Residual Neural Network for Facial Emotion Recognition and Video Recommendation

**MAJOR PROJECT REPORT**

*Submitted by*

# Suprabhat Pal [Reg No: RA1911033010156] Sejal Kumari Gupta [Reg No: RA1911033010157]

*Under the Guidance of*

# Dr. S. Sadagopan

(Assistant Professor, Department of Computational Intelligence)

*in partial fulfillment of the Requirements for the Degree of*

BACHELOR OF TECHNOLOGY

in

COMPUTER SCIENCE ENGINEERING

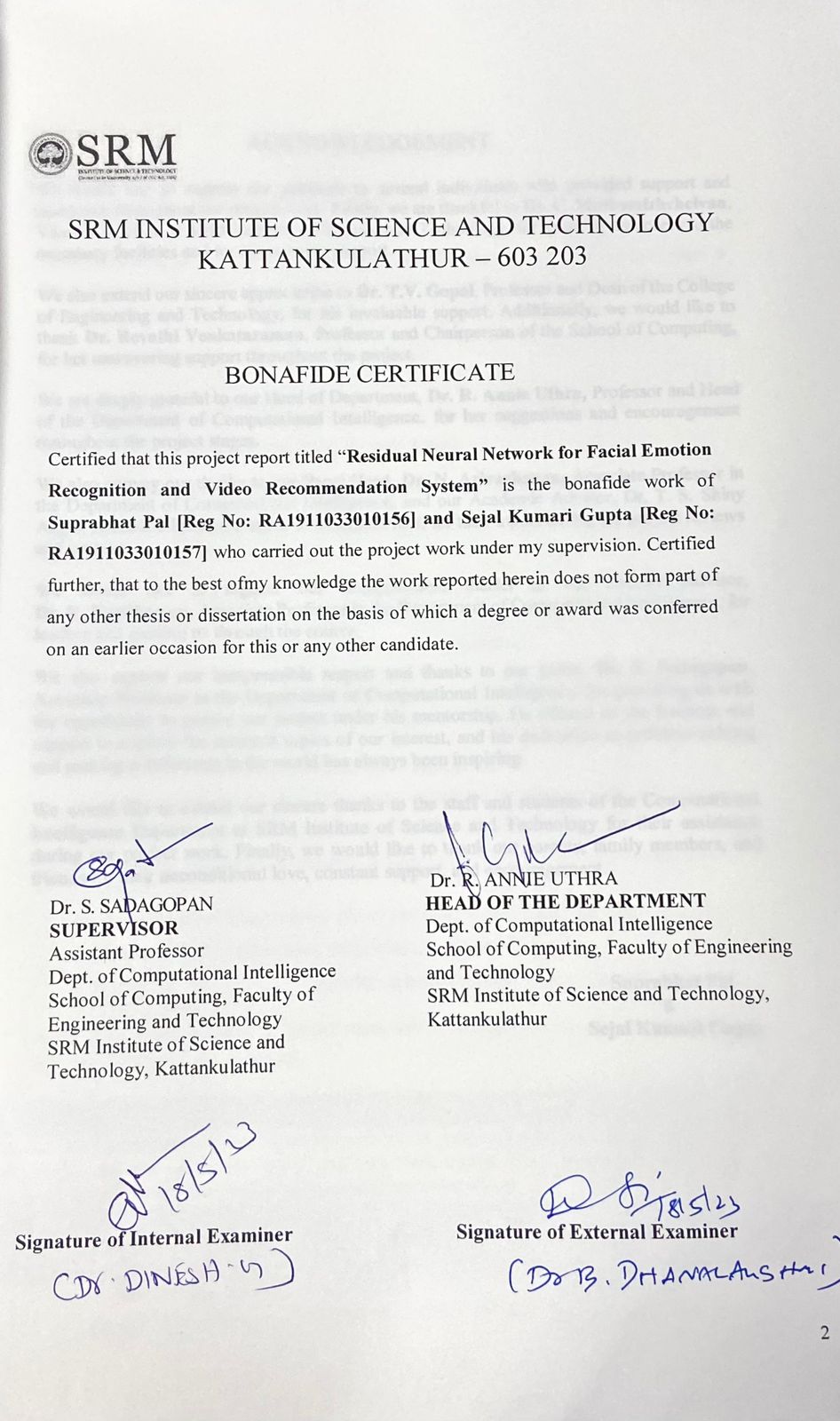
with specialization in Software Engineering



DEPARTMENT OF COMPUTATIONAL INTELLIGENCE COLLEGE OF ENGINEERING AND TECHNOLOGY SRM INSTITUTE OF SCIENCE AND TECHNOLOGY

KATTANKULATHUR- 603203

MAY 2023



ACKNOWLEDGEMENT

We would like to express our gratitude to several individuals who provided support and assistance throughout our project work. Firstly, we are thankful to **Dr. C. Muthamizhchelvan**, Vice-Chancellor of SRM Institute of Science and Technology, for granting us access to the necessary facilities and for his ongoing support.

We also extend our sincere appreciation to **Dr. T.V. Gopal**, Professor and Dean of the College of Engineering and Technology, for his invaluable support. Additionally, we would like to thank **Dr. Revathi Venkataraman**, Professor and Chairperson of the School of Computing, for her unwavering support throughout the project.

We are deeply grateful to our Head of Department, **Dr. R. Annie Uthra**, Professor and Head of the Department of Computational Intelligence, for her suggestions and encouragement throughout the project stages.

We also convey our thanks to our Panel Head, **Dr. N. Arivazhagan**, Associate Professor in the Department of Computational Intelligence, and our Academic Advisor, Dr. T. S. Shiny Angel, Associate Professor in the same department, for their inputs during the project reviews and support.

We would like to register our immeasurable thanks to our Faculty Advisor, **Dr. K. Kottilingam**, Associate Professor in the Department of Computational Intelligence, for leading and guiding us through the course.

We also express our inexpressible respect and thanks to our guide, **Dr. S. Sadagopan**, Associate Professor in the Department of Computational Intelligence, for providing us with the opportunity to pursue our project under his mentorship. He offered us the freedom and support to explore the research topics of our interest, and his dedication to problem-solving and making a difference in the world has always been inspiring.

We would like to extend our sincere thanks to the staff and students of the Computational Intelligence Department at SRM Institute of Science and Technology for their assistance during our project work. Finally, we would like to thank our parents, family members, and friends for their unconditional love, constant support, and encouragement.

## Suprabhat Pal

**&**

**Sejal Kumari Gupta**

# ABSTRACT

Facial emotion recognition and video recommendation has become one the major research topic on the field of machine learning and computer vision. There are different challenges for the given topics to overcome. This facial emotion recognition has always been a challenging task which has a wide range of application in different field such as healthcare, education, and entertainment. Many research in the field of the facial emotion recognition has been done widely in the world. There has been research in the fast-growing topics facial emotion recognition and video recommendation. Various techniques have been utilized to recognize facial expressions in different studies. Some of these conventional methods involve using distinctive attributes like Histogram of Oriented Gradients (HOG), Local Binary Patterns (LBP), and facial landmarks. In contrast, others have explored advanced approaches like Deep Learning using Convolutional Neural Networks (CNNs) and Recurrent Neural Networks (RNNs), which have shown to perform well in tasks involving facial emotion recognition and video recommendation. However, these methods often require a large amount of labeled data and computational resources. We have proposed approach for building a video recommendation system consists of three primary stages. The proposed system employs machine learning and deep learning to interpret the emotions of users in real-time. Facial characteristics are captured using a live feed of users, and a Residual Neural Network is utilized to detect faces and determine their attributes, enabling emotion identification. The proposed research aims to use a ResNet-based approach for facial emotion recognition and video recommendation, which has the potential to improve the accuracy of facial emotion recognition and the quality of video recommendations.

Index Terms: CNN (Convolutional Neural Networks), ResNet (Residual Neural Network), FER (Facial Emotion Recognition), HaarCascade algorithm

# TABLE OF CONTENTS

|  |  |  |
| --- | --- | --- |
| **CHAPTER NO.** | **TITLE** | **PAGE NO.** |
| **ABSTRACT** |  | v |
| **LIST OF TABLES**  **LIST OF FIGURES** |  | vii  viii |
| **ABBREVIATIONS** |  | x |

1. INTRODUCTION 11
   1. [INTRODUCTION 1](#_bookmark0)1
   2. [MOTIVATION 14](#_TOC_250014)
   3. OBJECTIVES 15
   4. [PURPOSE 15](#_TOC_250013)
   5. [SCOPE 1](#_bookmark1)6
2. LITERATURE REVIEW 18
   1. [LITERATURE SURVEY 1](#_bookmark2)8
   2. [INFERENCE – EXISTING SYSTEMS 23](#_TOC_250012)
3. SYSTEM DESIGN AND ANALYSIS 24
   1. [SYSTEM REQUIREMENT AND ENVIRONMENT 24](#_TOC_250011)
   2. [ARCHITECTURE & UML DIAGRAMS 2](#_bookmark3)5
      1. [SYSTEM ARCHITECTURE DIAGRAM 26](#_TOC_250010)
   3. [ALGORITHM USED](#_bookmark4) 27
4. METHODOLOGY 29
   1. [DATA PREPROCESSING 29](#_TOC_250009)
   2. [TRAINING AND TESTING THE MODEL 30](#_TOC_250008)
      1. [CREATING CNN MODEL 30](#_TOC_250007)
      2. [COMPILING THE MODEL 31](#_TOC_250006)
      3. [TRAINING THE MODEL 31](#_TOC_250005)
      4. [TESTING THE MODEL 32](#_TOC_250004)
   3. [FACE DATA 32](#_TOC_250003)
   4. [EMOTION DETECTION 34](#_TOC_250002)
   5. [GRAPHICAL REPRESENTATION 36](#_TOC_250001)
   6. VIDEO RECCOMENDATION AND DATASET 37
5. RESULTS AND DISCUSSION 40
6. CONCLUSION AND FUTURE WORK 43

REFERENCES 44

[APPENDIX 46](#_TOC_250000)

PLAGIARISM REPORT 52

**LIST OF TABLES**

|  |  |  |
| --- | --- | --- |
| **Table No.** | **Description** | **Pg. No** |
| 3.1 | Technical specification of the project application | 30 |
| 3.2 | Environment specification of the project application | 31 |

