MADHAV INSTITUTE OF TECHNOLOGY & SCIENCE GWALIOR

(A Govt. Aided UGC Autonomous Institute Affiliated to RGPV, Bhopal)

NAAC Accredited with A++ Grade

Project Report

on

REAL TIME FACIAL EMOTION RECOGNITION

Submitted By:

KRISHNA MALVIYA (0901AM211027)

SANSKAR HARDAHA (0901AM211050)

Faculty Mentor:

Dr. Bhagat Singh Raghuwanshi

Assistant Professor

CENTRE FOR ARTIFICIAL INTELLIGENCE

MADHAV INSTITUTE OF TECHNOLOGY & SCIENCE

GWALIOR - 474005 (MP) est. 1957

JULY-DEC. 2023

MADHAV INSTITUTE OF TECHNOLOGY & SCIENCE GWALIOR

(A Govt. Aided UGC Autonomous Institute Affiliated to RGPV, Bhopal)

NAAC Accredited with A++ Grade

CERTIFICATE

This is certified that **KRISHNA MALVIYA** (0901AM211027) **SANSKAR HARDAHA** (0901AM211050) has submitted the project report titled REAL TIME FACIAL EMOTION RECOGINITION under the mentorship of **Dr. Bhagat Singh Raghuwanshi**, in partial fulfilment of the requirement for the award of degree of Bachelor of Technology in **AIML** from Madhav Institute of Technology and Science, Gwalior.

23/11/23

Dr. Bhagat Singh Raghuwanshi
Faculty Mentor
Assistant Professor
Centre for Artificial Intelligence

28/11/23

Dr. R. R. Singh
Coordinator
Centre for Artificial Intelligence

MADHAV INSTITUTE OF TECHNOLOGY & SCIENCE GWALIOR

(A Govt. Aided UGC Autonomous Institute Affiliated to RGPV, Bhopal)

NAAC Accredited with A++ Grade

DECLARATION

I hereby declare that the work being presented in this project report, for the partial fulfilment of requirement for the award of the degree of Bachelor of Technology in **AIML** at Madhav Institute of Technology & Science, Gwalior is an authenticated and original record of my work under the mentorship of **Dr. Bhagat Singh Raghuwanshi,** Assistance Professor, Centre for Artificial Intelligence.

I declare that I have not submitted the matter embodied in this report for the award of any degree or diploma anywhere else.

KRISHNA MALVIYA (0901AM211027)
3rd Year 5th sem,
Centre for Artificial Intelligence

SANSKAR HARDAHA (0901AM211050)
3rd Year 5th sem,
Centre for Artificial Intelligence

MADHAV INSTITUTE OF TECHNOLOGY & SCIENCE GWALIOR
(A Govt. Aided UGC Autonomous Institute Affiliated to RGPV, Bhopal)
NAAC Accredited with A++ Grade

ACKNOWLEDGEMENT

The full semester project has proved to be pivotal to my career. I am thankful to my institute, **Madhav Institute of Technology and Science** to allow me to continue my disciplinary/interdisciplinary project as a curriculum requirement, under the provisions of the Flexible Curriculum Scheme (based on the AICTE Model Curriculum 2018), approved by the Academic Council of the institute. I extend my gratitude to the Director of the institute, **Dr. R. K. Pandit** and Dean Academics, **Dr. Manjaree Pandit** for this.

I would sincerely like to thank my department, **Centre for Artificial Intelligence,** for allowing me to explore this project. I humbly thank **Dr. R. R. Singh**, Coordinator, Centre for Artificial Intelligence, for his continued support during the course of this engagement, which eased the process and formalities involved.

I am sincerely thankful to my faculty mentors. I am grateful to the guidance of **Dr. Bhagat Singh Raghuwanshi**, assistance professor and **Dr. Anshika Shrivastava**, Assistant professor, Centre of Artificial Intelligence, for his continued support and guidance throughout the project. I am also very thankful to the faculty and staff of the department.

KRISHNA MALVIYA (0901AM211027)
3rd Year 5th sem,
Centre for Artificial Intelligence

SANSKAR HARDAHA (0901AM211050)
3rd Year 5th sem,
Centre for Artificial Intelligence

ABSTRACT

Abstract – An important aspect of human beings is their ability to display emotions. Emotions form the basis of intimate means of communications between one another. Recognition of emotions displayed via facial expression withthe help of a computer could prove to be a very powerful too. In recent years, an increased number of intelligent systems are using facial emotion recognition to improve human interaction. These systems cause constant changes in their operation based on the emotion of humans. In this paper, we propose an architecture based on the convolutional neural network (CNN) for the facial recognition of emotions. In the implementation, we use the Facial Expression Recognition 2013 dataset (FER- 2013). Through behavioural analysis, we show how differentemotions seem to be sensitive to different parts of the face.

