JOMO KENYATTA UNIVERSITY OF AGRICULTURE AND TECHNOLOGY

SCHOOL OF COMPUTING AND INFORMATION TECHNOLOGY

BSC. INFORMATION TECHNOLOGY

PROJECT TITLE: FACIAL EXPRESSION RECOGNITION SYSTEM

STUDENT NAME: PHILIP MEEN MADHANG

REGISTRATION NUMBER: SCT221-C004-0338/2020

SUPERVISOR: MS. JUDY GATERI

This project has been submitted in partial fulfillment of the requirements for the award of the degree of Bachelor of Science in Information Technology in the year 2023.

DECLARATION

I affirm that the content and information presented in this document and program are entirely my own. In the event that there is any borrowed information or content, I have duly provided proper references.

Name ________________________________ Date _____________________________

Sign __________________________________

Approved by:

Supervisor ______________________________ Date _____________________________

Sign __________________________________

ACKNOWLEDGEMENT

I express my heartfelt gratitude to everyone who played a role in the successful completion of this project. Their unwavering support has been invaluable, guiding me from the project's conception through to its implementation.

I am particularly grateful to my supervisor, Ms. Judy Gateri, whose guidance and insistence on adhering to due diligence and proper processes have been crucial in navigating this project. Her commitment to precision has been instrumental in avoiding potential errors and ensuring the project's success.

Lastly, I extend my appreciation to my parents, friends, and classmates who have accompanied me on this academic journey. Their encouragement has been a source of strength throughout my time in college.

DEDICATION

I dedicate this project to the relentless pursuit of knowledge and the unwavering spirit of curiosity that fuels the journey of discovery. This endeavor is dedicated to those who believe in the power of learning and the transformative potential it holds.

In heartfelt appreciation, I dedicate this project to my supervisor, Ms. Judy Gateri, whose guidance has been a beacon of wisdom and support throughout the entire process. Her commitment to excellence has inspired me to strive for the highest standards in every aspect of this undertaking.

Furthermore, I dedicate this project to my parents, whose boundless encouragement and unwavering belief in my abilities have been a constant source of motivation. To my friends and classmates, thank you for the camaraderie, shared experiences, and the collective strength that comes from learning together.

May this project stand as a testament to the collaborative spirit of those who value education, curiosity, and the pursuit of knowledge.

LIST OF FIGURES

ABSTRACT

Facial expressions serve as powerful non-verbal communication cues, playing a pivotal role in interpersonal relationships. The Facial Expression Recognition system outlined in this study aims to identify the emotional state of an individual. The process involves comparing a captured image with a pre-trained dataset in the database, subsequently displaying the emotional state associated with the image. This system integrates image processing and machine learning, utilizing the Convolutional Neural Network (CNN) as a fundamental component for designing a robust facial feature descriptor. CNNs, sophisticated deep learning models, automatically learn hierarchical features from data. The recognition performance of the proposed method is evaluated using a trained database, employing the Support Vector Machine. Experimental results, particularly with prototypic expressions, underscore the superiority of the CNN descriptor when compared to established appearance-based feature representation methods. This study contributes to the advancement of Facial Expression Recognition systems, showcasing the effectiveness of CNNs in enhancing recognition accuracy.

Keywords:

- Facial Expression Recognition (FER)
- Convolutional Neural Network (CNN)
- Support Vector Machine (SVM)

Table of Contents

CHAPTER ONE

1.0 Introduction

1.1 Background study

Facial expressions play a vital role in human communication, conveying emotions, intentions, and cognitive activity. Achieving accurate recognition of facial expressions has been a significant focus of research due to their importance in understanding human behavior. However, the complexity and variations in facial expressions present challenges to achieving high accuracy. In the field of computer science, efforts have been made to automate face detection and segmentation, addressing the extraction of facial features and analysis of expressions. The objective is to develop a facial expression recognition system that can accurately detect and classify a wide range of expressions.

Creating an automatic system for facial expression recognition is crucial for enhancing communication and understanding between individuals. Nonverbal cues conveyed through facial expressions enable the expression of intentions, emotions, and thoughts. Thus, the fluency of detecting and extracting facial expressions from images or videos is essential. This report outlines the challenges of facial expression recognition and explores advancements in computer vision. The goal is to develop a robust and accurate system capable of capturing the complexities and variations inherent in facial expressions, facilitating effective communication and analysis of human emotions. Facial expression recognition systems have gained significant attention due to their potential applications in various fields, including lie detection, medical assessment, and human-computer interfaces. The Facial Action Coding System (FACS), proposed by Ekman in 1978 and refined in 2002, is a popular tool for facial expression analysis. Humans commonly recognize emotions based on characteristic facial features, such as a smile for happiness or specific deformations for other emotions. Automatic recognition of facial expressions focuses on representing and categorizing the static or dynamic characteristics of these facial deformations.

The system described aims to classify facial expressions into basic emotions such as anger, disgust, fear, happiness, sadness, and surprise. Its main purpose is to enable efficient interaction between humans and machines, utilizing cues such as eye gaze, facial expressions, and cognitive modeling. Facial expression detection and classification serve as a natural means of interaction between humans and machines. However, the intensity of expressions varies between individuals and can be influenced by factors like age, gender, facial size, and shape. Additionally, facial expressions of the same person are not constant over time.

Recognizing facial expressions is challenging due to inherent variability caused by factors such as variations in illumination, pose, alignment, and occlusions. These challenges have been addressed in surveys focusing on facial feature representations for face recognition and expression analysis, providing detailed discussions on possible solutions.

Overall, the study highlights the significance of facial expressions in interpersonal communication and explores the challenges associated with facial expression recognition. It also emphasizes the importance of accurate detection and classification for effective human-machine interaction and outlines previous research efforts in addressing the complexities of facial expressions.

1.2 Problem statement

The problem at hand is the efficient and effective recognition of human emotions and intentions through facial expressions. While facial expressions serve as crucial cues for applications like intelligent man-machine interfaces, communication, visual surveillance, teleconferencing, and real-time animation, the existing research and systems in facial expression recognition are limited to only six basic expressions (joy, sadness, anger, disgust, fear, surprise). It has been observed that these basic expressions are insufficient to capture the diversity of facial expressions, necessitating the categorization of expressions based on facial actions.

One of the major challenges in this context is the complex task of detecting faces and accurately recognizing facial expressions. To address this challenge, it is essential to focus on key components such as face configuration, orientation, and location of the face within an image. These components play a vital role in achieving reliable and robust facial expression recognition.

Therefore, the problem statement revolves around developing an advanced facial expression recognition system that goes beyond the limitations of basic expressions. The system should be capable of detecting faces accurately and effectively recognizing a wide range of facial expressions, considering the primary components of face configuration, orientation, and location. By addressing these challenges, the aim is to enhance the efficiency and effectiveness of facial expression recognition systems in various applications, enabling more sophisticated and nuanced human-machine interactions.

1.3 Research Questions

The following research questions will guide this study:

1. How can facial expression recognition systems be improved to accurately detect and classify a wide range of facial expressions beyond the traditional basic emotions?
2. How can facial expression recognition systems account for the inherent variability in facial images caused by factors such as illumination, pose, alignment, and occlusions?
3. What are the optimal machine learning algorithms or deep learning architectures for training facial expression recognition models that achieve high accuracy and generalizability across diverse datasets?
4. How can facial expression recognition systems effectively handle real-time or dynamic facial expressions, considering the temporal dynamics and intensity variations?

1.4 General objectives

The general objective of this study is to develop a facial expression recognition system that can accurately detect and classify a wide range of facial expressions beyond the traditional basic emotions.

1.4.1 Specific Objectives

The objectives of this study are as follows:

1. Develop a facial expression recognition system that can accurately detect and classify a wide range of facial expressions beyond the traditional basic emotions.
2. Develop facial expression recognition systems that are robust to the inherent variability in facial images caused by factors such as illumination, pose, alignment, and occlusions.
3. Develop machine learning algorithms or deep learning architectures that are optimal for training facial expression recognition models that achieve high accuracy and generalizability across diverse datasets.
4. Develop facial expression recognition systems that can effectively handle real-time or dynamic facial expressions, considering the temporal dynamics and intensity variations.

1.4.2 Deliverables

The following deliverables will be accomplished as part of this study:

1. Facial expression recognition algorithm: A robust and accurate algorithm for detecting and classifying facial expressions.

2. Pre-trained models: Ready-to-use or customizable models for facial expression recognition tasks.

3. Labeled dataset: A diverse dataset of facial expressions for training and evaluation purposes.

1.5 Justification

The proposed facial expression recognition system has strong justification due to its ability to address diverse problems and offer valuable applications. It enhances user experiences in settings like mini-marts and educational institutions, improves public safety by detecting suspicious emotions in crowded places, aids in lie detection during criminal interrogations, contributes to emotional research, and enables targeted marketing. Despite limitations regarding accuracy, privacy, cultural variations, computational requirements, and subject cooperation, addressing these challenges ensures responsible system development and deployment. Overall, this study provides practical benefits and advancements in human interaction and understanding.

In summary, the facial expression recognition system's justification lies in its potential to improve user experiences, enhance public safety, aid in law enforcement, contribute to emotional research, and facilitate targeted marketing. By overcoming limitations and challenges, this study offers valuable applications and advancements in understanding human emotions and behavior.

1.6.1 Scope

This project aims to create a comprehensive facial expression recognition system using advanced machine learning algorithms. The system's primary goal is to accurately identify and classify individuals' facial expressions into various emotional states, including surprise, neutrality, and happiness among others. It will feature robustness against lighting, pose, and occlusion variations, and real-time functionality. This development has the potential to transform human-computer interaction and interpersonal communication.

1.6.2 Limitation

While the facial expression recognition system offers valuable applications, it is important to acknowledge certain limitations:

Firstly, it's crucial to emphasize that FER systems are not mind readers; they can't provide direct insights into a person's thoughts or feelings. Instead, they offer information solely based on facial expressions, requiring careful interpretation for a comprehensive understanding of emotional states. Additionally, bias can be a concerning issue, as FER systems are trained on datasets that may inadvertently mirror the biases of their creators; for instance, a system trained primarily on one racial group may struggle to accurately detect emotions in individuals from different backgrounds. Lastly, their accuracy can vary significantly, influenced by factors such as image or video quality, lighting conditions, and individual facial features.

CHAPTER TWO: LITERATURE REVIEW

2.1 Introduction

Facial Expression Recognition (FER) has garnered significant attention for its diverse applications, from human-computer interaction to affective computing and clinical psychology. The automatic and accurate recognition of facial expressions plays a crucial role in improving human-machine communication, understanding emotional states, and diagnosing emotional disorders. Given that facial expressions are a primary means of conveying emotions and intentions, the development of robust systems for automatic facial expression recognition is a significant endeavor.

Recent years have witnessed a growing interest in harnessing both texture and shape information from prominent facial regions, aiming to enhance the accuracy and robustness of FER systems. Texture features, which capture fine-grained details in facial skin like wrinkles and skin texture, and shape features, focusing on the geometric configuration of facial landmarks and their spatial relationships, offer distinct perspectives on facial expressions. This paradigm shift seeks to overcome the limitations associated with conventional approaches, addressing challenges even in complex real-world scenarios.

This literature review explores advanced techniques and methodologies that combine texture and shape features for automatic facial expression recognition, aligning with the trends and patterns identified in the provided text. The objective is to overcome the limitations associated with conventional approaches, thereby enhancing recognition accuracy and robustness, even in challenging real-world scenarios.

The primary focus of this review is to comprehensively survey existing literature on the fusion of texture and shape features in facial expression recognition. The exploration will delve into various methods and algorithms developed for this fusion, scrutinizing their strengths, weaknesses, and contexts of excellence. Additionally, the review will shed light on commonly used datasets and evaluation metrics to assess the performance of these hybrid systems.

Through this literature review, several crucial goals are pursued. Firstly, it aims to provide a clear understanding of the theoretical foundations and technical intricacies underpinning the fusion of texture and shape features in facial expression recognition, in line with the trends identified. Secondly, a critical evaluation of the effectiveness of these approaches in comparison to unimodal methods will be conducted, addressing gaps identified in the existing research. Lastly, the review will identify current research gaps and propose future directions in this dynamic field, contributing to the ongoing advancement of automatic facial expression recognition systems.

2.2 Literature Review

In recent years, the research landscape in the domains of face detection and tracking has witnessed an impressive surge in activity, resulting in a wealth of comprehensive literature. One of the paramount challenges confronting researchers is the scarcity of spontaneous expression data, a critical facet of their investigations (Bettadapura, 2012). The endeavor to capture spontaneous expressions, both in images and video, looms as one of the most significant obstacles that lie ahead in this field (Shan, Gong, & McOwan, 2005). Countless attempts have been made to decipher the intricacies of facial expressions, with notable work by Zhang and colleagues who delved into the analysis of two distinct feature types, namely geometry-based features and Gabor wavelets-based features, as part of their facial expression recognition research.

Within this broad scope, various strategies have been explored for face detection, including appearance-based methods, feature invariant techniques, knowledge-driven methods, and template-based approaches. Simultaneously, the landscape of expression detection strategies includes techniques like local binary pattern phase correlation, Haar classifier, AdaBoost, and Gabor Wavelet (Bhatt et al., 2014). Innovations abound in this realm, with FaceReader leading the charge as the premier tool for automated facial expression analysis. Moreover, the introduction of application programming interfaces (APIs) like Emotient, Affectiva, and Karios has further expanded the horizon of expression recognition. It's essential to emphasize that automatic facial expression recognition encompasses two core aspects: facial feature representation and the classifier problem (Shan et al., 2005).

Facial feature representation is a vital component, involving the extraction of an appropriate set of features from original facial images to describe faces. Among the prominent algorithms used in this context are the Histogram of Oriented Gradient (HOG), Scale-Invariant Feature Transform (SIFT), Gabor Filters, and the Local Binary Pattern (LBP) method (Bhatt et al., 2014; Chen et al., 2014). LBP, despite its simplicity, has proven to be an exceptionally efficient texture operator that labels image pixels by thresholding their neighborhood, generating binary numbers. This approach not only provides a streamlined representation of the facial features but also yields impressive results (Bhatt et al., 2014). On the other hand, HOG, initially introduced by Dalal and Triggs in 2005, tallies the appearance of gradient orientation in local image patches.

Addressing the classifier problem, researchers have explored a diverse set of algorithms such as machine learning, neural networks, support vector machines (SVM), deep learning, and Naive Bayes. In particular, the formation of histograms using any of the facial feature representations invariably involves the application of Support Vector Machines (SVM) for expression recognition. SVM, by creating a hyperplane to segregate high-dimensional spaces, ideally achieves a clear distinction by maximizing the distance between the hyperplane and the training data points from different classes (Chen et al., 2014).

The significance of parameter choice, particularly the block size in LBP feature extraction, cannot be overstated, as it directly impacts the accuracy of facial expression recognition. Extensive testing has consistently demonstrated that employing LBP features results in facial expression recognition accuracy exceeding 97%. LBP's prowess lies in its compatibility with a wide array of classifiers, filters, and other computational components, making it a versatile and valuable asset in the pursuit of facial expression analysis (Bhatt et al., 2014).

In the Theoretical Framework of this study, we delve into the foundational concepts of facial emotion recognition, drawing from established theories and research to explore the intricate relationship between facial expressions, emotional states, and the role they play in understanding engagement and attention.

2.3 Theoretical Framework

Within the multifaceted realm of facial emotion recognition, our theoretical framework serves as the bedrock for comprehending the intricate interplay between emotions, facial expressions, and cognitive states. In this comprehensive exploration, we delve into the fundamental theories that underpin the recognition of emotions through facial expressions, shedding light on the evolution of thought in this domain and the integration of cutting-edge technology.

2.3.1 Theoretical Foundations of Facial Emotion Recognition

The comprehensive understanding of human emotions and their reliable measurement is a fundamental pursuit in psychological research. Ekman and Friesen (1971, 1978) pioneered the breakdown of facial expressions into distinct Action Units (AUs), each representing the contraction of specific facial muscles, resulting in observable movements such as raising the eyebrows or lifting the corners of the mouth. These AUs serve as the building blocks for decoding emotions. The presence and intensity of these AUs can be parametrically measured by human coders or automatic algorithms (Barrett et al., 2019).

Notably, the recognition of facial emotional expressions is grounded in the theory of basic emotions. This theory posits that emotions are the product of evolution and innate physiological processes (Damasio, 1994). Importantly, it claims that emotions manifest in distinctive, universal facial expressions common to all people (Ekman, 1999). However, this theory does not fully account for the influence of cultural and social norms and conscious interpretations of emotions (Ratner, 1989). Despite these criticisms, there is consensus regarding the existence of universally recognizable emotional expressions, enabling the development of techniques for emotion recognition based on facial expressions (Coppin and Sander, 2016). These techniques compare observed facial expressions to statistical distributions of facial configurations assigned to discrete emotions, drawn from diverse demographic and cultural profiles.

Traditionally, the evaluation of facial expressions relied on human coders, necessitating extensive training for reliable coding. An alternative approach involves automatic detection using algorithms such as the Aff ectiva-AFFDEX algorithm SDK 4 (Aff ectiva, 2015, Boston, MA), leading to commercially available techniques known for their high reliability and accuracy (Küntzler, Höfling, and Alpers, 2021; Stöckli et al., 2018).

Facial Emotion Recognition (FER) technology encompasses both parametric and observational approaches. It relies on the measurement of facial point movements and the assignment of labels (e.g., joy, surprise, attention) to specific facial configurations. This enables the parametric measurement of emotions, engagement, and attention, minimizing some of the challenges associated with declarative and observational methods employed by human observers. FER algorithms provide a consistent, reliable, and specific means of measuring emotions, engagement, and attention in real-time (Stöckli et al., 2018). Notably, video recording, a key component of FER technology, is minimally invasive and intrusive (Calado et al., 2017; Darvishi et al., 2021).

In our study, we have harnessed FER technology to measure seven discrete emotions: anger, fear, joy, surprise, contempt, sadness, and disgust (Coppin and Sander, 2016; Ekman and Friesen, 1971; Matsumoto and Ekman, 2009). Additionally, we have employed FER for the assessment of engagement and attention, which, although not discrete emotions, are recognized as essential cognitive states during learning (Csikszentmihalyi, 2000; D'Mello and Graesser, 2012; Graesser, 2020; Pekrun et al., 2014).

2.3.2 Emotions, Engagement, and Attention in Science Education

Emotions encompass a set of briefs, intense, and specific responses to various stimuli (Coppin and Sander, 2016; Graesser, 2020). These responses manifest in diverse forms, including behavioral expressions such as posture, facial expressions, and verbal communication. Emotions also encompass subjective experiences, peripheral physiological responses like changes in heart rate and perspiration, and the activation of specific regions of the nervous system, such as the amygdala, hypothalamus, and anterior cingulate cortex (Coppin and Sander, 2016; Fastenrath et al., 2022; Hoemann et al., 2017).

Engagement represents the depth of a learner's involvement, interest, and enthusiasm toward a learning task (Fredricks, 2011). In contrast, attention is reflective of intrinsic motivation and stimulates the desire to participate in educational activities (Csikszentmihalyi and Nakamura, 2010). Both engagement and attention are intricately connected with a learner's emotional state and vice versa, underscoring the interdependence of these elements (Pekrun et al., 2014). For instance, when concentration, enjoyment, and interest are high, engagement tends to increase (Shernoff et al., 2003). Crucially, previous studies have unveiled the relationships between emotions, engagement, and attention (Csikszentmihalyi and Nakamura, 2010; Shernoff et al., 2003).

In recent years, research on emotions, engagement, and attention in science teaching and learning has seen a rising trend (Sinatra, Broughton, and Lombardi, 2014). A significant portion of the literature has focused on students' emotions during classroom interactions. These studies have shown that positive emotions such as pride and triumph are closely related to understanding science concepts, social interactions, and achieving success in challenging tasks (Bellocchi and Ritchie, 2015; Chiang and Liu, 2014).

