A close-up of a logo

Description automatically generated

**Facial Emotion Recognition Using Machine Learning**

Pawan Agrahari

Submitted in accordance with the requirements for the degree of   
Master of Science (Level 7)

York St John University

***Department of Computer Science***

Supervisor's Name:

**Dr. Soonleh Ling**

Date of Submission:

28/01/2025

**ABSTRACT**

Facial emotion detection is a developing field within artificial intelligence (AI) and computer vision, concentrating on recognizing and classifying human emotions based on facial expressions. This dissertation investigates a machine learning approach to identify emotions from facial images, leveraging advanced machine learning techniques that have become essential in numerous applications. By employing cutting-edge machine learning algorithms to analyse and interpret facial expressions, our research presents a novel method for emotion recognition. The main goal is to create a reliable system that can accurately recognize human emotions through facial analysis, offering valuable insights into individuals' emotional states. We assess various machine learning models, including Support Vector Machines (SVM), Logistic Regression, and Random Forest (RF), using datasets such as FER2013 and a custom-made dataset. The proposed system captures images through a webcam, processes them to extract key facial features, and classifies emotions using the most effective algorithms. The model demonstrates consistent performance by detecting subtle shifts in facial expressions, enabling real-time emotion recognition from both images and video streams. OpenCV is utilized for face detection and real-time emotion classification from live video feeds. This methodology holds potential applications in areas such as human-computer interaction, mental health monitoring, and customer sentiment analysis. The study also tackles the challenges associated with emotion recognition, including factors like inconsistent lighting, obstructions, and the subjective nature of emotions.

**Keywords**

Face Recognition, Feature extraction, SVM, Random Forest, Logistic Regression

**Declaration**

I hereby declare that this dissertation titled "Facial Emotion Recognition" is my original work and has been carried out under the supervision of Dr. Soonleh Ling at York St. John University London Campus. I confirm that all sources of information and material used in this dissertation have been fully acknowledged and cited according to academic standards. This work has not been submitted, either in whole or in part, for the award of any other degree or qualification at any other institution.

Signed: \_\_pawan\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

Name: Pawan Agrahari

Date: 28th January 2025

**Acknowledgement**

I would like to express my sincere gratitude to all those who have supported and guided me throughout the course of my dissertation project.

First and foremost, I wish to thank my supervisor, Dr. Soonleh Ling, for his invaluable guidance, expertise, and encouragement. His unwavering support and insightful feedback have been instrumental in shaping the direction of this research.

My family and friends have been a constant source of love, support, and motivation. I am thankful to everyone for their patience and understanding throughout this journey. Their encouragement has been invaluable.

Finally, I would like to express my appreciation to different databases for the resources and financial support provided during the course of my studies.

Thank you to everyone who contributed to this research in any way, your help has been indispensable.

**Table of Contents**

[**Keywords** 2](#_Toc186049373)

[**1.** **INTRODUCTION** 9](#_Toc186049374)

[**1.1** **Background** 9](#_Toc186049375)

[**1.2** **Problem Statement and Understanding** 10](#_Toc186049376)

[**1.3** **Research Aims/Objectives** 11](#_Toc186049377)

[**1.3.1** **Specific Objectives** 11](#_Toc186049378)

[**1.3.2** **General Objectives** 11](#_Toc186049379)

[**1.3.3** **Academic Objectives** 11](#_Toc186049380)

[**1.4** **Rationale of Research** 12](#_Toc186049381)

[**1.5** **Scope** 13](#_Toc186049382)

[**1.6** **Feasibility Study** 13](#_Toc186049383)

[**1.6.1** **Technical Feasibility** 13](#_Toc186049384)

[**1.6.2** **Operational Feasibility** 13](#_Toc186049385)

[**1.6.3** **Economic Feasibility** 14](#_Toc186049386)

[**1.6.4** **Schedule Feasibility** 14](#_Toc186049387)

[**1.7** **Dissertation Structure** 14](#_Toc186049388)

[**2.** **LITERATURE REVIEW** 15](#_Toc186049389)

[**2.1** **Advances in Machine Learning Algorithms for FER** 15](#_Toc186049390)