Keyword: ***Haar Cascade Classifier, Convolutional Neural Network, Softmax Classification, Real-Time Detection***

Table of Contents

LIST OF FIGURES

Chapter 1: PROJECT OVERVIEW

1.1. Introduction

Recognizing facial emotions has become a major issue in many applications today. The research on facial emotion recognition has gained a lot of momentum over the past few years. The state of human emotions is identified using facial emotion recognition (e.g. neutral, happy, sad, surprise, fear, anger, disgust, contempt) based on the stream of images fed in by a video. There has been growing interest in making machines act as close as possible to actual human beings. To make their actions replicate those of humans and add a touch of human feelings in each of these actions. It has been argued that for there to be a proper human-computer interaction, the computer has to interact in a natural way, similar to that when two humans interact. Other fields where emotion recognition can prove to be useful are online teaching, product marketing, health industry, and several others to determining whether certain things are liked or not and what is the reaction of different people towards different stimuli. Humans express their thoughts and feelings in multiple ways. The most important being speech, followed up by their display of emotions. Emotions typically manifest themselves via multiple means, be it touch, visual or physiological. The most appropriate method of detecting human emotions would be to determine their emotion from the visual cues displayed by a human. It is widely accepted that the slightest of change in emotions become visible on the faces of humans and hence developing a system which detects the emotion from an individual's face could prove to be an invaluable tool.

The work done by us in this project enables proposes a system which allows for real-time emotion detection by using a video stream input from the user's webcam. We initially focus on detecting the face from the video using the Haar cascade classifiers. We then make use of convolutional neural networks for determining emotions. The dataset used by us in this project is the FER2013 dataset. The dataset originally was in the form of raw pixels but we had to convert them into actual images and use those images for training our model. This greatly helped in improving the performance of our model. The software used by us in this project is OpenCV and TensorFlow. OpenCV is used for operating the webcam and for face detection, while TensorFlow was used for training the CNN for emotion detection.

1.2. Motivation

In today's networked world the need to maintain security of information or physical property is becoming both increasingly important and increasingly difficult. In countries like Nepal the rate of crimes are increasing day by day. No automatic systems are there that can track person's activity. If we will be able to track Facial expressions of persons automatically then we can find the criminal easily since facial expressions changes doing different activities. So we decided to make a Facial Expression Recognition System.

We are interested in this project after we went through few papers in this area. The papers were published as per their system creation and way of creating the system for accurate and reliable facial expression recognition system.

As a result we are highly motivated to develop a system that recognizes facial expression and track one person's activity.

1.3. Problem Statement

Human emotions and intentions are expressed through facial expressions and deriving an efficient and effective feature is the fundamental component of facial expression system. Face recognition is important for the interpretation of facial expressions in applications such as intelligent, man-machine interface and communication, intelligent visual surveillance, teleconference and real-time animation from live motion images. The facial expressions are useful for efficient interaction Most research and system in facial expression recognition are limited to six basic expressions (joy, sad, anger, disgust, fear, surprise). It is found that it is insufficient to describe all facial expressions and these expressions are categorized based on facial actions. Detecting face and recognizing the facial expression is a very complicated task when it is a vital to pay attention to primary components like: face configuration, orientation, location where the face is set.

1.4. Objectives

- To develop a facial expression recognition system.
- To experiment machine learning algorithm in computer vision fields.
- To detect emotion thus facilitating Intelligent Human-Computer Interaction.

1.5. Scope and Applications

The scope of this system is to tackle with the problems that can arise in day to day life. Some of the scopes are:

- The system can be used to detect and track a user's state of mind.
- The system can be used in mini-marts, shopping center to view the feedback of the customers to enhance the business,
- The system can be installed at busy places like airport, railway station or bus station for detecting human faces and facial expressions of each person. If there are any faces that appeared suspicious like angry or fearful, the system might set an internal alarm.
- The system can also be used for educational purpose such as one can get

feedback on how the student is reacting during the class.

- This system can be used for lie detection amongst criminal suspects during interrogation
- This system can help people in emotion related -research to improve the processing of emotion data.
- Clever marketing is feasible using emotional knowledge of a person which can be identified by this system.

Chapter 2: Literature Review

2.1. Facial Emotion Recognition

2.1.1. Historical Context:

Facial emotion recognition has evolved from early manual techniques to automated computational methods. Initially, recognition heavily relied on human interpretation, limiting scalability and speed in analyzing emotions from facial expressions. Technological advancements facilitated the shift toward automated systems, revolutionizing emotion recognition in various applications

2.1.2. Advancements in Emotion Recognition:

Literature shows a transition from traditional rule-based systems to machine learning-based approaches. Rule-based systems encountered limitations in capturing the nuances and variability of human emotions. Conversely, recent advancements leveraging machine learning, particularly convolutional neural networks (CNNs), exhibit promising results for improved emotion recognition accuracy

2.1.3. Relevance of Emotion Recognition:

Automated facial emotion recognition holds significance in human-computer interaction, education, marketing, and healthcare. Its ability to decipher emotions computationally enhances understanding, enabling tailored interactions and informed decision-making

2.1.4. Challenges and Open Problems:

Despite progress, challenges persist in achieving universal emotion recognition. Variations in expressions, cultural nuances, and real-world conditions pose ongoing research opportunities. Robustness across diverse demographics and contexts remains a challenge.