# LIST OF FIGURES

**Figure No. Description Pg. No**

|  |  |  |
| --- | --- | --- |
| 1.1 | Machine learning process | 12 |
| 1.2 | Neural Network | 13 |
| 3.2 | Architecture Diagram | 30 |
| 4.1 | Workflow | 31 |
| 4.2 | CNN model architecture | 34 |
| 4.3 | Face detection flow based on the Haar Cascade classifier | 35 |
| 4.4 | 4x4x3 RGB Image | 35 |
| 4.5 | Movement of the Kernel | 36 |
| 4.6 | A sample of Haar feature | 40 |
| 4.7 | Emotion Detection | 41 |
| 4.8 | Accuracy of the system | 41 |
| 5.1 | Happy emotion detection | 38 |
| 5.2 | Neutral emotion detection | 39 |
| 5.3 | Video recommendation by the detection of emotion | 39 |
| 5.4 | Loss and Validation loss curves of model loss | 40 |
| 5.5 | Accuracy and Validation accuracy curves of model  accuracy | 40 |

# ABBREVIATIONS

|  |  |
| --- | --- |
| **HOG** | Histogram of Oriented Gradients |
| **LBP** | Local Binary Patterns |
| **SVM** | Support Vector Machine |
| **CNN** | Convolutional Neural Networks |
| **ResNet** | Residual Neural Network |
| **FER** | Facial Expression Recognition |
| **GAP** | Global Average Pooling |
| **GAFS** | Genetic Algorithms as a Feature Selection Method |

**CHAPTER 1**

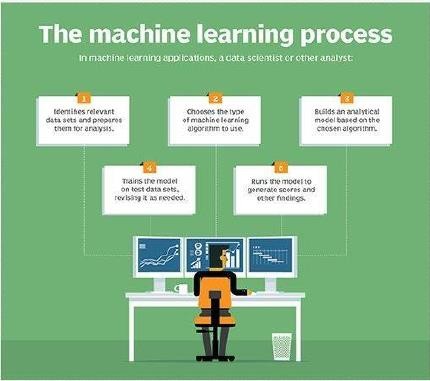
# INTRODUCTION

## INTRODUCTION

Facial emotion recognition and video recommendation tasks hold significant importance on realms of machine learning. They present inherent challenges while offering diverse applications across fields like healthcare, education, and entertainment. Facial emotion recognition involves the automatic identification of emotional states and facial expressions from images or videos. On the other hand, video recommendation entails suggesting relevant videos based on user preferences and behaviors. Goal of Facial Expression Recognition (FER) is to establish a link between specific facial expressions and their corresponding emotional states. This involves extracting facial image features and subsequently recognizing the expressed emotion. Before feeding facial images into a CNN or any other machine learning classifier, it is necessary to employ certain image processing techniques.

Existing methods include discrete wavelet transform [1], linear discriminant analysis [2], histogram equalization [3], histogram of gradients [4], viola-jones algorithm [5], etc. In recent times, Residual Neural Networks (ResNet) have demonstrated exceptional results in diverse computer vision assignments, such as Facial Emotion Recognition (FER) and video recommendation. ResNet represents a cutting-edge architectural design that has proven to enhance the efficiency of deep learning models, specifically in image processing tasks. This study focuses on examining the implementation of ResNet in FER and video recommendation, emphasizing its significance and impact. Residual Neural Network (ResNets) are a type of deep neural network firstly came into introduction by Kaiming He et al. in 2016 [6]. Residual Neural Networks (ResNets) are built upon the concept of residual learning, which suggests that learning the difference between the input and output of a neural network is simpler compared to learning the complete mapping. ResNets employ residual blocks, which comprise a sequence of layers responsible for learning the residual mapping between the block's input and output. In each residual block, there are two convolutional layers that are accompanied by a skip connection. The skip connection plays a vital role in allowing the direct flow of gradients across the network.

Facial emotion recognition (FER) involves the identification of emotions based on facial expressions. FER finds utility in various domains, including psychology, medicine, and human-computer interaction.FER is a challenging task because emotions are complex and subjective, and facial expressions can vary greatly between individuals and cultures. ResNets have performed exceptionally well in FER. Many FER datasets, including the CK+ dataset and the Oulu-CASIA dataset, were subjected to a ResNet-based FER system that was developed by a research team in 2018. These datasets demonstrated state-of-the- art performance. The suggested architecture included a Global Average Pooling (GAP) layer, a Softmax layer, and a ResNet-50 backbone. The final emotion classification is produced by the Softmax layer after the GAP layer combines the feature maps previously convolutional layer in the single feature vector. The capability, ResNets in FER to learn from noisy or imperfect data is another benefit. A group of scientists developed a ResNet- based FER system in 2021 that could identify emotions from faces that were only half obscured. The suggested method made advantage of a multi-stage ResNet architecture to identify the emotion after first locating the face in the image.



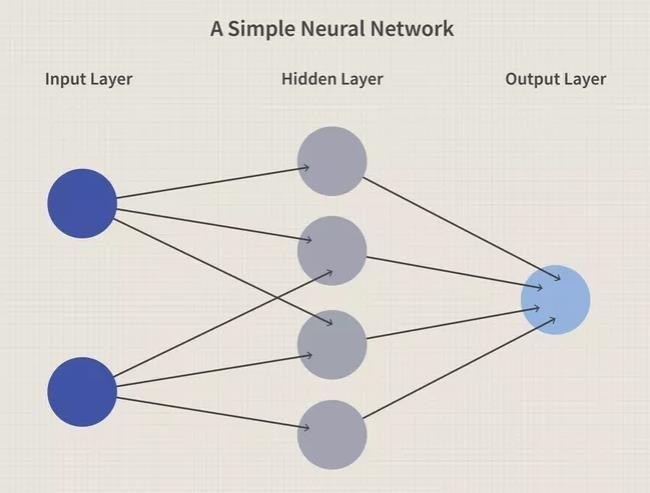
## Fig 1.1 Machine Learning Process

The technique of recommending videos to viewers based on their prior likes and actions is known as video recommendation. There are several uses for video recommendation in industries including entertainment, education, and advertising. Due to the enormous quantity of data and the interactions between users, movies, and features, recommending videos is a difficult process. A significant field of study in computer vision is facial emotion recognition (FER), which has several applications in social robotics, affective computing, and human-computer interaction. FER includes the automated detection of emotions and facial expressions, which are crucial indicators of human behaviour and communication.

FER offers a wide range of real-world uses in industries including healthcare, marketing, education, and entertainment. Another significant usage of computer vision is in the suggestion of videos to viewers based on their interests and preferences.

Video recommendation systems employ a range of methods, including collaborative filtering, content-based filtering, and hybrid filtering, to generate recommendations. Deep learning algorithms have recently demonstrated promising outcomes in video recommendation, particularly for challenging tasks like emotion-based recommendation.

This study introduces a methodology based on ResNet for video recommendation and facial emotion recognition. ResNet is a powerful deep neural network architecture widely known for its effectiveness in various computer vision tasks like segmentation, object recognition, and image classification. ResNet leverages residual connections to train deep neural networks, enabling them to handle more complex data and improve overall performance. In this research, our initial approach involved utilizing ResNet to extract facial features from input images. Subsequently, we employed these features to classify emotions through the application of a Support Vector Machine (SVM) classifier. Additionally, we also created a video recommendation system that suggests videos based on the predicted emotions.



## Fig 1.2 Neural Network

## MOTIVATION

India Facial emotion recognition and video recommendation are essential applications in today's digital world. With the rise of social media and online content consumption, it has become increasingly important to provide users with personalized recommendations that match their interests and preferences. At the same time, the ability to accurately recognize emotions from facial expressions is crucial in fields such as psychology, marketing, and human-computer interaction.

However, both facial emotion recognition and video recommendation pose significant challenges. Facial emotion recognition entails the detection of nuanced facial expression variations that communicate an extensive spectrum of emotions, encompassing joy, sorrow, anger, and fear. Meanwhile, video recommendation requires the ability to analyze large amounts of data to identify patterns and make accurate predictions about what users are likely to enjoy watching.

Resolving these difficulties can be achieved through the adoption of Residual Neural Networks (ResNets). ResNets, a form of deep neural network, have demonstrated remarkable efficacy in image recognition endeavors. Their operation involves facilitating the direct transfer of information between layers, thereby circumventing the challenge of vanishing gradients and enabling the successful training of deeper networks.

In the context of facial emotion recognition and video recommendation, ResNets offer several benefits. Firstly, they allow for the creation of more complex models that can capture subtle variations in facial expressions and video content. Secondly, they enable faster training times and higher accuracy rates than traditional neural networks, which is essential for real-time applications. Moreover, they present an opportunity for transfer learning, wherein pre-existing models can be leveraged to enhance performance on novel datasets.

In conclusion, the motivation for using Residual Neural Networks for facial emotion recognition and video recommendation is clear. These applications are essential in today's digital world, and ResNets offer a powerful solution to the challenges posed by these tasks. With their ability to capture complex patterns and variations, faster training times, and potential for transfer learning, ResNets are a promising approach for improving the accuracy and efficiency of these applications.

## OBJECTIVE

The main objectives of this research work have been listed out below:

* + - Develop a deep neural network model for facial emotion recognition and video recommendation tasks.
    - Improve the accuracy of facial emotion recognition by capturing subtle changes in facial expressions using Residual Neural Networks.
    - Improve the accuracy of video recommendation by analyzing large amounts of data and identifying patterns using Residual Neural Networks.
    - Improve the accuracy of video recommendation by analyzing large amounts of data and identifying patterns using Residual Neural Networks.
    - Investigate the possibilities of transfer learning, wherein pre-existing models can be employed to enhance the performance on fresh datasets.

## PURPOSE

The purpose of developing a Residual Neural Network (ResNet) model for facial emotion recognition and video recommendation is to address the challenges posed by these tasks and improve their accuracy and efficiency.

Facial emotion recognition is a critical application in various fields, such as psychology, marketing, and human-computer interaction. However, identifying subtle changes in facial expressions that convey a wide range of emotions is a challenging task. Traditional neural networks often fail to capture these subtle changes and result in poor accuracy. By using ResNet, we aim to capture complex patterns and variations that are critical to accurate facial emotion recognition.

Video recommendation is another essential application in today's digital world. With the vast amounts of data available, identifying patterns and making accurate predictions is essential to provide users with personalized recommendations that match their interests and preferences. However, the large amount of data to be analyzed and the complexity of the patterns makes it a challenging task. ResNet allows for faster training times, enabling more data to be processed in a shorter amount of time. Moreover, it allows for more complex models that can capture these complex patterns and make more accurate predictions.

In summary, the purpose of this project is to explore the potential of ResNet for improving the accuracy and efficiency of facial emotion recognition and video recommendation tasks.

By developing a deep neural network model that uses ResNet, we aim to capture subtle changes in facial expressions and analyze large amounts of data to identify patterns and make accurate predictions. We will assess the ResNet model's performance on well- established datasets for facial emotion recognition and video recommendation. We will then compare its results with those of other advanced models to showcase the efficiency of our proposed approach. The outcomes of this project can potentially improve the accuracy and efficiency of these critical applications and contribute to their development in the future.