[**2.2** **Inputs of Multimodal Inputs** 17](#_Toc186049391)

[**2.3** **Real World Applications and Challenges** 18](#_Toc186049392)

[**4.** **METHODOLOGY** 20](#_Toc186049393)

[**4.1** **Project Methodology Overview** 20](#_Toc186049394)

[**4.2** **Implementation Tools** 21](#_Toc186049395)

[**4.2.1** **Programming Language and Libraries** 21](#_Toc186049396)

[**4.3** **Development Environment** 22](#_Toc186049397)

[**4.4** **System Design:** 23](#_Toc186049398)

[**4.4.1** **System Diagram** 23](#_Toc186049399)

[**4.4.2** **System Flowchart (Training and Testing of Model)** 24](#_Toc186049400)

[**4.5** **Data collection** 27](#_Toc186049401)

[**4.5.1** **FER-2013 dataset** 27](#_Toc186049402)

[**4.5.2** **Self-Created dataset** 28](#_Toc186049403)

[**4.6** **Data Pre-processing:** 29](#_Toc186049404)

[**4.6.2** **FER-2013 Dataset Challenges and Modifications** 30](#_Toc186049405)

[**4.7** **Modules Used in Our System** 31](#_Toc186049406)

[**4.7.2** **Face Detection Module** 31](#_Toc186049407)

[**4.7.3** **Feature Extraction Module** 33](#_Toc186049408)

[**4.7.4** **Emotion Classification Module (Model Building)** 35](#_Toc186049409)

[**4.8** **Hyperparameter Tuning** 39](#_Toc186049410)

[**4.8.2** **Hyperparameter for SVM** 39](#_Toc186049411)

[**4.8.3** **Hyperparameter for Random Forest** 40](#_Toc186049412)

[**4.8.4** **Hyperparameter for Logistic Regression** 40](#_Toc186049413)

[**4.9** **System Evaluation** 40](#_Toc186049414)

[**5.** **Results, Analysis and Discussion** 42](#_Toc186049415)

[**5.1** **Experiments and Results** 42](#_Toc186049416)

[**5.1.1** **Experiments Based on Emotion Detection Module** 42](#_Toc186049417)

[**5.2** **Emotion Detection Module Output** 45](#_Toc186049418)

[**5.2.1** **Emotion Detection Module Output** 46](#_Toc186049419)

[**5.3** **Analysis of Experiment Results** 47](#_Toc186049420)

[**6.** **Discussion and Conclusion** 48](#_Toc186049421)

[**6.1** **Discussion and Future Enhancement** 48](#_Toc186049422)

[**6.1.1** **Implications and Applications** 48](#_Toc186049423)

[**6.1.2** **Challenges and Limitations** 48](#_Toc186049424)

[**6.1.3** **Future Enhancements** 49](#_Toc186049425)

[**6.2** **Conclusion** 49](#_Toc186049426)

[**References** 51](#_Toc186049427)

**List of Figures**

[Figure 1: System Diagram 24](#_Toc186049428)

[Figure 2: Model Training Flowchart 25](#_Toc186049429)

[Figure 3: Model Testing Flowchart 26](#_Toc186049430)

[Figure 4: Image Example of FER-2013 Dataset 27](#_Toc186049431)

[Figure 5: Image Distribution in FER-2013 Dataset 28](#_Toc186049432)

[Figure 6: Images in Our Self-Created Dataset 28](#_Toc186049433)

[Figure 7: Images Distribution in Self-Created Dataset 29](#_Toc186049434)

[Figure 8: Original Image Figure 9: Cropped Image 30](#_Toc186049435)

[Figure 10: Code Snippet to Crop the Image to get Face Data 31](#_Toc186049436)

[Figure 11: Face Detection Using Haar-Cascade Classifier Algorithm 32](#_Toc186049437)

[Figure 12: Code Snippet Showing Face Detection 33](#_Toc186049438)

[Figure 13: Grayscale Image 34](#_Toc186049439)

[Figure 14: Code to Change BGR image to Grayscale 34](#_Toc186049440)

[Figure 15: Code Snippet to Convert RGB Image to Grayscale 35](#_Toc186049441)

