THE COMPARATIVE STUDY OF DIFFERENT TYPES OF MODEL FOR RECOGNIZING FACIAL EXPRESSION

NURUL ISLAM

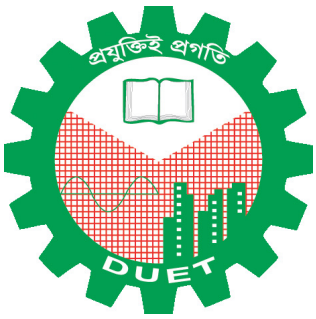
MD. AHASANUL ISLAM

MD. JANTO RAHMAN

A thesis submitted for the degree

of

Bachelor of Science in Computer Science and Engineering



Department of Computer Science and Engineering

Dhaka University of Engineering & Technology, Gazipur

June, 2023

RECOGNIZING FACIAL EXPRESSION USING CONVOLUTIONAL NEURAL NETWORK AND LOCAL BINARY PATTERN

NURUL ISLAM
Student no: 174017
Reg: 10318, Session: 2017-2018

MD. AHASANUL ISLAM
Student no: 174018
Reg: 10319, Session: 2017-2018

MD. JANTO RAHMAN
Student no: 174028
Reg: 10329, Session: 2017-2018

Supervisor: Dr. Fazlul Hasan Siddiqui

Professor

A thesis submitted for the degree of

BACHELOR OF SCIENCE IN COMPUTER SCIENCE AND ENGINEERING

Department of Computer Science and Engineering

Dhaka University of Engineering & Technology, Gazipur

June, 2023

Declaration

It is thereby declared that the work presented in this thesis or any part of this thesis has not been submitted elsewhere for the award of any degree or diploma.

Signatures:

.....

(NURUL ISLAM)

.....

(MD. AHASANUL ISLAM)

.....

(MD. JANTO RAHMAN)

Signature of the Thesis Supervisor

.....

Dr. Fazlul Hasan Siddiqui

Professor

Department of Computer Science & Engineering

Dhaka University of Engineering & Technology, Gazipur.

Acknowledgements

In the name of the Almighty Allah, the Most Merciful and Gracious, We would like to extend our heartfelt appreciation for the help and encouragement we have received from numerous individuals during the course of our research project.

Firstly, we would like to extend our sincere appreciation to our thesis supervisor, **Dr. Fazlul Hasan Siddiqui**, Professor in the Department of Computer Science and Engineering at DUET. We are truly grateful for his valuable time, dedication, guidance, and advice, which were instrumental in the progress of this work.

We would also like to thank **Dr.Md. Shafiqul Islam**, Professor, Department of Computer Science and Engineering, DUET, Gazipur for graciously evaluating our thesis as an external examiner.

Additionally, we extend our appreciation to all the teachers who have directly or indirectly contributed to our work, as well as our thesis mates and other members of our thesis group with whom we collaborated. Their contributions and support have been invaluable in shaping our research.

We acknowledge and offer our heartfelt thanks to all those who have assisted us, and we recognize that their contributions have played a significant role in the completion of this research.

Abstract

Facial expression recognition plays a crucial role in various applications such as emotion analysis, human-computer interaction, and facial recognition systems. The study compares different types of models, including Convolutional Neural Networks (CNN), Local Binary Patterns (LBP), VGG, ResNet, and Inception, to evaluate their performance in recognizing facial expressions.This thesis explores the recognition of facial expressions using the FER2013 dataset.The FER2013 dataset, widely used in the field, provides a diverse collection of facial images labeled with seven different emotional expressions. The dataset enables the investigation of model performance across various facial expressions, encompassing happiness, sadness, anger, disgust, fear, surprise, and neutral.

Comparing different models for recognizing facial expressions is a valuable approach in our research.We utilized two datasets in our research: the FER2013 dataset, a commonly used and effective dataset for facial expression recognition, served as the training dataset. The second dataset consisted of real-life scenarios and was specifically collected for testing purposes.

This research contributes to the field of facial expression recognition by providing a comparative study of different models using the FER2013 dataset. The findings serve as a valuable reference for researchers and practitioners interested in developing accurate and efficient systems for facial expression recognition, ultimately enhancing applications involving emotion analysis and human-computer interaction.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

In this chapter, we provide an overview of facial expression recognition, its importance in various applications, and the motivation behind our research. We outline the aim and objectives of this study, which involve comparing different models for recognizing facial expressions.we highlight the contribution of this thesis in terms of providing insights into model performance and recommendations for facial expression recognition systems.

1.1 Facial Expression Recognition

Facial Expression Recognition (FER) is a field of research in computer vision and affective computing that focuses on analyzing and interpreting human emotions based on facial expressions. Facial expressions serve as important nonverbal cues, conveying a wide range of emotions such as happiness, sadness, anger, fear, surprise, disgust, and neutrality. Being able to automatically detect and recognize these facial expressions has numerous applications in various domains, including psychology, human-computer interaction, healthcare, and entertainment.It is widely acknowledged that facial expressions provide valuable insights into understanding an individual's true emotions [1].

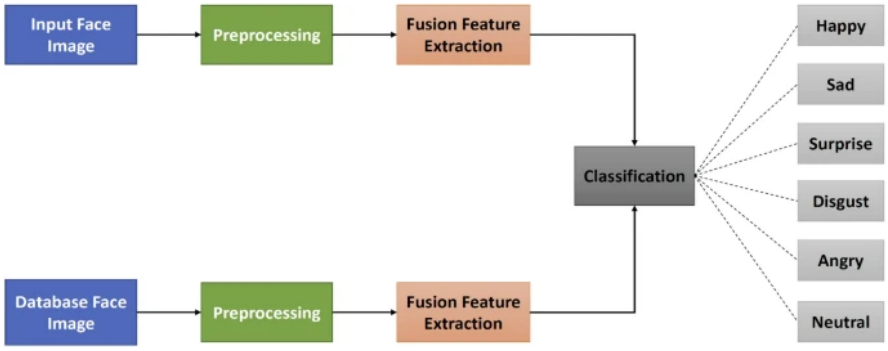


Figure 1.1 Face expression recognition system's structure.

Facial Expression Recognition plays a vital role in understanding human emotions and facilitating human-machine interactions. It involves extracting facial features and employing machine learning or deep learning models to classify facial expressions into specific emotion categories. The advancements in deep learning, particularly CNNs, have significantly contributed to the progress of FER, opening up opportunities for various practical applications. [2] human emotional feelings and desires is facial expression recognition .

1.2 Motivation

Facial expression recognition has become a common practice in human-computer interaction systems. Over the years, researchers have proposed various feature descriptors, employed different partitioning algorithms, and conducted numerous experiments using diverse datasets to achieve accurate facial expression identification. However, a majority of these approaches have relied on 2D static images or 2D video sequences for recognition purposes. Unfortunately, the limitations of such 2D-based analysis include challenges related to variations in posture and lighting conditions.

The ability to recognize emotions holds significant importance in camera surveillance systems, as it enables the identification of potential suspects. For instance, an

alarm system can be triggered when a person displaying fear is detected. Emotion recognition systems can also serve as sub-modules in various applications, including music recommendation systems and camera surveillance systems.

1.3 Aim and Objective

The Facial Emotion Recognition (FER) system is designed to analyze body language in both still photographs and video streams in order to provide insights into an individual's emotional state. This project aims to enhance comparing different type of model such as CNN,VGG,ResNet,Inception with their accuracy of the system in categorizing seven distinct emotions: anger, disgust, fear, happiness, sadness, surprise, and neutrality, surpassing previous performance levels. [3]. Furthermore, the primary objective of this study is to explore and comprehend the advantages associated with the utilization of convolutional neural network models compared to other learning models.

1.4 Our Contribution

We conducted an extensive review of various papers focusing on facial expression recognition systems and neural network techniques that have the potential to be effective approaches for emotion recognition.The former challenges have served as strong motivations for us to develop a precise, secure, and efficient method for facial expression recognition within computer vision systems. In this paper we compare the various model such as Convolutional Neural Network (CNN) model ,Visual Geometric Group (VGG) model, Residual model and Inception model.To observe which model is better among these models and to determine the features of the models.Determine the loss, value loss, value accuracy, accuracy among these models and determine which model is better among them.

In addition to utilizing the FER2013 dataset for training purposes, we have made a

significant contribution by collecting a real-time dataset specifically for testing our trained models. This real-time dataset consists of facial expression samples captured in real-life scenarios, providing a more realistic and diverse set of images to evaluate the performance of the trained models. By incorporating this real-time dataset into our testing process, we aim to assess the models' ability to accurately recognize facial expressions in real-world environments and validate their generalizability beyond the FER2013 dataset.

1.5 Summary

The first chapter of our thesis provides an overview of the research, outlining the motivations behind selecting this particular field of study. We also define our goals, objectives, and the specific contribution we intend to make through our research.

Chapter 2

Literature Review

The literature review section of a thesis aims to provide a comprehensive summary of existing research and relevant publications in the field. In the context of facial expression recognition, the literature review focuses on investigating various learning algorithms, including machine learning and deep learning, as well as specific models such as CNN, LBP, VGG, ResNet, and Inception. The literature review critically analyzes and synthesizes previous studies to understand the current state of knowledge, identify research gaps, and establish the need for further investigation.

2.1 Background Study

Facial Emotion Recognition is a field of research that aims to identify emotions based on an individual's facial expressions. Emotion Recognition involves studying the connection between emotions and the techniques and methods used to detect them. Emotions can be discerned through facial expressions, verbal cues, and other indicators [4]. To deduce emotions, various techniques such as machine learning, neural networks, and artificial intelligence have been extensively utilized. Recognizing emotions from facial expressions is a challenging task within the domain of emotional intelligence, with photographs serving as the input for analysis [5].There are three crucial challenges that need

to be addressed by a facial expression analysis approach: face identification, extraction of facial features, and recognition of facial expressions from static images or sequences of images. [6]. As depicted in Figure 2.1, the conventional composition of Automatic Facial Expression Analysis (AFEA) involves three main steps: face acquisition, extraction and representation of facial data, and facial expression recognition. Face feature extraction techniques can be broadly classified into two categories: approaches based on geometric or predictive features, and methods based on visual appearances. [7].