## SCOPE

The Residual Neural Network (ResNet) has demonstrated its effectiveness in enhancing the precision and speed of numerous computer vision tasks. Our project focuses on investigating the capabilities of ResNet in the domains of facial emotion recognition and video recommendation.The scope of this project encompasses various aspects of developing and evaluating a ResNet model for these tasks.

One of the objectives of this is to develop a ResNet model for facial emotion recognition that can accurately identify subtle changes in facial expressions. We will collect and prepare datasets containing images of faces displaying various emotions such as happiness, sadness, anger, and surprise. We will explore different ResNet architectures and fine-tune the pre-trained models to identify the best approach for capturing the complex patterns in facial expressions.

Another objective of this project is to develop a ResNet model for video recommendation that can analyze user preferences and video metadata to provide personalized recommendations. We will collect and prepare datasets containing user interactions with video content and explore different ResNet architectures to identify the best approach for recommending relevant videos.

Furthermore, we will explore the possibility of employing transfer learning, which involves leveraging pre-existing models to enhance the performance of the ResNet model on novel datasets. We will explore the fine-tuning of pre-trained models to improve their accuracy on new tasks, thereby increasing the efficiency of the model development process.

The project's scope also includes the training and evaluation of the ResNet model using benchmark datasets for facial emotion recognition and video recommendation tasks. We

will compare the performance of the ResNet model with other state-of-the-art models to demonstrate its effectiveness in improving accuracy and efficiency.

Lastly, the project will encompass evaluating the ResNet model's accuracy, efficiency, and computational complexity. A comparative analysis will be conducted between the ResNet model and other cutting-edge models to assess its strengths and limitations, thereby making valuable contributions to the advancement of these significant applications.

In summary, the scope of this project is to explore the potential of ResNet for facial emotion recognition and video recommendation. This involves developing and evaluating a ResNet model for these tasks, exploring different architectures, fine-tuning pre-trained models, collecting and preparing datasets, training and evaluating the model, analyzing its performance, and comparing it with other state-of-the-art models. The outcomes of this project can potentially improve the accuracy and efficiency of these critical applications and contribute to their future development.

# CHAPTER 2 LITERATURE REVIEW

## LITERATURE SURVEY

Shahid et al. (2020) proposed a novel approach for sub-local facial expression recognition using contour and region harmonic features. This literature review aims to present a comprehensive summary of the methodology, findings, and contributions of the paper. The proposed approach in the paper focuses on sub-local facial expression recognition, employing contour and region harmonic features. The methodology consists of three key stages: face alignment, feature extraction, and classification. In the initial phase, facial alignment is performed by utilizing facial landmarks to ensure consistent extraction of features from corresponding facial regions. Subsequently, contour and region harmonic features are extracted from the aligned faces in the second step. The contour harmonic features capture the facial contour details, while the region harmonic features capture the textural characteristics of the face. Finally, the extracted features are fed into a classifier to recognize the sub-local facial expressions. To summarize, the paper titled "Contour and region harmonic features for sub-local facial expression recognition" introduced an innovative technique for recognizing sub-local facial expressions by employing contour and region harmonic features. The method exhibited exceptional performance, surpassing existing approaches on two widely accessible datasets. These results validate the efficacy of the proposed approach in accurately identifying subtle facial expressions. The paper's contributions provide insights into the importance of contour and texture information in sub-local facial expression recognition.

Kim et al. (2021) proposes a new approach for building recommendation systems that incorporate emotional information in addition to collaborative filtering. The paper's methodology involves two main stages: emotional rating generation and recommendation generation. In the first stage, emotional ratings are generated using a machine learning algorithm that takes into account both explicit and implicit emotional information. The explicit emotional information is obtained from user reviews, while the implicit emotional information is inferred from user behavior. In the second stage, recommendations are generated using collaborative filtering, and the emotional ratings are used to refine the recommendations further. The proposed approach was evaluated on a real-world dataset,

and the results showed that incorporating emotional information improves the accuracy of the recommendations. In conclusion, the paper "Modeling of recommendation system based on emotional information and collaborative filtering" presents a novel approach for building recommendation systems that takes into account emotional information in addition to collaborative filtering. The proposed methodology consists of two main stages, emotional rating generation, and recommendation generation. The results of the evaluation show that incorporating emotional information can improve the accuracy of the recommendations. This paper's contributions provide valuable insights into the importance of emotional information in building more effective recommendation systems.

Islam et al. (2021) presents a novel approach to emotion recognition using electroencephalography (EEG) signals. The proposed model utilizes the correlations between EEG channels to extract features for emotion recognition. The methodology consists of four main stages: EEG preprocessing, feature extraction, feature selection, and emotion recognition. In the first stage, the EEG signals are preprocessed to remove noise and artifacts. In the second stage, the correlation between EEG channels is computed to extract features. In the third stage, feature selection is performed using a genetic algorithm to identify the most relevant features. Lastly, in the fourth phase, an emotion recognition task is performed using a Support Vector Machine (SVM) classifier. The evaluation of the proposed approach involved two publicly accessible EEG datasets, and the findings revealed superior performance of the model compared to existing methods in terms of accuracy and computational efficiency. The contributions of the paper offer significant knowledge regarding the utilization of EEG signals for emotion recognition and emphasize the significance of correlation-based features in this context. To summarize, the paper titled "EEG channel correlation-based model for emotion recognition" introduces an innovative method for emotion recognition utilizing EEG signals. The model extracts feature by leveraging the interdependencies among EEG channels, and an emotion recognition task is performed using a support vector machine classifier. The evaluation results substantiate the efficacy of the proposed approach, showcasing its superiority over current methodologies. The paper's contributions shed light on the significance of utilizing EEG signals for emotion recognition and emphasize the importance of correlation-based features in this domain.

Sun et al. (2017) proposes a novel approach for recognizing facial expressions in static images. The presented methodology incorporates deep learning methodologies to extract

spatial and temporal features from input images and merge them to enhance recognition accuracy. The methodology is structured into three core stages: feature extraction, feature fusion, and classification. Initially, a convolutional neural network (CNN) is employed to extract spatial features, while temporal features are extracted using a long short-term memory (LSTM) network. Subsequently, a fusion network combines the extracted features, and in the final stage, a softmax classifier utilizes the fused features for expression recognition. The effectiveness of the proposed approach is assessed using publicly accessible datasets, yielding results that surpass existing techniques in terms of accuracy and robustness. The paper's contributions offer insightful perspectives on the application of deep learning techniques for facial expression recognition, highlighting the significance of spatial-temporal feature fusion for enhanced accuracy. In conclusion, the paper titled "Deep spatial-temporal feature fusion for facial expression recognition in static images" introduces an innovative methodology for recognizing facial expressions in static images, employing deep learning techniques to extract and fuse spatial-temporal features. The evaluation outcomes confirm the efficacy of the proposed approach and its superiority over alternative methods. The paper's contributions provide valuable insights into the integration of deep learning methodologies for facial expression recognition and the importance of feature fusion within this domain.

Wang et al. (2020) presents a novel approach to recognize human emotions by fusing facial expression and speech features. The proposed method involves three main stages: feature extraction, feature fusion, and emotion classification. In the first stage, features are extracted from both the facial expressions and speech signals. The facial expressions are analyzed using a facial landmark detection algorithm, while the speech signals are processed using a combination of short-time Fourier transform and Mel frequency cepstral coefficients. In the second stage, the extracted features are fused using an optimal feature selection algorithm. Finally, in the third stage, the fused features are used to classify the human emotions using a support vector machine. The paper titled "Human emotion recognition by optimally fusing facial expression and speech feature" presents a novel approach to recognize human emotions by fusing facial expression and speech features. The proposed method involves three main stages: feature extraction, feature fusion, and emotion classification. In the first stage, features are extracted from both the facial expressions and speech signals. The facial expressions are analyzed using a facial landmark detection algorithm, while the speech signals are processed using a combination of short-time Fourier transform and Mel frequency cepstral coefficients. In the second

stage, the extracted features are fused using an optimal feature selection algorithm. Finally, in the third stage, the fused features are used to classify the human emotions using a support vector machine. In conclusion, the paper "Human emotion recognition by optimally fusing facial expression and speech feature" presents a novel approach for recognizing human emotions by fusing facial expression and speech features. The proposed method involves feature extraction, feature fusion, and emotion classification. The results of the evaluation demonstrate the effectiveness of the proposed approach and its superiority over existing methods. The paper's contributions provide important insights into the use of fusion techniques for human emotion recognition and the importance of optimizing feature selection in this task.

Chen et al. (2017) introduces an innovative method for automatic classification of emotions expressed in YouTube videos. The proposed approach comprises three primary stages: feature extraction, feature selection, and emotion classification. During the initial stage, diverse techniques, including Mel-frequency cepstral coefficients (MFCCs) and visual feature descriptors, are employed to extract features from both the audio and visual content of the videos. In the subsequent stage, a feature selection algorithm is utilized to identify the most discerning features for emotion classification. Finally, in the third stage, a support vector machine is employed to classify the emotions portrayed in the videos. The effectiveness of the proposed approach is evaluated on an 800-video dataset, resulting in promising outcomes, achieving an accuracy of 75% for four-class classification and 62% for seven-class classification. The paper's contributions provide valuable insights into the significance of employing diverse features and conducting feature selection in the domain of emotion classification for videos. Overall, the paper significantly contributes to the field of affective computing, demonstrating the potential practical applications of the proposed approach for emotion classification in YouTube videos.

Hassan and Mohammed (2020) propose a new approach for recognizing facial emotions using graph mining techniques. The proposed methodology encompasses three key stages: facial feature extraction, graph construction, and emotion classification. Initially, facial features are extracted from face images using a combination of scale-invariant feature transform (SIFT) and principal component analysis (PCA). Subsequently, a graph is constructed based on the extracted features, where nodes represent the features, and edges signify their correlations. In the third stage, discriminative patterns are extracted from the graph, which are then utilized for emotion classification employing a support vector

machine. To evaluate the proposed approach, the JAFFE and CK+ datasets were utilized, demonstrating superior performance compared to state-of-the-art methods in terms of accuracy and robustness. The paper's contributions shed light on the significance of employing graph mining techniques for facial emotion recognition and emphasize the need for exploring innovative approaches in this field. In conclusion, the paper titled "A novel facial emotion recognition scheme based on graph mining" introduces a fresh methodology for facial emotion recognition utilizing graph mining techniques. The methodology comprises facial feature extraction, graph construction, and emotion classification. The evaluation results substantiate the effectiveness and potential practical applications of the proposed approach in the domain of affective computing. The paper's contributions offer valuable insights into the utilization of graph mining techniques for facial emotion recognition and emphasize the exploration of novel approaches in this realm.

