

In conclusion, this project demonstrates the critical role of model architecture in the performance of emotion recognition systems. The ResNet50 model’s near-perfect accuracy highlights its potential for practical applications, while the challenges faced by VGGNet19 and VGGNet16 emphasize the need for careful model selection and regularization. Moving forward, enhancing data diversity, leveraging ensemble methods, and exploring advanced techniques will be key to further improving emotion recognition capabilities. The findings from this project pave the way for more effective and robust emotion recognition systems, with significant implications for fields such as therapy, counseling, and mental health. Through continued research and innovation, these systems can become powerful tools for understanding and responding to human emotions in diverse real-world scenarios.

10. REFERENCES

[1] M. S. Bartlett, G. Littlewort, M. Frank, C. Lainscsek, I. Fasel and J. Movellan, "Recognizing facial expression: machine learning and application to spontaneous behavior," 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05), San Diego, CA, USA, 2005, pp. 568-573 vol. 2, doi: 10.1109/CVPR.2005.297.

[2] LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. Nature, 521(7553), 436-444

[3] P. Lucey, J. F. Cohn, T. Kanade, J. Saragih, Z. Ambadar and I. Matthews, "The Extended Cohn-Kanade Dataset (CK+): A complete dataset for action unit and emotion-specified expression," *2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition - Workshops*, San Francisco, CA, USA, 2010, pp. 94-101, doi: 10.1109/CVPRW.2010.5543262.

[4] P. Viola and M. Jones, "Rapid object detection using a boosted cascade of simple features," *Proceedings of the 2001 IEEE Computer Society Conference on Computer Vision and Pattern Recognition. CVPR 2001*, Kauai, HI, USA, 2001, pp. I-I, doi: 10.1109/CVPR.2001.990517.

[5] K. He, X. Zhang, S. Ren and J. Sun, "Deep Residual Learning for Image Recognition," *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Las Vegas, NV, USA, 2016, pp. 770-778, doi: 10.1109/CVPR.2016.90.

[6] R. Cowie *et al*., "Emotion recognition in human-computer interaction," in *IEEE Signal Processing Magazine*, vol. 18, no. 1, pp. 32-80, Jan 2001, doi: 10.1109/79.911197.

[7] Ekman, P., & Friesen, W. V. (1971). Constants across cultures in the face and emotion. *Journal of Personality and Social Psychology, 17*(2), 124–129.

[8] Russell, J. A. (1980). A circumplex model of affect. *Journal of Personality and Social Psychology, 39*(6), 1161–1178.

[9] K. He, X. Zhang, S. Ren and J. Sun, "Deep Residual Learning for Image Recognition," *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Las Vegas, NV, USA, 2016, pp. 770-778, doi: 10.1109/CVPR.2016.90.

[10] Hao, Man & Cao, Weihua & Liu, Zhen-Tao & Wu, Min & Xiao, Peng. (2020). Visual-Audio Emotion Recognition Based on Multi-Task and Ensemble Learning with Multiple Features. Neurocomputing. 391. 10.1016/j.neucom.2020.01.048.

[11] M. Bouaziz, M. Morchid, R. Dufour and G. Linarès, "Improving multi-stream classification by mapping sequence-embedding in a high dimensional space," *2016 IEEE Spoken Language Technology Workshop (SLT)*, San Diego, CA, USA, 2016, pp. 224-231, doi: 10.1109/SLT.2016.7846269.

[12] R. GUETARI, A. CHETOUANI, H. TABIA and N. KHLIFA, "Real time emotion recognition in video stream, using B-CNN and F-CNN," *2020 5th International Conference on Advanced Technologies for Signal and Image Processing (ATSIP)*, Sousse, Tunisia, 2020, pp. 1-6, doi: 10.1109/ATSIP49331.2020.9231902.

[13] E. Pranav, S. Kamal, C. Satheesh Chandran and M. H. Supriya, "Facial Emotion Recognition Using Deep Convolutional Neural Network," *2020 6th International Conference on Advanced Computing and Communication Systems (ICACCS)*, Coimbatore, India, 2020, pp. 317-320, doi: 10.1109/ICACCS48705.2020.9074302.

[14] Jogin, Manjunath & ., Mohana & Madhulika, M & Divya, G & Meghana, R & Apoorva, S. (2018). Feature Extraction using Convolution Neural Networks (CNN) and Deep Learning. 2319-2323. 10.1109/RTEICT42901.2018.9012507.

[15] M. Soleymani, S. Asghari-Esfeden, Y. Fu and M. Pantic, "Analysis of EEG Signals and Facial Expressions for Continuous Emotion Detection," in *IEEE Transactions on Affective Computing*, vol. 7, no. 1, pp. 17-28, 1 Jan.-March 2016, doi: 10.1109/TAFFC.2015.2436926.

[16] A. Zadeh, R. Zellers, E. Pincus and L. -P. Morency, "Multimodal Sentiment Intensity Analysis in Videos: Facial Gestures and Verbal Messages," in *IEEE Intelligent Systems*, vol. 31, no. 6, pp. 82-88, Nov.-Dec. 2016, doi: 10.1109/MIS.2016.94.