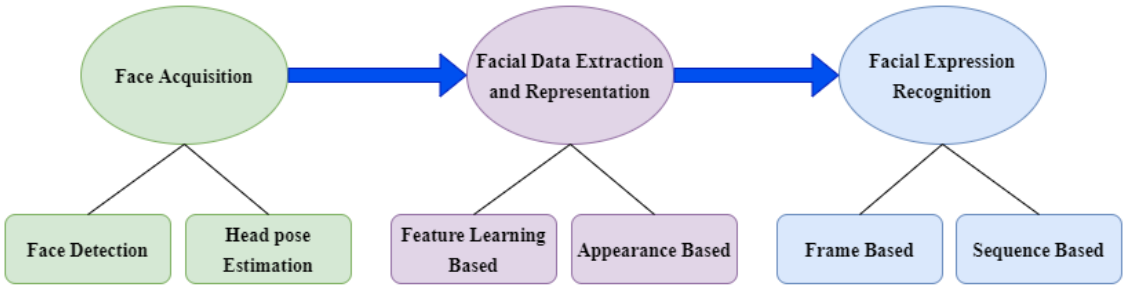


Figure 2.1 Automatic Facial Expression Analysis framework

2.2 Deep Learning

Deep learning is considered as one of the most effective approaches to address the challenges of feature extraction. This is because deep learning models have the ability to autonomously learn and focus on relevant features, which helps overcome the difficulties associated with manual feature extraction. [8].Deep learning in neural networks has the capability to yield outcomes across a wide range of mediums, encompassing photographs, textbooks, audio, and more. It enables the performance of diverse tasks such as image identification, classification, decision-making, and pattern recognition. [9]. Neural networks strive to emulate the functioning of the human brain in order to generate outcomes that resemble those of the human mind. [10].

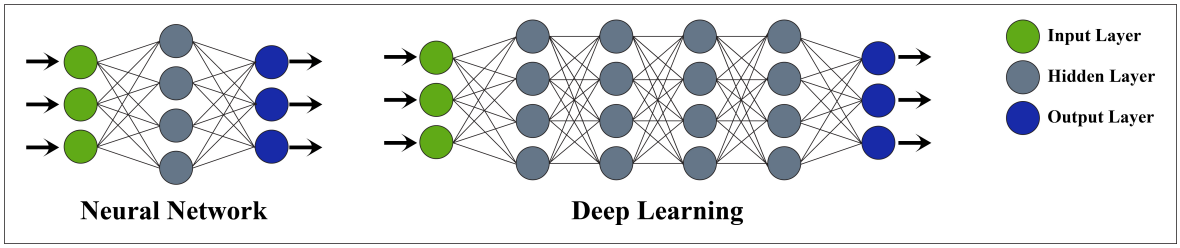


Figure 2.2 Deep learning Architecture

2.3 Machine Learning

Machine learning is a field within the realm of artificial intelligence (AI) and computer science that emphasizes the utilization of data and algorithms to imitate the learning process observed in humans, progressively enhancing its ability to perceive and understand information. [11]. Deep learning is a specific form of machine learning that delves into advanced technical aspects. To gain a comprehensive understanding of deep learning, it is necessary to grasp the foundational concepts of machine learning. Essentially, machine learning can be seen as an applied statistics field that utilizes computer algorithms to estimate complex functions statistically, rather than focusing on demonstrating confidence intervals around these functions. [12].

2.3.1 Learning Algorithm

The concept of "learning" in machine learning refers to the ability of machines to acquire knowledge from data. Mitchell (1997) provides a concise definition stating that a computer program is considered to learn from experience E in a specific class of tasks T, measured by performance measure P, if its performance in tasks from T, as evaluated by P, improves with accumulated experience " [12].

Machine learning can be categorized into three main categories. [13].

1. Supervised Learning,
2. Unsupervised Learning,
3. Reinforcement Learning

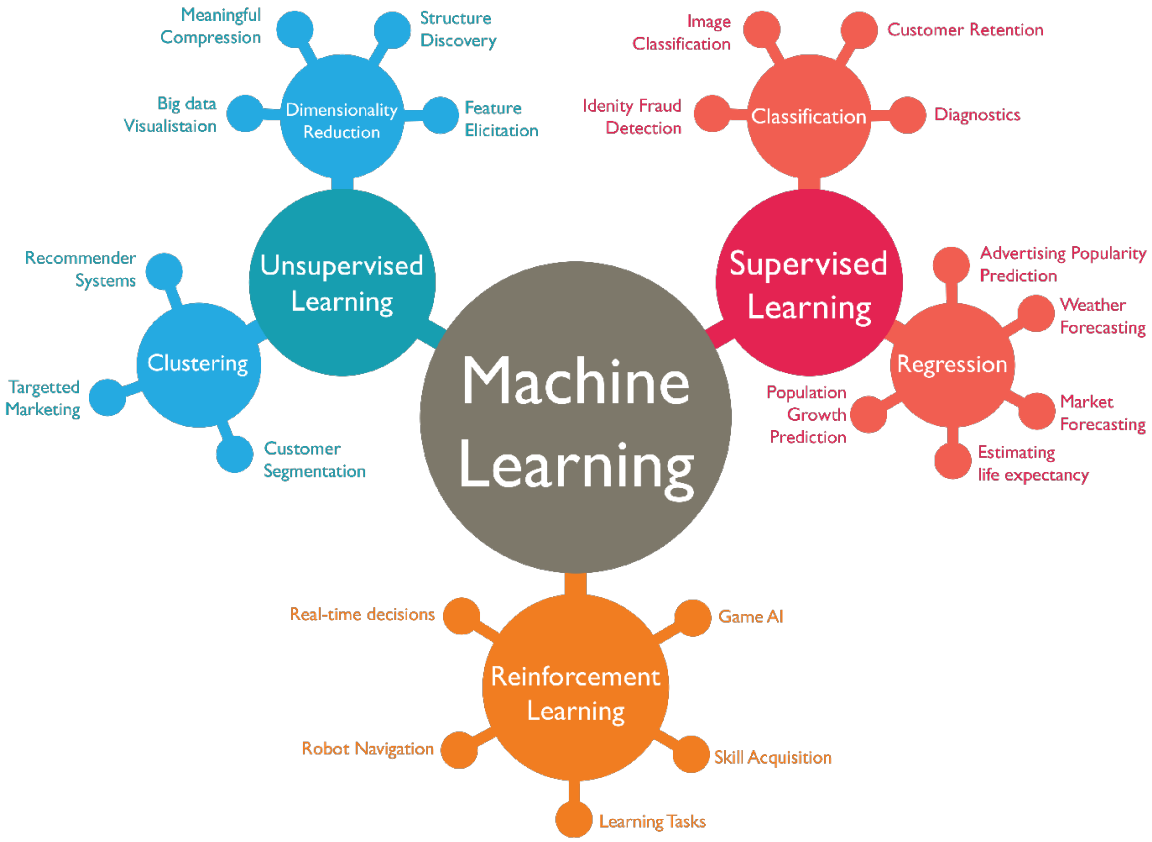


Figure 2.3 Classification of machine learning

2.3.1.1 Supervised learning

Supervised learning algorithms leverage datasets with labeled information to generate predictions. This learning approach is advantageous when there is a clear understanding of the desired outcome or result.

2.3.1.2 Unsupervised learning

Unsupervised learning algorithms operate on unlabeled data, and they assign labels or categorize the data based on its inherent characteristics, structure, or patterns. This

technique proves useful when the specific type or nature of the desired outcome is unknown.

2.3.1.3 Reinforcement learning

Reinforcement learning is a distinctive form of machine learning where the model learns through trial and error. The model receives rewards for making correct decisions and is penalized for incorrect ones, enabling it to learn patterns and improve its accuracy in making decisions on unfamiliar data.

2.3.2 Convolutional Neural Network (CNN)

A fundamental template for a Convolutional Neural Network (CNN) consists of various components that are straightforward to understand and relate to the suggested CNN model. Illustrated in Figure 2.4, a simple CNN comprises three types of layers: input, hidden, and output. The input layer receives data into the CNN, which then flows through multiple hidden layers before reaching the output layer. The output layer represents the network's prediction. The network's output is compared to the actual labels using metrics like loss or error.

The hidden layers in the network form the core of data processing. Each layer can be divided into functions such as convolution, pooling, normalization, and activation. These subsequent layers contribute to the structure of a convolutional neural network. [14]. In the environment of our exploration, we also address the CNN model parameters.

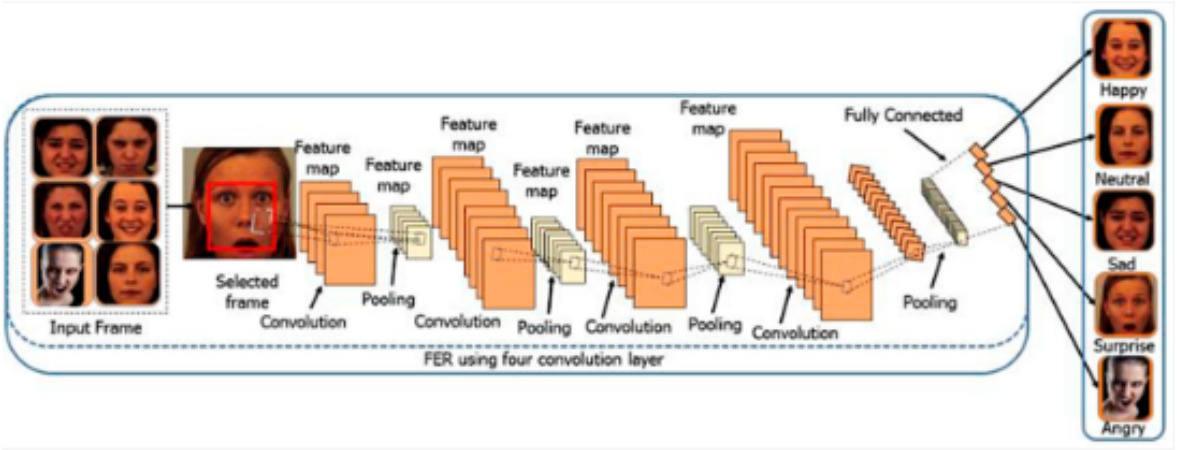


Figure 2.4 Convolutional Neural Network (CNN) framework

2.3.2.1 Convolution layer

The primary challenge in a Convolutional Neural Network lies in extracting features from input images. Convolution, depicted in Figure 2.5, preserves the spatial relationship between pixels by learning image characteristics through filtering small regions of input data. The convolutional layer is characterized by terms such as 'filter,' 'kernel,' and 'feature detector.' When the filter slides across the image and computes the dot product, it generates a matrix known as the 'Convolved Feature,' 'Activation Map,' or 'Feature Map.' This layer allows for customization of the filter size and stride (the number of pixels after which the filter moves). The output of the convolutional layer is depicted in Equation 2.1.

$$\frac{W - F + 2P}{S} + 1 = O \tag{2.1}$$

where F = Spatial extents, P = Padding, S = Stride

If the stride is not appropriately selected, neurons do not align smoothly and evenly across the input. To address this issue, zero padding is employed throughout the image to ensure that neurons fit properly.