[Figure 16: Graph for SVM (Sumbatilinda, 2024) 36](#_Toc186049442)

[Figure 17: Graph for Random Forest Classifier (Khushaktov, 2023) 38](#_Toc186049443)

[Figure 18: Graph for Logistic Regression 39](#_Toc186049444)

[Figure 19: Confusion Matrix Heatmap using SVM model 44](#_Toc186049445)

[Figure 20: Confusion Matrix Heatmap using SVM in Self-Created Dataset 45](#_Toc186049446)

[Figure 21: GUI for our Study 46](#_Toc186049447)

[Figure 22: Emotion Recognition Using Webcam 47](#_Toc186049448)

**List of Tables**

[Table 1: Dataset Collection and Description 29](#_Toc186049449)

[Table 2: Model, Accuracy score and Hyperparameters (FER-2013) 43](#_Toc186049450)

[Table 3: Performance Evaluation Metrics of SVM on FER-2013 dataset 43](#_Toc186049451)

[Table 4: Model, Accuracy Score and Hyperparameters (Self-Created Dataset) 44](#_Toc186049452)

[Table 5: Performance Evaluation Metrix using SVM in Self-Created Dataset 44](#_Toc186049453)

# **INTRODUCTION**

## **Background**

Facial emotion recognition (FER) is an interdisciplinary domain that blends computer vision, psychology, and artificial intelligence (AI). It is dedicated to recognizing and analysing human emotions through facial expressions. Emotions play a crucial role in human communication, impacting decisions, interactions, and relationships. In the modern digital age, the ability to automatically detect and interpret emotions holds transformative possibilities across various sectors, including healthcare, education, security, marketing, and human-computer interaction.

Understanding human emotions through facial expressions enables machines to respond more intelligently and empathetically to user needs, enhancing user experiences across diverse applications (Zeng, et al., 2009). Emotions are crucial in communication, and facial expressions are among the clearest signs of a person’s emotional condition. Common human emotions include anger, disgust, fear, joy, neutrality, sadness, and surprise. Additionally, more nuanced emotions, such as cheerfulness (a variation of joy) and contempt (a variation of disgust), fall under this umbrella. These subtle emotions are challenging to detect due to the faint facial muscle contortions involved, as even minor variations can result in distinct expressions (Kret, 2015). Furthermore, emotions are highly context-dependent, leading to differences in how individuals express the same feeling. Although specific facial regions, like the eyes and mouth, are commonly emphasized for emotion detection, the techniques used to capture and categorize these expressions continue to pose a major challenge.

Historically, emotion analysis depended on manual observations or self-reported data, both of which are subjective, time-intensive, and difficult to scale. With the rise of AI, machine learning (ML) has become a robust alternative, allowing for the automated and objective detection of emotions from facial expressions. Deep learning methods, particularly convolutional neural networks (CNNs), have demonstrated impressive potential. Nonetheless, traditional ML algorithms such as Support Vector Machines (SVM), Logistic Regression, and Random Forest continue to be valuable due to their simplicity, interpretability, and efficiency, especially in resource-limited settings.

Facial emotion recognition, a subset of affective computing, involves analysing facial expressions to identify emotional states. This ability is crucial for improving human-computer interactions, advancing mental health assessments, and tailoring user experiences in various sectors. With advancements in AI, particularly in machine learning and deep learning, automated systems now achieve accurate and real-time emotion recognition (Gera & Balasubramanian, 2022). These breakthroughs have unlocked opportunities in healthcare, education, security, and customer service. The potential applications of FER are extensive and impactful. In healthcare, emotion recognition systems can assist in mental health monitoring by detecting signs of depression or anxiety from facial cues. In customer service, such systems analyse customer satisfaction in real time by interpreting facial reactions. Similarly, educational tools can adapt content or teaching methods based on students' emotional engagement. Neural networks and ML models have demonstrated strong performance in these areas, showcasing their vast potential.