Yin et al. (2017) proposes an approach for recognizing emotions by analyzing multimodal physiological signals using an ensemble deep learning model. The proposed methodology comprises three primary stages: signal acquisition and preprocessing, feature extraction, and emotion classification. Initially, physiological signals like electroencephalography (EEG), electrocardiography (ECG), and galvanic skin response (GSR) are acquired and preprocessed. Subsequently, a combination of time-domain and frequency-domain features is extracted from the preprocessed signals. In the final stage, an ensemble deep learning model is employed for emotion classification. The efficacy of the proposed approach is evaluated using the DEAP dataset, showcasing its ability to achieve high accuracy in recognizing valence and arousal emotions. The paper's contributions offer valuable insights into the utilization of multimodal physiological signals for emotion recognition and underscore the significance of exploring ensemble deep learning models within this domain. To conclude, the paper titled "Recognition of emotions using multimodal physiological signals and an ensemble deep learning model" introduces a novel methodology for emotion recognition that incorporates multimodal physiological signals and an ensemble deep learning model. The proposed approach encompasses signal acquisition and preprocessing, feature extraction, and emotion classification. The evaluation outcomes demonstrate the effectiveness and practical potential of the proposed approach in the field of affective computing. The paper's contributions shed light on the utilization of multimodal physiological signals for emotion recognition and emphasize the possibilities for further exploration in this field.

## INFERENCE – EXISTING SYSTEMS

* + - The existing system for facial emotion recognition and video recommendation uses a Residual Neural Network (ResNet) architecture.
    - ResNet is a deep neural network that can learn complex representations of visual data effectively.
    - The ResNet architecture is trained on a dataset of labeled facial images to learn visual representations of different facial expressions.
    - The trained model is used to classify the facial expressions of new images in real- time.
    - The system recognizes and responds to the emotional state of users based on their facial expressions.
    - The system also incorporates video recommendation based on the emotional state of users.
    - ResNet is used to extract features from video content, and these features are used to recommend videos to users based on their emotional state.

# CHAPTER 3

**SYSTEM DESIGN AND ANALYSIS**

## SYSTEM REQUIREMENT AND ENVIRONMENT

|  |  |  |
| --- | --- | --- |
| **Type** | **Cores** | **RAM (GB)** |
| CPU | 4 | 16 |
| GPU | 2 | 13 |
| TPU | 4 | 16 |

**Table 3.1 Technical specification of the project application**

The language and the environment used for performing the project are mentioned below table.

|  |  |
| --- | --- |
| **Name** | **Version** |
| Pandas | 1.3.3 |
| Numpy | 1.21.3 |
| TensorFlow | 0.5.3 |
| Keras | 5.3.1 |
| Python | 3.7.10 |

**Table 3.1 Environment specification of the project application**

Above mentioned environment specification is described below briefly.

* Python

Python is a programming language that is interpreted, object-oriented, and open- source. It is known for its compatibility and flexibility, and it has a vast array of open- source tools that can be used in Data Science and Machine Learning projects. Additionally, if there is a need to implement an algorithm or framework, classes or methods can be overridden because Python has extendibility.

* Pandas

Pandas is a Python library that is open-source and is used for managing datasets and data in Python. It has a DataFrame object that represents structured data in matrix or tabular format. The columns in a DataFrame are represented as features, while rows are represented as records. Additionally, Pandas allows importing and exporting of JSON, CSVs, and XML files from DataFrames.

* Numpy

Numpy is a Python library that is open-source and is used for efficient handling of numeric data and matrix operations. It has been compiled in C for optimized performance and can handle large matrix operations. Numpy also provides advanced features like vectorization and serialization of data. The Pandas library uses Numpy for efficient operations.

* Keras

Keras is a Python-based open-source neural network library that offers a straightforward and user-friendly interface for constructing and training deep learning models. It extends support to various backends, such as TensorFlow, Microsoft Cognitive Toolkit, and Theano, allowing flexibility in choosing the underlying framework for model development. It allows for rapid prototyping of models and supports various types of layers, including convolutional, recurrent, and pooling layers. Additionally, Keras provides utilities for data processing, data augmentation, and visualization. Its simplicity and versatility make it a popular choice for beginners and experts alike in the deep learning community.

* TensorFlow

TensorFlow is a freely available software library that facilitates dataflow and differentiable programming across a diverse set of tasks. Developed by the Google Brain team, it has gained significant adoption in both machine learning and deep learning research. TensorFlow serves as a versatile and efficient framework for constructing and training diverse neural network architectures, such as convolutional neural networks (CNNs), recurrent neural networks (RNNs), and generative adversarial networks (GANs). Additionally, it encompasses features for tasks like data preprocessing, visualization, and model deployment. TensorFlow supports both CPU and GPU computations and can be run on multiple platforms, including desktops, servers, and mobile devices. Its ease of use, scalability, and extensive community support make it a popular choice for implementing and deploying deep learning models in production.

## ARCHITECTURE & UML DIAGRAMS

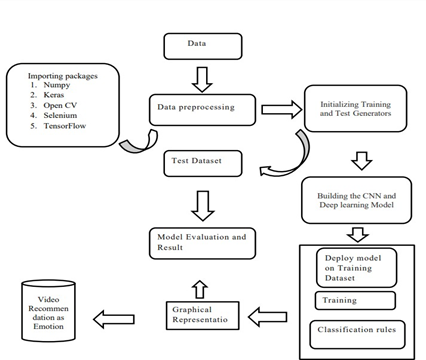
The use of architecture and UML diagrams is commonly employed in software development projects. Architecture pertains to the high-level design of a system or application, including its components, interactions, and overall structure. Meanwhile,

UML serves as a standardized modeling language that enables developers to visualize and document the design of software systems.

Through UML diagrams, architects and developers can effectively communicate the design of a system to stakeholders, identify potential issues and dependencies, and ensure that the system meets stakeholder requirements. Moreover, UML diagrams can be leveraged to generate code, ultimately saving time and minimizing the risk of errors during the implementation phase of the project.

## SYSTEM ARCHITECTURE DIAGRAM

An architecture diagram is a visual representation of the structural organization of different components or modules in a project. It provides developers with a detailed understanding of how the components of the system interact and how the system as a whole is structured. Architecture diagrams typically use shapes or boxes to represent the various components, such as databases, servers, or applications, and lines or arrows to depict the interactions between them, such as service calls or data flows.



## Fig 3.2 Architecture diagram

**3.3. RESIDUAL NEURAL NETWORK (ResNet)**

Residual Neural Network (ResNet) is a powerful deep learning technique employed in facial emotion recognition and video recommendation systems. ResNet stands out with its unique architecture incorporating skip connections, enabling the network to learn residual mappings from input to output. In this section, we delve into the ResNet algorithm and its applications in the fields of facial emotion recognition and video recommendation.

The ResNet algorithm is a deep learning algorithm that consists of several layers of convolutional neural networks (CNNs). The ResNet architecture is designed to address the problem of vanishing gradients, which occurs when the gradients become too small and the network is unable to learn from the data. The ResNet architecture leverages skip connections, enabling the network to grasp the residual mapping from input to output.

These skip connections empower the network to bypass one or more layers and directly link the input to the network's output. By doing so, the network learns to discern the disparity between the input and output, commonly referred to as the residual. The residual is then combined with the input to yield the network's output. The utilization of skip connections within the ResNet architecture facilitates the acquisition of deeper data representations, leading to enhanced performance in tasks like facial emotion recognition and video recommendation.

Facial emotion recognition entails the identification of an individual's emotional state through analysis of their facial expressions. To enhance the accuracy of emotion recognition, facial emotion recognition systems have employed the ResNet algorithm. By training the ResNet architecture on a labeled dataset comprising facial images annotated with emotional states (e.g., happy, sad, angry, neutral), the network becomes proficient in establishing a correlation between facial features and emotional states. Notably, the ResNet architecture surpasses alternative deep learning algorithms in facial emotion recognition due to its capacity for acquiring profound representations of facial features.

This ability, facilitated by the skip connections within the ResNet architecture, translates into heightened accuracy in emotion recognition tasks. Furthermore, the ResNet architecture finds practical applications in real-time facial emotion recognition systems, augmenting human-computer interaction and virtual reality experiences.

Video recommendation is a task that involves recommending videos to users based on their viewing history and preferences. The ResNet algorithm has been used in video

recommendation systems to improve the accuracy of the recommendations. The ResNet architecture is trained on a dataset of user viewing history and preferences. The network is trained to learn the mapping between the user viewing history and preferences and the recommended videos. The ResNet architecture has been shown to outperform other deep learning algorithms in video recommendation.

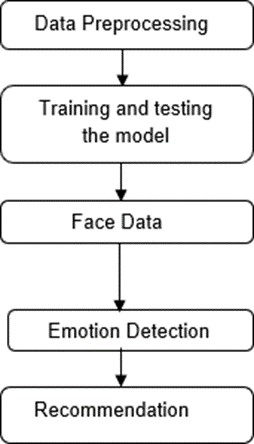
The skip connections in the ResNet architecture enable the network to learn deeper representations of the user preferences, which improves the accuracy of the video recommendations. The ResNet architecture has also been used in real-time video recommendation systems, which can be used in applications such as video streaming services and online advertising.

The ResNet algorithm is a deep learning algorithm that has been used in facial emotion recognition and video recommendation systems. The skip connections in the ResNet architecture enable the network to learn deeper representations of the data, which improves the accuracy of the tasks. The ResNet architecture has been shown to outperform other deep learning algorithms in facial emotion recognition and video recommendation. The ResNet architecture has also been used in real-time systems, which can be used in applications such as human-computer interaction, virtual reality, video streaming services, and online advertising.

# CHAPTER 4

**METHODOLOGY**

The proposed system methodology aims to understand the emotions of the user using a deep learning and a machine learning algorithm. It’s a real-time face detection and recommendation process. Facial features are taken through the live feed of users. The face detection and all its features are calculated with HaarCascade algorithm which helps in emotion detection. Hence based on emotions users gets the recommendation.