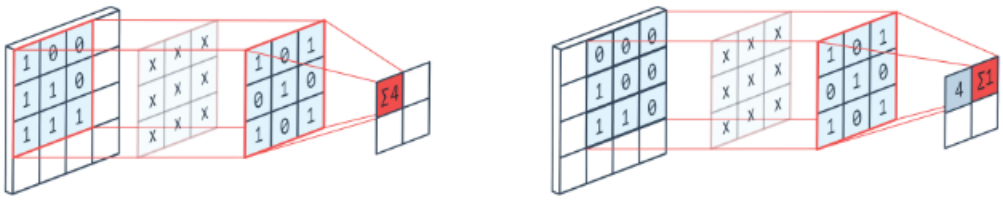


Figure 2.5 Outline of convolution layer

2.3.2.2 ReLu layer

In this layer, negative values in the filtered image are eliminated and replaced with zero. This is done to ensure that the overall sum does not approach zero. When the input exceeds a specific threshold, the Rectified Linear Unit (ReLU) activation function activates a node, resulting in a non-zero output. If the input does not reach the threshold, the output remains zero. The key benefit of ReLU's activation function is that its gradient is consistently equal to 1, as indicated in Equation 2.2. This means that during back-propagation, a significant portion of the error is effectively propagated back through the network. [15] [16].

$$f(x) = max(0, x) \tag{2.2}$$

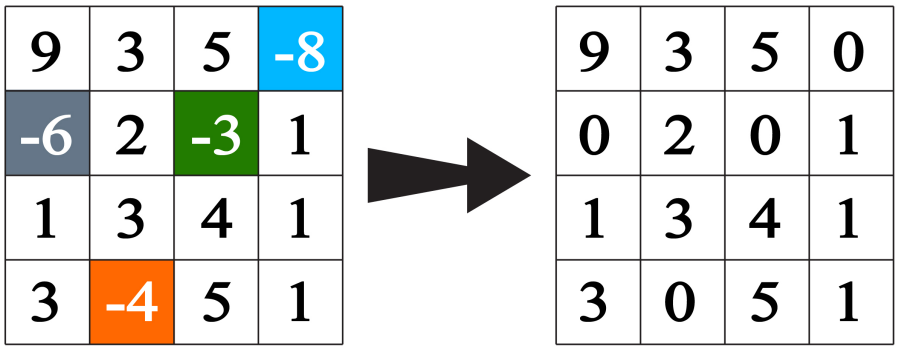


Figure 2.6 Outline of ReLU layer

2.3.2.3 Pooling Layer

The objective of a pooling layer is to decrease the spatial size of the processed feature map, leading to a more compact representation of features. The outcome of this layer is a pooled feature map. The most commonly utilized pooling algorithms are maximum pooling and mean pooling. In the case of maximum pooling, the maximum value within each region is retained while discarding the other values. This downsampling process reduces the overall size of the network. The formula used for calculating the pooling layer is as follows:

$$\frac{I + 2P - F}{S} + 1 = O \tag{2.3}$$

where I = input matrix, S= stride, P=padding and F=filter

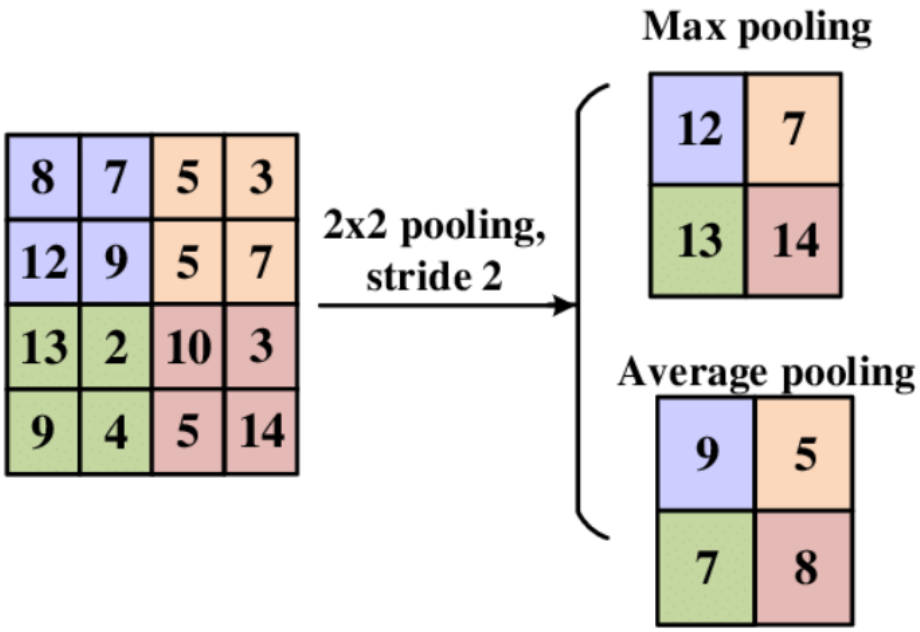


Figure 2.7 Outline of Pooling layer

2.3.2.4 Fully connected layer

The pooled feature map undergoes a transformation in the Fully Connected Layer, where it is converted from a two-dimensional structure to a one-dimensional vector known as a feature vector. This process essentially "flattens" the pooled feature map. The resulting feature vector is then used as input in a standard Fully Connected Layer for classification purposes.

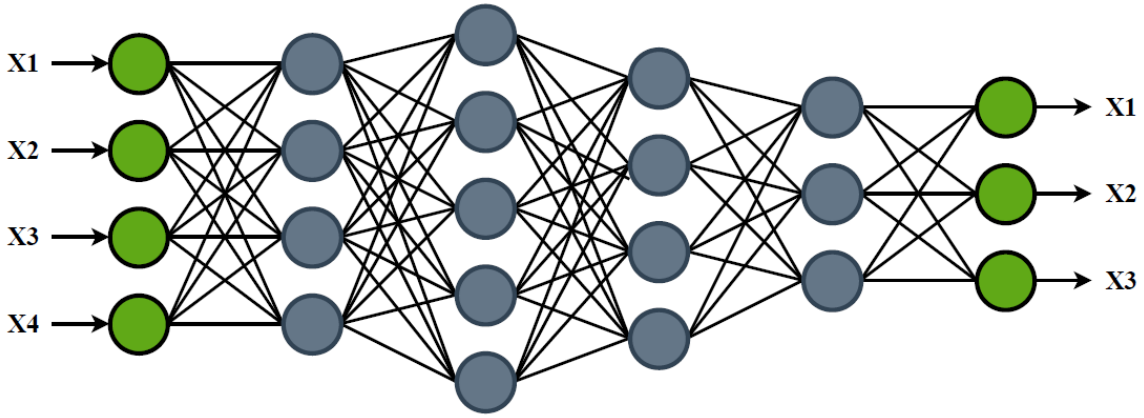


Figure 2.8 Outline of Fully Connected layer

2.3.2.5 Softmax

This parameter specifies the activation function to be used in the dense layer. The softmax activation function is commonly used in multi-class classification problems as it normalizes the outputs, producing a probability distribution over the classes. It ensures that the sum of the probabilities for all classes is equal to 1, allowing us to interpret the output as class probabilities.

2.3.2.6 Batch normalization

It calculates the mean and standard deviation of the input data and uses them to normalize the inputs. It also learns and applies scale and offset factors (gamma and beta) to further adjust the normalized values. Batch normalization expedites the learning

process by incorporating a mechanism that maintains the mean activation close to 0 and the standard deviation of activation close to 1

2.3.3 Visual Geometry Group(VGG)

The VGG model, also known as the Visual Geometry Group model, is a deep convolutional neural network architecture introduced by researchers from the University of Oxford in 2014. Developed by Karen Simonyan and Andrew Zisserman's team, VGG was primarily created for image classification tasks and achieved remarkable performance in the 2014 ImageNet Large Scale Visual Recognition Challenge (ILSVRC). The fundamental concept behind the VGG model lies in its consistent architecture. It comprises a sequence of convolutional layers using small 3x3 filters, followed by max-pooling layers for downsampling. The network is relatively deep, with either 16 or 19 layers containing trainable parameters, and it applies the rectified linear unit (ReLU) activation function after each convolutional layer. The uniform architecture of the VGG model, along with the utilization of small filters, enables it to learn diverse image features at various spatial scales. The depth of the network contributes to its ability to capture intricate hierarchical patterns within images.

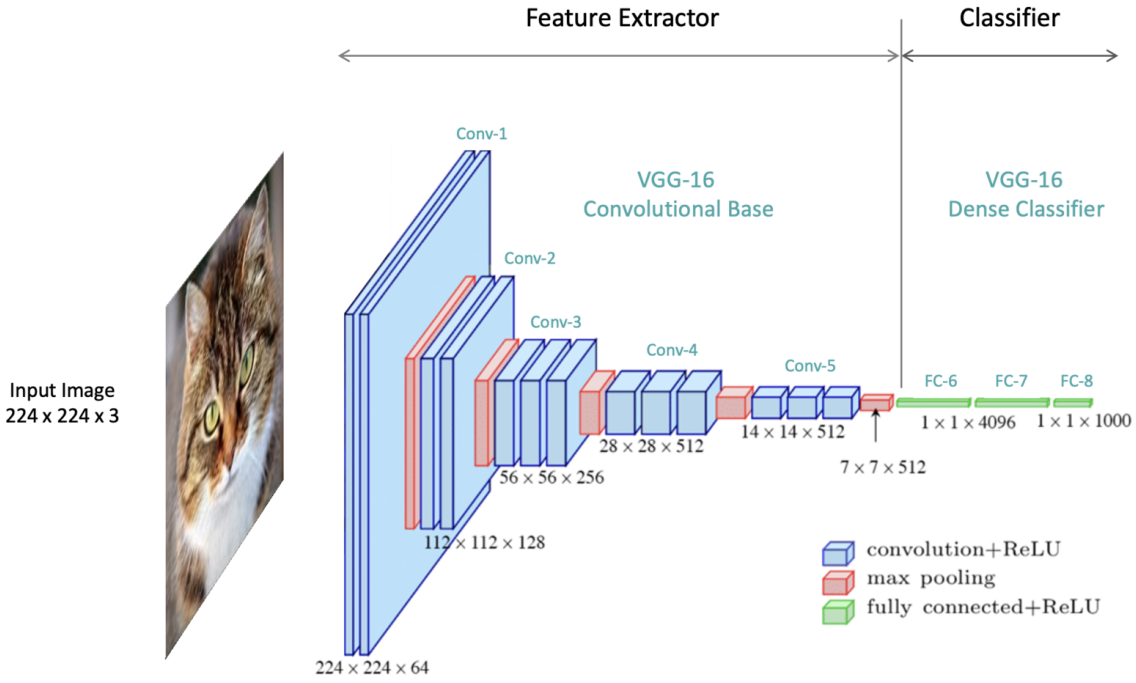


Figure 2.9 Visual Geometry Group(VGG) framework

2.3.3.1 Convolutional Layers

The image is passed through a series of convolutional layers. Each convolutional layer consists of multiple filters that convolve across the image, extracting various features through a dot product operation. The VGG model uses small 3x3 filters, which helps capture local patterns and details in the image. The rectified linear unit (ReLU) activation function is applied after each convolution to introduce non-linearity.

2.3.3.2 Max-Pooling Layers

After each convolutional layer, max-pooling layers are employed for spatial downsampling. Max-pooling reduces the spatial dimensions of the feature maps while retaining the most salient information. It achieves this by partitioning the feature maps into non-overlapping regions and selecting the maximum value within each region.