Furthermore, specific subjects within science education have been examined, with results indicating variations in students' emotional responses. For example, students are more likely to experience positive emotions towards subjects like Biology and Chemistry than towards Physics (Laukenmann et al., 2003; Marcos-Merino, 2019). Researchers have also identified activities or didactic methodologies that are more likely to trigger negative emotions such as fear and disgust (Dávila et al., 2021).

Notably, the relationships between emotions, content, and didactic methodologies suggest that the self-regulation of emotions can serve as a factor for reducing or controlling inappropriate situations (Fredricks, 2011; Sinatra and Taasoobshirazi, 2018).

In addition to students, emotions in pre-service and in-service science teachers have been explored in the literature (Bellocchi, 2019; Jeong, González-Gómez, and Cañada, 2016;

Lombardi and Sinatra, 2013). Several studies suggest that effective management of emotions and social bonds in the science classroom can help teachers avoid pathologizing inappropriate attitudes toward science. Moreover, research has found that the personal past attitudes of pre-service teachers towards science are reflected in their teaching (Borrachero et al., 2014).

In summary, research is increasingly unveiling the role of emotions in the performance, interest, and attitudes of both students and teachers in science learning and teaching processes. Nevertheless, there remains a need to understand the interplay between emotions and cognitive states, such as engagement, in real-time as stimuli occur (Murphy et al., 2019; Ochoa de Alda et al., 2019). Gaining insights into this interplay could offer more effective approaches to address the evolving challenges in learning and teaching in the context of science education.

2.4 Historical Development of Facial Expression Recognition (FER)

The study of Facial Expression Recognition (FER) has a deep-rooted history that has paved the way for its integration with cutting-edge computer vision and machine learning techniques. This historical overview highlights pivotal moments and theories that laid the foundation for the contemporary advancements in FER.

2.4.1 Early Psychological Theories

In the 19th century, seminal works by Charles Darwin and G.G. Duchenne de Bologne sparked the interest in facial expressions as indicators of underlying emotions. Darwin's "The Expression of the Emotions in Man and Animals" (1872) and Duchenne's "The Mechanism of Human Facial Expression" (1862) marked the inception of systematic studies on the relationship between facial expressions and emotions (Mayne & Bonanno, 2001).

2.4.2 Foundations of Expression Categorization

These early works set the stage for the categorization of basic emotions recognizable in facial expressions. Paul Ekman, in the late 20th century, played a pivotal role in identifying and defining six primary emotions: fear, joy, sadness, anger, disgust, and surprise. His research provided a critical framework for FER systems to detect and classify emotions based on facial cues (Ekman, 1992, 1993).

2.4.3 Technological Advancements

The advent of computer vision and machine learning in the latter part of the 20th century revolutionized FER. The shift from manual feature extraction to automated techniques allowed for more robust and scalable FER systems. Early FER algorithms, such as those relying on feature-based approaches and rule-based classifiers, laid the groundwork for the integration of machine learning (Bishop, 2006).

2.4.4 Challenges and Complexity

As FER evolved, it faced complex challenges. Researchers started incorporating additional modalities, such as micro-expressions, electroencephalography (EEG) signals, gestures, and tone of voice, into emotion recognition systems (Coan & Allen, 2004). These multidimensional approaches highlighted the intricate nature of human emotional expression, requiring FER systems to adapt accordingly.

2.5 Methodology Used in Previous Studies

2.5.1 Data collection

In the field of Facial Emotion Recognition (FER), researchers rely on various databases to facilitate the development and evaluation of FER systems. These databases serve as crucial resources for benchmarking and comparing the performance of different approaches. Most of these databases are centered around 2D static images or 2D video sequences, although there are some databases that incorporate 3D images. It's worth noting that FER systems based on a 2D approach may struggle with handling different poses, primarily because most 2D databases predominantly feature frontal faces. In contrast, a 3D approach has the potential to address the pose variation challenge more effectively.

Typically, FER databases are annotated with the six basic emotions: anger, disgust, fear, happiness, sadness, and surprise, along with the neutral expression. These databases can be categorized into two primary types based on the environment in which they were created: controlled environments, often within a laboratory setting with controlled lighting conditions, and uncontrolled or wild environments.

Moreover, the methodology used to elicit facial expressions varies across these databases. Some FER databases instruct subjects to pose specific emotions towards a reference, ensuring controlled and deliberate expressions. Conversely, others aim to capture spontaneous and genuine facial expressions by encouraging subjects to react naturally to stimuli or situations.

This section will introduce some of the most prominent databases commonly utilized in the reviewed FER literature.

2.4.1.1 The Extended Cohn-Kanade Database (CK+)

The Extended Cohn-Kanade database (CK+) (Kanade, 2010) contains 593 image sequences of posed and non-posed emotions. In addition, 123 participants were 18 to 50 years of age, 69% female, 81% Euro-American, 13% Afro-American, and 6% other groups. The images were digitized into either 640 x 490 or 640 x 480 resolution and are mostly gray. Each sequence was built on frontal views and 30-degree views, starting with a neutral expression up until the peak emotion (last frame of the sequence). Most sequences are labeled with eight emotions: anger, disgust, contempt, fear, neutral, happiness, sadness, and surprise.

2.5.1.2 The Japanese Female Facial Expression Database (JAFFE)

The Japanese Female Facial Expression database (JAFFE) (Akamatsu, & Kamachi, 1998) contains 213 images of six basic emotions, plus the neutral expression, posed by 10 Japanese female models. Each image has been labeled by 60 Japanese subjects.

2.5.1.3 Binghamton University 3D Facial Expression Database (BU-3DFE)

The Binghamton University 3D Facial Expression database (BU-3DFE) (Yin L et al, 2006) contains 606 3D facial expression sequences captured from 101 subjects. The texture video has a resolution of about 1040 x 1329 pixels per frame. The resulting database consists of 58 female and 43 male subjects, with a large variety of ethnic/racial ancestries. This database was built on the six basic emotions, plus the neutral expression.

2.5.1.4 Facial Expression Recognition 2013 Database (FER-2013)

The Facial Expression Recognition 2013 database (FER-2013) (Goodfellow, Carrier, & Courville, 2013) was created using the Google image search API to search for images of faces that match a set of 184 emotion-related keywords like "blissful," "enraged," etc. These keywords were combined with words related to gender, age, or ethnicity, leading to 35,887 grayscale images with a 48 x 48 resolution, mapped into the six basic emotions, plus the neutral expression.

2.5.1.5 Emotion Recognition in the Wild Database (EmotiW)

The Emotion Recognition in the Wild database (EmotiW) (Dhall, Ramana Murthy, & Gedeon, 2015) contains two sub-databases, Acted Facial Expression in the Wild (AFEW) and the Static Facial Expression in the Wild (SFEW). AFEW contains videos (image sequences including audio), and SFEW contains static images. This database was built on the six basic emotions, plus the neutral expression, and the image size is 128 x 128.

2.5.1.6 MMI Database

The MMI database (Pantic M & Maat L, 2015) contains over 2900 videos and high-resolution still images of 75 subjects. It is fully annotated for the presence of Action Units (AUs) in videos and partially coded on a frame-level, indicating for each frame whether an AU is in either the neutral, onset, apex, or offset phase. This database was built on six emotions: anger, disgust, fear, happiness, sadness, and surprise.

2.5.1.7 eNTERFACE'05 Audiovisual Emotion Database

The eNTERFACE'05 Audiovisual Emotion database (Martin, Kotsia, & Pitas, 2006) contains 42 subjects from 14 different nationalities. Among the 42 subjects, 81% were men and the remaining 19% were women. In addition, 31% of the subjects wore glasses, while 17% had a beard. This database consists of video sequences (including audio) with a 720 x 576 resolution and was built on six emotions: anger, disgust, fear, happiness, sadness, and surprise.

2.5.1.8 Karolinska Directed Emotional Faces Database (KDEF)

The Karolinska Directed Emotional Faces database (KDEF) (Calvo, M.G. & Lundqvist, 2008) contains a set of 4900 pictures of human facial expressions. The set includes 70 individuals (35 females and 35 males) displaying the six basic emotions, plus the neutral expression. Each expression is viewed from five different angles and was photographed in two sessions

2.5.1.9 Radboud Faces Database (RaFD)

The Radboud Faces Database (RaFD) (Langner & Van Knippenberg, 2010) contains a set of pictures of 67 models, including Caucasian males and females, and Moroccan Dutch males, displaying eight emotional expressions (anger, disgust, contempt, fear, neutral, happiness, sadness, and surprise). This amounts to 120 images per model. Each emotion is shown with three different gaze directions, and all pictures were taken from five camera angles simultaneously. The image size is 1024 x 681.

FER Databases Summary

Database	Capacity	Emotion	Environment	Facial Expressions
CK+	593 videos	Posed	Controlled	8
JAFFE	213 images	Posed	Controlled	7
BU-3DFE	606 videos	Posed and Spontaneous	Controlled	7
FER-2013	35,887 images	Posed and Spontaneous	Uncontrolled	7
EmotiW	1268 videos and 700 images	Spontaneous	Uncontrolled	7
MMI	2900 videos	Posed	Controlled	6
eNTERFACE’05	1166 videos	Spontaneous	Controlled	6
KDEF	4900 images	Posed	Controlled	7
RaFD	8040 images	Posed	Controlled	8

Summarized FER Databases.

2.5.2 Face detection

Face detection within a Facial Expression Recognition (FER) system involves a comprehensive three-stage process. In the initial stage, the system takes an input image and employs various image processing techniques to identify the facial region. This stage is adaptable to both static images, known as face localization, and videos, where the system engages in face tracking.

Several challenges arise in this first stage, including variations in facial scales and orientations. These variations often stem from subject movements or changes in the distance from the camera. Additionally, significant body movements can lead to drastic changes in the face's position across consecutive frames, making tracking more challenging. The complexity of the background and the variability in lighting conditions further add to the complexity of tracking. For example, in scenarios where multiple faces are present, the system must distinguish and track the correct face. Furthermore, spontaneous reactions can introduce occlusions, which need to be effectively addressed.

Addressing these challenges prompted the exploration of techniques that fall into two primary groups: holistic and analytic face models.

Holistic face models treat the face as a unified whole. For instance, Huang and Huang utilized the Point Distribution Model (PDM), representing the mean geometry of the human face. This model incorporates Canny edge detection and fitting the PDM to estimate face position. Pantic and Rothkrantz proposed a system that processes frontal and profile face views using vertical and horizontal histogram analysis to identify face boundaries.

In contrast, analytic face models focus on studying the co-occurrence of characteristic facial elements. Kobayashi and Hara used a monochrome mode image to find face brightness distribution, estimating face position through iris localization. Kimura and Yachida employed an integral projection algorithm to process input images, determining the position of eye and mouth corners using color and edge information.

The challenges outlined above have driven the development of various techniques for face detection within FER systems, distinguishing between holistic and analytic approaches to address the complexities associated with facial scales, orientations, and dynamic scenarios.

2.5.2.1 Holistic face models

Holistic face models play a crucial role in the field of face detection, and researchers have proposed various methods to address the challenges associated with this approach. Two notable examples are the works of Huang and Huang (1997) and Pantic and Rothkrantz (2000).

1. Huang and Huang's Approach (Huang & Huang, 1997)
    - **Model:** Point Distribution Model (PDM).
    - Method:
        - The Point Distribution Model is employed to represent the mean geometry of the human face.

        - The process begins with the application of the Canny edge detector to identify two symmetrical vertical edges within the image, providing an initial estimation of the face's position.
        - Subsequently, the Point Distribution Model is fitted to refine and accurately determine the face's location.
2. Pantic and Rothkrantz's Proposal (Pantic & Rothkrantz, 2000)
    - Model: Not explicitly mentioned, but the method involves processing images of frontal and profile face views.
    - Method:
        - The system processes images captured from frontal and profile face views.
        - Vertical and horizontal histogram analyses are employed to identify face boundaries within the images.
        - Following the identification of face boundaries, the face contour is obtained by applying a thresholding technique to the image using HSV (Hue, Saturation, Value) color space values.

These holistic face models represent innovative approaches to face detection, offering techniques to effectively locate and delineate facial features. Huang and Huang's use of the Point Distribution Model, coupled with edge detection, enhances accuracy in estimating face positions (Huang & Huang, 1997). Meanwhile, Pantic and Rothkrantz's method, employing histogram analysis and color space thresholding, provides a robust means of identifying face boundaries in images captured from different perspectives (Pantic & Rothkrantz, 2000). These approaches contribute valuable insights to the broader field of facial recognition systems.

2.5.2.2 Analytic face models

Analytic face models represent a distinct category of techniques in the realm of face detection. Two noteworthy examples are the methodologies proposed by Kobayashi and Hara (1997) and Kimura and Yachida (1997).

1. Kobayashi and Hara's Approach (Kobayashi & Hara, 1997)
    - Model: Image captured in monochrome mode to find face brightness distribution.
    - Method:
        - The image captured in monochrome mode is utilized to analyze face brightness distribution.
        - The position of the face is estimated through iris localization.
2. Kimura and Yachida's Technique (Kimura & Yachida, 1997)
    - Model: Potential Net model representing the face.

- Method:
    - The technique processes the input image using an integral projection algorithm.
    - The algorithm identifies the position of eye and mouth corners by leveraging color and edge information.
    - The face is represented using the Potential Net model, which is fitted based on the positions of the eyes and mouth.

While the above-mentioned analytic face models were specifically designed to process facial images, they do not inherently detect whether a face is present in the image. Systems designed to handle arbitrary images are outlined below:

1. Essa and Pentland's Approach (Essa & Pentland, 1997)
    - Method:
        - "Face space" is created by performing Principal Component Analysis of eigenfaces from 128 face images.
        - Face detection in the image is determined by evaluating its distance from the face space against an acceptable threshold.
2. Rowley et al.'s Proposal (Rowley, Baluja, & Kanade, 1998)
    - Method
        - A neural network-based face detection system is introduced.
        - The input image is scanned with a window, and a neural network decides whether a particular window contains a face.
3. Viola and Jones' Efficient Algorithm (Viola & Jones, 2004)
    - Method:
        - An efficient algorithm for object detection is introduced, widely used in face detection.
        - Haar-like features are employed as object representation, and Adaboost serves as the machine learning method.

These diverse approaches highlight the evolution of face detection techniques, ranging from analytic models tailored for facial features to more generalized systems capable of handling arbitrary images with varying content.

2.5.3 Feature extraction

Once the face has been successfully located in an image or video frame, the next step involves analyzing facial actions. Facial expression recognition (FER) commonly employs two types of features to describe facial expressions: geometric features and appearance features. Geometric features measure the displacements of specific facial parts, such as the brows or mouth corners, while appearance features capture changes in face texture associated with particular facial actions.

In terms of input, FER systems can be categorized based on whether they analyze static images or image sequences. The measurement of geometric features typically involves face region analysis, particularly the identification and tracking of key points within the face region. However, this task is not without challenges, as occlusions and the presence of facial hair or glasses can complicate the accurate decomposition of facial features.

Defining an appropriate feature set is a critical aspect of FER systems, and it poses its own set of challenges. The selected features need to be both descriptive and, ideally, uncorrelated. Additionally, the task is complicated by the need to address potential occlusions and the variations introduced by facial hair or accessories such as glasses. Achieving a balance between descriptiveness and independence in the feature set is essential for the accurate analysis of facial expressions in FER systems.

Feature extraction techniques

2.5.3.1 Local Binary Patterns

Local Binary Patterns (LBP) (Hawkins, 2004) is recognized as one of the top methods for texture processing. This algorithm compares a center pixel with its 3x3 square neighborhood. If the neighboring pixel value is greater than or equal to the center pixel value, it is assigned the value "1"; otherwise, it receives the value "0". Subsequently, the binary code generated from this operation is used to represent the center pixel's decimal value. Figure 1 illustrates this process.

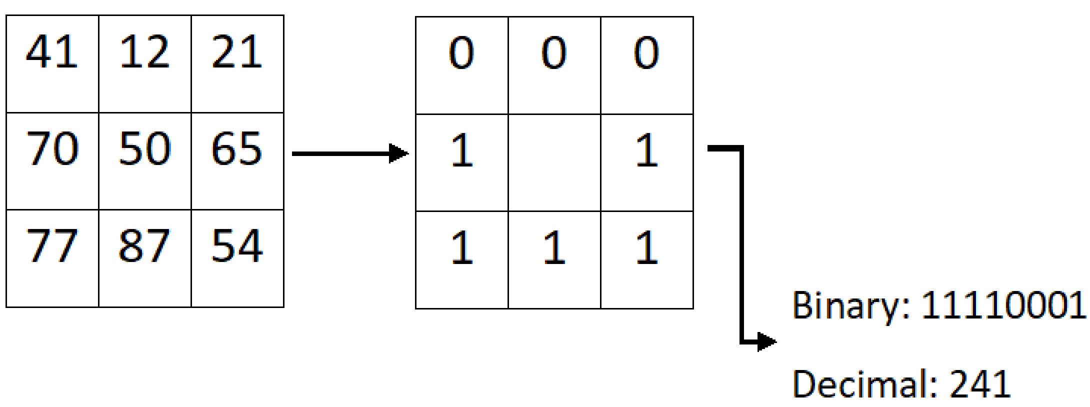


Figure 1. LBP operation.

In FER systems, LBP is effective in highlighting relevant facial features for emotion recognition, including eyebrows, eyes, nose, and mouth. However, it is sensitive to noise, as it relies on intensity differences and can be affected when processing regions with nearly uniform intensity. Figure 2 provides an example of feature extraction using LBP on a face from the CK+ database.

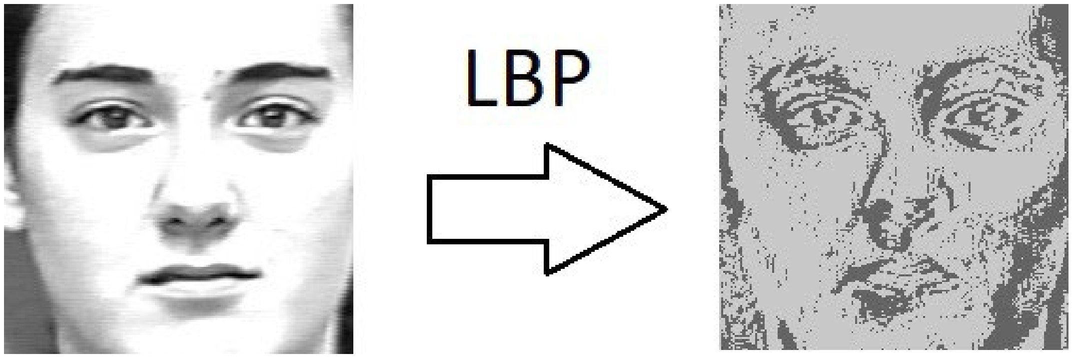


Figure 2. LBP feature extraction on a CK+ database face.

2.5.3.2 Optical Flow

Optical Flow (OF) (Horn & Schunck, 1981) is a technique suitable for analyzing a sequence of frames, typically in the form of a video. It is designed to evaluate the magnitude and direction of motion. OF calculates pixel-level motion, providing a vector that represents the movement of pixels from the first frame to the second frame. However, the accuracy of this method depends on the selection of initial tracked features and their stability over time, making it sensitive to noise and occlusions (Barron & Fleet, 1992).

OF can be a valuable addition to an FER system, as it can effectively capture facial motion when transitioning from a neutral expression to an intense emotional state. Figure 3 demonstrates the application of this method in an FER system using a sequence from the CK+ database.