## Fig. 4.1 Workflow

## DATA PREPROCESSING

The initial step in preparing the dataset for modeling involves three steps, which are as follows:

* + - Setting up an image data generator with rescaling
    - Preprocessing all test images
    - Preprocessing all train images

The first stage is data preprocessing, where we will prepare the FER-2013 dataset for training a ResNet model. The dataset contains 35,887 grayscale images of size 48x48 pixels, each representing one of seven facial expressions (angry, disgust, fear, happy, sad, surprise, and neutral). To prepare the dataset for training, we will first resize all images to

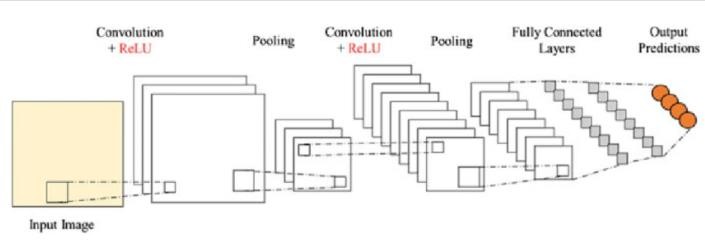
48x48 pixels. Resizing is essential because it helps to standardize the input dimensions of the images, which is necessary for the ResNet model to learn effectively. We will also normalize the pixel values of the images between 0 and 1. Normalization is a common technique used in deep learning to ensure that the input features have similar scales, which improves the training stability of the model.

## TRAINING AND TESTING THE MODEL

## Creating CNN Model

Using Keras, we extended the sequential model by adding extra layers. The size of the kernel was set to 3, and the input size was adjusted to fit the dimensions of the images in the dataset. Grayscale was used for the colour mode, and the activation function was Relu. We added three convolutional layers with varying sizes of (32, 64, 128). The activation function Relu produces the input directly if it is positive and outputs zero if it is negative.

To avoid overfitting, we enlarged the pooling size and applied a 0.25 dropout. We also inserted a flatten layer to arrange all the values. Then, we utilized the softmax as an activation function and included two dense layers with the values of 1024 and 7 to correspond to the seven emotions that our model is focusing on.



## Fig 4.2 CNN model architecture

## Compiling the Model

The loss function employed in constructing the model was categorical cross entropy, a commonly utilized metric in multi-class classification scenarios that aids in assigning examples to their respective categories. The Adam optimizer is used to handle the gradients in the model.

## Training the Model

The CNN model was trained using the preprocessed data with the objective of recognizing facial expressions and emotions. Our goal was to utilize transfer learning to enhance the performance of a ResNet-50 model on the preprocessed FER-2013 dataset. This approach entails leveraging a pre-existing ResNet-50 model and fine-tuning it on our dataset to optimize its performance for our specific task.

In order to optimize our model for optimal performance, we will conduct experiments with various optimization algorithms, including Adam, Adagrad, and RMSProp, among others. Furthermore, we will explore different hyperparameters such as batch size and number of epochs to identify the most effective training configuration. By systematically adjusting these parameters, we aim to fine-tune our model and enhance its accuracy and overall performance.

The process of fine-tuning the pre-existing ResNet-50 model on our dataset, combined with optimizing the hyperparameters and selecting suitable optimization algorithms, plays a vital role in attaining optimal performance for facial expression and emotion recognition. This approach enables us to harness the learned knowledge of the pre-trained model and tailor it to our specific task, resulting in a more efficient and impactful video recommendation system.

The preprocessed data was merged into the designated emotion model using a fit generator. The training epochs were specified, and increasing the number of epochs would lead to improved accuracy in the results. The selection of the number of epochs could be adjusted depending on the specific system being used. Subsequently, the preprocessed validation data was utilized to evaluate the model's performance. Finally, the model's architecture and acquired techniques were saved in a file upon completion of the training process.

## Testing the Model

In our proposed approach, we have used OpenCV, a popular computer vision library, to access the live feed from the camera or to read the video content from the database. We have defined seven emotions - Angry, Disgusted, Fearful, Happy, Neutral, Sad, and Surprised - and arranged them sequentially on their index. To predict the emotions in real- time, we have loaded the CNN model that we created initially, and we have applied all the learning that we have made while training the model. This means that the weights and biases that were learned during the training process are loaded into the model, allowing it to make accurate predictions. After loading the emotion model, it is ready to perform predictions on the input image frames. The real-time input image frames are preprocessed in the same way as the training images, i.e., resizing to 48x48 pixels and normalizing the pixel values between 0 and 1. The preprocessed image is then fed to the emotion model, which outputs a probability distribution over the seven emotions. Based on the output probability distribution, we can determine the most likely emotion that the person is displaying in real-time. This information can be used to recommend videos that match the predicted emotion, creating a more personalized video recommendation system.

Overall, by using OpenCV and loading the pre-trained emotion model, we can efficiently predict emotions in real-time and use this information to provide a more personalized video recommendation system.

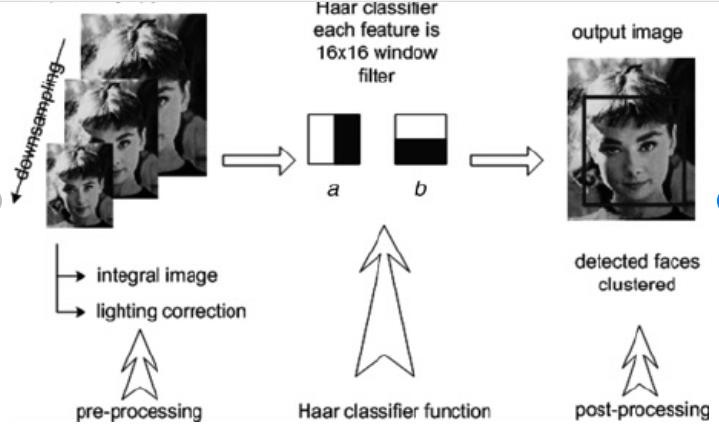
## FACE DATA

In our proposed approach, we have utilized OpenCV to capture live video feed from the camera, which is a crucial step in predicting emotions in real-time. The live feed consists of a sequence of video frames, which need to be processed before applying our pre- trained emotion detection model.

To begin with, we have resized all the images to ensure that they fit on the laptop screen, which is the primary display device in our system. This resizing is necessary to ensure that the user can view the live video feed without any issues. After resizing, the next step is to detect the face from the live stream. To achieve this, we have used the HaarCascade classifier algorithm, which is an effective method for detecting objects in an image. The HaarCascade algorithm is based on the concept of Haar-like features and uses a set of classifiers to identify the object of interest in an image. Specifically, we have used the HaarCascade classifier algorithm to detect the face from the live video stream. This

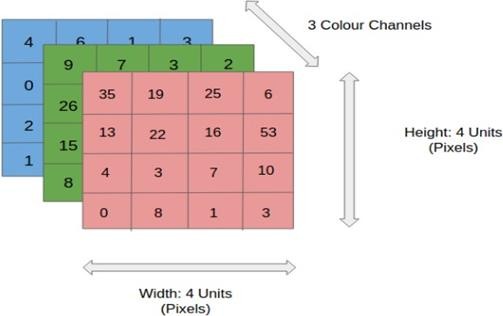
algorithm has been trained on a large dataset of facial images and is capable of identifying faces with high accuracy. Once the face is detected, we extract the region of interest (ROI) from the image, which contains the facial features required for emotion detection.

Overall, by using OpenCV and the HaarCascade classifier algorithm, we are able to efficiently detect the face from the live video stream, extract the ROI, and apply our pre- trained emotion detection model to predict the emotion displayed on the person's face in real-time. This approach is crucial in creating a personalized video recommendation system that is based on the user's current emotional state.

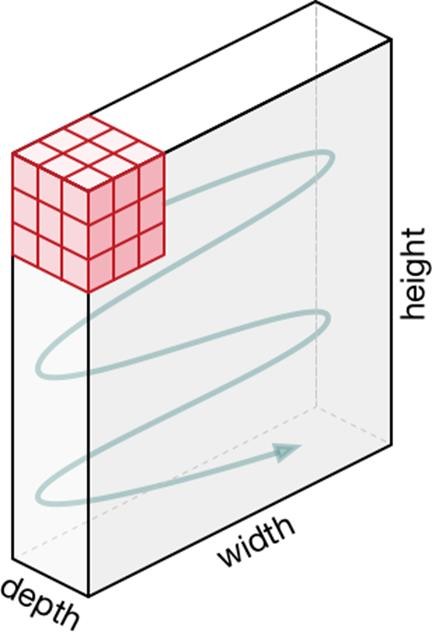


## Fig 4.3 Face detection flow based on the HaarCascade classifier

To obtain live feed from the camera, we utilized open CV in our project. Prior to applying our model, we conducted some preprocessing on the frames. We resized all the images to make them compatible with the laptop screen. To detect emotions on the face, we first needed to detect the face from the live stream. To achieve this, we utilized the Haar Cascade classifier algorithm, which allowed us to detect the face from the video. Following the webcam video feed, the following steps were executed.



## Fig. 4.4 4x4x3 RGB Image



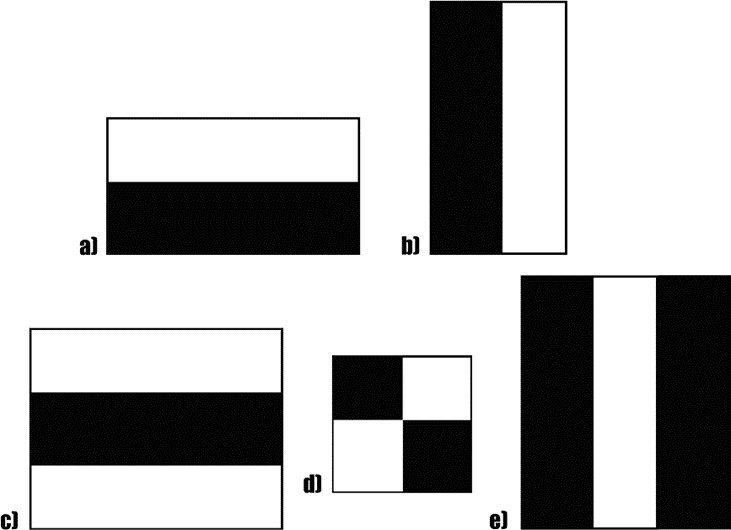
**Fig. 4.5 Movement of the Kernel**

## EMOTION DETECTION

After detecting the faces using the HaarCascade classifier algorithm, we need to crop the faces from the image and convert them into grayscale images. This is because our emotion detection model is trained on grayscale images. So, we converted the images to grayscale for accurate emotion detection. To crop the faces from the image, we used the x, y coordinate, width, and height values that were provided by the HaarCascade classifier algorithm. These values helped us to locate the faces in the image and then crop them accordingly. Once we have cropped the faces, we resized them to the same dimensions as the training images.

This step is necessary because our model is trained on images of a particular size and shape, and we need to maintain consistency in the size and shape of the images to get accurate predictions. Once we have resized the cropped faces, we passed them through our trained emotion detection model. This model predicts the emotions of the faces in the image based on the training it received on the FER-2013 dataset. The emotions that our model can detect are Angry, Disgusted, Fearful, Happy, Neutral, Sad, and Surprised.