2.3.3.3 Fully Connected Layers

The output from the last max-pooling layer is flattened into a vector and fed into a series of fully connected layers. These layers perform high-level reasoning and classification based on the learned features. The fully connected layers have a large number of parameters, allowing the model to capture complex relationships between features.

2.3.3.4 Softmax Activation

The final fully connected layer is followed by a softmax activation function, which converts the outputs into probabilities representing the likelihood of the image belonging to different classes. This enables the VGG model to perform multi-class classification.

2.3.3.5 Training and Evaluation

The VGG model is trained on a large dataset with labeled images, such as the ImageNet dataset, using the aforementioned steps. During training, the model learns to extract meaningful features from the input images and make accurate predictions. The model's performance is evaluated on a separate validation or test set to assess its generalization ability.

2.3.4 Residual model

A residual model, also known as ResNet or residual network, is a deep learning architecture designed to tackle the challenges of training deep neural networks. It was introduced by Microsoft Research in 2015 and has since gained widespread adoption in computer vision tasks.The main idea behind a residual model is residual learning, which involves learning the residual or the difference between the input and output rather than learning the entire mapping directly. This is achieved using skip connections, also called shortcut connections or identity mappings. In a residual model, the

network is composed of layers or blocks, each consisting of multiple convolutional layers followed by batch normalization and activation functions like ReLU. The skip connections allow information to bypass certain layers and flow directly from earlier layers to later layers. By incorporating these connections, a residual model can effectively propagate information through the network, even in the presence of deep architectures.

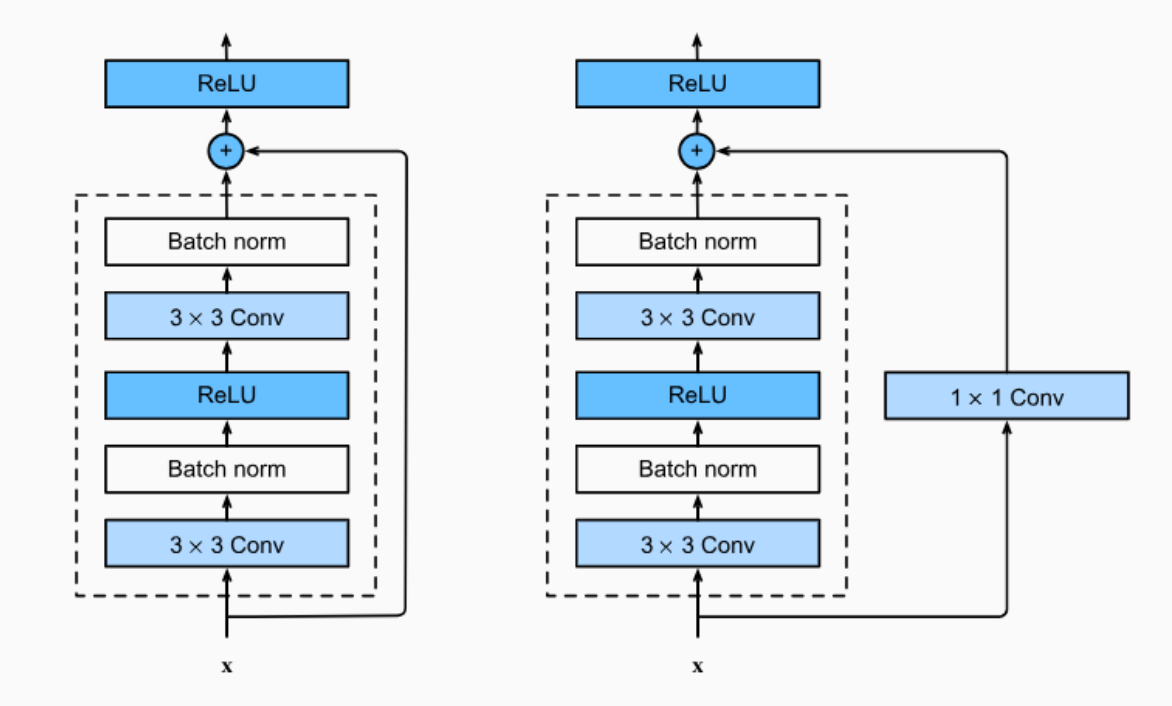


Figure 2.10 Residual model framework

2.3.4.1 Batch Normalization

Batch normalization is a technique applied after the convolutional layer in which the outputs of the previous layer are normalized. This normalization process makes the network more resilient to variations in input data and speeds up the training process.

2.3.4.2 Convolutional Layer

After obtaining the activated outputs from the previous step, they are passed through an additional convolutional layer. This layer plays a crucial role in further enhanc-

ing and refining the features that were learned within the block. By applying this additional layer, the network can continue to transform and improve the extracted features, resulting in the creation of more intricate representations and the incorporation of richer information into the model's learning process.

2.3.4.3 Activation Function (ReLU)

The batch-normalized outputs go through an activation function, commonly the Rectified Linear Unit (ReLU). ReLU introduces non-linearity, enabling the network to capture intricate relationships among features and learn complex representations.

2.3.5 Inception model

The Inception model, also known as GoogLeNet, is a convolutional neural network (CNN) architecture developed by Google researchers. It gained considerable attention for its exceptional performance in image classification tasks when it was introduced in 2014. The Inception model stands out due to its unique architecture, which incorporates multiple layers of convolutional filters with different sizes. This design enables the network to capture features at various scales and levels of abstraction, encompassing both local and global information. At the core of the Inception model is the Inception module, which consists of parallel convolutional layers with different filter sizes. The outputs of these layers are then concatenated. An important feature of the Inception model is its efficient utilization of computational resources. It achieves this by incorporating 1x1 convolutions for dimensionality reduction and reducing the overall number of parameters in the network.

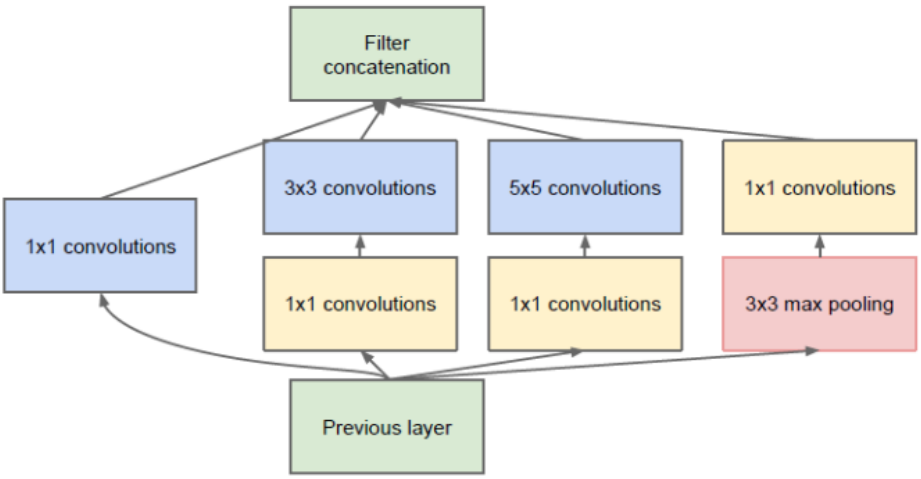


Figure 2.11 Inception model framework

2.3.5.1 Convolutional Layers

The Inception model works by applying a sequence of convolutional layers to the input image. Each layer utilizes filters to extract features from the input, focusing on specific patterns or features. What distinguishes the Inception model is its approach to capturing features at various scales. This is accomplished by employing filters of different sizes within each layer. By incorporating filters with different receptive fields, the model can capture both detailed local information and broader contextual information simultaneously. This ability to extract features at multiple scales contributes to the Inception model's effectiveness in tasks like image classification.

2.3.5.2 Inception Modules

An Inception module consists of multiple branches of convolutional layers running in parallel. Each branch employs filters of different sizes to capture features at various levels of abstraction. The outputs of these parallel branches are then merged together through concatenation, resulting in the final output of the Inception module. This merging process combines the diverse set of features extracted by each branch, creating a more comprehensive and holistic representation of the input. By incorporating

features from multiple scales and levels of abstraction, the Inception model excels at capturing intricate patterns and relationships within the data.

2.3.5.3 Max Pooling

Once the input undergoes the Inception modules, the feature maps are then subjected to pooling layers for downsampling. Pooling layers serve to decrease the spatial dimensions of the features while preserving the most relevant information. Common pooling operations, such as max pooling or average pooling, are employed for this purpose.

2.3.6 Related Work

Facial expressions are crucial for recognizing emotions and are used in non-verbal communication and identifying individuals. Since the face is the most prominent and visible part of the body, computer vision systems, often using cameras, can analyze facial images to detect emotions [17].

Social interaction refers to the act of engaging in communication, whether verbal or nonverbal, to participate in social exchanges. This can involve various forms of expression such as speech, body language, gestures, eye contact, and facial expressions, which are employed to initiate and respond to interactions with others. [18]. It has a significant impact on contemporary society in terms of diagnosing mental disorders and identifying human social and physiological interactions. [19].Emotionally intelligent systems are essential for the effective development and endurance of a range of expressions, including engagement, disappointment, sorrow, fear, and happiness. [20].A typical automated facial recognition framework consists of three main stages: face tracking, feature extraction, and expression classification. [21].The primary theories for comprehending how humans perceive and categorize emotional facial expressions have been the continuous and categorical models. In the continuous model, each emotional facial expression is defined as a feature vector in face space. On the other hand,

the categorical model consists of multiple classifiers, with each classifier specifically designed for a distinct emotion category. [22]. There are seven basic emotions angry, disgust, fear, happy, sad, surprise, neutral, may be detected by looking at a person's face [23]. The objective of distinguishing emotions is to contribute to existing research on the capacity to identify the seven fundamental emotions based on facial cues. [24].

2.4 Summary

The literature review conducted in this study aimed to provide a comprehensive understanding of facial expression recognition by examining various research papers and publications. The review focused on different learning algorithms, including machine learning and deep learning, as well as specific models such as CNN, LBP, VGG, ResNet, and Inception.Deep learning algorithms, particularly Convolutional Neural Networks (CNNs), were also extensively studied. The architectures of various CNN models, such as VGG, ResNet, and Inception, were analyzed in relation to their performance in facial expression recognition tasks. The methodologies, training procedures, and performance evaluation metrics used in these studies were critically evaluated.The literature review synthesized the existing knowledge on different learning algorithms and models in facial expression recognition.

Chapter 3

Methodology

In this chapter, we provide a detailed explanation of how we implemented and tested our model using the fer2013 dataset. We discuss the steps involved in collecting the required images, as well as the processes of preprocessing, feature extraction, and classification. Additionally, we highlight the main differences between the mathematical models used in our study.