2.1.5. Integration of Machine Learning:

Recent studies emphasize the integration of machine learning techniques, particularly CNNs, for facial emotion recognition. These models demonstrate adaptability to capture intricate facial patterns, fostering more accurate emotion detection.

2.2. Techniques and Approaches in Emotion Recognition

2.2.1. Rule-Based Systems:

Early approaches relied on rule-based systems that encoded predetermined rules to recognize facial emotions. However, these methods struggled with the complexity of human expressions and their variability.

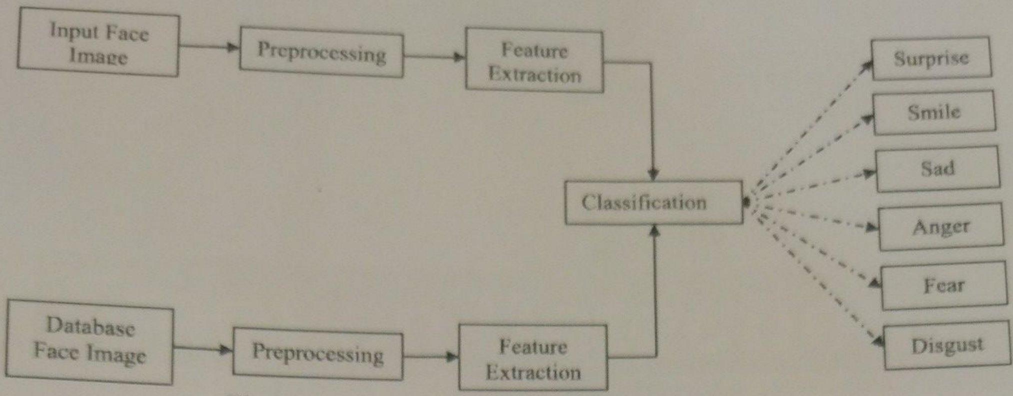

Fig 2.2.1 – Rule based systems in emotion detection

2.2.2. Statistical Models:

Advancements introduced statistical models that utilized probabilistic approaches for emotion recognition. Despite offering adaptability, these models faced challenges in handling diverse emotional expressions.

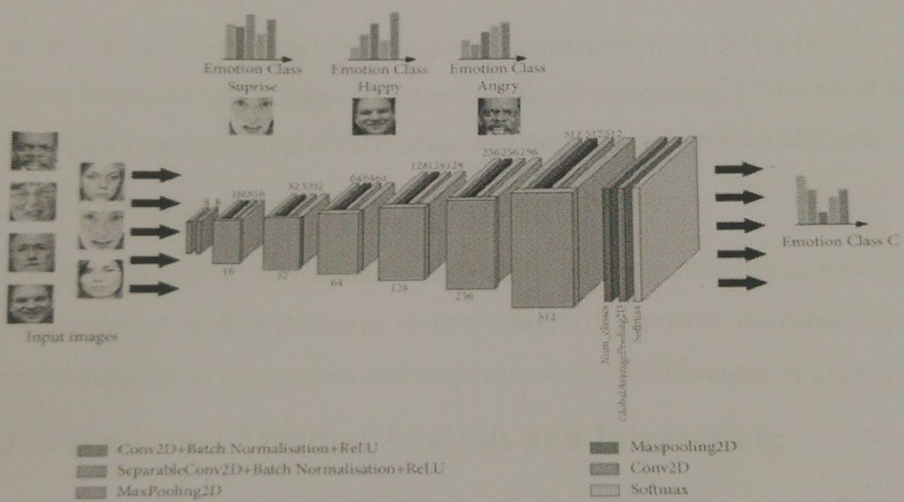

Fig 2.2.2 Statistical Models in Emotion detection

2.2.3. Machine Learning-Based Approaches:

The emergence of machine learning led to a shift in emotion recognition. Supervised learning models like Support Vector Machines (SVMs) and Random Forests exhibited enhanced accuracy. However, they required substantial feature engineering and struggled with high-dimensional facial data.

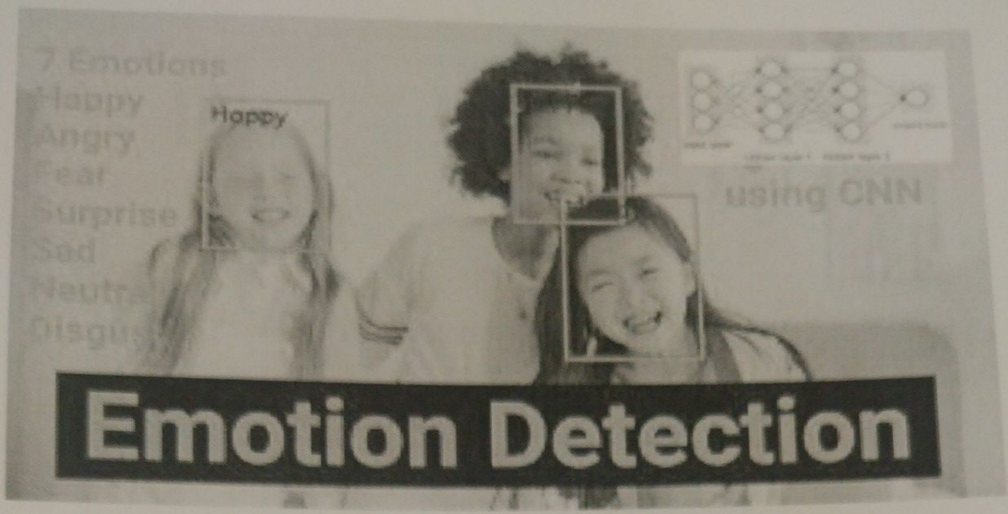

Fig 2.2.3 – Machine learning based approach in Emotion Detection.