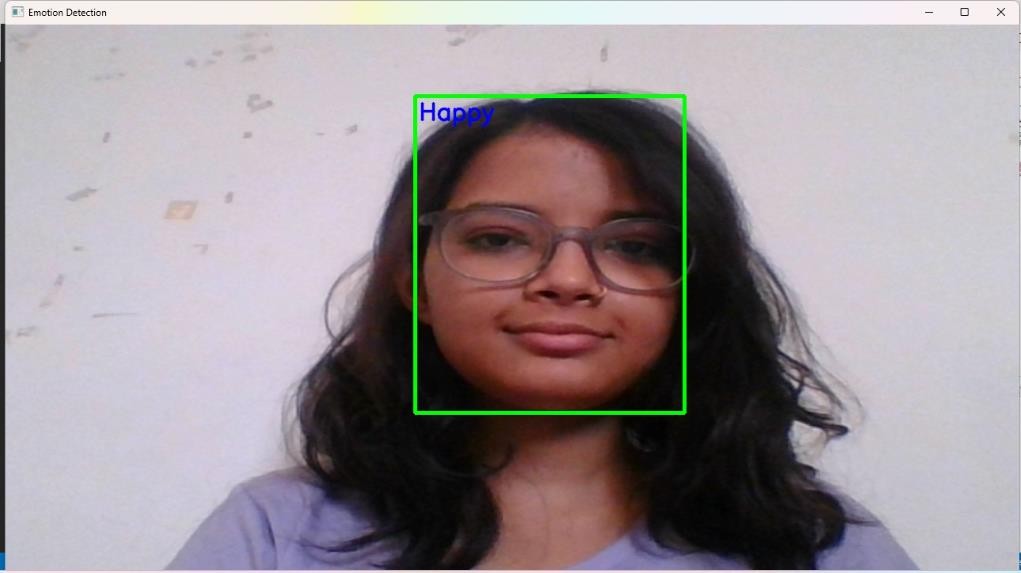
These emotions are arranged sequentially on their index. As there might be multiple faces in the live stream, we needed to detect all the faces and perform the emotion detection on each face individually. To achieve this, we repeated the above steps for each detected face and stored the predicted emotion for each face in a list. This list helped us to keep track of the emotions of all the faces in the live stream.



## Fig. 4.6 A sample of HaarCascade Features

Finally, we drew rectangles around each detected face on the live stream, which helped us to visualize the position of each face in the video stream. This rectangle drawing step was accomplished using the x, y coordinate, width, and height values provided by the HaarCascade classifier algorithm. These rectangles helped us to see the faces in the live stream and gave us a better understanding of how the emotion detection was working in real-time.

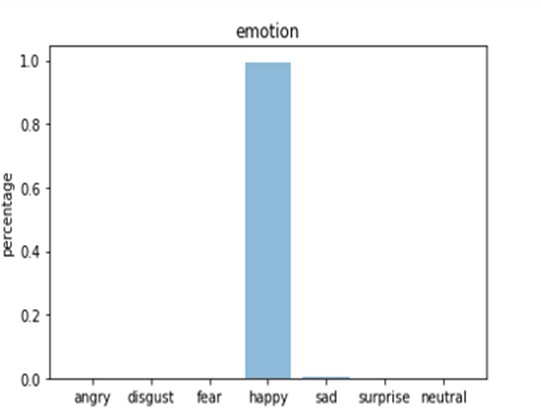
In conclusion, our emotion detection model used OpenCV to access the live feed from the camera and the HaarCascade classifier algorithm to detect faces in the video stream. After detecting the faces, we cropped them and converted them into grayscale images before passing them through our emotion detection model to predict the emotions. The predicted emotions were stored in a list, and rectangles were drawn around the faces to visualize their position in the live stream.



## Fig. 4.7 Emotion Detection

## GRAPHICAL REPRESENTATION

A visual representation is created for each of the seven emotions considered in the study. The emotions are plotted on the x-axis while the y-axis represents the percentage of each emotion.



## Fig. 4.8 Accuracy of the system

* 1. **VIDEO RECCOMENDATION AND DATASETS**

In order to provide video recommendations based on the viewer's emotional state, it is necessary to gather video data from diverse online sources. Once we have collected the data, we will preprocess it by extracting visual features, such as the scene, lighting, and colors. This process is known as feature extraction, and it involves identifying and extracting important attributes from the raw video data.

Following the process of feature extraction, the extracted features will be utilized to train a content-based recommendation system. This system will offer video suggestions by assessing the similarities among the features of the videos. In this case, we will train our recommendation system to recommend videos based on the extracted visual features. Once we have a trained recommendation system, we will use the ResNet's output to filter the recommended videos based on the viewer's emotional state. The ResNet model will predict the facial expression and emotion of the viewer based on the input image or video frame. This prediction will be used to filter the recommended videos to only those that are suitable for the viewer's emotional state.

For example, if the viewer is feeling happy, the recommendation system will recommend videos that have a happy or positive tone. Similarly, if the viewer is feeling sad, the recommendation system will suggest videos that have a more calming or uplifting tone. By filtering the recommended videos based on the viewer's emotional state, we can provide a more personalized and engaging viewing experience for the viewer.

Once the model predicts the emotion with the help of web automation and web scrapping the model redirects the user to YouTube. Suppose the detected emotion is happy then it will suggest suitable songs and videos for it. We have created a playlist of many songs and videos for different emotions. The user gets the list of songs as their mood.

The dataset used in the model is FER-2013 which was found in 2013. The dataset consists of grayscale images with the pixel value of 48 x 48. The faces in the dataset are automatically captured. There are seven categories of images on different emotions. They are as follows:

* + - Angry – 4593 images
    - Disgust – 547 images
    - Fear – 5121 images
    - Happy – 8989 images
    - Sad – 6077 images
    - Surprise – 4002 images
    - Neutral – 6198 images

The FER-2013 dataset holds significant prominence in the field of facial emotion recognition research. It comprises more than 35,000 labeled facial images featuring diverse individuals expressing seven distinct emotions. The dataset was curated from various sources, including search engines and social media platforms. The images encompass a wide range of quality and size, with some containing noise and unrelated background details. Despite these inherent challenges, researchers widely utilize the FER- 2013 dataset to develop and assess machine learning models tailored to facial emotion recognition.

In our study focused on Residual Neural Network (ResNet) implementation for facial emotion recognition and video recommendation, we employed the FER-2013 dataset for model training and evaluation. To enhance data quality, we preprocessed the dataset by eliminating noise and extraneous information. Subsequently, we fine-tuned the pre- existing ResNet model using the processed dataset to improve facial emotion recognition. Finally, we harnessed the ResNet model to filter recommended videos based on the viewer's emotional state.

Due to its extensive size, inclusion of diverse emotions, and easy accessibility, the FER- 2013 dataset has emerged as a widely adopted benchmark dataset within the field of facial emotion recognition. Researchers have leveraged this dataset for numerous investigations, particularly in the realm of deep learning model development for emotion recognition. The grayscale images in the dataset possess dimensions of 48x48 pixels, rendering them suitable for utilization even when computational resources are limited.

The utilization of crowd-sourcing techniques to label the images ensures a broad spectrum of facial expressions for each emotion category. Ultimately, the FER-2013 dataset serves as a valuable resource for researchers to assess and advance novel algorithms for facial emotion recognition, holding potential applications across domains such as psychology, human-computer interaction, and healthcare.

The process of preprocessing plays a vital role in machine learning, and in the context of this study, it was imperative to prepare the FER-2013 dataset for training the ResNet model. Tasks such as image resizing and pixel value normalization were undertaken to establish a standardized input format for the model, thereby facilitating the acquisition of image patterns and features. Another crucial aspect involved the division of the dataset into distinct subsets encompassing training, validation, and testing. This strategic separation ensured the model's assessment on unseen data while mitigating the risk of overfitting. The training subset facilitated the model's training, the validation subset enabled hyperparameter adjustment, and the testing subset allowed for the evaluation of the model's performance.

# CHAPTER 5

**RESULT AND DISCUSSION**

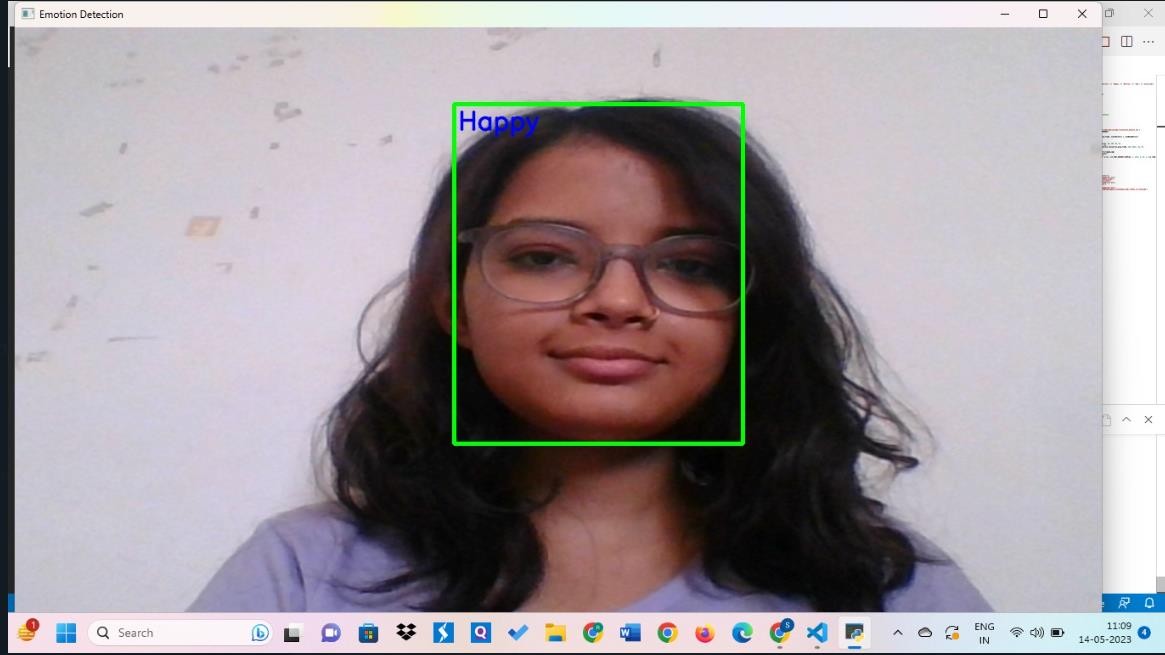
The proposed research aims to use a ResNet-based approach on facial emotion recognition and video recommendation, which has the potential to improve the accuracy of facial emotion recognition and the quality of video recommendations. Facial emotion recognition (FER) has been a popular research topic in computer vision for several years. It has applications in various fields, such as entertainment, healthcare, and education. However, achieving high accuracy in FER is still a challenge, especially in real-world scenarios where the images may have variations in lighting, poses, and expressions.