3.1 Introduction

In our study comparing different models, such as CNN, we provide a detailed description of how these models function.Once the Emotion Recognition Model is established, it needs to undergo testing to evaluate its performance. This involves detecting faces, preprocessing the images, applying LBP for feature extraction, and obtaining the feature vectors. The extracted features are then compared with the model to determine the output of the image, which represents the recognized emotion. Here is a block diagram of the procedure of our working environment.

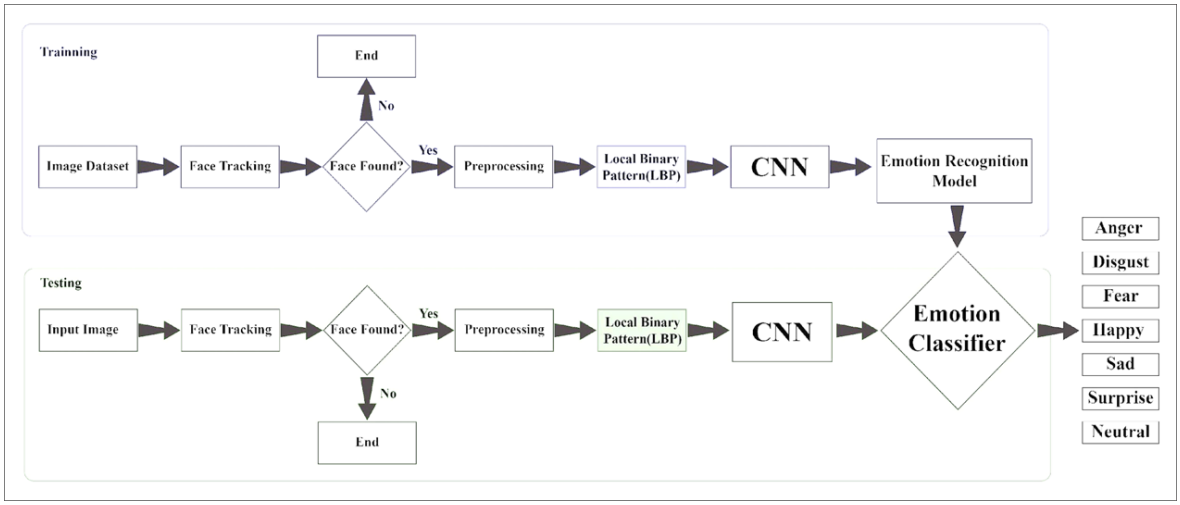


Figure 3.1 Flow chart of the phases of the Facial Expression Recognition (FER) method

3.2 Dataset

In our research, we made use of two distinct datasets. The first dataset we utilized for training our facial expression recognition system was the FER2013 dataset, which is widely recognized and proven to be effective for this purpose. As for the second dataset, it was specifically collected to represent real-life scenarios and was exclusively employed for testing the performance of our system.It should be noted that the FER2013 dataset has a lower resolution of 48x48 pixels, which makes it particularly useful for assessing performance in low-light scenarios. These datasets encompass a combination of publicly validated and tested datasets, as well as private dataset.

3.2.1 FER2013 Dataset

The FER2013 dataset from the FER2013 Kaggle Challenge [25] The FER2013 dataset is a commonly used dataset in the field of facial expression recognition. It contains a large collection of facial images labeled with corresponding emotion categories. The dataset consists of approximately 35,887 grayscale images, each sized at 48x48 pixels

which are further categorized into 3,589 experiments and 28,709 training images. The dataset also encompasses 3,589 private test photos specifically utilized for final evaluation.. These images represent various facial expressions, including anger, disgust, fear, happiness, sadness, surprise, and neutral. The FER2013 dataset has been widely utilized for training and evaluating facial expression recognition models, enabling researchers to develop and benchmark their algorithms in this domain.The distribution of expressions within the FER2013 dataset is visualized in Figure 3.2.

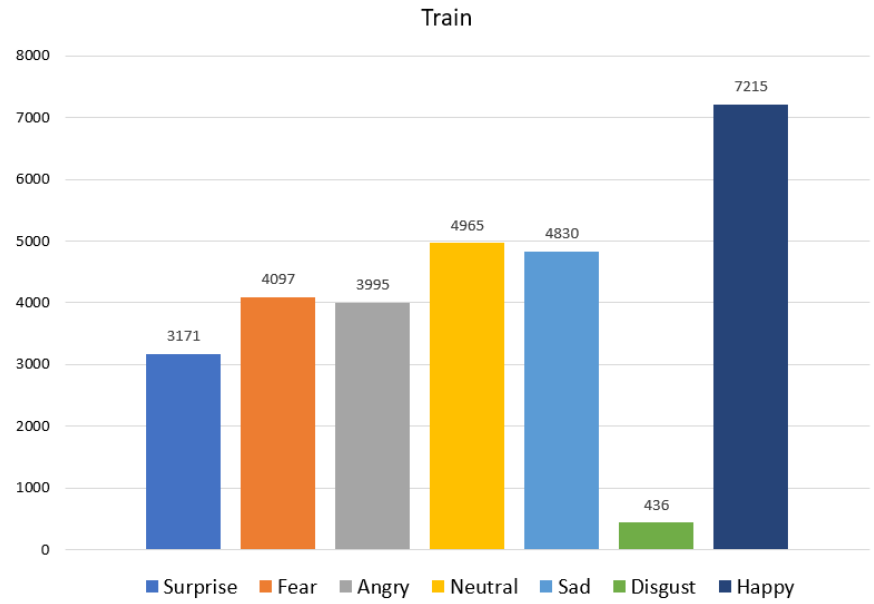


Figure 3.2 FER2013 dataset distribution of seven face expressions

For the sake of our thesis, we developed several test photos. Figure 3.3 shows an example of seven fundamental emotions from our own test picture database.

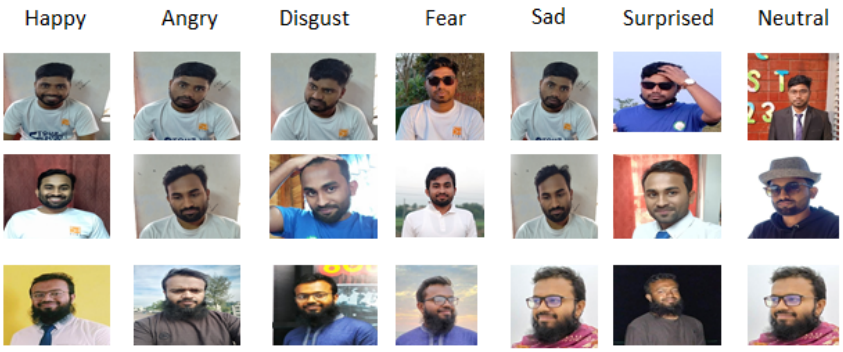


Figure 3.3 Seven basic emotional examples for verification purposes

3.3 Face recognition

Face recognition is a technology that involves identifying and verifying individuals based on their unique facial features.Face recognition technology finds applications in various domains, including security systems (such as access control and surveillance), identity verification for authentication purposes (e.g., unlocking smartphones).CNNs are capable of automatically learning discriminative features from raw facial images, enabling more accurate and efficient face recognition algorithms. However, it is important to note that face recognition technology raises privacy and ethical concerns. Here an illustrated in Figure 3.4, where rectangular features are employed in this approach to detect facial components within an image. [26].

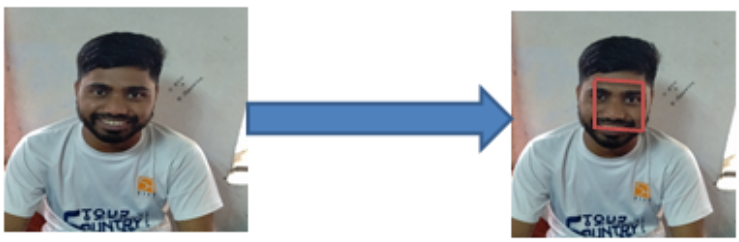


Figure 3.4 Selection of different face components

3.4 Pre-data processing

Prior to training our facial expression recognition system, we performed data preprocessing to align and normalize the facial images. This entails utilizing face detection and alignment algorithms to crop the faces and resize them to a consistent size.The facial images were preprocessed to enhance their quality and remove any noise or artifacts. Common preprocessing techniques include resizing, normalization, and histogram equalization.

3.5 Feature Extraction

Feature extraction involves converting raw data into numerical features that can be processed while preserving the original information in the dataset. This technique enhances the application of machine learning algorithms directly on raw data. An effective method for visual feature extraction is Local Binary Pattern (LBP), which is a straightforward yet powerful descriptor. LBP reduces the computational load on Convolutional Neural Networks (CNN) and improves the accuracy of classification tasks [27].Use LBP to extract texture features from the images or videos. This can involve dividing the image into small patches and computing the LBP descriptor for each patch.

3.5.1 Local Binary Pattern (LBP)

Local Binary Patterns (LBP) based image filter operators are employed to describe textures and extract features. The input image is divided into a grid of cells or patches.For each cell, a center pixel is selected, and the pixel intensities of its neighboring pixels are compared with the center pixel's intensity. Typically, a circular neighborhood of pixels is considered.Based on the intensity comparisons, a binary pattern is generated, where each neighbor pixel is assigned a value of 1 if its intensity is greater or equal

to the center pixel's intensity, and 0 otherwise. This results in a binary pattern representation for the center pixel's neighborhood. The LBP algorithm captures local texture patterns in an image and has been proven effective in various computer vision tasks, including facial expression recognition. By encoding local texture information, LBP provides discriminative features that can be used to distinguish different facial expressions.

3.6 Experimental Procedure

The proposed methodology in this research focuses on several key processes, including face detection, feature extraction, and facial expression categorization. The goal is to develop an approach that utilizes a Convolutional Neural Network (CNN) for accurate emotion detection by extracting meaningful features from facial images.Face recognition is a computer vision technique that involves identifying human faces in digital photos. In this research, the Local Binary Pattern (LBP) method is employed for face recognition due to its high accuracy and real-time performance.