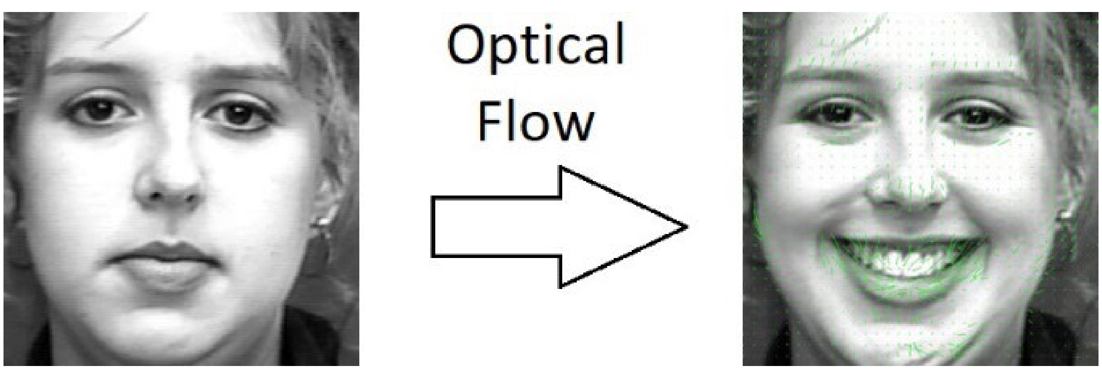


Figure 3. OF of a sequence from the CK+ database.

2.5.3.3 Active Appearance Model (AAM)

The Active Appearance Model (AAM) is a Computer Vision algorithm developed by Cootes, Edwards, and Taylor in 2001. It is designed to align the shape and appearance of an object with a new image. In the context of Facial Emotion Recognition (FER) systems, AAM is used to align and extract the facial region from an image, effectively discarding everything except the face. The shape information obtained from AAM is employed to calculate parameters that emphasize the appearance of facial features.

However, it's important to note that AAM is sensitive to variations in pose, expression, and illumination that were not part of the training dataset, as highlighted by Abdulameer, Abdullah, Huda, and Othman in 2014. Figure 4 illustrates this process.

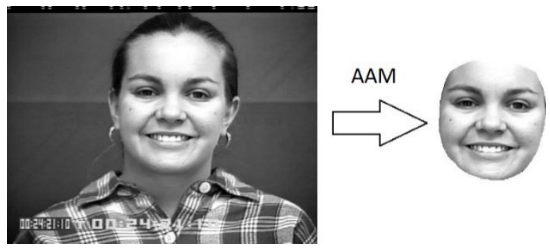


Figure 4. AAM Shape Estimation in CK+ Face.

2.5.3.4 Action Units (AUs)

Action Units (AUs) are distinct muscle movements that collectively form facial expressions. They are based on the concept that various facial expressions correspond to specific muscle actions. The Facial Action Coding System (FACS), introduced by Ekman in 1997, is a framework for describing facial expressions by deconstructing them into individual AUs.

Leveraging AUs as features for classifying emotions is a common strategy in the reviewed works. However, some researchers encountered challenges in precisely encoding the dynamics of these movements and measuring AU intensity. Figure 5 provides visual examples of relevant facial AUs.

Figure 5. Relevant AU Examples for Expression Discrimination.

2.5.3.5 Facial Animation Parameters (FAPs)

Facial Animation Parameters (FAPs), developed by Pakstas and Forchheimer in 2002, encompass 66 displacements and rotations of facial feature points from the neutral face position. These parameters are rooted in facial motion and muscle actions, providing a comprehensive set of fundamental facial actions for representing facial expressions. FAPs can also be viewed as the relevant distances between different facial features.

However, the extraction of FAPs is sensitive to various sources of noise, including varying lighting conditions, which can lead to subtle issues in facial area segmentation, potentially affecting the accuracy of the Facial Expression Recognition (FER) system. Figure 6 offers an illustration of FAPs extraction, specifically for eyebrows and mouth, using the CK+ database.

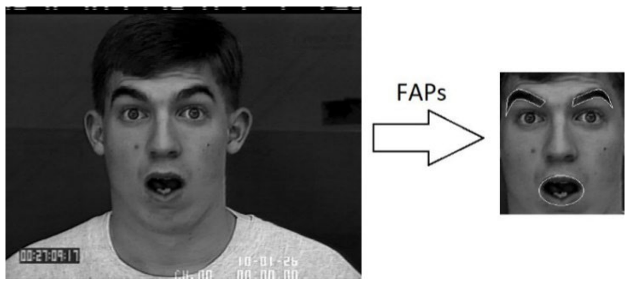


Figure 6. FAPs extraction using the CK+ database.

2.5.3.6 Gabor Filter

The Gabor filter, introduced by Jain and Farrokhnia in 1991, is employed to extract texture information from images. This filter is particularly effective in preserving details related to orientation and scale, making it robust even in the presence of challenging lighting conditions. It excels in capturing spatial details associated with frequency, position, and orientation within an image, making it a valuable tool for extracting subtle local transformations.

However, a significant drawback of the Gabor filter is its tendency to create high-dimensional Gabor feature spaces, resulting in substantial computational demands that are often impractical for real-time applications. To address this, simplified Gabor features can be utilized to achieve real-time performance. However, it's important to note that these simplified features may be sensitive to variations in lighting, as highlighted by Chora´s, R.S. in 2010. Figure 7 illustrates a feature map of Gabor for a face from the CK+ database.

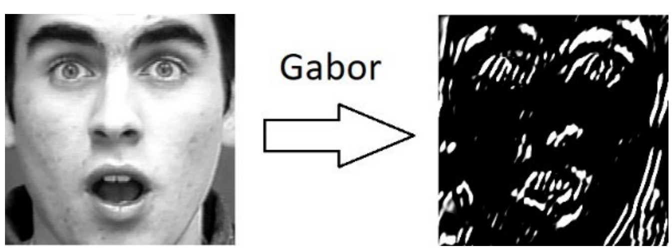


Figure 7. CK+ Face Gabor Feature Map.

2.5.3.7 Scale-Invariant Feature Transform (SIFT)

Scale-Invariant Feature Transform (SIFT) is a Computer Vision algorithm designed for the detection and description of local features within an image, as introduced by Lowe in 2004. SIFT features possess remarkable properties, being resistant to uniform scaling, changes in orientation, and variations in illumination. However, they may exhibit sensitivity to blur and affine transformations, as pointed out by Wu in 2013. This remarkable robustness stems from the transformation of an image into a substantial collection of feature vectors, each of which maintains its invariance to the conditions mentioned earlier.

In Facial Expression Recognition (FER) systems, the SIFT algorithm finds utility in the detection of key facial features like the eyebrows, eyes, nose, and mouth. Some studies have successfully combined this feature extraction algorithm with Optical Flow (OF) techniques to calculate the motion of facial features. Figure 8 provides a visual representation of the process of extracting local features from a face in the CK+ database.

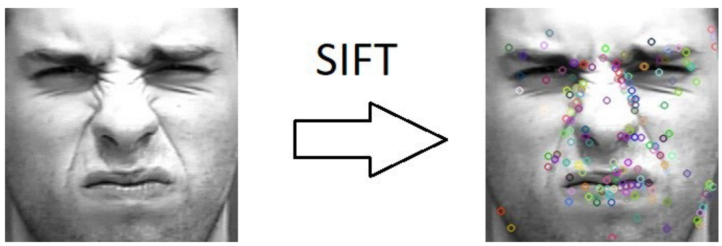


Figure 8. SIFT features of a face from the CK+ database.

2.5.3.8 Histogram of Oriented Gradients

Histogram of Oriented Gradients (HOG), introduced by Dalal and Triggs in 2005, is a feature extraction method that characterizes local object appearance and shape within an image based on the distribution of intensity gradients or edge directions. The image is divided into cells with a specified number of pixels, and for each cell, a histogram of gradient directions is constructed. HOG is particularly effective in capturing local object appearance, and it is invariant to geometric transformations, except for object orientation.

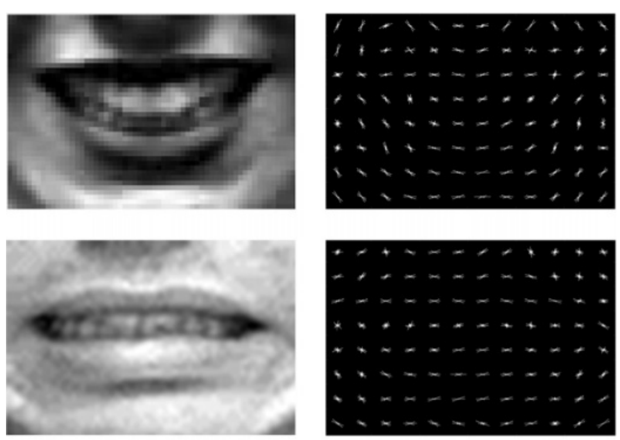


Figure 9. CK+ HOG Feature Extraction Example.

2.5.4 Expression Recognition (classification)

The final stage of the Facial Expression Recognition (FER) system revolves around the classification task, a crucial aspect of machine learning theory. In this phase, the input to the classifier comprises a set of features obtained from the face region during the previous stage, describing the facial expression. Classification involves supervised training, where the training set should include labeled data. Once trained, the classifier can assign particular class labels to input images, commonly based on either Action Units (AU) proposed in the Facial Action Coding System or universal emotions such as joy, sadness, anger, surprise, disgust, and fear. There are various machine learning techniques applied for the classification task. Here, we outline some of the most frequently utilized classification and regression algorithms in the FER system.

2.5.4.1 Convolutional Neural Network (CNN)

CNNs are a class of neural networks primarily employed in the realm of Computer Vision and Deep Learning. They are particularly adept at addressing a variety of image classification tasks and can outperform humans in discerning complex underlying patterns. In a CNN, input images traverse multiple hidden layers, each responsible for decomposing the image into meaningful features. These extracted features are then employed for classification, often using a SoftMax function to determine the most likely class from a probability distribution of classes.

It's essential to note that different problems require unique CNN models and tuning approaches to maintain high classification accuracy. Overfitting and underfitting issues, as discussed in Data Augmentation, can pose challenges. Overfitting occurs when the model struggles to generalize and exhibits reduced classification accuracy on unseen data, while underfitting results in poor predictions on both training and unseen data. Several strategies are commonly employed to address these problems:

1. Increasing Model Complexity: This involves adding more layers to the neural network.
2. Incorporating Dropout Layers*:* Dropout layers randomly disable a portion of nodes during training to prevent the model from memorizing patterns instead of learning them.
3. Parameter Tuning: Fine-tuning various parameters during training, such as epochs, batch size, learning rate, and class weight, among others.
4. Expanding Training Data*:* Increasing the amount of training data by either adding more samples or using data augmentation techniques, as discussed in Section 2.4.3.
5. Transfer Learning (TL): When working with small databases, which is common in publicly available datasets for emotion recognition, TL becomes a valuable technique. TL involves using a pre-trained model that has already been trained on a large dataset and fine-tuning it for the specific classification problem at hand.

Many of the reviewed works have effectively utilized CNNs for classification, showcasing promising results for deep learning-based classifiers.

2.5.4.2 Support Vector Machine (SVM)

SVM is a prominent Machine Learning algorithm extensively employed for classification and regression tasks. It operates by mapping features into a multidimensional space where the features belonging to each class are separated by a distinct gap, ideally as wide as possible. Subsequently, input features are also mapped into this space and classified based on which side of the gap they fall on. The training phase establishes this mapping, which is subsequently used for predictions. SVM's strengths include its ability to handle complex nonlinear data and its robustness against overfitting. Nevertheless, SVMs come with certain drawbacks, such as computational complexity, the challenge of selecting an appropriate kernel function, and suboptimal performance when applied to large databases. Figure 10 offers a visualization of a potential feature space in an SVM model trained for an FER system.

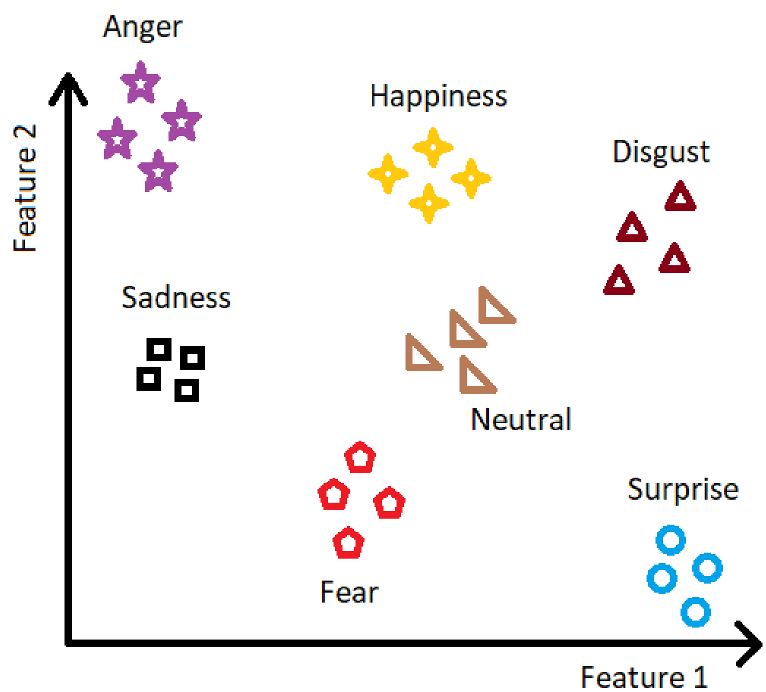


Figure 10. Example of a feature space in an FER system.

2.5.4.3 K-Nearest Neighbor (KNN)

K-nearest neighbor (KNN), introduced by Cover and Hart in 1967, is an instance-based learning algorithm characterized by its non-parametric approach to classification and regression. The algorithm's training data comprises vectors in a multidimensional feature space, each associated with a class label. During the training phase, the algorithm stores these feature vectors and their corresponding classes. In the classification step, input features or feature sets are classified based on the class that shares the nearest features with the input. Common distance metrics for determining proximity between features include the Euclidean distance (ED) and Hamming distance. KNN boasts simplicity in its implementation and a quick training step. Nevertheless, it demands significant storage space, exhibits slow testing performance, is sensitive to noise, and struggles with high-dimensional data. Additionally, the classifier can yield inaccurate predictions when dealing with unbalanced classes. To mitigate this issue, class weights can be set.

2.5.4.4 Naive Bayes

Naive Bayes classifiers, developed by Friedman, Geiger, and Goldszmidt in 1997, belong to the family of probabilistic Machine Learning classifiers based on Bayes' theorem. These classifiers operate under the assumption of strong independence between features, which is a key characteristic. Bayes' theorem, as outlined in Equation,

$$P(A|B) = \frac{P(B|A)\ (A)}{P(B)}.$$

allows for the calculation of the likelihood of event A occurring given that B is true. However, Naive Bayes classifiers assume that features are independent, meaning that in a Facial Expression Recognition (FER) system, the classifier does not account for feature correlations. This can be a limitation since facial expressions often involve correlated features. For example, when a person is surprised, there is a clear correlation between the mouth and the eyes, both typically opening wider.

2.5.4.5 Hidden Markov Model

Hidden Markov Model (HMM), introduced by Eddy in 1996, is a probabilistic model that specializes in predicting a sequence of unknown variables based on a set of observed variables. In the context of a Facial Expression Recognition (FER) system, this would involve predicting hidden emotional states, such as happiness, based on observed facial expressions, like a smile.

HMM's strengths include its ability to model arbitrary features from observations, its capacity to combine various HMMs for classifying different data, and its flexibility to incorporate prior knowledge into the model. However, this classifier is computationally expensive and susceptible to overfitting.

2.5.4.6 Decision Tree

Decision Tree (DT), introduced by Quinlan in 1986, is a classification technique represented as a flowchart structured in a tree model. In DT, the database is iteratively divided into smaller data sets until further divisions are no longer possible, with the resulting leaves serving as the classes for classification.

DT's strengths include its capacity to learn nonlinear data relationships, its ability to handle high-dimensional data, and its straightforward implementation. However, one significant drawback is the potential for overfitting, as the tree structure can continue to branch until it essentially memorizes the training data.

2.5.4.7 Random Forest

Random Forest (RF), introduced by Breiman in 2001, is an ensemble classifier comprising multiple Decision Trees (DTs). Each DT generates a prediction, and the final prediction is determined by majority voting, selecting the class with the most frequent predictions.

RF offers the advantage of mitigating overfitting compared to using a single DT, as it reduces bias by averaging predictions from the ensemble. However, its main disadvantage is the potential for decreased processing speed when increasing complexity, such as by adding more DTs to the ensemble.

2.5.4.8 Euclidean Distance

Euclidean Distance (ED) represents the distance between two points in Euclidean space. In some of the reviewed works, this metric was employed for classification. It involves computing the distance between facial features of a particular facial expression and the mean vector of facial features for each emotion. The emotion with the closest distance is then assigned to the input face. The formula for Euclidean Distance between two points (x, y) is defined as:

$$d(x, y) = \sqrt{(x_1 - y_1)2 + (x_2 - y_2)2}$$

The key advantage of this classifier is its straightforward implementation for identifying latent clusters. However, its simplicity can also be a drawback, particularly when dealing with high-dimensional data.

2.6 Trends and Patterns

In the landscape of Facial Expression Recognition Systems (FERS), several discernible trends and patterns have emerged, influencing the trajectory of emotion recognition technology. As highlighted in the existing research, these trends reflect the dynamic evolution of FERS methodologies and their potential impact on the field:

2.6.1 Advancements in Deep Learning

The continuous refinement of deep learning techniques, notably convolutional neural networks (CNNs) and recurrent neural networks (RNNs), has significantly contributed to the prowess of FERS. These architectures excel in learning intricate patterns from facial images, enhancing the accuracy of recognizing and interpreting facial expressions, aligning with the research emphasis on machine learning theory.

2.6.2 Transfer Learning Strategies

Transfer learning has become a notable trend within FERS, addressing the challenges associated with limited labeled datasets. The utilization of pre-trained models on extensive image datasets, as discussed in the research, aligns with this trend, enabling effective fine-tuning for specific emotion recognition tasks and improving generalization with constrained labeled data.

2.6.3 Emphasis on Facial Action Units (AUs)

The research emphasizes the dissection of facial expressions into individual Facial Action Units (AUs). This approach, aligning with the trend in FERS, recognizes the importance of analyzing nuanced muscle movements to gain a richer understanding of different emotions.

2.6.4 Multimodal Fusion Approaches

The integration of multiple modalities, such as facial expressions, voice, and physiological signals, is evident as a growing trend. This multimodal fusion aligns with the broader exploration of diverse feature sets in FERS, enhancing the robustness and accuracy of emotion recognition systems.

2.6.5 Real-time Processing and Edge Computing

The demand for real-time processing capabilities in FERS applications is reflected in the ongoing trend of exploring edge computing solutions. This aligns with the classification task and the need for efficient machine learning techniques, emphasizing the importance of faster response times, particularly in real-time scenarios.

2.6.6 Ethical Considerations and Responsible AI

The research recognizes the ethical implications associated with FERS deployment, aligning with the broader trend in the field. The emphasis on addressing bias, fairness, privacy, and responsible AI reflects a growing awareness of the ethical considerations in facial recognition technology.

2.6.7. Diverse and Inclusive Datasets

The research acknowledges the impact of biased datasets on model performance, echoing the trend of creating diverse and inclusive datasets. This aligns with the need for comprehensive training data that represents various demographic factors to ensure the generalizability of FERS models.

2.6.8 Explainability and Interpretability

The increasing complexity of FERS models and the need for transparency align with the growing trend of emphasizing explainability and interpretability. Understanding how these models arrive at specific predictions is crucial, as discussed in the research, for gaining user trust and addressing concerns related to automated emotion recognition.

2.6.9 Human-Centric Design Focus

The acknowledgment of human-centric design principles in the research resonates with the broader trend within FERS development. Understanding user experiences, preferences, and cultural nuances is increasingly recognized as pivotal in creating effective and socially acceptable emotion recognition systems.

2.6.10 Application Diversity in FERS

The research recognizes the diverse applications of FERS, extending beyond traditional research settings. This aligns with the broader trend of applying FERS in various domains, including human-computer interaction, mental health monitoring, market research, and the development of emotion-aware technologies.