This dissertation focuses on developing a reliable Facial Emotion Recognition (FER) system utilizing conventional machine learning algorithms, including SVM, Logistic Regression, and Random Forest. The research addresses key components such as data pre-processing, feature extraction, and training the models, while also highlighting the advantages and disadvantages of each algorithm. The study assesses these models using recognized FER datasets, addressing challenges such as variations in facial expressions, lighting conditions, and occlusions. Furthermore, it evaluates the performance of these algorithms in terms of accuracy, precision, recall, and computational efficiency. Despite significant progress, FER systems face several challenges. Emotional expressions vary across cultural, gender, and age groups, introducing variability that complicates model accuracy. Real-world environments pose additional difficulties, such as variations in lighting, occlusions (e.g., glasses, masks), and dynamic backgrounds. Moreover, training datasets are often imbalanced, with certain emotions underrepresented, leading to biased detection outcomes. Addressing these challenges requires robust pre-processing techniques, balanced and diverse datasets, and state-of-the-art ML architectures.

By leveraging classical ML techniques, this research aims to provide insights into their applicability in FER and their effectiveness in real-world scenarios. The findings are expected to contribute to the development of accessible and efficient emotion recognition systems, paving the way for empathetic and intelligent technologies. Using a webcam interface, the system captures facial images and processes them through advanced image segmentation and feature extraction techniques. These features are analysed using ML models to classify the emotions being expressed. By improving detection accuracy, the system can be applied across contexts, from enhancing security systems to creating personalized user experiences. The integration of such technologies into daily life promises to revolutionize human-machine interactions, fostering more intuitive and empathetic connections.

## **Problem Statement and Understanding**

Understanding and interpreting human emotions through facial expressions is inherently complex due to the subtlety, variability, and context-dependent nature of these expressions. Even the same emotion may manifest differently across individuals or situations, making the task challenging. For instance, happiness can range from a slight smile to broad laughter, depending on the context. Such variations necessitate the use of sophisticated machine learning algorithms capable of generalizing across diverse datasets while accurately identifying minute details. In today’s digital era, machines equipped with emotional understanding can significantly enhance human-computer interactions. By bridging the gap between human intuition and artificial intelligence, these systems make interactions more responsive and adaptive. For example, customer service chatbots can analyse emotional cues to tailor responses, creating a more personalized and satisfying experience. Similarly, healthcare systems that monitor emotional states over time can aid clinicians in diagnosing and tracking mental health conditions like depression or anxiety.

Developing a reliable facial emotion detection system involves addressing several challenges. Variations in lighting, facial orientations, and occlusions such as glasses, masks, or facial hair can greatly influence the accuracy of an emotion recognition system. Changes in lighting conditions can cause shadows or highlights on the face, making it difficult to accurately detect facial features. Similarly, differences in facial orientations—such as turning the head or tilting the face—can distort key features, affecting the system's ability to correctly interpret expressions. Additionally, occlusions like glasses, masks, or facial hair can obscure important facial landmarks (like the eyes, mouth, and nose), leading to partial or incorrect emotion detection. To ensure reliability in real-world scenarios, these factors must be carefully considered during development. Emotions are often conveyed through specific facial landmarks, such as the eyes, eyebrows, and mouth, which require precise detection and analysis. The process typically includes several steps: detecting the face within an image, extracting essential features from facial landmarks, and using machine learning algorithms to classify these features into defined emotion categories. By training models on large, labelled datasets, systems can learn to recognize subtle patterns in facial features, enabling them to differentiate emotions like anger, joy, or surprise effectively. This project employs advanced image processing techniques and machine learning to overcome these challenges, creating a system that adapts to diverse environments and user profiles.

The proposed application leverages cutting-edge facial expression recognition algorithms to accurately capture and analyse users' emotions. It begins by isolating the face from input images or videos and extracting key features related to emotional expressions. These features are then processed by machine learning models to classify emotions into distinct categories. This technology offers potential benefits across various fields, such as enhancing security systems by identifying suspicious behaviour, tailoring educational content based on student engagement, and providing valuable insights in therapeutic settings. By accurately detecting emotions, the system delivers real-time feedback and improves user experiences, highlighting the transformative impact of integrating emotional intelligence into artificial intelligence solutions.