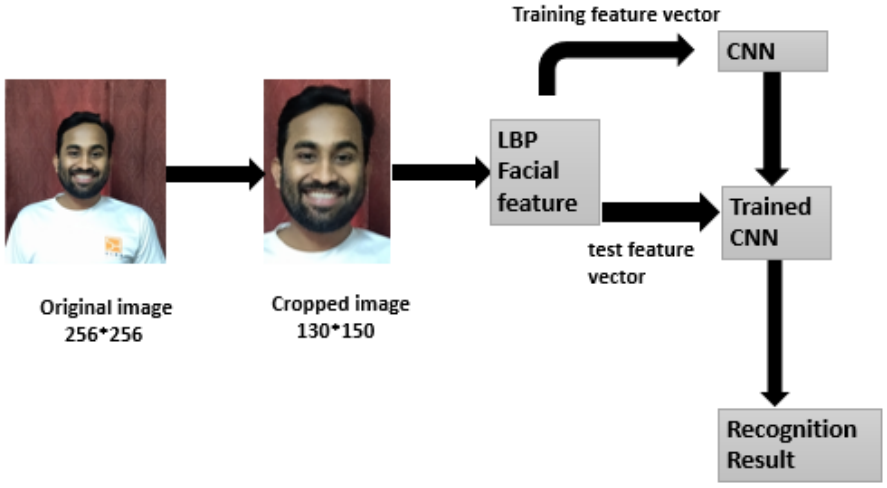


Figure 3.5 Architecture of our proposed system

The training process involves iteratively adjusting the network weights through forward and backward passes.The datasets used in this research include the FER2013 and Own datasets. The FER2013 dataset comprises grayscale face images with a resolution of 48x48 pixels. The Own dataset contains grayscale images of ten Bangladeshi male participants with a resolution of 256x256 pixels. Both datasets are labeled with seven different emotions, and each emotion is represented by a numerical label. By leveraging the CNN architecture, integrating LBP, and utilizing these datasets, the proposed methodology aims to improve facial expression recognition by accurately detecting and categorizing emotions.

3.7 Summary

The methodology followed a sequential process, starting from preprocessing the input images, extracting relevant features, classifying the facial expressions, updating the model's weights based on the evaluation results, and finally obtaining the predicted desired output for each facial image.

Note: The above description provides a general outline of the methodology section. The specific techniques and algorithms used in each step may vary depending on the research approach and the chosen models for facial expression recognition.

Chapter 4

EXPERIMENTAL RESULT AND DISCUSSION

In this chapter, we have described our data set collection, required tools, experimental results, and discussion of them.

4.1 Data Set

In our system, we have used FER2013 datasets.

4.2 Environmental Equipment

We are utilizing the Google Colab environment for our coding tasks. Google Colab is an online platform provided by Google that offers a Jupyter Notebook environment for writing and executing code. It provides a convenient and powerful environment, especially for machine learning and data analysis tasks, with features like pre-installed libraries, access to cloud-based computing resources, collaboration capabilities, integration with Google Drive, and free usage.We've used Windows 10 OS.Windows 10 provides a highly customizable command interface and offers a robust platform for

resolving various issues. To ensure fast processing, we utilized a system with 4GB RAM and an Intel Core i3-4200U CPU, operating at 1.60GHz (with a maximum turbo frequency of 2.30 GHz). This configuration helped us achieve efficient and speedy processing for our implementations.

4.3 Result Analysis and Discussion

In this section, we conducted an evaluation of several methods, including Convolutional neural network(CNN), Visual Geometry Group (VGG), Inception, and Residual model, using the Facial Expression Recognition (FER2013) dataset. The FER2013 dataset is commonly used for training and evaluating facial expression recognition models. By comparing the results obtained from the CNN, VGG, Inception, and Residual models, we gained insights into their respective strengths and weaknesses in facial expression recognition. This analysis helped us understand the relative effectiveness of each method and provided valuable information for selecting the most suitable model for facial expression recognition tasks.The databases consist of three parts 80 percentages training data, 10 percentages public testing data, and another 10 percentages of images for other purposes.

In the evaluation of different models, we compare their accuracy, loss, value loss, and value accuracy to determine the best-performing model. By comparing the accuracy, loss, value loss, and value accuracy across different models, we can identify the model that achieves the highest accuracy, lowest loss, and best generalization performance. This comparison allows us to select the best model for facial expression recognition based on its ability to accurately classify facial expressions, minimize errors, and generalize well to unseen data.

4.3.1 Confusion Matrix

Comparing the confusion matrices across the different models allows us to determine which model achieves the best performance in terms of accurately classifying facial expressions. The model with the highest overall accuracy, precision, recall, or F1 score, as derived from the confusion matrix, can be considered the best performing model in this context.

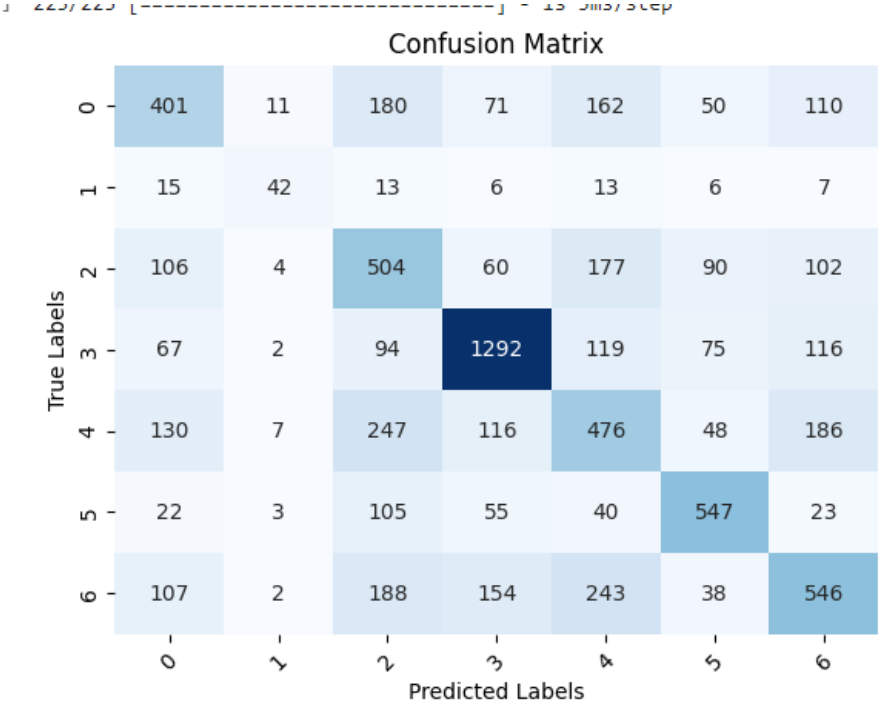


Figure 4.1 Confusion matrix of facial emotion recognition results on the FER2013 dataset from CNN model

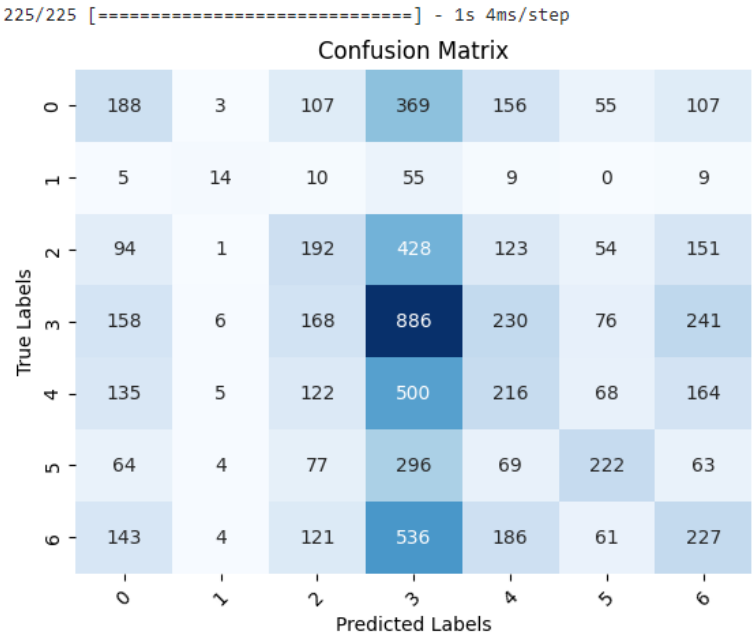


Figure 4.2 Confusion matrix of facial emotion recognition results on the FER2013 dataset from Residual Model

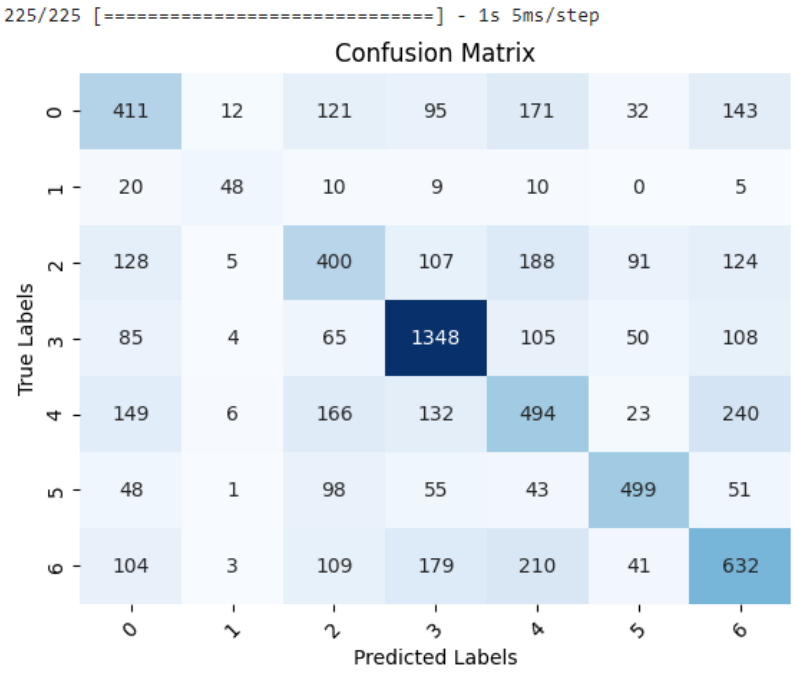


Figure 4.3 Confusion matrix of facial emotion recognition results on the FER2013 dataset from VGG model

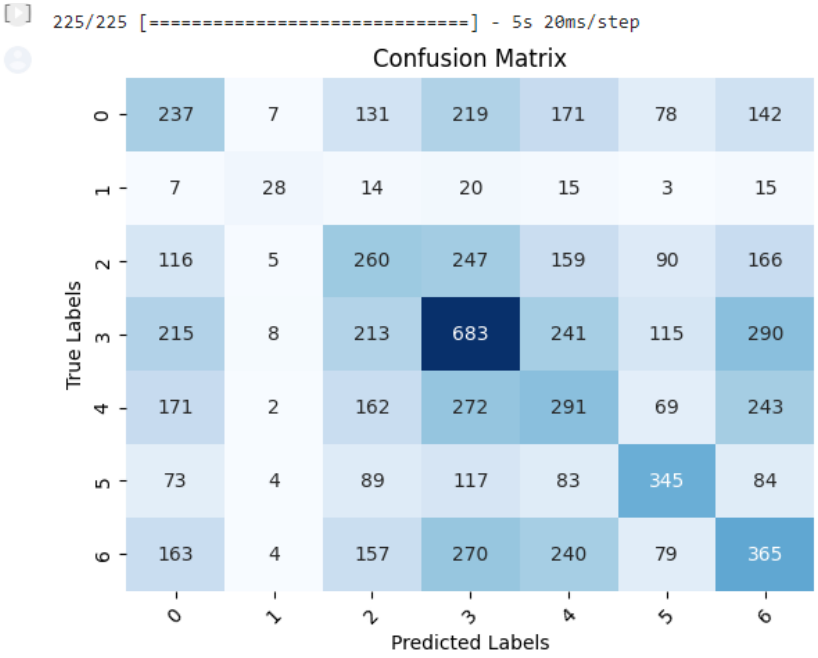


Figure 4.4 Confusion matrix of facial emotion recognition results on the FER2013 dataset from Inception model

4.3.2 Classification report

A classification report is a useful tool for evaluating the performance of a classification model. It provides a comprehensive summary of various evaluation metrics, such as accuracy, precision, recall, and F1 score, for each class in a multi-class classification problem.

	precision	recall	f1-score	support
0	0.48	0.52	0.50	498
1	0.62	0.44	0.52	52
2	0.56	0.43	0.48	545
3	0.77	0.84	0.80	881
4	0.50	0.54	0.52	588
5	0.75	0.76	0.75	414
6	0.60	0.57	0.58	611
accuracy			0.62	3589
macro avg	0.61	0.58	0.59	3589
weighted avg	0.62	0.62	0.62	3589

Figure 4.5 Classification report of CNN model

4.3.3 Loss,value loss,accuracy,value accuracy every epoch

1. Loss:- Loss refers to the objective function or the measure of how well the model is performing during training. It quantifies the error between the predicted facial expression labels and the ground truth labels. The goal of training is to minimize this loss value, indicating that the model is becoming more accurate in its predictions.