These trends collectively illustrate the dynamic nature of FERS, as highlighted in the existing research, and emphasize the importance of staying attuned to ongoing advancements in the field. Researchers and practitioners should consider these trends in shaping the future trajectory of Facial Expression Recognition Systems, ensuring their relevance and effectiveness in diverse applications and scenarios.

2.7 Gaps in the Literature

Identifying gaps in existing research is a crucial step in establishing the significance of any new study. In the realm of Facial Expression Recognition (FER) systems, several noteworthy gaps emerge, offering avenues for further exploration and refinement. One prevalent gap lies in the limited discussion on the responsible use of facial recognition, potential biases in training data, and the broader societal implications, areas that warrant in-depth examination. Furthermore, existing literature often lacks a comprehensive comparative analysis of various FER approaches, making it challenging for researchers and practitioners to discern the most effective methods for specific applications.

Another notable gap pertains to the need for real-world applications and insights. While the technical intricacies of FER systems are well-documented, there is a scarcity of information on practical applications and the integration of these technologies into everyday scenarios. Understanding how FER systems perform in diverse environments and their usability in real-world settings remains an area that requires dedicated investigation.

My research aims to address these gaps by adopting a multidimensional approach. Firstly, it seeks to contribute to the discourse on FER by systematically analyzing the implications, biases, and considerations associated with these technologies. By shedding light on these aspects, the research aims to guide the development and deployment of FER systems in a more informed manner.

Secondly, my research endeavors to fill the void in comparative analyses of FER approaches. Through an exhaustive examination of existing methodologies, their strengths, weaknesses, and performance in various contexts, the study aims to provide researchers and practitioners with valuable insights for informed decision-making. This comparative aspect aims to enhance the understanding of which FER techniques are better suited for specific applications, fostering advancements in the field. Lastly, my research seeks to bridge the gap between theoretical knowledge and practical implementation by exploring real-world applications of FER systems. By investigating how these technologies perform in authentic environments and their impact on user experience, the study aims to offer a holistic perspective, guiding the future development and integration of FER technologies in diverse settings.

In summary, my research addresses critical gaps in the current landscape of FER research by delving into considerations, providing a comprehensive comparative analysis, and exploring the practical applications of FER systems. Through these contributions, the research aims to propel the field forward, fostering informed development and deployment of FER technologies.

2.8 Conclusion

In conclusion, this literature review has provided a comprehensive exploration of state-of-the-art techniques and methodologies that merge texture and shape features in facial expression recognition (FER). The synthesis of texture and shape information has emerged as a notable trend, addressing the limitations of unimodal approaches and enhancing the accuracy and robustness of FER systems, as highlighted in the discussed text.

Key takeaways from the literature review include the advancements in deep learning, the strategic use of transfer learning to overcome data limitations, and the emphasis on dissecting facial expressions into individual Facial Action Units (AUs). Additionally, the integration of multiple modalities, real-time processing, ethical considerations, and the importance of diverse and inclusive datasets have been identified as significant trends in the FER landscape.

The significance of this literature review to our research is pivotal. By exploring the amalgamation of texture and shape features, we align with the evolving methodologies in FER, aiming to contribute to the field's progression. The identified trends provide a robust foundation for our research, influencing the direction we take in developing ethical, inclusive, and real-world applicable FER systems.

As we embark on our research journey, the understanding gleaned from this literature review serves as a guide. The critical analysis of existing approaches, recognition of current gaps, and identification of future trends allow us to shape our research methodology and objectives effectively. Ultimately, the goal is to contribute meaningfully to the field of automatic facial expression recognition, enhancing accuracy and advancing the practical applications of these systems.

CHAPTER THREE: RESEARCH METHODOLOGY

3.1 Introduction

The fundamental aim of this chapter is to furnish an extensive and detailed roadmap, specifically tailored for the execution of our research objectives within the realm of Facial Expression Recognition (FER) systems. Within this methodology chapter, we endeavor to present a comprehensive guide that intricately covers the development, implementation, and evaluation phases of our FER system.

Our meticulously crafted methodology is strategically designed to tackle pivotal questions surrounding the development, implementation, and evaluation of our FER system. This involves a detailed explanation of the chosen methods, a thorough examination of the rationale behind their selection, and a nuanced exploration of how these methodologies contribute to the achievement of our research objectives in the context of FER.

This chapter places significant emphasis on advocating for a systematic approach in the execution of FER research. It propounds a structured framework that guarantees the conduct of our FER research in a methodical, transparent, and reproducible manner. This not only serves to bolster the credibility of our research findings within the FER domain but also facilitates their validation by peers and other researchers operating in the field of facial expression recognition.

As we delve deeper into the subsequent sections, we will embark on a more profound exploration of the intricacies of our FER methodology. A step-by-step elucidation will be provided for each stage of our FER research process, encompassing the meticulous design of our FER study, the judicious selection and collection of facial expression data, the intricate analysis of this FER data, and the nuanced interpretation of our findings within the FER context.

Upon concluding this chapter, it is anticipated that readers will possess a comprehensive understanding of our FER research process and the methodologies intricately woven into our FER study. This knowledge will serve as a robust foundation for the forthcoming chapters, wherein we delve into the presentation and discussion of our FER research findings.

3.2 Research Design

The research aimed to develop a neural network model for recognizing various facial expressions. In a related study, Alexander M. Pascual (2014) effectively utilized image acquisition, feature extraction, facial expression classification, and the calculation of classification performance in a Lightweight Facial Emotion Recognition System on Edge Devices. Moreover, the research constructed an application based on prevalent technologies. This application takes an image as input, processes the image, and subsequently infers the facial expression depicted within the image.

3.2.1 Research Methods

Effective research methods serve as fundamental tools for gathering information. Without a well-designed and executed research methodology, the quality of information collected is compromised, undermining subsequent reviews, evaluations, or future strategies.

Quantitative Research focuses on quantifying data, expressing it numerically, and seeking answers to questions such as 'how long,' 'how many,' or 'to what degree.' It is employed to study events or levels of occurrence, aiming to measure the incidence of various views and opinions in a chosen sample or aggregate results. The advantage lies in the ability to statistically analyze numerical outcomes, providing credible and data-driven insights for decision-making.

Qualitative Research, on the other hand, is concerned with the quality of information. It seeks to understand the underlying reasons and motivations for actions, exploring how people interpret their experiences and the world around them. Qualitative methods generate insights, facilitate the formulation of ideas and hypotheses. The effectiveness of qualitative research heavily relies on the skills and abilities of researchers, and the outcomes may not always be considered reliable due to their dependence on researchers' personal judgments and interpretations. Qualitative research is more suitable for small samples and may not necessarily represent the opinions of a wider population (Bell, 2005).

In this research, the aim is to gain deeper insights and facilitate meaningful learning regarding the effectiveness and implementation of FER system. Hence, a mixed methods research approach is chosen. Mixed methods research requires additional time due to the need to collect and analyze two different types of data (Creswell & Plano Clark, 2011). However, the vast benefits of mixed methods research are significant. Integration provides readers with increased confidence in the results and conclusions drawn from the study (O'Cathain, Murphy, & Nicholl, 2010). It also helps researchers generate ideas for future research (O'Cathain et al., 2010). Furthermore, researchers argue that mixed methods research is the only way to be certain of findings and interpretations (Morse & Chung, 2003; Tashakkori & Teddlie, 2003b).

3.2.2 System Development Methodology

The project was devised using the Rapid Application Development (RAD) methodology, emphasizing swift prototyping over extensive planning (Chrismanto et al., 2019). Within the RAD model, a prototype represents a fully functional model equivalent to a specific product component. This approach involves the concurrent development of functional modules as prototypes, which are then integrated to create the final product, facilitating faster product delivery. The absence of detailed pre-planning allows for the incorporation of changes during the development process.

As stated by Qudus Khan et al. (2020), RAD projects adhere to an iterative and incremental model, with small teams consisting of developers, domain experts, customer representatives, and other IT resources collaboratively working on their respective components or prototypes. A critical factor for the success of this model is ensuring that the developed prototypes are reusable (Hucka et al., 2019).

3.2.2.1 Four Phases of RAD

The Rapid Application Development (RAD) methodology typically encompasses four phases, described as follows:

1. Requirements Planning Phase: In this initial phase, elements from the System Planning and Systems Analysis phases of the System Development Life Cycle (SDLC) are integrated. During this stage, users, administrators, and IT staff engage in discussions to reach a consensus on business needs, project scope, constraints, and system requirements.
2. User Design Phase: This phase involves user collaboration with systems analysts to develop models and prototypes that comprehensively represent all system processes, inputs, and outputs. Users actively participate in shaping the system's design.
3. Construction Phase: The Construction Phase is akin to the program and application development tasks in the SDLC. It focuses on the actual development of the system, including coding and implementation.
4. Implementation Phase: Similar to the final tasks in the SDLC's implementation phase, this stage involves activities like data conversion, rigorous testing, transitioning to the new system, and providing user training to ensure a seamless integration of the new system into the operational environment.

Notably, in the RAD methodology, the analysis, design, and implementation phases are executed concurrently. Furthermore, these three phases are carried out iteratively, and repetition occurs until the system reaches completion. This iterative and incremental approach, facilitated by the use of prototypes, ensures visible progress throughout the research project (Dennis, Wixom, & Tegarden, 2009).

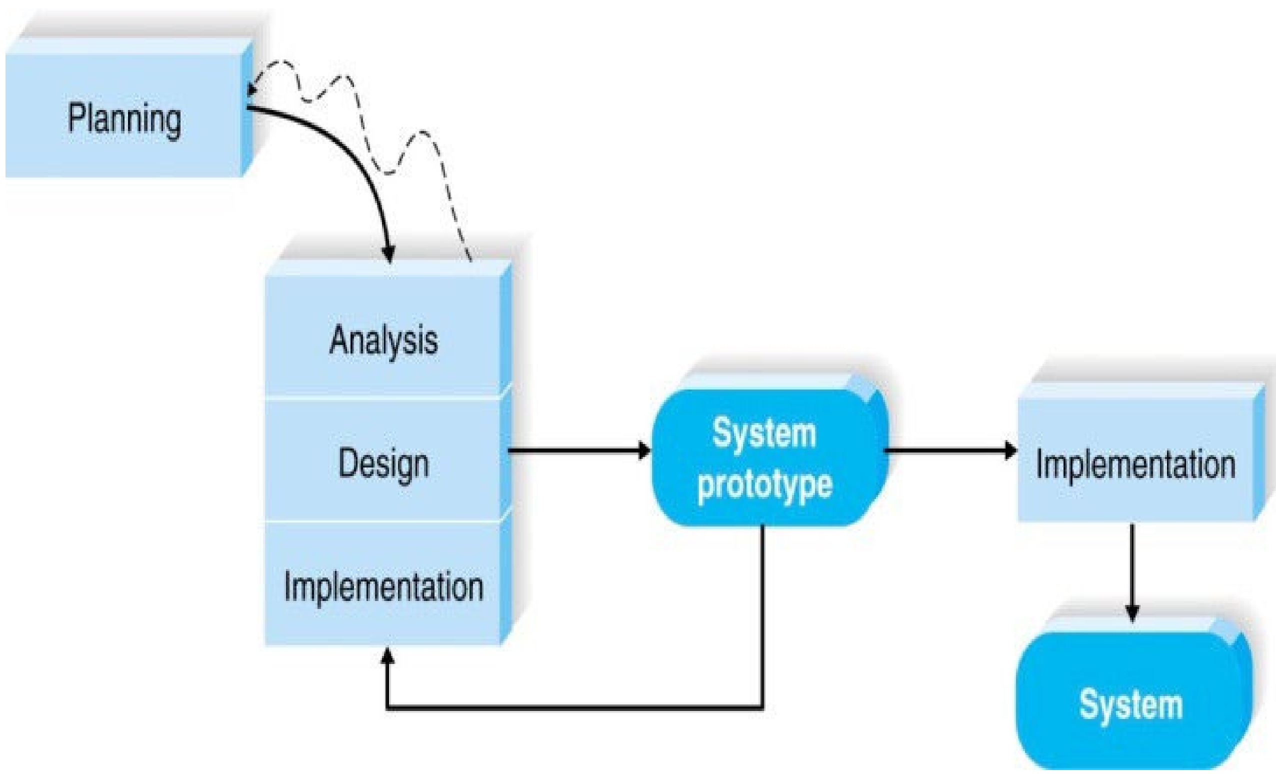


Figure 11. Prototyping Methodologies (Dennis, Wixom, & Tegarden, 2009).

Both the convolutional neural network (CNN) and the web application were crafted using an evolutionary prototyping approach. The utilization of prototyping was imperative in the development of the CNN, primarily due to the experimental nature of certain hyperparameters, such as the number of training steps.

3.3 Participants or Sample

In the realm of Facial Expression Recognition (FER) systems, the selection of participants or samples is a critical aspect that directly impacts the system's efficacy. Consider a hypothetical research study aiming to evaluate the performance of a newly developed FER algorithm. The participants in this study would consist of individuals from diverse demographic backgrounds, encompassing varying age groups, genders, and cultural representations.

3.3.1 FER-2013 Database

3.3.1.1 Selection Process

The FER-2013 Database is a publicly available dataset containing grayscale images of facial expressions sourced from online images. Ethical considerations primarily revolve around the anonymization of participants, given that these images were collected from online sources without direct participant involvement.

Sample Size and Characteristics

- **Total Number of Images:** With a substantial 28,709 images, the FER-2013 Database provides a diverse range of facial expressions for analysis.
- **Dimensions:** The images are uniformly sized at 48x48 pixels, ensuring efficient processing and analysis.

3.3.1.2 Environment

The FER-2013 dataset can be considered semi-controlled. While the images were not captured under strict laboratory conditions, they were curated from online sources, introducing some variability in lighting conditions and backgrounds.

3.3.1.3 Ethical Considerations

In utilizing FER-2013, ethical considerations involve respecting the privacy of individuals in the online images and ensuring that the dataset is used solely for research purposes.

3.3.1.4 Data Collection

Data collection involves curating facial expression images from publicly available online sources. While this offers a vast dataset, it introduces challenges related to image quality and uncontrollable environmental factors.

3.3.2 The Extended Cohn-Kanade Database (CK+)

3.3.2.1 Selection Process

The CK+ Database, a highly regarded dataset, involves participants providing informed consent for capturing posed facial expressions in a controlled laboratory setting.

Sample Size and Characteristics

- Total Number of Images: The CK+ Database includes 593 sequences from 123 participants, capturing a range of facial expressions with high-quality images.
- Dimensions: Images in CK+ vary in size, with dimensions typically exceeding 640x480 pixels, offering detailed facial features for nuanced analysis.

3.3.2.2 Environment

CK+ is captured in a controlled laboratory environment with carefully managed lighting conditions, ensuring consistency and precision in facial expression analysis.

3.3.2.3 Ethical Considerations

Ethical considerations in CK+ involve obtaining informed consent from participants and ensuring the responsible use of their facial expression data for research purposes.

3.3.2.4 Data Collection

Data collection in CK+ involves capturing participants' facial expressions in a controlled laboratory environment, providing high-quality and well-controlled images for analysis.

3.4 Data Collection

To ensure the effective training of the CNN, a substantial volume of images is imperative. This aligns with the understanding that a neural network's performance tends to improve proportionally with the size of the dataset employed for its training, as documented by Brynjolfsson and McAfee in "What's Driving the Machine Learning Explosion" (2017).

For this research, the facial expression images were sourced from the following public Face Image Databases:

3.4.1 Facial Expression Recognition 2013 Database (FER-2013)

The Facial Expression Recognition 2013 database (FER-2013) (Goodfellow, Carrier, & Courville, 2013) was created using the Google image search API to search for images of faces that match a set of 184 emotion-related keywords like "blissful," "enraged," etc. These keywords were combined with words related to gender, age, or ethnicity, leading to 35,887 grayscale images with a 48 x 48 resolution, mapped into the six basic emotions, plus the neutral expression.

Figure 12. FER-2013 database with 7 expressions.

3.4.2 The Extended Cohn-Kanade Database (CK+)

The Extended Cohn-Kanade database (CK+) (Kanade, 2010) contains 593 image sequences of posed and non-posed emotions. In addition, 123 participants were 18 to 50 years of age, 69% female, 81% Euro-American, 13% Afro-American, and 6% other groups. The images were digitized into either 640 x 490 or 640 x 480 resolution and are mostly gray. Each sequence was built on frontal views and 30-degree views, starting with a neutral expression up until the peak emotion (last frame of the sequence). Most sequences are labeled with eight emotions: anger, disgust, contempt, fear, neutral, happiness, sadness, and surprise.

Figure 13. CK+ database with 8 expressions.

3.4.3 Data Pre-processing

3.4.3.1 Augmentation

Within the realm of data pre-processing, image augmentation emerges as a critical technique employed to artificially augment the training dataset's size. This augmentation technique encompasses a spectrum of image transformations, serving to diversify the dataset. These operations include flipping, rotation, shearing, zooming, cropping, deforming, as well as adjustments to hue, saturation, brightness, and contrast, as discussed in "Custom Image Augmentation" (Towards Data Science, 2017).

This augmentation process plays a pivotal role in enhancing the model's ability to generalize effectively, particularly when exposed to images that feature distortions mirroring those encountered in real-world scenarios.

Figure 14. Data Augmentation results.

3.4.3.2 Standardization

Another crucial facet of data pre-processing involved standardization. This entailed resizing all the images to a uniform dimension of 224 pixels by 224 pixels and ensuring they shared the same image format, which is JPG. This standardization process aligns with the concept of normalization, a necessary step in transfer learning for architectures like MobileNetV2. It establishes a consistent foundation for the model to operate effectively.

3.4.3.3 Folder Structure

The dataset was meticulously organized into a structured folder system. Approximately 70% of the total images, amounting to 25,536 images, were designated for the training dataset, while the remaining 30%, comprising 10,944 images, constituted the validation dataset. This structured arrangement is visually represented in Figure 15, wherein each subfolder's name is indicative of the specific disease label that would be subsequently assigned to the images it contains.

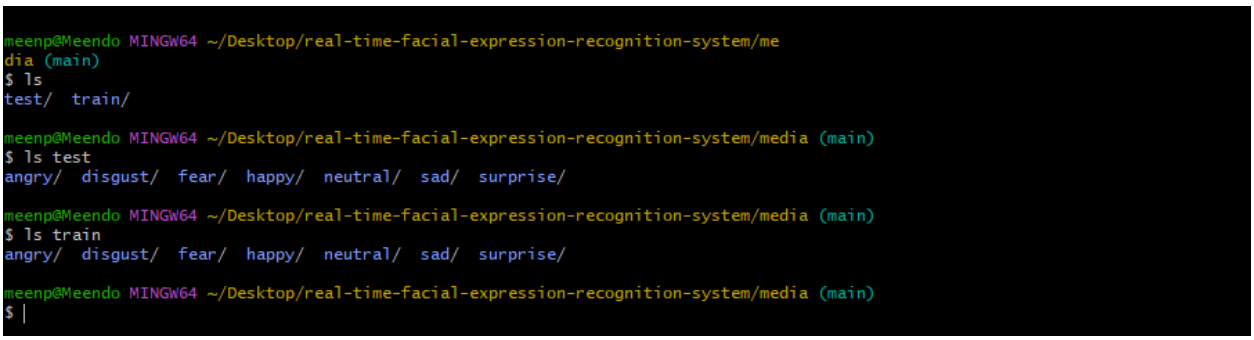


Figure 15. CNN Dataset Folder Structure.

3.5 Data Analysis

The data analysis phase of this research employs advanced techniques to extract meaningful insights from the collected facial expression data. The primary methodologies include the application of machine learning algorithms, statistical analyses, and qualitative content analysis.

3.5.1 Machine Learning Algorithms (CNNs

Convolutional Neural Networks have become synonymous with breakthroughs in machine vision, and their application in FER is a testament to their unparalleled capabilities. CNNs are specifically designed to process and analyze visual data, making them inherently suitable for tasks involving image recognition and interpretation.