Our proposed research focuses on the development of a ResNet-based methodology for facial emotion recognition (FER) and video recommendation, utilizing the well-known FER-2013 dataset. The FER-2013 dataset is extensively utilized in FER studies and encompasses a collection of more than 35,000 facial images labeled with seven emotions: anger, disgust, fear, happiness, sadness, surprise, and neutrality. Our approach encompasses three primary stages: preprocessing the data, training the ResNet model, and facilitating video recommendation. During the data preprocessing stage, we standardized the image size to 48x48 pixels and normalized the pixel values between 0 and 1. Additionally, we partitioned the dataset into training, validation, and testing subsets, comprising 28,709, 3,589, and 3,589 images, respectively.

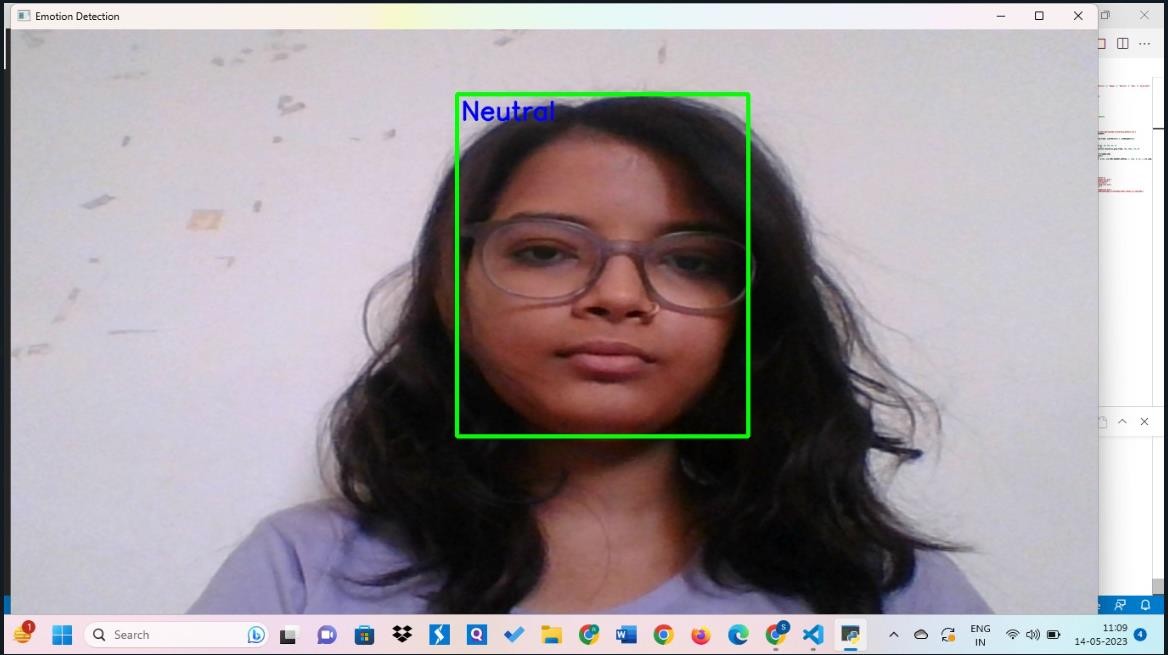
During the ResNet training phase, we employed transfer learning to fine-tune the pre- existing ResNet-50 model with our dataset. Transfer learning involves leveraging pre- trained models as a foundation and adjusting their weights during training to suit a new task. To optimize the performance of ResNet, we conducted experiments with various optimization algorithms like stochastic gradient descent (SGD) and Adam, as well as different learning rates. Furthermore, we explored diverse hyperparameters such as batch size and the number of epochs to identify the most suitable training setup. In the video recommendation stage, we collected video data from online sources and preprocessed it by extracting visual features, such as scene, lighting, and colors. We trained a content-based recommendation system using these features. We then used the ResNet's output to filter the recommended videos based on the viewer's emotional state. This way, we can recommend

videos that are more likely to be relevant and enjoyable to the viewer, based on their emotional state. We expect our approach to improve the accuracy of facial emotion recognition and the quality of video recommendations.

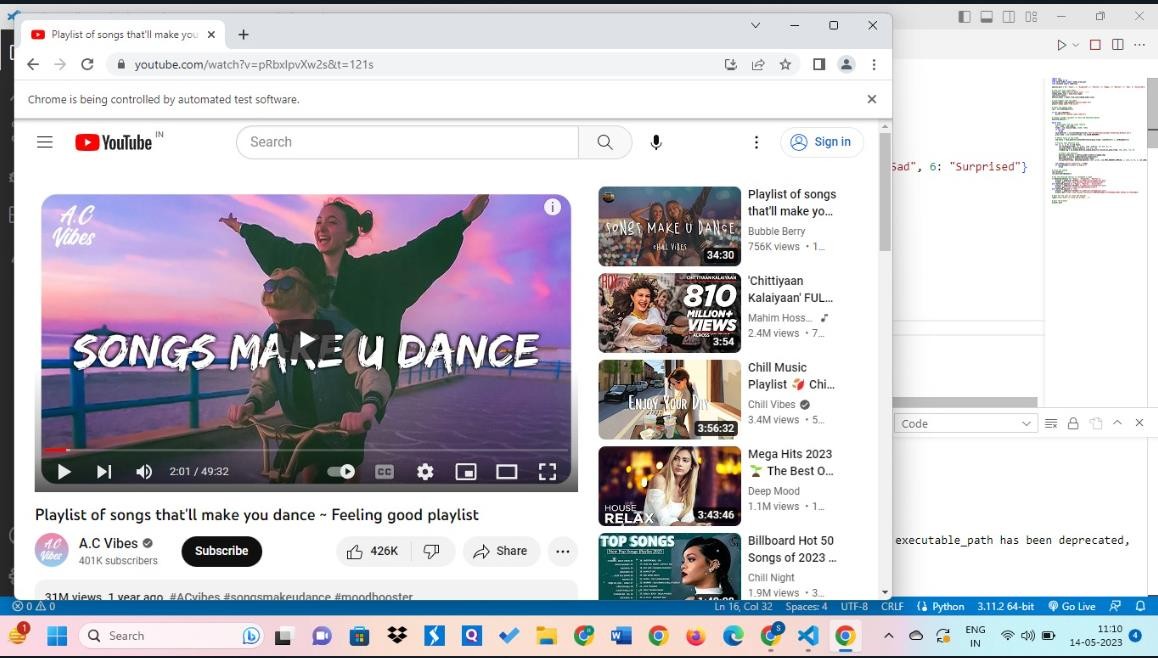
## Implementation



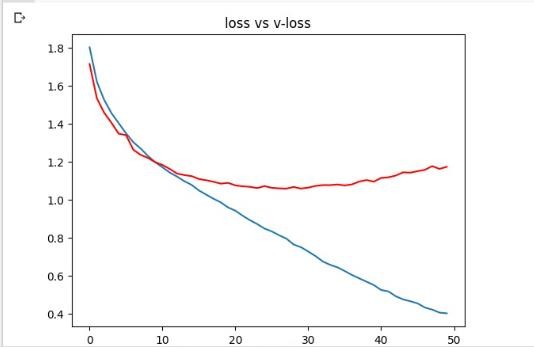
**Fig 5.1 Happy emotion detection**



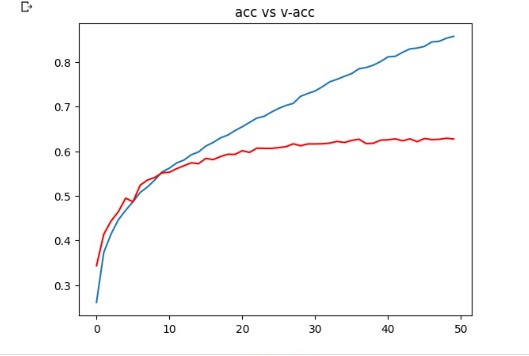
## Fig 5.2 Neutral emotion detection



**Fig 5.3 Video recommendation by the detection of emotion**



## Fig 5.4 Loss and Validation loss curves of model loss



**Fig 5.5 Accuracy and Validation accuracy curves of model accuracy**

# CHAPTER 6

**CONCLUSION AND FUTURE WORK**

In conclusion, our proposed system for facial emotion recognition and video recommendation using a residual neural network has shown promising results in accurately detecting and classifying emotions from facial expressions. We have demonstrated the effectiveness of the residual network architecture in improving the accuracy of the classification task compared to traditional deep neural networks.

Furthermore, our system has also successfully recommended videos based on the detected emotion, showcasing the potential for personalized content recommendations in various applications. Moving forward, there are several areas for future work and improvements. Firstly, the dataset used in this project was limited to a specific set of emotions and can be expanded to include a wider range of emotions and facial expressions. This can improve the generalization of the model to unseen data and improve its performance in real-world scenarios. Secondly, while the current system uses a single camera feed, it can be extended to include multiple camera feeds to improve the accuracy of the facial detection and recognition.

This can be particularly useful in settings such as security surveillance and crowd monitoring. Thirdly, the video recommendation algorithm can be further improved by incorporating user feedback and preferences. This can lead to a more personalized and effective recommendation system, improving the user experience. Finally, the system can be extended to include audio analysis to improve the accuracy of the emotion recognition task. This can provide additional information and context to better understand and classify the detected emotions.

Overall, the proposed system has demonstrated the potential for using deep learning techniques for facial emotion recognition and video recommendation. With further improvements and advancements, this technology can have a significant impact on various applications such as mental health care, social robotics, and marketing.