2. Value Loss :- Value loss is a specific type of loss that is calculated during the validation or testing phase of the model. It represents the loss value obtained on a separate dataset that was not used for training. Value loss helps evaluate the generalization ability of the model and provides an indication of its performance on unseen data.

3. Accuracy:- Accuracy is a metric that measures the percentage of correctly classified facial expressions by the model. It compares the predicted labels with the ground truth labels and determines how many predictions are correct. Higher accuracy values indicate better performance.

4. Value Accuracy:- Similar to value loss, value accuracy is the accuracy calculated on a separate validation or testing dataset. It represents the model's performance on unseen data and provides an estimate of its ability to generalize.

Epoch	Loss	Value Loss	Accuracy	Value Accuracy
1	1.5221	1.6978	0.4092	0.3848
81	0.0401	2.9951	0.9897	0.6183

Table 4.1 Example table for CNN Model

Epoch	Loss	Value Loss	Accuracy	Value Accuracy
1	1.8532	1.7207	0.2680	0.3400
30	0.2751	3.1489	0.9129	0.5155

Table 4.2 Example table for VGG Model

Epoch	Loss	Value Loss	Accuracy	Value Accuracy
1	0.9205	11.7199	0.6475	0.2557
4	0.8334	12.5358	0.6805	0.2529

Table 4.3 Example table for Residual Model

4.3.4 Loss,value loss,accuracy,value accuracy Curve

By examining these curves, researchers and practitioners can evaluate how the model is progressing during training, detect any problems related to overfitting or underfitting, and make informed choices regarding model selection and optimization.

Epoch	Loss	Value Loss	Accuracy	Value Accuracy
1	.023	1.32	.23	2.3
6	0.9229	1.2693	0.6544	0.5413

Table 4.4 Example table for Iception Model

```
[ ]  plt.plot(history_residual.history['accuracy'], 'b')
     plt.plot(history_inception.history['accuracy'], 'g')
     plt.plot(history_vgg.history['accuracy'], 'r')
     plt.plot(history_cnn.history['accuracy'], 'c')
```

```
[<matplotlib.lines.Line2D at 0x7f7ce4e1d990>]
```

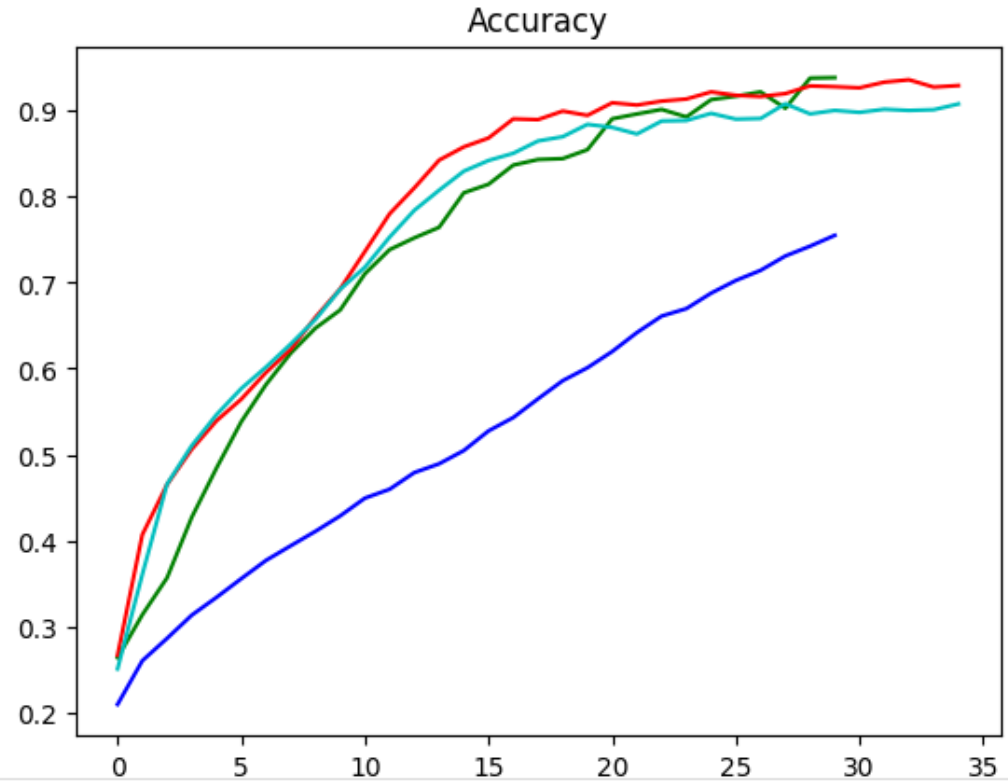


Figure 4.6 FER2013 dataset Accuracy Curve for CNN model,VGG model,Inception model,Residual model

```
plt.plot(history_residual.history['val_accuracy'], 'b')
plt.plot(history_inception.history['val_accuracy'], 'g')
plt.plot(history_vgg.history['val_accuracy'], 'r')
plt.plot(history_cnn.history['val_accuracy'], 'c')
```

```
[<matplotlib.lines.Line2D at 0x7f7ce4bece50>]
```

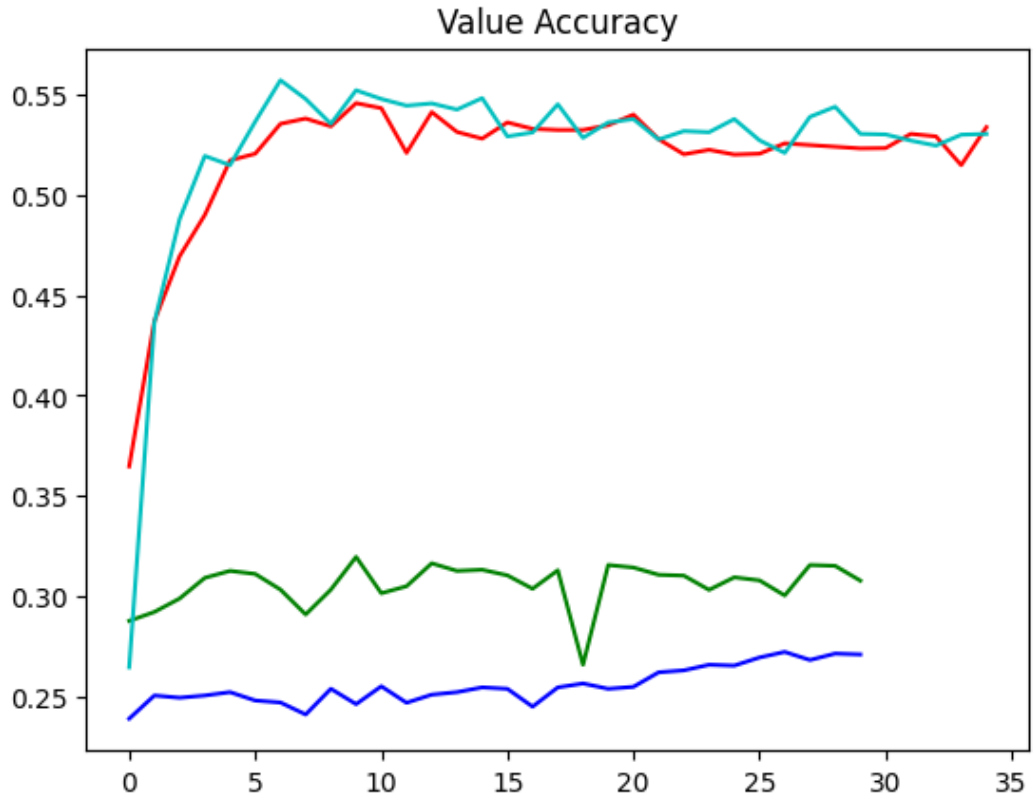


Figure 4.7 FER2013 dataset Value Accuracy Curve for CNN model,VGG model,Inception model,Residual model