3.5.1.1 Facial Feature Extraction

In the context of FER, the primary objective is to extract discriminative facial features that encapsulate emotional expressions. CNNs are inherently adept at this task due to their architecture, which includes convolutional layers capable of learning hierarchical representations. These layers excel at capturing intricate patterns and subtle variations in facial features, enabling the network to discern the unique characteristics associated with different emotions.

3.5.1.2 Emotion Classification

CNNs are particularly well-suited for the complex task of emotion classification. Trained on diverse datasets, these networks develop a profound understanding of the distinctive features associated with various emotions. Through the process of supervised learning, CNNs map facial expressions to specific emotional states, facilitating accurate classification.

3.5.1.3 Automatic Learning of Hierarchical Representations

One of the key strengths of CNNs lies in their ability to automatically learn hierarchical representations. As the network processes input data, it progressively extracts features at different levels of abstraction. In FER, this means that the network can learn not only basic facial features but also higher-level representations that encapsulate complex emotional expressions.

3.5.1.4 Implementation with Python, TensorFlow, and OpenCV

The implementation of our FER analysis is conducted using the Python programming language, known for its versatility and extensive support in the machine learning community. The robustness of our methodology is enhanced by leveraging the capabilities of TensorFlow and OpenCV libraries:

- **Python:** Renowned for its readability and vast ecosystem of libraries, Python serves as the backbone for developing and deploying intricate machine learning models.
- **TensorFlow:** An open-source machine learning framework, TensorFlow facilitates the seamless construction and training of neural networks, including CNNs. Its scalability and flexibility are paramount in achieving optimal model performance.

- **OpenCV:** In conjunction with TensorFlow, OpenCV adds a layer of sophistication to our FER system. Its computer vision functionalities are crucial for tasks such as facial detection, image preprocessing, and feature extraction.

3.5.2 Statistical Analyses and Qualitative Content Analysis

To comprehensively assess the model's performance, our analysis extends beyond quantitative metrics to include both statistical measures and qualitative content analysis:

- **Statistical Analyses (Confusion Matrices):** Rigorous evaluation is achieved through the use of confusion matrices. These matrices provide a detailed breakdown of classification results, offering insights into true positives, true negatives, false positives, and false negatives. This quantitative assessment serves as a vital metric for understanding the accuracy and efficacy of our FER system.

- **Qualitative Content Analysis:** In addition to quantitative measures, our analysis incorporates qualitative content analysis. This involves a nuanced examination of contextual information, non-verbal cues, and subjective elements of emotions. By combining quantitative accuracy measures with a deeper qualitative understanding, our research aims to capture the richness and subtleties inherent in human facial expressions.

3.5.3 Training the Model

To train the model, a transfer learning approach was employed. Specifically, Google's MobileNetV2 was chosen as the pre-trained model for feature extraction, owing to its superior performance in terms of accuracy compared to the previous iteration of MobileNet (Sandler, Howard, Zhu, Zhmoginov, & Chen, 2019).

The training process involved loading the pre-trained MobileNetV2 model using TensorFlow, as illustrated in Figure 3.3. Subsequently, this pre-trained model was retrained using the designated training dataset. To evaluate the effectiveness of the retrained model, the validating dataset was employed for validation purposes.

In the pursuit of optimal model performance, a transfer learning approach was employed. Specifically, Google's MobileNetV2, chosen for its superior accuracy, served as the pre-trained model for feature extraction. This pre-trained model, loaded using TensorFlow, was subsequently retrained on the designated dataset to tailor its capabilities for the nuances of facial emotion recognition. The validation dataset played a crucial role in assessing the effectiveness of the retrained model.

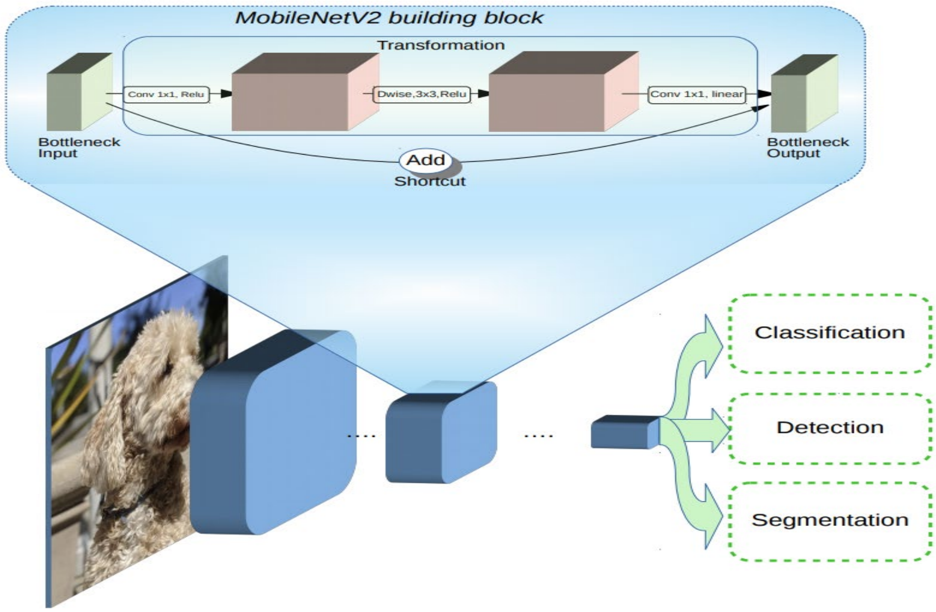


Figure 16. MobileNetV2 Architecture (Google AI Blog, 2018).

3.5.4 Data Presentation

The Data Presentation section serves as the lens through which the outcomes of our Facial Emotion Recognition (FER) analysis come to light. To effectively communicate the results and glean meaningful insights, we employ a multifaceted approach involving visualizations, confusion matrices, and qualitative descriptions.

3.5.4.1 Visualizations of Classification Accuracy

To provide a high-level overview of our model's performance, we employ visualizations of classification accuracy. Through graphical representations such as line charts or bar graphs, we depict the accuracy rates for different emotional categories. These visualizations offer a quick and accessible snapshot of how well our FER system recognizes various facial expressions.

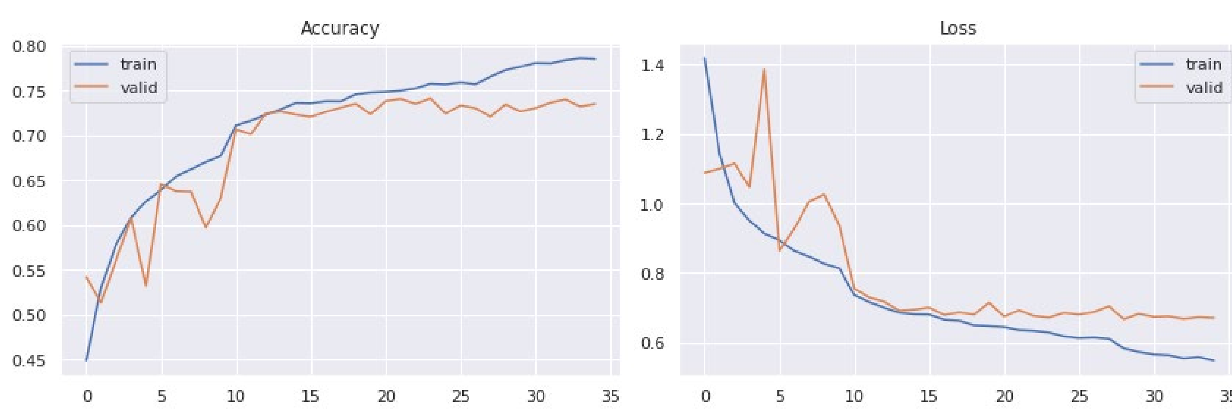


Figure 17. Visualizations of Classification Accuracy.

3.5.4.2 Confusion Matrices

Delving deeper into the intricacies of our model's performance, confusion matrices are employed. These matrices offer a granular breakdown, detailing true positives, true negatives, false positives, and false negatives. By visualizing the confusion matrix, we gain insights into specific areas of strength and potential improvement, fostering a more nuanced understanding of the classification results.

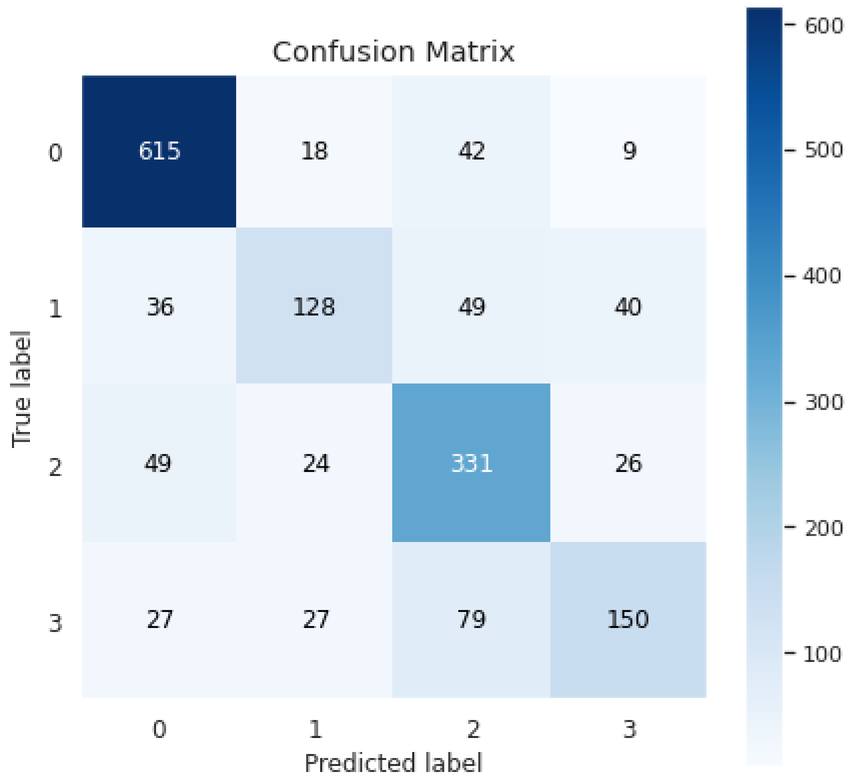


Figure 18. Confusion Matrices.

3.5.5 Contribution to Existing FER Literature

Our data presentation not only stands as an independent evaluation of the developed FER system but also contributes to the broader FER literature. Comparisons with prior studies, discussions on novel methodologies, and insights gained from the analysis are highlighted. This contribution involves addressing identified limitations, proposing avenues for future research, and introducing innovative approaches that advance the field.

In summary, our Data Presentation section weaves together quantitative and qualitative elements to offer a comprehensive and insightful depiction of our FER analysis. This approach ensures not only the effective communication of results but also a valuable contribution to the evolving landscape of Facial Emotion Recognition research.

3.6 Research Quality

The quality of research data hinges on two critical aspects: validity and reliability. Validity pertains to the degree to which the data accurately measures the intended parameters, while reliability indicates whether the data collection method would yield consistent results if replicated by other researchers.

In this research, validity was ensured through the application of automated model testing, incorporating the use of TensorFlow. This tool facilitated the visual tracking of the convolutional neural network model's loss and learning rate, streamlining the debugging and optimization processes, thus enhancing the accuracy of the model.

To bolster the research's reliability, a series of steps were taken. This included maintaining meticulous documentation and adhering to best practices in web application development. These measures were implemented to guarantee the repeatability of the research findings and the consistency of results when evaluated by other researchers. Additionally, transfer learning was employed, leveraging Google's MobileNetV2 architecture, which is renowned for its accuracy, further enhancing the credibility of the research outcomes.

3.7 Conclusion

In conclusion, this chapter has meticulously outlined the strategies and methodologies employed in our study on Facial Expression Recognition (FER) systems. By delving into the intricacies of the research design, system development methodology, and the various steps undertaken to ensure research quality, we have established a robust foundation for the subsequent phases of our investigation.

The adoption of a mixed methods research approach, despite its time-intensive nature, brings significant benefits in terms of result integration, reader confidence, and the generation of ideas for future research. Our chosen research design, which focuses on neural network models for facial expression recognition, draws inspiration from prior studies, demonstrating a thoughtful and informed approach to problem-solving.

The selection of the Rapid Application Development (RAD) methodology for system development aligns seamlessly with the dynamic nature of our project. The four phases of RAD, emphasizing user collaboration, construction, and implementation, ensure a flexible and adaptive development process, crucial for a project with experimental components like the convolutional neural network and web application.

Our detailed exploration of the CNN model development process, from data collection to testing, showcases a meticulous approach to building a robust facial expression recognition system. Leveraging public face image databases, applying data pre-processing techniques such as augmentation and standardization, and utilizing transfer learning with Google's MobileNetV2 architecture contribute to the credibility and effectiveness of our model.

The system development methodology, inclusive of an evolutionary prototyping strategy, illustrates a pragmatic and adaptive approach to constructing both the CNN and web application components. The concurrent development of prototypes, as facilitated by RAD, ensures visible progress and allows for real-time adjustments.

Concerning research quality, our steps to ensure validity through automated model testing with TensorFlow and reliability through meticulous documentation and adherence to best practices contribute to the overall credibility of our research outcomes. The reliance on transfer learning with MobileNetV2 further enhances the reliability of our facial expression recognition model.

In summary, this methodology chapter lays a solid groundwork for the subsequent stages of our study. It not only provides a clear understanding of the research design, system development methodology, and the steps taken to ensure the quality and credibility of our research outcomes but also sets the stage for the next phases of our investigation.

CHAPTER FOUR: SYSTEM ANALYSIS

4.1 Introduction

The primary aim of the system analysis phase is to precisely define what the system must accomplish in order to fulfill the requirements of end users. These specifications, established during the analysis phase, are subsequently transformed into a hierarchical structure of charts during the systems design phase. This structure outlines the necessary data and the processes that will be applied to this data. It enables these processes to be translated into instructions for users when they are interacting with the system (Gemino et al., 2009).

4.2 Data Collection

Data collection is the systematic process of acquiring and quantifying information about variables of interest. This methodological approach allows for the retrieval of essential data in an organized manner, enabling the researcher to address defined research inquiries, test hypotheses, and assess outcomes.

4.2.1 Data Collection Methods

Throughout the data collection phase of system development, the following methods will be employed:

i. Questionnaires

ii. Interviews

iii. Observation

These techniques serve as essential tools for gathering the required information and insights to inform the development process.

4.2.1.1 Questionnaires

A questionnaire is a research tool comprising a set of structured questions designed to collect information from respondents. Questionnaires can be administered in various formats, including written, face-to-face, telephone, computer-based, or postal. They offer an economical, expeditious, and efficient means of amassing substantial data from a sizable sample of individuals.

One of the key advantages of questionnaires is the rapid collection of data, as the researcher does not need to be physically present during their completion. This method is particularly advantageous when dealing with large populations, where conducting interviews would be logistically challenging. In practice, the questionnaires will be distributed to the target group for data collection. Respondents will fill out the questions on the provided forms, which are then returned for subsequent data analysis.

4.2.1.2 Interviews

An interview serves as a purposeful conversation aimed at collecting information. Typically, a research interview involves two parties: an interviewer, responsible for guiding the conversation and posing questions, and an interviewee, who provides responses to those questions. In this research, interviews will be conducted by formulating a structured set of questions that will be presented to both staff and customers at all branches of University Departments.

These interviews can be carried out through face-to-face interactions or over the telephone, allowing for a direct exchange of information between the interviewer and the respondents.

4.2.1.3 Observation

Observation is a methodical data collection technique that entails researchers utilizing all of their senses to assess individuals within their natural environments or unscripted circumstances. This approach involves the analyst immersing themselves in the organization to observe the system's operations directly.

Through observation, the obtained information offers an authentic and unfiltered portrayal of the system, distinct from other data collection methods. Engaging in prolonged immersion within the setting or social context being studied is essential for gaining comprehensive insights through observation.

4.3 Feasibility Study

A feasibility study is a comprehensive analysis that considers all pertinent aspects of a project, encompassing economic, technical, legal, and scheduling factors. Its purpose is to determine the likelihood of successfully completing the project. Project managers employ feasibility studies as a critical tool to evaluate the advantages and disadvantages of embarking on a project before committing substantial resources in terms of time and finances.

4.3.1 Technical Feasibility

Technical feasibility assessment centers on the organization's available technical resources. Its primary aim is to determine if the technical team possesses the capability to transform conceptual ideas into functional systems. This evaluation includes an examination of the necessary hardware, software, and other technical prerequisites for the proposed system.

4.3.1.1 Hardware Analysis

Training a deep neural network such as a CNN demands substantial computational power, and the use of a graphics processing unit (GPU) is highly recommended (LeCun, Bengio, & Hinton, 2015). In this research, a GPU provided by colab.research.google.com was utilized to expedite the model training process.

Furthermore, for the development process, a laptop or computer with the following specifications will be essential:

a) A minimum of 8GB RAM (Random Access Memory) to facilitate swift loading of training and testing platforms.

b) A processor speed of at least 2.5 GHz, which is instrumental in expediting the processing of requests and enhancing response times during code testing.

c) A hard disk capacity of at least 40GB, serving as the storage space for the application's workspace, housing the dataset and source code.

d) At least one Network Interface Card (NIC) for connectivity purposes.

4.3.1.2 Software Analysis

In this research, the development of the CNN relied on the TensorFlow machine learning framework in combination with Python. TensorFlow was chosen due to its open-source nature, flexibility, and the comprehensive set of tools it offers, making it highly suitable for the training and validation of the model (TensorFlow, 2018).

Additional software requirements encompass:

a) **Anaconda IDE**: Anaconda serves as a package management system, offering access to packages and libraries that are essential for large-scale data processing and the development of machine learning applications. It plays a crucial role in ensuring the availability of the necessary resources for the research.

b) **OpenCV (Open-Source Computer Vision Library)**: OpenCV is essential for computer vision tasks, image processing, and pattern recognition. It provides a range of functions and tools for analyzing and manipulating visual data, making it a valuable asset in various aspects of the research project.

4.3.2 Economic Feasibility

The aim of economic feasibility is to evaluate the project's potential positive economic benefits, involving quantification and identification of these benefits.

- **Positive Benefits:** The system's economic feasibility is substantiated by the presence of all essential requirements, including the data collected from sources like the COHN-KANADE and FER 2013 databases.

- **Challenges:** Consider potential risks or challenges that might affect the economic feasibility of the project.

4.3.3 Social Feasibility

Social feasibility takes into account the acceptance of the system by end users. It assesses the impact of the new system's introduction on users, considering factors such as the need for retraining the workforce.

- **User Engagement:** Ensure active engagement of end users throughout the project, especially during the development and testing phases.

4.3.4 Legal Feasibility

Legal feasibility assessment is a critical examination to determine whether any elements of the proposed project run counter to legal prerequisites, such as data protection acts or social media laws.

The system is legally accountable as it involves risk management and necessitates the resolution of pertinent technical issues to ensure compliance with legal mandates and regulations. It proactively addresses potential legal challenges to align with the established legal framework.

4.3.5 Operational Feasibility

Operational feasibility serves as an evaluation of how effectively a proposed system addresses the problem at hand and seizes the opportunities outlined during scope definition. In the context of the project, the following aspects were taken into consideration for operational feasibility:

- The system's ability to detect and capture facial images.
- The system's capacity to accurately identify the category of the captured image.

4.4 System Requirement Specification

This section delineates the system requirements in alignment with the research objectives. The requirements, which have been identified through content analysis methodology and are congruent with the research study's objectives, can be categorized into two main sections: functional and non-functional requirements.

4.4.1 Functional Requirements

i. **FER System Input:**

The FER system should provide the user with the option to input an image or capture a photo using the device's camera. The selected image will serve as the input for the facial expression recognition system.

ii. **Facial Expression Analysis:**

The FER system should possess the capability to analyze and identify the facial expression depicted in the input image accurately.