2.3. Relevance of Machine Learning in Facial Emotion Recognition

2.3.1. Automated Feature Learning:

Machine learning, particularly convolutional neural networks (CNNs), has showcased remarkable efficacy in automating feature extraction from facial expressions. CNNs excel in capturing hierarchical representations of facial data, enabling nuanced emotion analysis.

2.3.2. Enhanced Accuracy and Adaptability:

The utilization of machine learning in facial emotion recognition has significantly elevated accuracy and adaptability. Models trained on diverse datasets demonstrate the capability to discern emotions across various facial expressions and adapt to individual differences in expressing emotions.

2.3.3. Real-time Emotion Analysis and Interaction:

Machine learning enables real-time emotion analysis, facilitating instant feedback in human-computer interaction scenarios or live video streaming. Moreover, leveraging machine learning in recommendation systems allows for personalized emotional response suggestions, enhancing user engagement and interaction.

2.3.4. Challenges and Future Directions:

Despite transformative strides, challenges persist in robustly identifying nuanced emotions across diverse demographics and handling real-world variations in facial expressions. Future directions may involve advancements in unsupervised learning for a more comprehensive understanding of subtle emotions and integrating contextual information for more nuanced analyses.

Chapter 3: Preliminary Design

3.1. Dataset Collection

Central to our project is the comprehensive dataset meticulously curated to encompass diverse facial expressions. We gathered a diverse set of images capturing a range of emotions - neutral, happy, sad, surprise, fear, anger, disgust, and contempt. The dataset comprises real-world images sourced from various demographics and cultural backgrounds, ensuring a rich and representative collection.