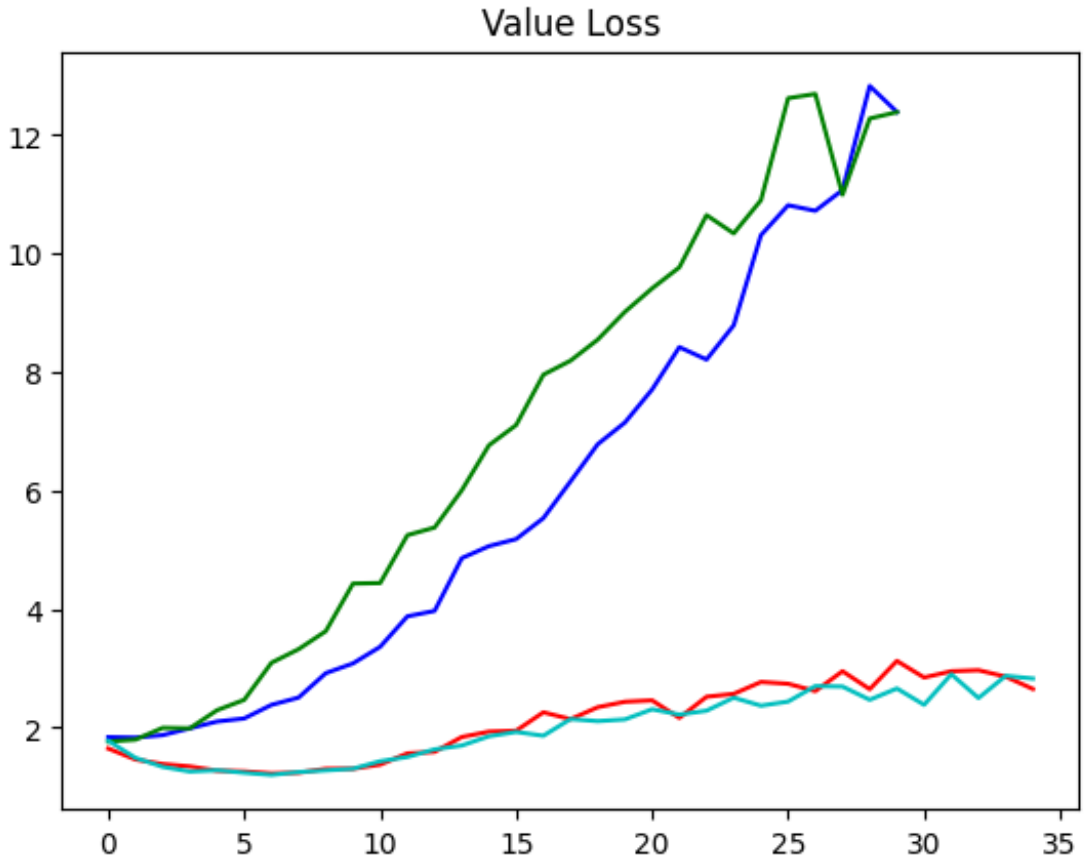


Figure 4.8 FER2013 dataset Loss Curve for CNN model,VGG model,Inception model,Residual model

```
plt.plot(history_residual.history['loss'], 'b')
plt.plot(history_inception.history['loss'], 'g')
plt.plot(history_vgg.history['loss'], 'r')
plt.plot(history_cnn.history['loss'], 'c')
```

```
[<matplotlib.lines.Line2D at 0x7f7ce4cbf340>]
```

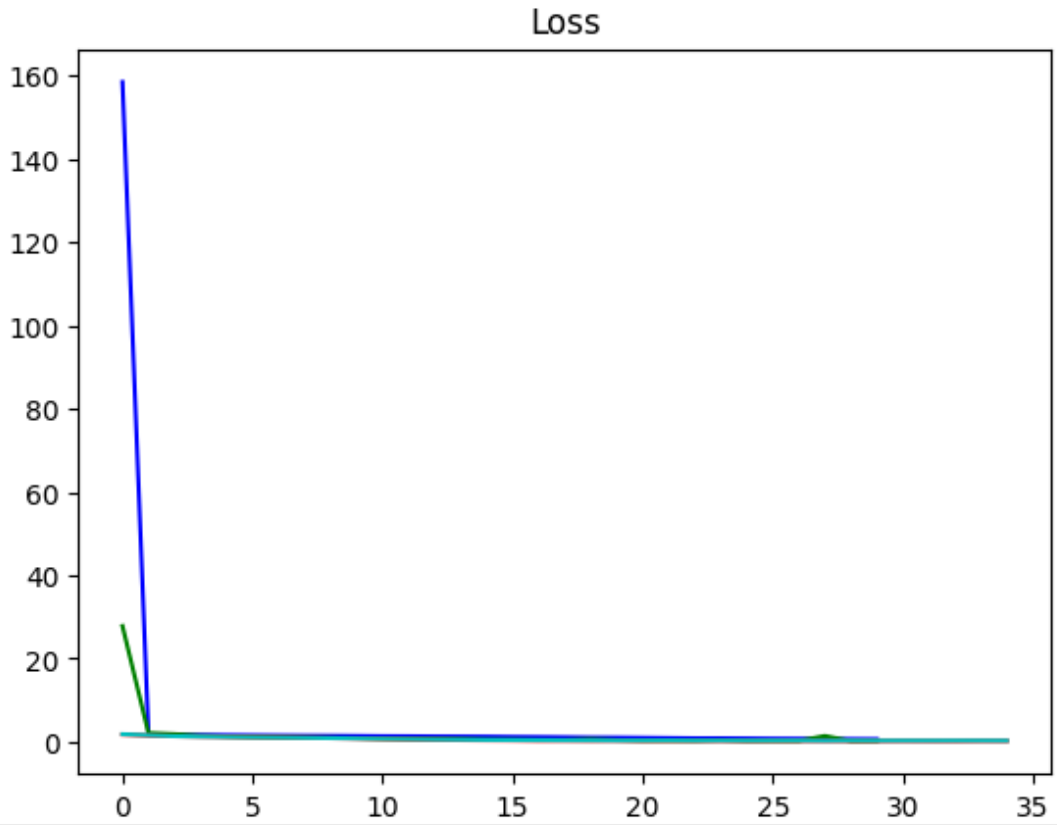


Figure 4.9 FER2013 dataset Value Loss Curve for CNN model,VGG model,Inception model,Residual model

Chapter 5

CONCLUSION AND FUTURE WORK

5.1 Conclusion

the objective of increasing accuracy with the trained model and testing it on another real-time generated dataset was not fully achieved.These results indicate that there is room for improvement in the performance of the model.The findings suggest that the current methodology may need to be revised to enhance accuracy and overall performance. Exploring alternative approaches such as refining the preprocessing techniques, fine-tuning the model architecture, or incorporating additional features could potentially lead to improved results. It is important to acknowledge that the challenges faced in achieving higher accuracy may stem from the limitations of the dataset or the complexity of the facial expression recognition task.

5.2 Future Work

In the future, researchers and practitioners in facial expression recognition using CNN and Local Binary Patterns can focus on several areas. Firstly, expanding the datasets used for training and evaluation to include more diverse and representative samples can enhance the model's performance in real-world scenarios. Secondly, exploring novel network architectures and optimization techniques can further improve the accuracy and efficiency of the models.

References

[1] Erika L Rosenberg and Paul Ekman. *What the face reveals: Basic and applied studies of spontaneous expression using the Facial Action Coding System (FACS).* Oxford University Press, 2020.

[2] Lucy Nwosu, Hui Wang, Jiang Lu, Ishaq Unwala, Xiaokun Yang, and Ting Zhang. Deep convolutional neural network for facial expression recognition using facial parts. *2017 IEEE 15th Intl Conf on Dependable, Autonomic and Secure Computing, 15th Intl Conf on Pervasive Intelligence and Computing, 3rd Intl Conf on Big Data Intelligence and Computing and Cyber Science and Technology Congress(DASC/PiCom/DataCom/CyberSciTech)*, pages 1318–1321, 2017.

[3] Chieh-En James Li and Lanqing Zhao. Emotion recognition using convolutional neural networks. 2019.

[4] Ji-Hae Kim, Byung-Gyu Kim, Partha Pratim Roy, and Da-Mi Jeong. Efficient facial expression recognition algorithm based on hierarchical deep neural network structure. *IEEE access*, 7:41273–41285, 2019.

[5] Nithya Roopa. S "emotion recognition from facial expression using deep learning". *International Journal of Engineering and Advanced Technology (IJEAT) ISSN*, pages 2249–8958, 2019.

[6] A.L.I. Ghali and Mohamad-Bassam Kurdy. Emotion recognition using facial expression analysis. *Journal of Theoretical and Applied Information Technology*, 96:6117–6129, 09 2018.

[7] Tanoy Debnath, Md Reza, Anichur Rahman, Amin Beheshti, Shahab S Band, Hamid Alinejad-Rokny, et al. Four-layer convnet to facial emotion recognition with minimal epochs and the significance of data diversity. *Scientific Reports*, 12(1):1–18, 2022.

[8] Ashish Bakshi. What is deep learning? https://www.edureka.co/blog/what-is-deep-learning, February 2001.

[9] g.sumalatha. Facial expression recognition using cnn. *IJIRT*, 8(11), 2022.

[10] Gourav Singh. Top 5 skills needed to be a deep learning engineer! https://www.analyticsvidhya.com/blog/2021/06/top-5-skills-needed-to-be-a-deep-learning-engineer/, 2021.

[11] IBM Cloud Education. Machine learning. https://www.analyticsvidhya.com/blog/2021/06/top-5-skills-needed-to-be-a-deep-learning-engineer/, 2020.

[12] Ian Goodfellow, Yoshua Bengio, and Aaron Courville. *Deep Learning*. MIT Press, 2016. http://www.deeplearningbook.org.

[13] Priya Pedamkar. Machine learning algorithms. https://www.educba.com/machine-learning-algorithms/.

[14] Ali Mollahosseini, David Chan, and Mohammad H. Mahoor. Going deeper in facial expression recognition using deep neural networks. *CoRR*, abs/1511.04110, 2015.

[15] Richard HR Hahnloser, Rahul Sarpeshkar, Misha A Mahowald, Rodney J Douglas, and H Sebastian Seung. Digital selection and analogue amplification coexist in a cortex-inspired silicon circuit. *nature*, 405(6789):947–951, 2000.

[16] Manasi Patil, Brijesh Iyer, and Rajeev Arya. *Performance Evaluation of PCA and ICA Algorithm for Facial Expression Recognition Application*, pages 965–976. 03 2016.

[17] Paweł Tarnowski, Marcin Kołodziej, Andrzej Majkowski, and Remigiusz Rak. Emotion recognition using facial expressions. *Procedia Computer Science*, 108:1175–1184, 12 2017.

[18] N Udayakumar. Facial expression recognition system for autistic children in virtual reality environment. *Int. J. Sci. Res. Publ*, 6(6):613–622, 2016.

[19] S Shaul Hammed, A Sabanayagam, and E Ramakalaivani. A review on facial expression recognition systems. *Journal of critical reviews*, 7(4):903–905, 2020.

[20] V. Hima Deepthi G. Sailaja. Facial expression recognition complications with the stages of face detection and recognition. *International Journal of Recent Technology and Engineering (IJRTE)*, 8(2), July 2019.

[21] Rizwan Khan. Detection of emotions from video in non-controlled environment. 11 2013.

[22] Aleix Martinez and Shichuan Du. A model of the perception of facial expressions of emotion by humans: Research overview and perspectives. *Journal of machine learning research : JMLR*, 13:1589–1608, May 2012.

[23] Maria Guarnera, Zira Hichy, Maura Cascio, Stefano Carrubba, and Stefania Buccheri. Facial expressions and the ability to recognize emotions from the eyes or mouth: A comparison between children and adults. *The Journal of Genetic Psychology*, 178:1–10, 10 2017.

[24] Elena Ryumina and Alexey Karpov. Facial expression recognition using distance importance scores between facial landmarks. *Proceedings of the 30th International Conference on Computer Graphics and Machine Vision (GraphiCon 2020). Part 2*, pages paper32–1, 12 2020.

[25] Kaggle. Fer-2013 dataset.

[26] A. Mohan, C. Papageorgiou, and T. Poggio. Example-based object detection in images by components. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 23(4):349–361, 2001.

[27] Xiangpo Wei, Xuchu Yu, Bing Liu, and Lu Zhi. Convolutional neural networks and local binary patterns for hyperspectral image classification. *European Journal of Remote Sensing*, 52(1):448–462, 2019.

Appendix

---

https://github.com/janturahaman/Facial$_e xpression$