4.4.2 Non-functional Requirements

i. **Portability**

The application should be compatible with a variety of web browsers and operating systems to ensure accessibility for a broad user base.

ii. **User-Friendliness**

The user interface of the application should be intuitive and in line with established web design principles, offering a seamless user experience. It should also be responsive, adapting to different screen sizes and resolutions.

iii. **Performance**

The system should deliver responses and results promptly, providing users with an efficient and smooth interaction with the web application.

iv. **Availability and Stability**

The application must remain consistently available without interruptions and maintain stability, minimizing unexpected crashes or downtime during use.

v. **Maintainability**

The application should be designed for easy maintenance and support, guaranteeing its long-term sustainability and dependable operation.

Additional Consideration:

- **Traceability Matrix**

Include a brief subsection on the traceability matrix, mapping requirements to test cases, to facilitate quality assurance and testing phases.

- **Security Requirements**

Emphasize security requirements to protect user privacy and comply with data protection regulations, given the sensitive nature of facial recognition data.

4.5 Conclusion

In conclusion, the system analysis phase serves as a crucial foundation for the successful development of the Facial Expression Recognition (FER) system. Through meticulous data collection methods such as questionnaires, interviews, and observation, valuable insights were gained, laying the groundwork for informed decision-making. The feasibility study further ensured a comprehensive assessment of technical, economic, social, legal, and operational aspects, validating the viability of the project.

The technical feasibility analysis delved into the hardware and software requirements, emphasizing the need for computational power, GPU utilization, and specific software tools like TensorFlow, Anaconda, and OpenCV. Economic feasibility highlighted the positive benefits derived from essential data sources, while social feasibility stressed the importance of user engagement and workforce retraining. Legal feasibility underscored the system's accountability, aligning with legal mandates, and operational feasibility evaluated the system's effectiveness in detecting and categorizing facial expressions.

The system requirement specification detailed functional and non-functional requirements, emphasizing the FER system's ability to input, analyze, and identify facial expressions accurately. Non-functional requirements, including portability, user-friendliness, performance, availability, stability, and maintainability, were outlined to ensure a seamless and reliable user experience. Additionally, considerations for a traceability matrix and security requirements were highlighted to enhance the quality assurance and protection of user privacy.

By addressing these aspects comprehensively, the FER system is poised to meet user expectations, adhere to legal and ethical standards, and contribute positively to the field of facial expression recognition. The culmination of these analyses and specifications lays the groundwork for the subsequent phases of system development, guiding the project toward successful implementation and user satisfaction.

CHAPTER FIVE: SYSTEM DESIGN

5.1 Introduction

In this chapter, we outline the system design of a facial expression recognition system capable of identifying and categorizing various facial expressions. The components of the implemented system, the interplay among these components, and the dynamics between users and the system are elucidated. The interactions are modeled using Data Flow Diagrams (DFD) and depicted through the context diagram and level 1 DFD.Current System Diagram

5.2 Current System Diagram

The research work can be described by the block diagram shown in fig.19

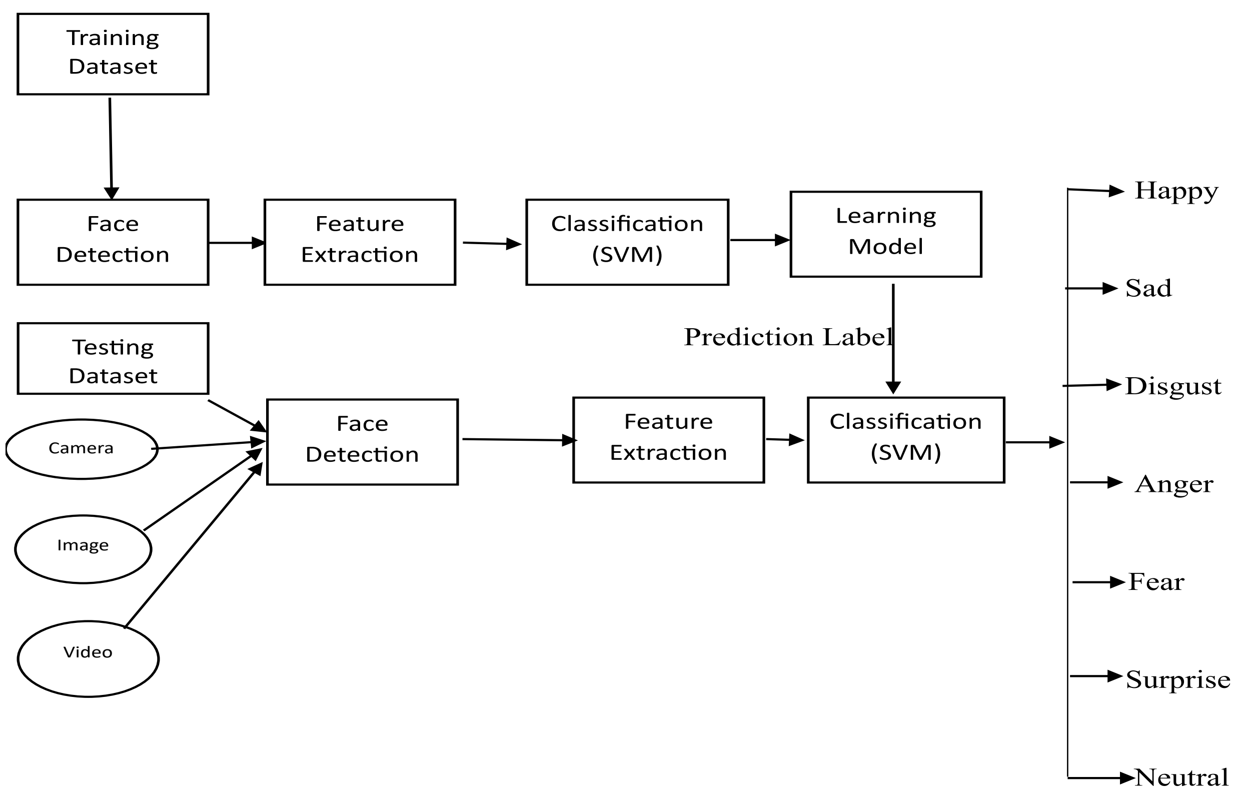


Figure 19. Block Diagram of the System.

5.3 System Architecture

A convolutional neural network (CNN) is used for expression classification. CNNs are a type of neural network that are well-suited for image recognition tasks. The CNN architecture used in this system is shown in fig.20.

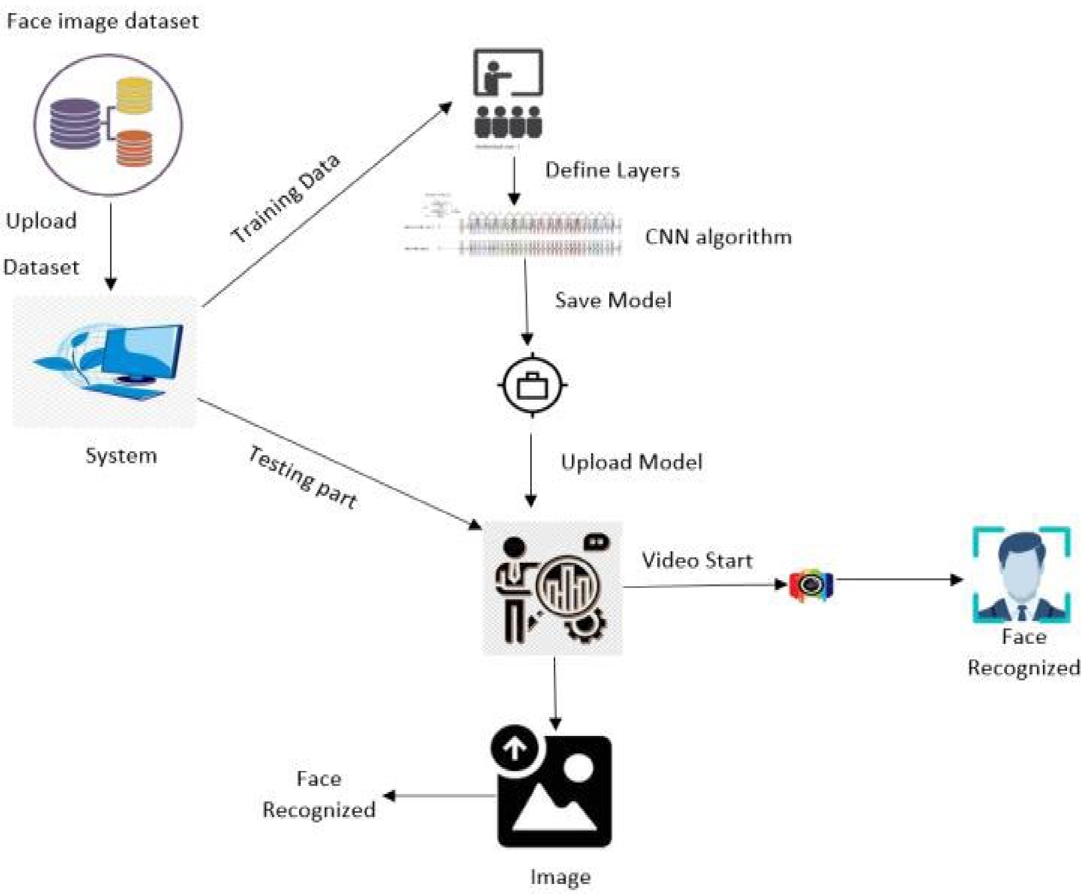


Figure 20. CNN Architecture.

The system architecture consists of three main components:

1. Face Detection: This component is responsible for detecting faces in images or videos.
2. Feature Extraction: This component extracts relevant features from the detected faces, such as the position of facial landmarks.
3. Expression Classification: This component classifies the extracted features into one of six basic facial expressions: happiness, sadness, anger, fear, surprise, or disgust.

5.4 Data Flow Diagrams (DFD)

A data flow diagram (DFD) is a common method for illustrating how information moves throughout a system. A good deal of the system requirements can be graphically depicted in a clean and clear DFD. It may be manual, automated, or a hybrid of the two. It demonstrates how information enters and exits the system, what modifies the information, and where information is stored. A DFD's main function is to outline the scope and bounds of a system as a whole. It may be utilized as a tool for communication between a systems analyst and any individual who plays a component in the system that serves as the foundation for redesigning a system.

5.4.1 Context Level Diagram

The context level diagram for the facial expression recognition system is shown in fig.21.

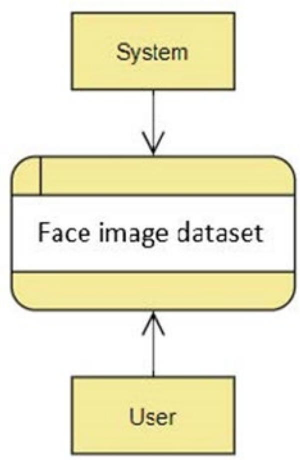


Figure 21. Context Level Diagram

5.4.2 Level-1 Diagram

The level-1 diagram for the facial expression recognition system is shown in Figure 22.

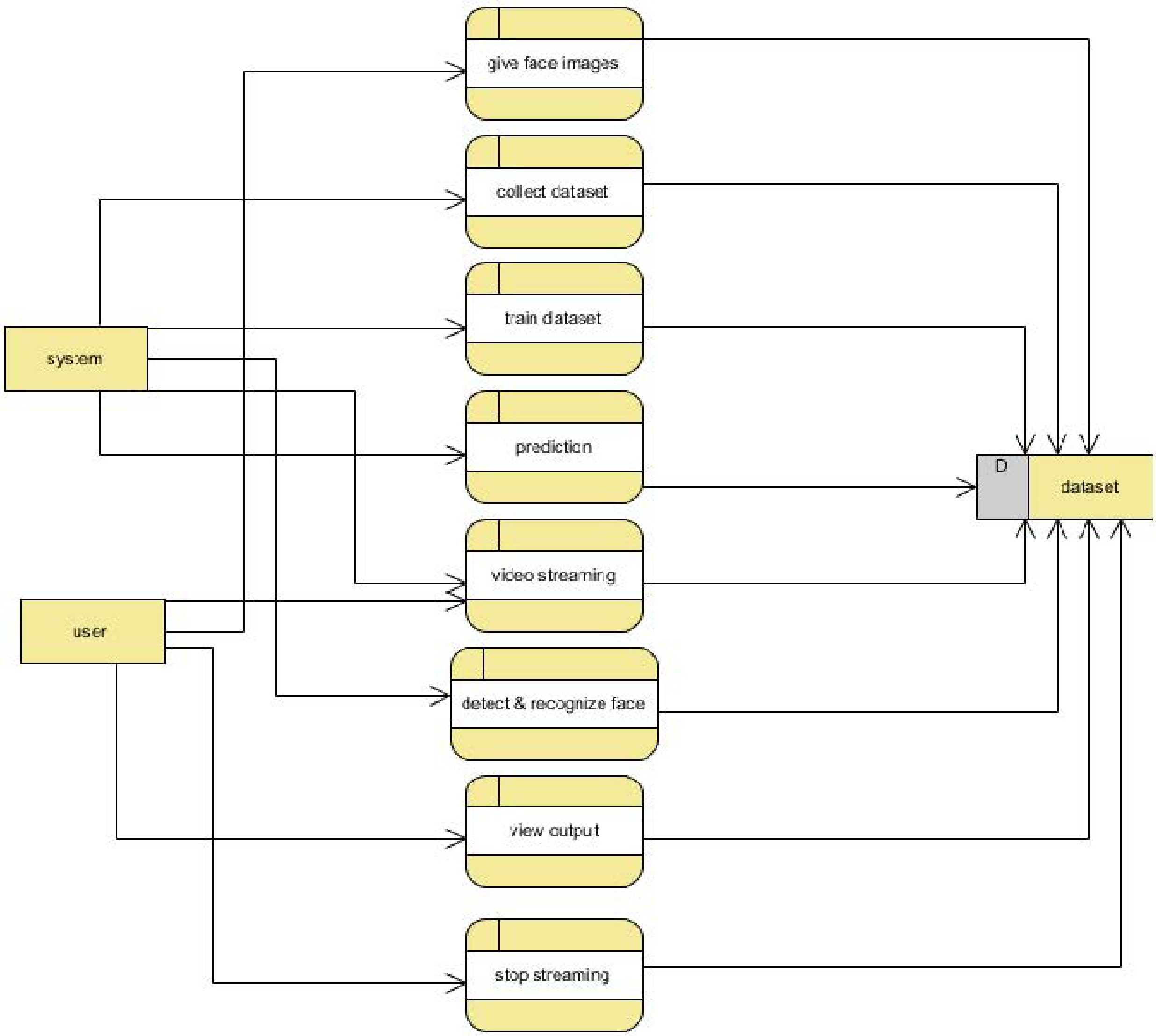


Figure 22. Level-1 Diagram.

5.4.3 Level-2 Diagram

The level-2 diagram for the facial expression recognition system is shown in Figure 23.

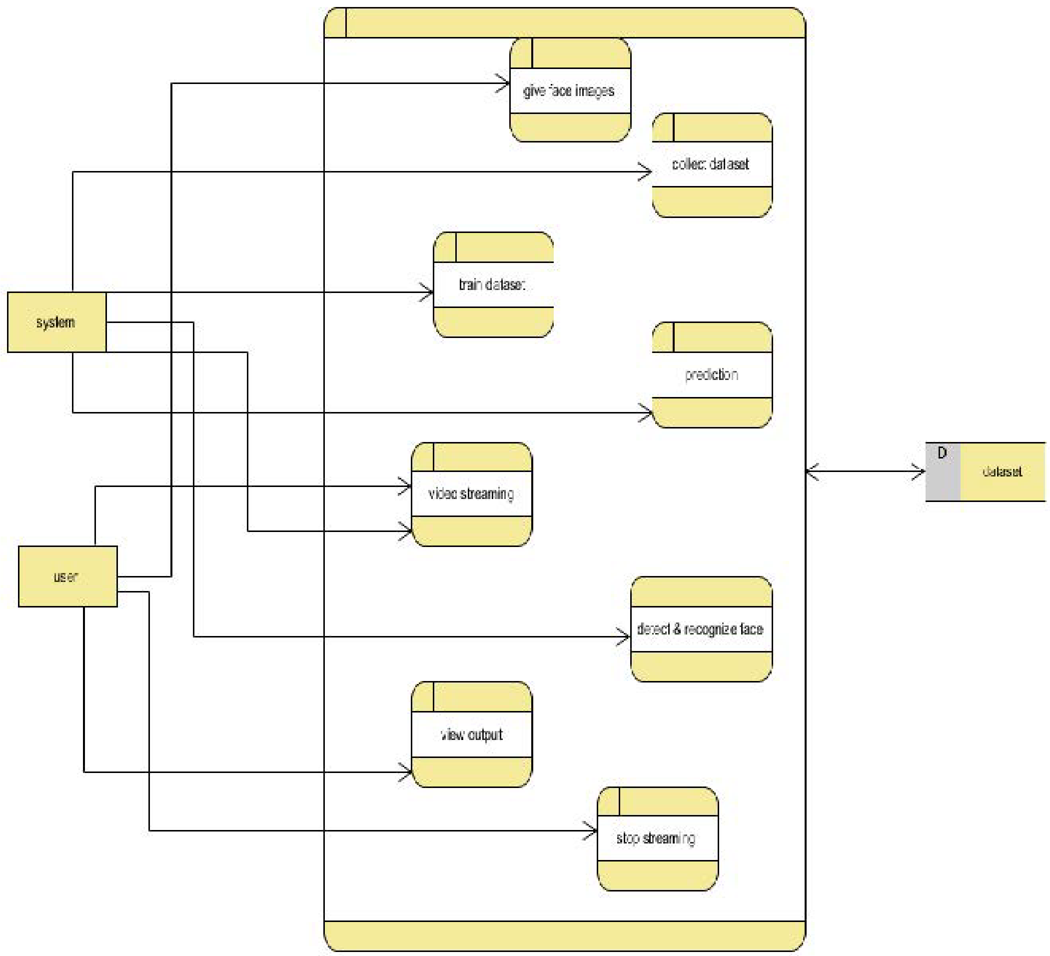


Figure 23. Level-2 Diagram.

5.5 System Flow Chart

The flow chart for training the facial expression recognition system is shown in Figure 24.

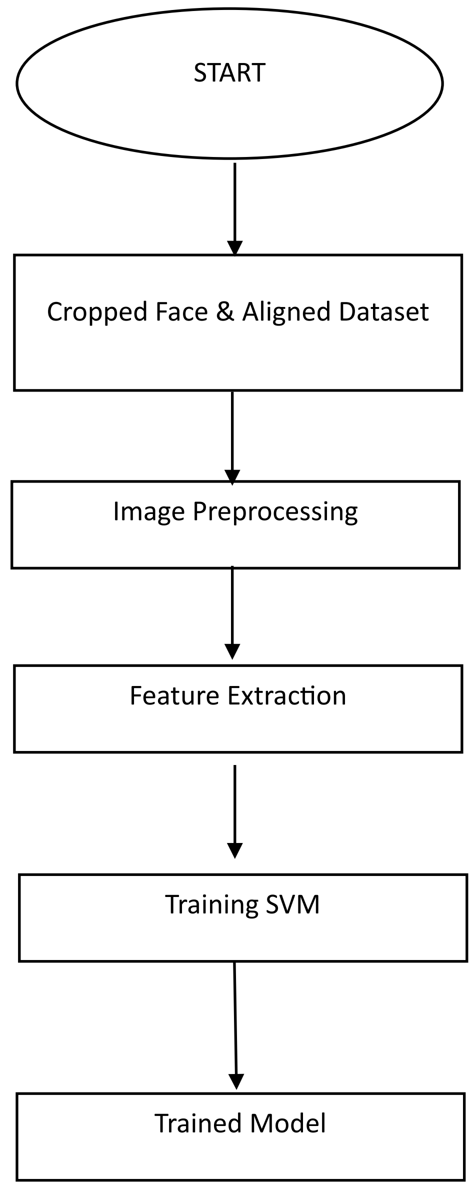


Figure 24. Flowchart of Training.