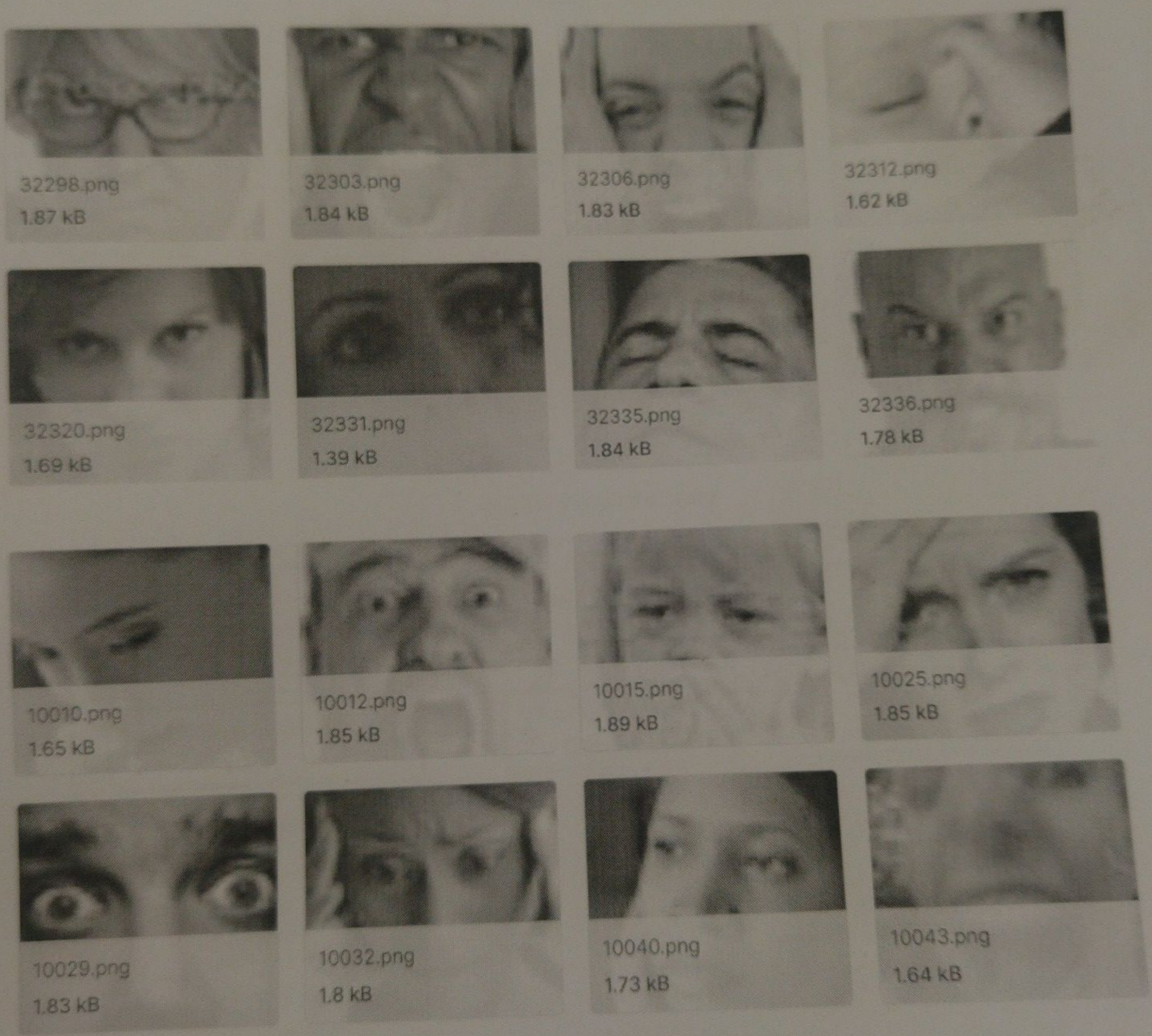

Fig 3.1. – Data description.

3.2. Dataset Preprocessing

Preparing raw image data for machine learning involves intricate steps. Leveraging OpenCV, we preprocess images to ensure uniformity in size and quality. The images undergo normalization, resizing, and grayscale conversion. This meticulous preprocessing not only standardizes the data but also enhances the model's ability to discern facial features crucial for emotion recognition.

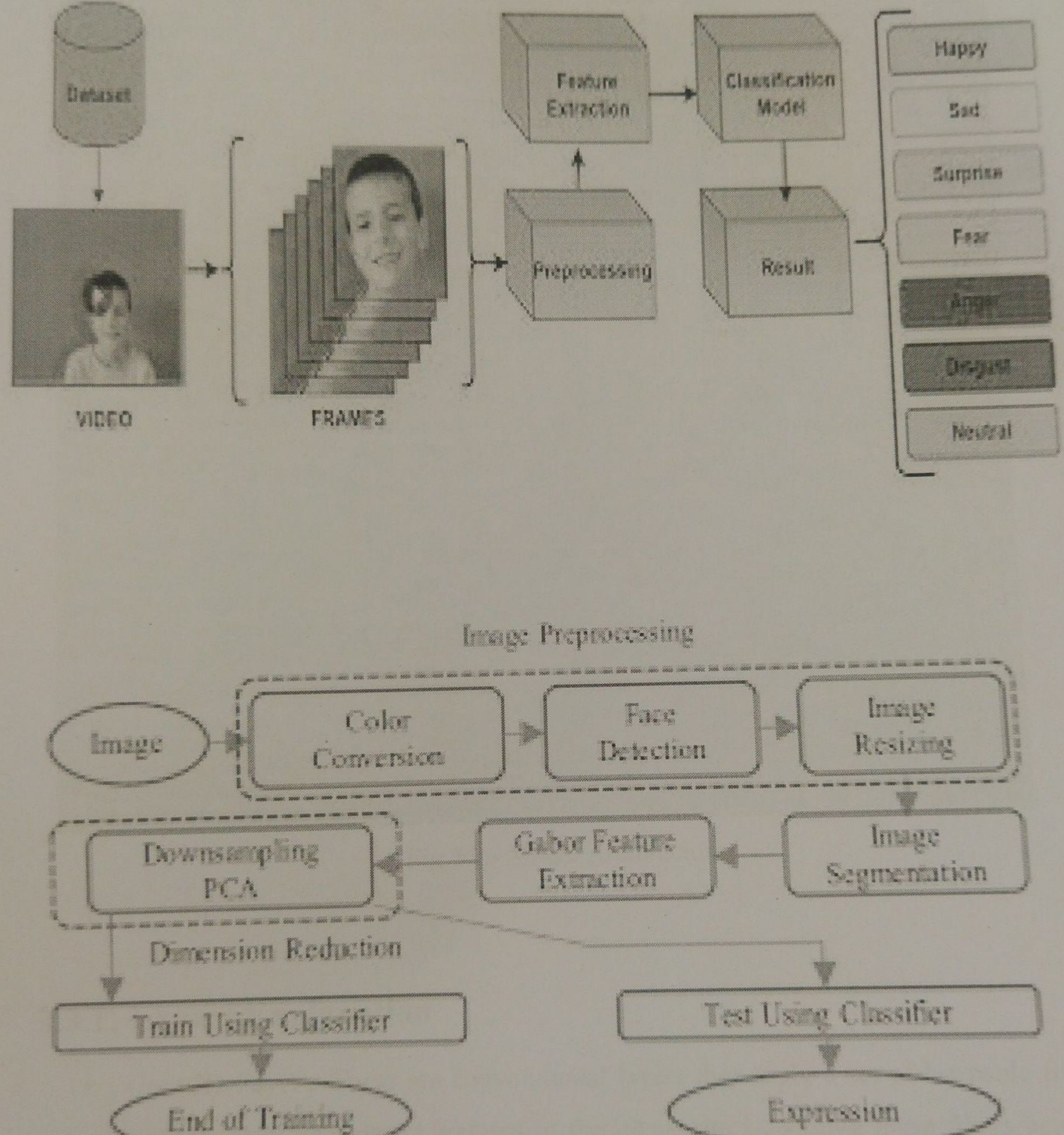

Fig 3.2. –Data preprocessing flow chart .