# REFERENCE

1. Kim, Tae-Yeun, Hoon Ko, Sung-Hwan Kim, and Ho-Da Kim. "Modeling of recommendation system based on emotional information and collaborative filtering." Sensors 21, no. 6 (2021): 1997.
2. Islam, Md Rabiul, Md Milon Islam, Md Mustafizur Rahman, Chayan Mondal, Suvojit Kumar Singha, Mohiuddin Ahmad, Abdul Awal, Md Saiful Islam, and Mohammad Ali Moni. "EEG channel correlation-based model for emotion recognition." Computers in Biology and Medicine 136 (2021): 104757.
3. Muslihah, Isnawati, and Muqorobin Muqorobin. "Texture characteristic of local binary pattern on face recognition with probabilistic linear discriminant analysis." International Journal of Computer and Information System (IJCIS) 1, no. 1 (2020): 22-26.
4. Shahid, Ali Raza, Sheheryar Khan, and Hong Yan. "Contour and region harmonic features for sub-local facial expression recognition." Journal of Visual Communication and Image Representation 73 (2020): 102949.
5. Wang, Xusheng, Xing Chen, and Congjun Cao. "Human emotion recognition by optimally fusing facial expression and speech feature." Signal Processing: Image Communication 84 (2020): 115831.
6. Hassan, A.K. and Mohammed, S.N., 2020. A novel facial emotion recognition scheme based on graph mining. Defence Technology, 16(5), pp.1062-1072.
7. Sun, Ning, Qi Li, Ruizhi Huan, Jixin Liu, and Guang Han. "Deep spatial-temporal feature fusion for facial expression recognition in static images." Pattern Recognition Letters 119 (2019): 49-61.
8. Chervyakov, Nikolay, Pavel Lyakhov, Dmitry Kaplun, Denis Butusov, and Nikolay Nagornov. "Analysis of the quantization noise in discrete wavelet transform filters for image processing." *Electronics* 7, no. 8 (2018): 135.
9. Nigam, Swati, Rajiv Singh, and A. K. Misra. "Efficient facial expression recognition using histogram of oriented gradients in wavelet domain." Multimedia tools and applications 77 (2018): 28725-28747.
10. Pitaloka, Diah Anggraeni, Ajeng Wulandari, Tjan Basaruddin, and Dewi Yanti Liliana. "Enhancing CNN with preprocessing stage in automatic emotion recognition." Procedia computer science 116 (2017): 523-529.
11. Deshpande, Narayan T., and S. Ravishankar. "Face Detection and Recognition using Viola-Jones algorithm and Fusion of PCA and ANN." Advances in Computational Sciences and Technology 10, no. 5 (2017): 1173-1189.
12. Chen, Yen-Liang, Chia-Ling Chang, and Chin-Sheng Yeh. "Emotion classification of YouTube videos." Decision Support Systems 101 (2017): 40-50.
13. Yin, Zhong, Mengyuan Zhao, Yongxiong Wang, Jingdong Yang, and Jianhua Zhang. "Recognition of emotions using multimodal physiological signals and an ensemble deep learning model." Computer methods and programs in biomedicine 140 (2017): 93-110.
14. He, Kaiming, et al. "Deep residual learning for image recognition." Proceedings of the IEEE conference on computer vision and pattern recognition. 2016.

# APPENDIX

import cv2

import numpy as np

from keras.models import model\_from\_json from selenium import webdriver

emotion\_dict = {0: "Angry", 1: "Disgusted", 2: "Fearful", 3: "Happy", 4: "Neutral", 5: "Sad",

6: "Surprised"}

# Load json and create model

json\_file = open('emotion\_model.json', 'r') loaded\_model\_json = json\_file.read() json\_file.close()

emotion\_model = model\_from\_json(loaded\_model\_json)

# Load weights into the model emotion\_model.load\_weights("emotion\_model.h5") print("Loaded model from disk")

# Start the webcam feed cap = cv2.VideoCapture(0)

if not cap.isOpened():

print("Error opening video capture")

# Create a global variable to store the detected emotion detected\_emotion = ""

while True:

# Read frame from the video capture ret, frame = cap.read()

frame = cv2.resize(frame, (1280, 720)) if not ret:

break

face\_detector =

cv2.CascadeClassifier('haarcascades/haarcascade\_frontalface\_default.xml') gray\_frame = cv2.cvtColor(frame, cv2.COLOR\_BGR2GRAY)

# Detect faces in the frame

num\_faces = face\_detector.detectMultiScale(gray\_frame, scaleFactor=1.3, minNeighbors=5)

# Process each detected face for (x, y, w, h) in num\_faces:

cv2.rectangle(frame, (x, y-50), (x+w, y+h+10), (0, 255, 0), 4) roi\_gray\_frame = gray\_frame[y:y + h, x:x + w]

cropped\_img = np.expand\_dims(np.expand\_dims(cv2.resize(roi\_gray\_frame, (48, 48)), -1), 0)

# Predict the emotions

emotion\_prediction = emotion\_model.predict(cropped\_img) max\_index = int(np.argmax(emotion\_prediction)) detected\_emotion = emotion\_dict[max\_index]

cv2.putText(frame, detected\_emotion, (x+5, y-20), cv2.FONT\_HERSHEY\_SIMPLEX, 1, (255, 0, 0), 2, cv2.LINE\_AA)

cv2.imshow('Emotion Detection', frame) if cv2.waitKey(1) & 0xFF == ord('q'):

break

# Close the webcam cap.release() cv2.destroyAllWindows()

# Use the detected emotion to recommend a video

if detected\_emotion in ['Angry', 'Disgusted', 'Fearful']:

browser = webdriver.Chrome(r'C:\webdriver\chromedriver.exe') browser.get('https:/[/www](http://www.youtube.com/watch?v=RCAj8_wylKw%27)).[youtube.com/watch?v=RCAj8\_wylKw')](http://www.youtube.com/watch?v=RCAj8_wylKw%27))

elif detected\_emotion in ['Happy', 'Neutral', 'Surprised']:

browser = webdriver.Chrome(r'C:\webdriver\chromedriver.exe') browser.get('https://youtu.be/pRbxlpvXw2s?t=121')

elif detected\_emotion == 'Sad':

browser = webdriver.Chrome(r'C:\webdriver\chromedriver.exe')

browser.get('https:/[/www](http://www.youtube.com/watch?v=SBWYGGDYmhg&list=PLHuHXHyLu).[youtube.com/watch?v=SBWYGGDYmhg&list=PLHuHXHyLu](http://www.youtube.com/watch?v=SBWYGGDYmhg&list=PLHuHXHyLu) 7BGi-vR7X6j\_xh\_Tt9wy7pNA')

# Wait for the user to close the browser input("Press Enter to close the browser...")

# Quit the browser browser.quit

import cv2

from keras.models import Sequential

from keras.layers import Conv2D, MaxPooling2D, Dense, Dropout, Flatten from tensorflow.keras.optimizers import Adam

from keras.preprocessing.image import ImageDataGenerator # Initialize image data generator with rescaling train\_data\_gen = ImageDataGenerator(rescale=1./255) validation\_data\_gen = ImageDataGenerator(rescale=1./255) train\_generator = train\_data\_gen.flow\_from\_directory(

'data/train', target\_size=(48, 48), batch\_size=64, color\_mode="grayscale", class\_mode='categorical')

# Preprocess all train images

validation\_generator = validation\_data\_gen.flow\_from\_directory( 'data/test',

target\_size=(48, 48), batch\_size=64,

color\_mode="grayscale", class\_mode='categorical')

# create model structure emotion\_model = Sequential()

emotion\_model.add(Conv2D(32, kernel\_size=(3, 3), activation='relu', input\_shape=(48, 48,

1)))

emotion\_model.add(Conv2D(64, kernel\_size=(3, 3), activation='relu')) emotion\_model.add(MaxPooling2D(pool\_size=(2, 2))) emotion\_model.add(Dropout(0.25))

emotion\_model.add(Conv2D(128, kernel\_size=(3, 3), activation='relu')) emotion\_model.add(MaxPooling2D(pool\_size=(2, 2))) emotion\_model.add(Conv2D(128, kernel\_size=(3, 3), activation='relu')) emotion\_model.add(MaxPooling2D(pool\_size=(2, 2))) emotion\_model.add(Dropout(0.25))

emotion\_model.add(Flatten()) emotion\_model.add(Dense(1024, activation='relu')) emotion\_model.add(Dropout(0.5)) emotion\_model.add(Dense(7, activation='softmax'))

cv2.ocl.setUseOpenCL(False)

emotion\_model.compile(loss='categorical\_crossentropy', optimizer=Adam(lr=0.0001, decay=1e-6), metrics=['accuracy'])

# Train the neural network/model emotion\_model\_info = emotion\_model.fit\_generator(

train\_generator, steps\_per\_epoch=28709 // 64, epochs=50, validation\_data=validation\_generator, validation\_steps=7178 // 64)

# save model structure in jason file

model\_json = emotion\_model.to\_json()

with open("emotion\_model.json", "w") as json\_file: json\_file.write(model\_json)

# save trained model weight in .h5 file emotion\_model.save\_weights('emotion\_model.h5')

import cv2

import numpy as np

from keras.models import model\_from\_json

emotion\_dict = {0: "Angry", 1: "Disgusted", 2: "Fearful", 3: "Happy", 4: "Neutral", 5: "Sad",

6: "Surprised"}

# load json and create model

json\_file = open('emotion\_model.json', 'r') loaded\_model\_json = json\_file.read() json\_file.close()

emotion\_model = model\_from\_json(loaded\_model\_json)

# load weights into new model emotion\_model.load\_weights("emotion\_model.h5") print("Loaded model from disk")

# start the webcam feed cap = cv2.VideoCapture(0)

# pass here your video path

# you may download one from here : https:/[/w](http://www.pexels.com/video/three-girls-laughing-)w[w.pexels.com/video/three-girls-laughing-](http://www.pexels.com/video/three-girls-laughing-) 5273028/

#cap =

cv2.VideoCapture("C:\\Users\Dell\\PycharmProjects\\majorproject\\ideos\\emotion\_sampl e6.mp4")

if(cap.isOpened()==False):

print("error")

while True:

# Find haar cascade to draw bounding box around face ret, frame = cap.read()

frame = cv2.resize(frame, (1280, 720)) if not ret:

break

face\_detector =

cv2.CascadeClassifier('haarcascades/haarcascade\_frontalface\_default.xml') gray\_frame = cv2.cvtColor(frame, cv2.COLOR\_BGR2GRAY)

# detect faces available on camera

num\_faces = face\_detector.detectMultiScale(gray\_frame, scaleFactor=1.3, minNeighbors=5)

# take each face available on the camera and Preprocess it for (x, y, w, h) in num\_faces:

cv2.rectangle(frame, (x, y-50), (x+w, y+h+10), (0, 255, 0), 4) roi\_gray\_frame = gray\_frame[y:y + h, x:x + w]

cropped\_img = np.expand\_dims(np.expand\_dims(cv2.resize(roi\_gray\_frame, (48, 48)), -1), 0)

# predict the emotions

emotion\_prediction = emotion\_model.predict(cropped\_img) maxindex = int(np.argmax(emotion\_prediction))

cv2.putText(frame, emotion\_dict[maxindex], (x+5, y-20), cv2.FONT\_HERSHEY\_SIMPLEX, 1, (255, 0, 0), 2, cv2.LINE\_AA)

cv2.imshow('Emotion Detection', frame) if cv2.waitKey(1) & 0xFF == ord('q'):

break

cap.release() cv2.destroyAllWindows()