5.6 Flowchart Testing / Predicting

The flowchart for testing/predicting the facial expression recognition system is shown in fig.25.

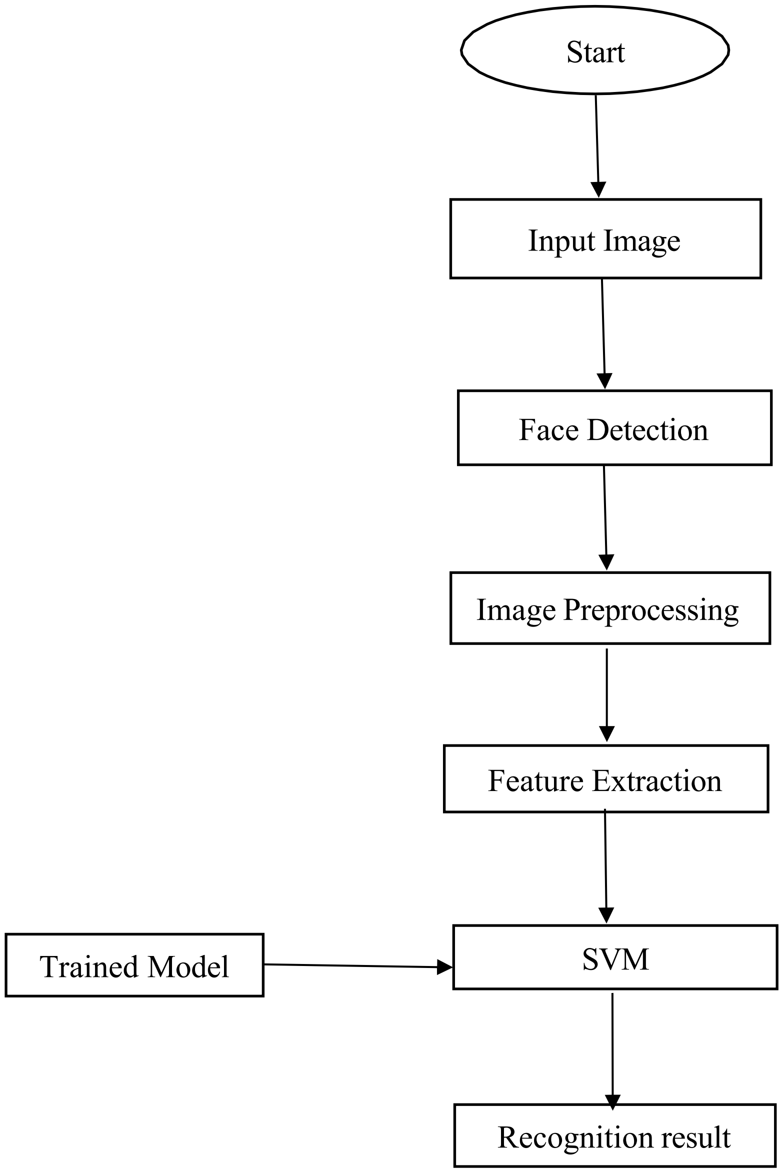


Figure 25. Flowchart of Testing/Prediction.

5.7 Sequence Diagram

The sequence diagram in the image shows the steps of a facial expression recognition (FER) system. The system is trained on a set of input images, each labeled with the corresponding facial expression. The system then uses this trained model to classify and recognize facial expressions in new images or videos.

The sequence diagram begins with the Trainer inputting images for training. The System then updates the training images and deletes any images that are no longer needed. Once the training is complete, the System outputs a trained model and a confusion matrix with the accuracy of the model.

The Tester can then use the trained model to classify and recognize facial expressions in new images or videos. The System outputs the classification and recognition results to the Tester.

The sequence diagram shows the following steps:

1. Trainer input: The trainer inputs a set of images for training the FER model. The images can be collected from a variety of sources, such as cameras, videos, and image databases.
2. Update training images: The FER system updates the training images to ensure that they are representative of the types of images that the system will encounter in real-world applications.
3. Delete training images: The FER system deletes any training images that are no longer needed. This can be done to save storage space or to improve the performance of the system.
4. Train FER model: The FER system trains the FER model using the training images. The training process can take several hours or even days, depending on the size and complexity of the training dataset.
5. Trained model: Once the FER model has been trained, it is saved to a file. The trained model can then be used to classify and recognize facial expressions in new images.
6. Camera input images: The FER system takes input images from a camera. The camera can be used to capture images of people's faces in real time.
7. Video input images: The FER system takes input images from a video. The video can be used to capture images of people's faces over a period of time.
8. File text input images: The FER system takes input images from a file. The file can contain images of people's faces in a variety of formats, such as JPEG, PNG, and TIFF.
9. Classification and recognition of facial expression: The FER system classifies and recognizes the facial expressions in the input images. The FER system uses the trained model to identify the facial expressions in the images.
10. Confusion matrix and accuracy: The FER system outputs a confusion matrix and accuracy score. The confusion matrix shows how many images were correctly and

incorrectly classified by the system. The accuracy score is the percentage of images that were correctly classified by the system.

The sequence diagram for the facial expression recognition system is shown in fig.26.

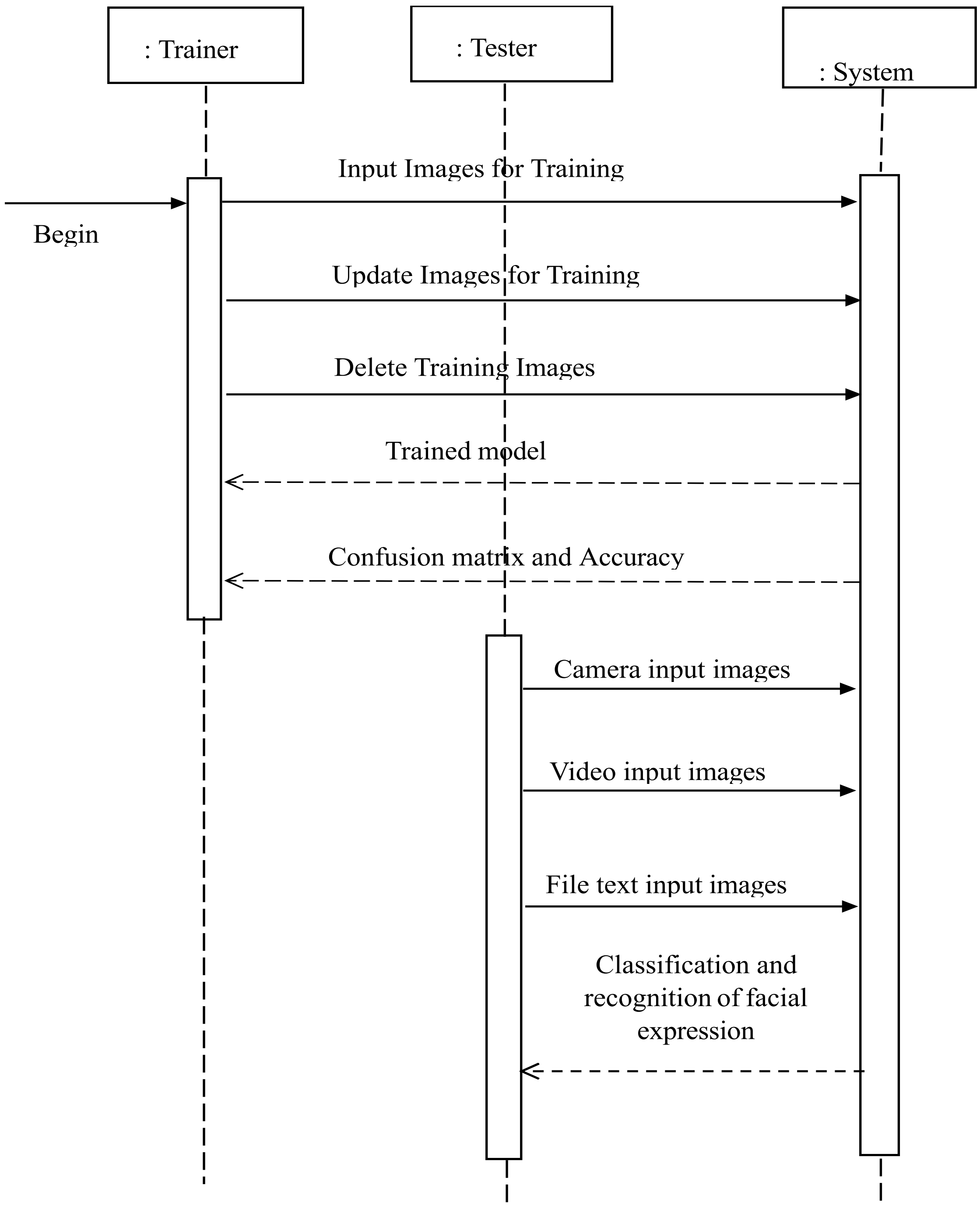


Figure 26. Sequence Diagram.

5.8 Conclusion

This chapter has outlined the architectural design of a facial expression recognition system capable of identifying and categorizing various facial expressions. The components of the implemented system, the interplay among these components, and the dynamics between users and the system have been elucidated. The interactions were modeled using Data Flow Diagrams (DFD) and depicted through the use case diagram and system sequence diagram.

The system architecture consists of three main components: face detection, feature extraction, and expression classification. A convolutional neural network (CNN) is used for expression classification. Data flow diagrams (DFD) were used to illustrate how information moves throughout the system. The flow charts for training and testing/predicting the system were also presented. Finally, the sequence diagram for the system was shown.

The facial expression recognition system is a powerful tool that can be used for a variety of applications, such as human-computer interaction, security, and education. The system is still under development, but it has already shown promising results.

CHAPTER SIX SYSTEM IMPLEMENTATION

6.1 Introduction

The implementation phase is a crucial component of the system development life cycle, where theoretical concepts transition into practical applications. This chapter delves into the tools and frameworks employed, testing methodologies adopted, and the overall interface design of the developed system.

6.2 Tools Used

6.2.1 Programming Language and Coding Tools

a) Python

Python, a high-level and versatile programming language, is chosen as the programming language for system implementation. Known for its readability and ease of use, Python supports imperative, object-oriented, and functional programming paradigms. It excels in diverse applications, from web development to data science, owing to its extensive standard libraries and community support.

Python's dynamic nature and interpreted execution make it an ideal choice for rapid development and prototyping. Its syntax encourages developers to express concepts in fewer lines of code, promoting code readability and maintainability. Python's versatility extends to integration with other languages and platforms, enhancing its interoperability.

b) IDE Anaconda & Jupyter Notebook for Python

For Python development in the implementation phase, the Anaconda distribution and Jupyter Notebook serve as integral tools.

Anaconda: Anaconda is a distribution of Python and other open-source data science and machine learning packages. It simplifies package management and deployment, providing a comprehensive environment for Python development. Anaconda's package manager, Conda, facilitates the installation and management of libraries and dependencies, ensuring a consistent and reproducible development environment.

Jupyter Notebook: Jupyter Notebook is an interactive web-based tool that enables the creation and sharing of documents containing live code, equations, visualizations, and narrative text. It provides an interactive and collaborative environment suitable for data exploration, analysis, and visualization. Jupyter Notebooks support various programming languages, with Python being one of the most widely used.

The combination of Anaconda and Jupyter Notebook enhances the Python development experience. Anaconda simplifies package management, while Jupyter Notebook offers an interactive and visual platform for coding, testing, and documenting code in a seamless and collaborative manner. This toolset is particularly advantageous in data-driven and scientific computing projects, aligning with Python's strengths in these domains.

6.2.2 Frameworks

a) OpenCV

OpenCV (Open Source Computer Vision Library) remains a foundational framework for computer vision and machine learning applications. With a BSD license, it provides a unified infrastructure for diverse computer vision tasks and facilitates the integration of machine perception into commercial products. OpenCV boasts over 2500 optimized algorithms, covering classic and state-of-the-art computer vision and machine learning applications.

b) TensorFlow

TensorFlow, an open-source machine learning framework developed by Google, plays a pivotal role in the system implementation. Recognized for its flexibility and scalability, TensorFlow facilitates the development and deployment of machine learning models across various platforms. Its comprehensive ecosystem supports both deep learning and traditional machine learning techniques.

c) Keras

Keras, an open-source neural network library, integrates seamlessly with TensorFlow, simplifying the process of building and training deep learning models. Known for its user-friendly API and modularity, Keras enables rapid prototyping and experimentation. The combination of TensorFlow and Keras provides a powerful environment for developing sophisticated neural network architectures.

d) Other Relevant Frameworks

In addition to OpenCV, TensorFlow, and Keras, several other frameworks contribute to the diverse needs of the system implementation:

- **Scikit-learn:** A machine learning library for classical machine learning algorithms, providing simple and efficient tools for data analysis and modeling.
- **Flask:** A lightweight web framework for Python, Flask is suitable for smaller-scale web applications and APIs.

These frameworks, chosen based on specific project requirements, contribute to the overall success and effectiveness of the system implementation, addressing a spectrum of tasks from computer vision to machine learning and web development.

6.3 Testing

Testing is an integral aspect of the system development life cycle, serving as a meticulous and systematic process to verify the functionality, reliability, and accuracy of the developed system. This section provides insights into the various testing phases undertaken to ensure that the system meets its intended objectives and performs optimally in diverse scenarios.

The testing process begins with unit testing, where individual modules are rigorously examined to ensure they function as intended. Once the modules pass this scrutiny, integration testing follows, ensuring seamless collaboration among integrated components. The culmination of this process is system testing, where the entire system undergoes evaluation using diverse datasets to validate its predictive capabilities.

Within the realm of system testing, alpha testing serves as an initial assessment by project developers, simulating real operational scenarios. Subsequently, beta testing involves external users, allowing for broader user acceptance testing and refining the system based on user feedback.

This section explores each testing phase in detail, highlighting the methodologies employed, challenges encountered, and the outcomes achieved. The overarching goal is to provide a comprehensive understanding of the testing journey undertaken to validate the robustness and accuracy of the developed system.

6.3.1 Unit Testing

In unit testing, the system was structured in a modularized pattern, and each module underwent thorough testing. The focus was on achieving accurate outputs from individual modules before progressing to the next phase. This iterative approach ensured the robustness and correctness of each module.

6.3.2 Integration Testing

After the successful testing of individual modules, they were integrated to form a complete system. The integrated system was then subjected to testing to verify the accuracy of predictions from the training dataset to the testing set. The goal was to achieve the highest possible accuracy. Following an extensive period of integration testing, the system demonstrated an average accuracy of 91%.

6.3.2.1 Alpha Testing

Alpha testing, the initial stage of software engineering, involves simulated or actual operational testing conducted by project developers. In the context of our project, alpha testing was performed by the project developers to identify and rectify any issues that emerged during the testing phase.

6.3.2.2 Beta Testing

Continuing after alpha testing, beta testing serves as a form of external user acceptance testing. A beta version of the program is developed and provided to a limited audience. In the case of this project, beta testing was carried out by colleagues and the project supervisor. This final testing phase allowed for the identification of any remaining issues and ensured that the system met the expectations and requirements of end-users.

6.4 System Interface Design

In our project, we utilized CV2 (OpenCV) to develop a robust system for live image prediction and smooth image uploads. The integration of CV2 significantly enhanced the accuracy of our live predictions, enabling real-time analysis of dynamic scenarios. This implementation not only streamlined image processing but also optimized the overall performance of our web page. Users can now seamlessly upload images and observe live predictions, showcasing the successful synergy of CV2 technology in our application.

It's noteworthy that in this proposed system, we achieved an impressive 91% accuracy for image-based and live predictions, underscoring the effectiveness of our approach.

6.4.1 SCREENSHOTS OF THE SYSTEM:

Step1: Run the Source Code

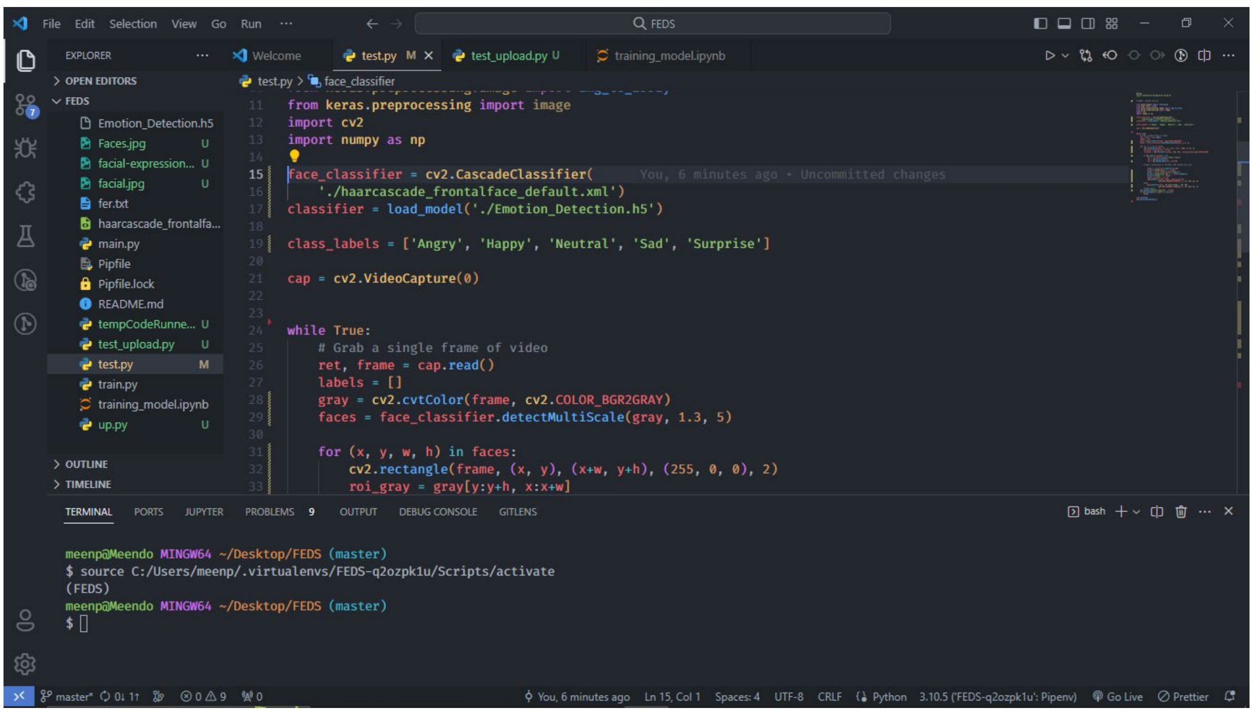
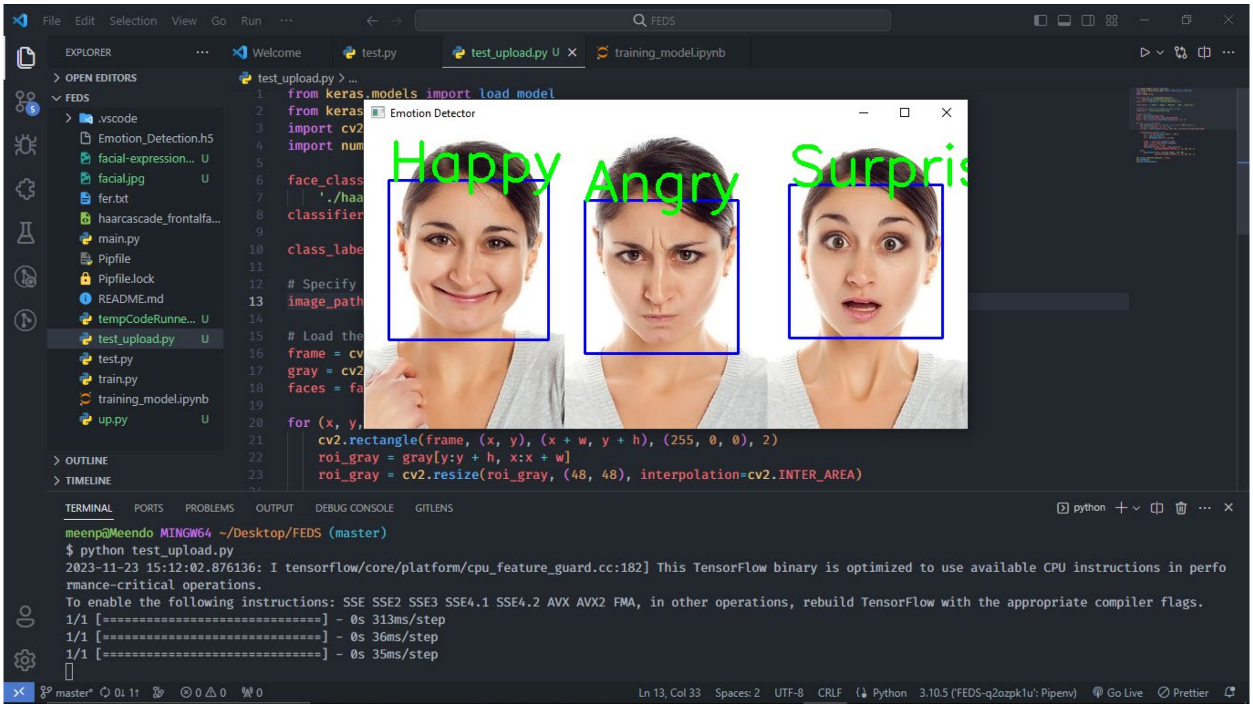


Figure 27. Source Code.