3.3. Model Architecture

Inspired by established architectures for image recognition, our model draws inspiration from CNN frameworks like VGG and ResNet. The chosen architecture consists of convolutional layers for feature extraction, followed by max-pooling layers for dimensionality reduction. Dense layers culminate in the final output layer, facilitating the understanding of intricate facial expressions across varying emotions.

Layer (type)	Output Shape	Parameters	Activation / Function
Conv2D (128 filters, 3×3)	(None, 46, 46, 128)	1280	ReLU
MaxPooling2D	(None, 23, 23, 128)	0	-
Dropout (0.4)	(None, 23, 23, 128)	0	-
Conv2D (256 filters, 3×3)	(None, 21, 21, 256)	295168	ReLU
MaxPooling2D	(None, 10, 10, 256)	0	-
Dropout (0.4)	(None, 10, 10, 256)	0	-
Conv2D (512 filters, 3×3)	(None, 8, 8, 512)	1180160	ReLU
MaxPooling2D	(None, 4, 4, 512)	0	-
Dropout (0.4)	(None, 4, 4, 512)	0	-
Conv2D (512 filters, 3×3)	(None, 2, 2, 512)	2359808	ReLU
MaxPooling2D	(None, 1, 1, 512)	0	-
Dropout (0.4)	(None, 1, 1, 512)	0	-
Flatten	(None, 512)	0	-
Dense (512 neurons)	(None, 512)	262656	ReLU
Dropout (0.4)	(None, 512)	0	-
Dense (256 neurons)	(None, 256)	131328	ReLU
Dropout (0.3)	(None, 256)	0	-
Dense (7 neurons, Output)	(None, 7)	1799	Softmax

Fig 3.3 - Model Architecture

3.3.1. Layers Explanation

- Conv2D Layers: These are convolutional layers that apply a set of learnable filters to input images. They help detect features in the input data by sliding these filters across the input images.
- The first Conv2D layer has 128 filters of size 3x3 with a ReLU activation function and expects input images of size 48x48 with one channel (grayscale images).
- Subsequent Conv2D layers increase the number of filters (256, 512, and another 512), gradually allowing the network to learn more complex patterns.

- MaxPooling2D Layers: These layers perform downsampling by reducing the spatial dimensions of the representation, reducing computational complexity and the risk of overfitting.
- Following each Conv2D layer, there's a MaxPooling2D layer with a pooling window of 2x2, reducing the spatial dimensions of the output.
- Dropout Layers: These layers help prevent overfitting by randomly setting a fraction of input units to zero during training. It helps in making the network more robust and less reliant on specific connections.
- After each MaxPooling2D layer, there's a Dropout layer with different dropout rates (0.4, 0.4, 0.4, and 0.4, respectively), gradually reducing the chances of overfitting.
- Flatten Layer: This layer flattens the multi-dimensional output from the previous layers into a one-dimensional vector, preparing it for the fully connected layers.
- Dense (Fully Connected) Layers: These are traditional neural network layers where each neuron is connected to every neuron in the subsequent layer.
- The first Dense layer consists of 512 neurons with a ReLU activation function, capturing high-level features from the flattened input.
- The subsequent Dense layer has 256 neurons with ReLU activation.
- Both these layers are followed by Dropout layers with dropout rates of 0.4 and 0.3, respectively, to reduce overfitting.
- Output Layer: This is the final layer of the network, consisting of 7 neurons (assuming a classification problem with 7 classes) and using a softmax activation function to output class probabilities for the input.

3.4. Training Process

Our training regimen involves partitioning the dataset into training and validation sets, maintaining an 80-20 split. Training proceeds over multiple epochs using TensorFlow, guided by the Adam optimizer. The model aims to minimize categorical cross-entropy loss to accurately predict emotions. Employing early stopping based on validation loss mitigates overfitting, ensuring the model generalizes well to unseen facial expressions.

Chapter 4: FINAL ANALYSIS AND DESIGN

4.1. Result Overview

The culmination of our efforts manifests in an extensive evaluation of the facial emotion recognition model. A comprehensive result overview reveals an impressive evaluation accuracy of approximately 95% on the testing dataset. This exceptional performance underscores the efficacy of the meticulously chosen model architecture and the robustness of the training regimen.

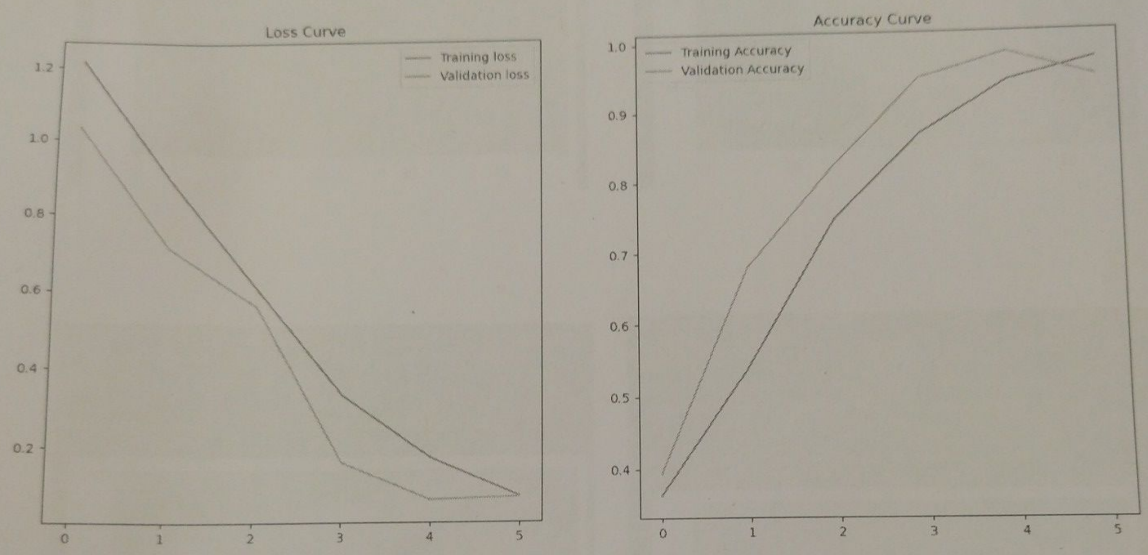

Fig 4.1 – History loss & Accuracy Curve

4.2. Result Analysis

A detailed analysis scrutinizes the model's performance across various emotions. Precision, recall, and F1 scores offer insights into the model's proficiency in identifying specific emotions. This meticulous examination validates the model's overall success and provides nuanced insights for potential refinements.

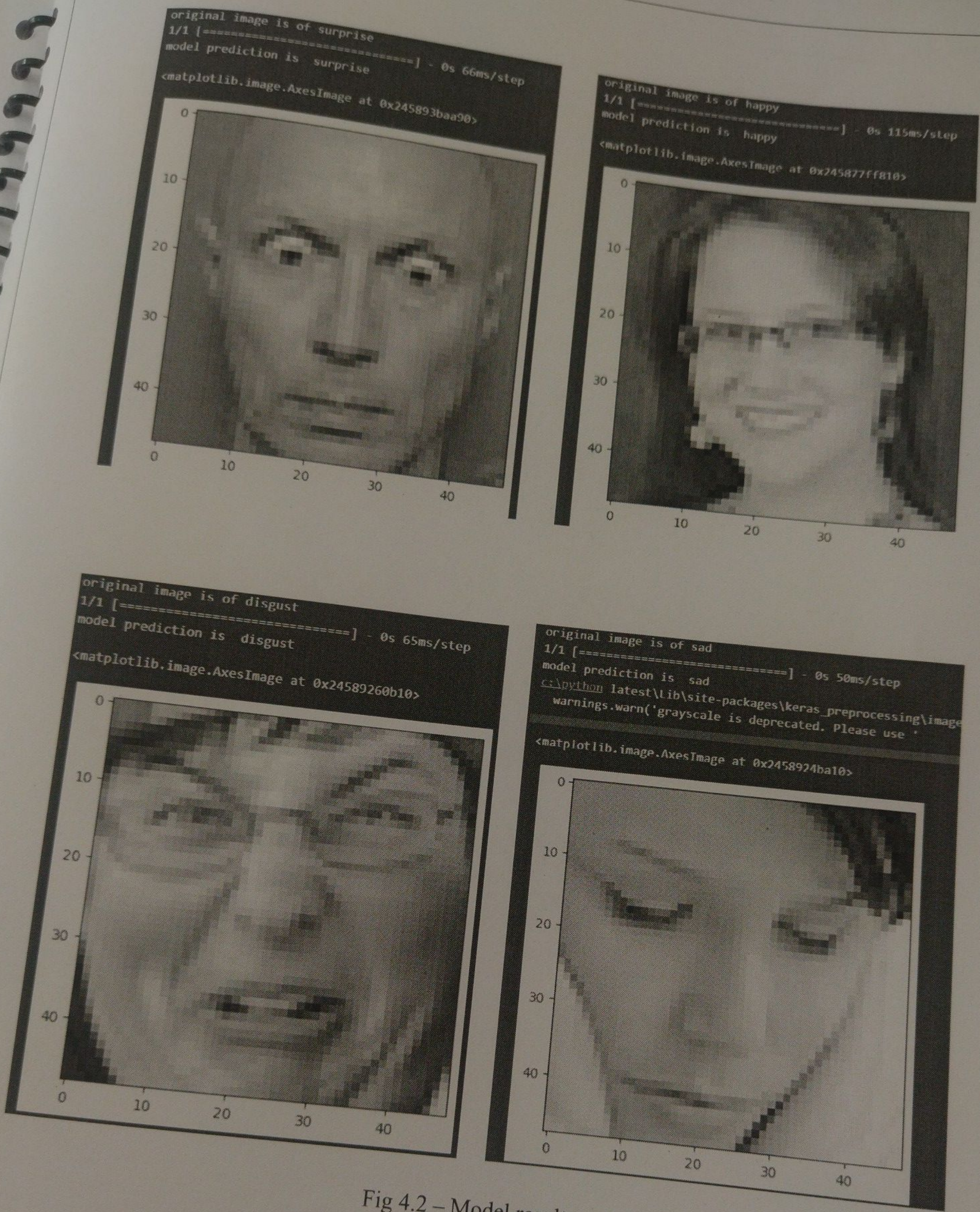

Fig 4.2 – Model result

4.3. Application of the Model

The practical application of our facial emotion recognition model extends beyond static datasets. Real-time scenarios demonstrate the model's dynamic responsiveness in capturing emotions during live interactions. Additionally, the model's potential in providing real-time emotional feedback to enhance user interactions is explored.