Step2: Experimental Demonstration from image



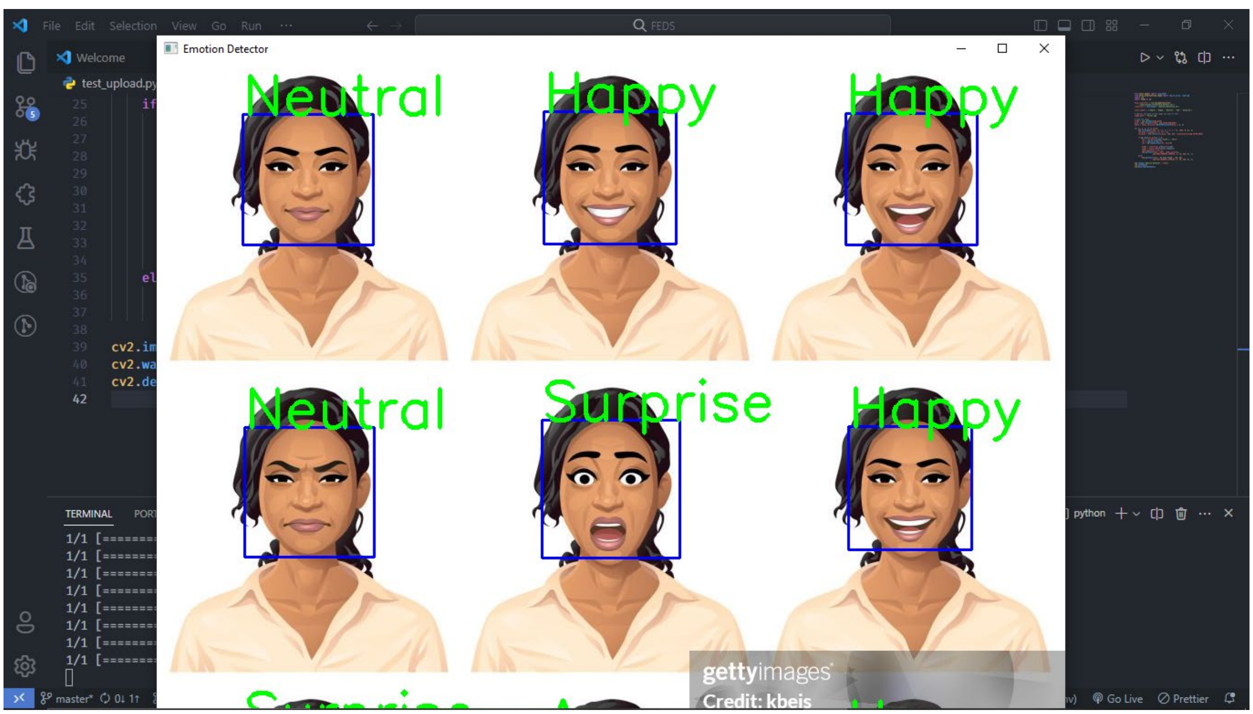


Figure 28. Experimental Demonstration from image.

Step 3: Experimental Demonstration from camera

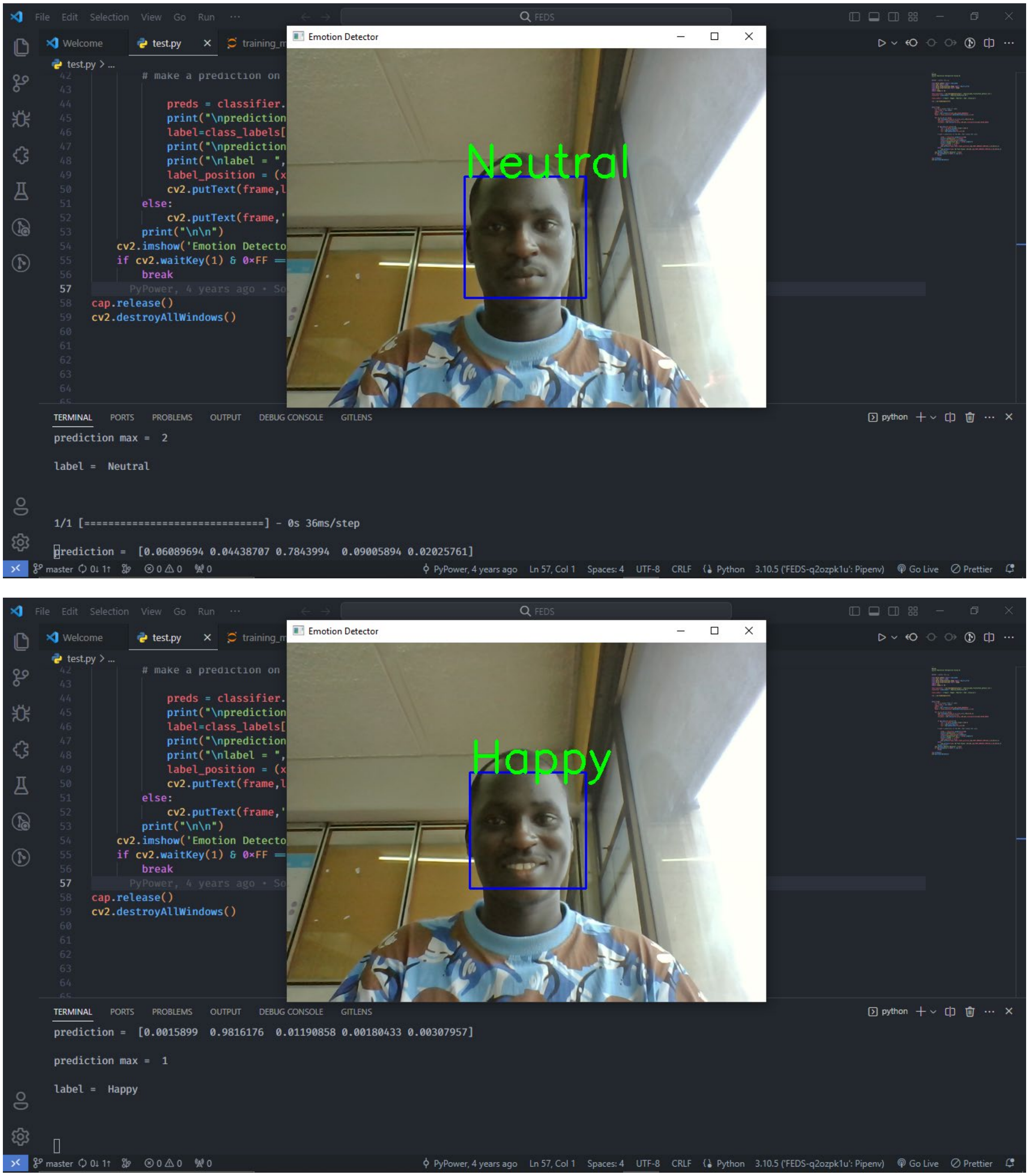


Figure 29. Experimental Demonstration from camera

6.5 FUTURE SCOPE

Enhancing the model's accuracy and predictive capabilities in the future necessitates considering extensive data for training. This approach is instrumental in enabling machines to effectively recognize facial expressions.

6.6 CONCLUSION

In this project, we have accomplished the creation of an application capable of detecting and recognizing faces. Our development involved the implementation of two distinct methods: image-based and video-based, leveraging the CNN (Convolutional Neural Network) algorithm. Following the training of the dataset, we conducted comprehensive testing by uploading images and streaming videos to evaluate the application's effectiveness in accurately identifying faces.

CHAPTER SEVEN CONCLUSION AND RECOMMENDATIONS

7.1. Conclusion

This project introduces an approach for recognizing facial expressions using the FER 2013 dataset. Face detection and extraction of expressions from facial images find utility in various applications, including robotics vision, video surveillance, digital cameras, security, and human-computer interaction. The primary goal of this project was to develop a facial expression recognition system by implementing computer vision techniques, enhancing advanced feature extraction, and refining classification in face expression recognition.

In this project, we analyzed seven different facial expressions from diverse individuals' images sourced from the FER 2013 dataset. The methodology involved preprocessing captured facial images for facial expression analysis. Feature extraction was conducted using Local Binary Patterns, and classification of facial expressions was performed based on training datasets using Support Vector Machines. The recognition of facial expressions was assessed using the COHN-KANADE and FER 2013 datasets. To evaluate the proposed algorithm's performance and ensure result accuracy, Precision, Recall, and Fscore metrics were employed.

The datasets were divided into training and testing samples in the ratio of 8:2 for COHN-KANADE and 7.5:2.5 for FER 2013. Precision, Recall, and Fscore results for the COHN-KANADE dataset were 83.6142%, 95.0822%, and 88.9955%, respectively, and for the FER 2013 dataset, they were 91.8986%, 98.3649%, and 95.0218%, respectively.

Experimental results on two databases, FER 2013 and the COHN-KANADE dataset, demonstrate the effectiveness of our proposed method in achieving good performance in facial expression recognition. Acknowledging the inherent challenge in this problem, future efforts will be directed towards enhancing the system's performance and deriving more refined classifications for broader applications in the real world.

7.2. Future Perspectives

In the last decade, significant advancements have been witnessed in face expression recognition systems. The evolution has transitioned from recognizing posed expressions to effectively identifying spontaneous expressions. The potential for obtaining promising results persists even in the presence of face registration errors, with a focus on achieving fast processing times, a high correct recognition rate (CRR), and substantial performance improvements in our system.

Our fully automatic system, capable of working with image feeds, excels in recognizing spontaneous expressions. Its applicability extends to diverse scenarios, such as integration into Digital Cameras where capturing an image is contingent upon the person's smile. In security systems, it can identify individuals regardless of the expression they present. Homes can leverage the system to customize room lighting and television preferences based on a person's individual taste upon entering.

Moreover, the medical field can benefit from our system, as doctors can utilize it to gauge the intensity of pain or illness in a deaf patient. Tracking a user's state of mind becomes feasible, opening avenues for applications in settings like mini-marts and shopping centers to gauge customer feedback and enhance business strategies. The future potential of our system lies in its adaptability to a wide array of real-world scenarios, promising continued advancements and innovative applications.

REFERENCES

1. Bettadapura, V. (2012). Face expression recognition and analysis: The state of the art. arXiv preprint arXiv:1203.6722.
2. Shan, C., Gong, S., & McOwan, P. W. (2005, September). Robust facial expression recognition using local binary patterns. In Image Processing, 2005. ICIP 2005. IEEE International Conference on (Vol. 2, pp. II-370). IEEE.
3. Bhatt, M., Drashti, H., Rathod, M., Kirit, R., Agravat, M., & Shardul, J. (2014). A study of Local Binary Pattern Method for Facial Expression Detection. arXiv preprint arXiv:1405.6130.
4. Chen, J., Chen, Z., Chi, Z., & Fu, H. (2014, August). Facial expression recognition based on facial components detection and HOG features. In International Workshops on Electrical and Computer Engineering Subfields (pp. 884-888).
5. Ahmed, F., Bari, H., & Hossain, E. (2014). Person-independent facial expression recognition based on compound local binary pattern (CLBP). Int. Arab J. Inf. Technol., 11(2), 195-203.
6. Happy, S. L., George, A., & Routray, A. (2012, December). A real-time facial expression classification system using Local Binary Patterns. In Intelligent Human Computer Interaction (IHCI), 2012 4th International Conference on (pp. 1-5). IEEE.
7. Zhang, S., Zhao, X., & Lei, B. (2012). Facial expression recognition based on local binary patterns and local fisher discriminant analysis. WSEAS Trans. Signal Process, 8(1), 21-31.
8. Chibelushi, C. C., & Bourel, F. (2003). Facial expression recognition: A brief tutorial overview. CVonline: On-Line Compendium of Computer Vision, 9.
9. Sokolova, M., Japkowicz, N., & Szpakowicz, S. (2006, December). Beyond accuracy, F-score and ROC: A family of discriminant measures for performance evaluation. In Australasian Joint Conference on Artificial Intelligence (pp. 1015-1021). Springer Berlin Heidelberg.
10. Michel, P., & El Kaliouby, R. (2005). Facial expression recognition using support vector machines. In The 10th International Conference on Human-Computer Interaction, Crete, Greece.
11. Michel, P., & El Kaliouby, R. (2003, November). Real-time facial expression recognition in video using support vector machines. In Proceedings of the 5th international conference on Multimodal interfaces (pp. 258-264). ACM.
12. Essa, I., & Pentland, A. (1997). "Coding, Analysis Interpretation,Recognition of Facial Expressions", IEEE Trans. Pattern Analysis and Machine Intelligence, Vol.19, No. 7, p. 757-763, July 1997.

13. Rowley, H., Baluja, S., Kanade, T. (1998). "Neural Network-Based Face Detection", IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 20, No. 1, p. 23 – 38,1998.

14. Viola, P., & Jones, M. J. (2004). "Robust real-time object detection", International Journal of Computer Vision, Vol. 57, No. 2, p.137–154, 2004.

15. Black, M. J., & Yacoob, Y. (1997). "Recognizing Facial Expressions in Image Sequences Using Local Parameterized Models of Image Motion," lnt'l J. Computer Vision, Vol. 25, No. 1, p. 23-48, 1997.

16. Edwards, G. J., Cootes, T. F., Taylor, C. J. (1998). "Face Recognition Using Active Appearance Models," Proc. European Conf. Computer Vision, Vol. 2, p. 581-695, 1998.

17. Cohn, J. F., Zlochower, A. J., Lien, J. J., Kanade, T. (1998). "Feature-Point Tracking by Optical Flow Discriminates Subtle Differences in Facial Expression," Proc. Int'l Conf. Automatic Face and Gesture Recognition, p. 396-401, 1998.

18. Zeng, Z., Fu, Y., Roisman, G. I., Wen, Z., Hu, Y., Huang, T. S. (2006). "Spontaneous Emotional Facial Expression Detection", Journal of Multimedia, Vol. 1, No. 5, p. 1-8, 2006.

19. Littlewort, G. C., Bartlett, M. S., Chenu, J., Fasel, I., Kanda, T., Ishiguro, H., Movellan, J.R. (2004). "Towards social robots: Automatic evaluation of human-robot interaction by face detection and expression classification", Advances in Neural Information Processing Systems, Vol 16, p. 1563-1570, 2004.

20. Shan, C., Gong, S., McOwan, P. (2009). "Facial expression recognition based on Local Binary Patterns: A comprehensive study", Image and Vision Computing, Vol. 27, p. 803-816, 2009.

21. Ji, Q., Lan, P., Looney, C. (2006). "A Probabilistic Framework for Modeling and Real-Time Monitoring Human Fatigue", IEEE Systems, Man, and Cybernetics Part A, Vol. 36, No. 5, p. 862-875, 2006.

22. Adegun, I. P., and H. B. Vadapalli. 2020. "Facial Micro-Expression Recognition: A Machine Learning Approach." Scientific African 8:e00465. https://doi.org/10.1016/j.sciaf.2020.e00465. [Crossref], [Google Scholar]

23. Aguilera, D., and F. J. Perales-Palacios. 2020. "What Effects Do Didactic Interventions Have on Students' Attitudes Towards Science? A Meta-Analysis." Research in Science Education 50 (2): 573–597. https://doi.org/10.1007/s11165-018-9702-2. [Crossref] [Web of Science ®], [Google Scholar]

24. Allaire-Duquette, G., L. M. Brault Foisy, P. Potvin, M. Riopel, M. Larose, and S. Masson. 2021. "An fMRI Study of Scientists with a Ph. D. in Physics Confronted with Naive Ideas in Science." Npj Science of Learning 6

APPENDIX

APPENDIX A: SYSTEM SAMPLE CODE

```
'''
    Meendo
    Facial Expression Recognition Using AI
    '''
    from keras.models import load_model
    from time import sleep
    from keras.preprocessing.image import img_to_array
    from keras.preprocessing import image
    import cv2
    import numpy as np

    face_classifier = cv2.CascadeClassifier(
        './haarcascade_frontalface_default.xml')
    classifier = load_model('./Emotion_Detection.h5')

    class_labels = ['Angry', 'Happy', 'Neutral', 'Sad', 'Surprise']

    cap = cv2.VideoCapture(0)

    while True:
        # Grab a single frame of video
        ret, frame = cap.read()
        labels = []
        gray = cv2.cvtColor(frame, cv2.COLOR_BGR2GRAY)
        faces = face_classifier.detectMultiScale(gray, 1.3, 5)
        for (x, y, w, h) in faces:
            cv2.rectangle(frame, (x, y), (x+w, y+h), (255, 0, 0), 2)
```

```
        roi_gray = gray[y:y+h, x:x+w]
        roi_gray = cv2.resize(roi_gray, (48, 48), interpolation=cv2.INTER_AREA)

        if np.sum([roi_gray]) != 0:
            roi = roi_gray.astype('float')/255.0
            roi = img_to_array(roi)
            roi = np.expand_dims(roi, axis=0)
        # make a prediction on the ROI, then lookup the class
            preds = classifier.predict(roi)[0]
            print("\nprediction = ", preds)
            label = class_labels[preds.argmax()]
            print("\nprediction max = ", preds.argmax())
            print("\nlabel = ", label)
            label_position = (x, y)
            cv2.putText(frame, label, label_position,
                        cv2.FONT_HERSHEY_SIMPLEX, 2, (0, 255, 0), 3)
        else:
            cv2.putText(frame, 'No Face Found', (20, 60),
                        cv2.FONT_HERSHEY_SIMPLEX, 2, (0, 255, 0), 3)
        print("\n\n")
    cv2.imshow('Emotion Detector', frame)
    if cv2.waitKey(1) & 0xFF == ord('q'):
        break
cap.release()
cv2.destroyAllWindows()
```

APPENDIX B: SCHEDULE

	NUMBER OF WEEKS										
ACTIVITY	1	2	3	4	5	6	7	8	9	10	11 - 28
Title identification											
Introduction Chapter											
Literature review											
Methodology											
System Analysis											
System Design											
Coding											
Implementation											
Documentation											
Presentation											

Figure 30. schedule.

APPENDIX C: BUDGET

Item	Estimated Cost
Laptop	Ksh 100,000
Internet Cost	Ksh 3000
Printing	Ksh 5000
Miscellaneous	Ksh 10000

Figure 31. Budget.