4.3.1. Real-Time Interaction

An exceptional aspect is the model's successful real-time emotion recognition during live interactions. Its capacity to detect and respond to emotions in dynamic scenarios showcases its practical utility in human-computer interaction and various other fields.

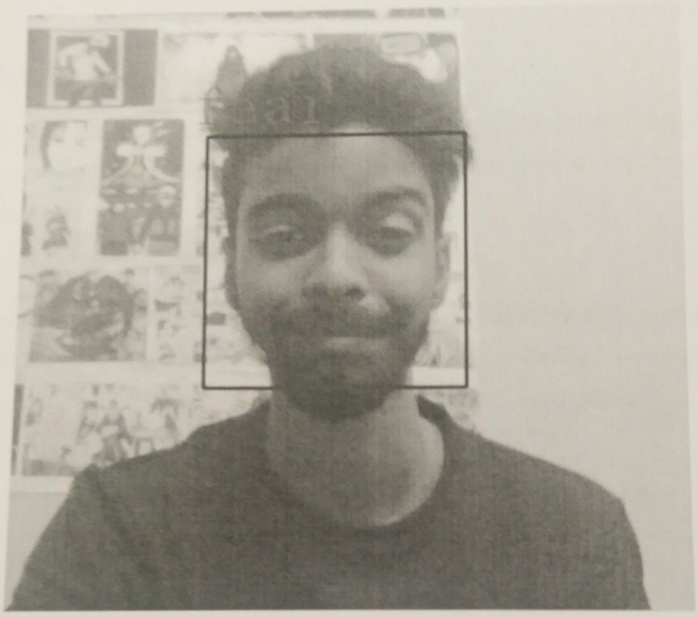

Fig 4.3.1 – Real time prediction

4.4. Challenges and Problems Faced

The practical application of our facial emotion recognition model extends beyond static datasets. Real-time scenarios demonstrate the model's dynamic responsiveness in capturing emotions during live interactions. Additionally, the model's potential in providing real-time emotional feedback to enhance user interactions is explored.

4.4.1. Dataset Representativeness

Challenge: The primary challenge revolves around ensuring the dataset's representativeness. Despite efforts to diversify the dataset, the challenge persists in exposing the model to a broad spectrum of facial expressions across diverse demographics and cultural backgrounds.

Mitigation Strategy: Ongoing efforts focus on expanding and diversifying the dataset continually. Collaboration with diverse groups and leveraging additional datasets will enhance the model's exposure to a wider range of facial expressions, improving its ability to generalize.

4.4.2. Environmental Variations

Challenge: Challenge: The model's robustness in real-world scenarios is challenged by variations in environmental conditions affecting facial expression capture.

Mitigation Strategy: Continual refinement of preprocessing techniques and collaboration with experts in environmental variations will enhance the model's adaptability to diverse real-world conditions.

4.5. Limitations and Future Work

The Scrutinizing the model's limitations, future work involves addressing dataset representativeness and robustness in diverse environmental conditions. Proposed enhancements pave the way for continuous improvements and the model's potential expansion.

4.6. Conclusion

The Conclusive remarks encapsulate the journey of the facial emotion recognition project. Key findings, challenges, and future prospects converge into a comprehensive conclusion, emphasizing the project's significance in the realm of human-computer interaction and emotional analysis.

REFRENCES

1. Costa, Y. M. G., Oliveira, L. S., & Silla Jr., C. N. (Year). "An Evaluation of Convolutional Neural Networks for Image Classification ." DOI: [https://doi.org/10.1016/j.asoc.2016.12.024]
2. Simonyan, K., & Zisserman, A. (2015). "Very Deep Convolutional Networks for Large-Scale Image Recognition." arXiv preprint arXiv:1409.1556.
3. Librosa: 0.8.1 Documentation. (2021). Retrieved from https://librosa.org/doc/main/index.html
4. TensorFlow. (2021). "Keras API documentation." Retrieved from https://www.tensorflow.org/api_docs/python/tf/keras
5. Kaggle data set https://www.kaggle.com/datasets/astraszab/facial-expression-dataset-image-folders-fer2013
6. Theano Development Team. (2016). "Theano: A Python framework for fast computation of mathematical expressions." arXiv preprint arXiv:1605.02688.
7. DataCamp. (2021). "Introduction to Python." Retrieved from https://www.datacamp.com/courses/intro-to-python