## **Research Aims/Objectives**

Facial emotion detection is a machine learning task that revolves around identifying and categorizing human emotions based on facial expressions. The main goal is to create a model that can automatically recognize and classify a variety of emotions—such as happiness, sadness, anger, surprise, fear, disgust, and neutral—using images or video frames of human faces. This technology holds vast potential across numerous fields, including psychology, where it can aid in understanding emotional states, human-computer interaction, by improving the responsiveness of systems to user emotions, healthcare, for better monitoring of mental health, and customer experience monitoring, helping businesses gauge consumer reactions and improve service quality.

### **Specific Objectives**

* Develop a system that accurately detects and classifies facial emotions using machine learning techniques.

### **General Objectives**

* Create a user-friendly application for real-time facial emotion detection.
* Implement and evaluate various machine learning models to enhance emotion recognition accuracy.

### **Academic Objectives**

This project also focuses on achieving academic goals, including gaining deeper insights into Big Data and Machine Learning. Key academic objectives are:

* **Practical Application of Theoretical Knowledge:** Apply classroom knowledge in a real-world project to bridge the gap between theory and practice.
* **Object-Oriented Development:** Use an object-oriented approach to design and implement the system.
* **Hands-on Machine Learning Experience:** Learn and apply supervised machine learning methods, gaining familiarity with concepts like:
* Accuracy
* Classifiers
* Precision
* Recall
* Hyperparameter tuning
* **Programming and Scripting Proficiency:** Enhance programming skills, particularly in Python, while implementing the project

## **Rationale of Research**

The rationale for researching facial emotion recognition (FER) with machine learning stems from its potential to bridge the gap between human intuition and artificial intelligence, allowing robots to engage with humans in a more intelligent, empathic, and responsive manner. Facial expressions are a global language of emotion that cross linguistic and cultural boundaries and play an important part in human communication. By creating systems that can comprehend these expressions, we can increase the quality of interactions in a variety of fields, including healthcare, education, security, and customer service. Machine learning provides the computational foundation for analysing complex patterns in face features, allowing automated systems to discern emotions with accuracy and consistency.

This research is particularly significant given the growing demand for emotion-aware systems in today’s increasingly digitized world. In healthcare, for example, FER systems can assist in mental health diagnostics by detecting subtle signs of distress, anxiety, or depression from facial expressions, providing valuable insights for clinicians. In education, emotion-aware tools can monitor student engagement and adapt teaching strategies to optimize learning outcomes. Similarly, in customer service, businesses can leverage FER to analyse customer satisfaction in real time, enhancing service quality and client retention. The integration of such technology into everyday applications underscores the practical relevance of this research and its ability to transform user experiences.

Additionally, the difficulties in creating reliable FER systems emphasize the need for more study. Accurate emotion recognition is made more difficult by the diversity of emotional manifestations throughout cultures, genders, and age groups as well as practical problems such occlusions, changing surroundings, and illumination. A promising answer to these problems is machine learning, which can generalize patterns and learn from big datasets. This study intends to overcome these constraints by furthering the development of FER systems, increasing the accuracy and generalizability of emotion recognition in a variety of settings. In the end, this research advances the larger objective of building emotionally intelligent and intelligent machines, ushering in a new era of human-computer connection.

## **Scope**

The potential for facial expression recognition through machine learning is vast and expanding with technological progress. This area has applications across various sectors, including healthcare, education, entertainment, and human-computer interaction. In healthcare, emotion detection can assist in tracking patients' emotional well-being, enabling early intervention for mental health issues. In education, it can provide teachers with immediate insights into student engagement and understanding, leading to more tailored learning experiences. This technology aids the entertainment business by tailoring content recommendations based on viewers' emotional responses, which increases user engagement. Furthermore, facial emotion recognition is crucial in the development of social robots and virtual assistants, allowing them to respond more naturally to human emotions and so improve the user experience. However, the scope also involves tackling issues such as cultural variances in emotional expression, ethical concerns about privacy, and the necessity for robust algorithms that can operate accurately across varied demographics and settings. As machine learning advances, particularly deep learning, the potential applications and effectiveness of facial emotion recognition systems are expected to grow, paving the door for creative solutions in both established and emerging sectors. Some of the application is listed below:

* **Human-Computer Interaction:** Enhances user experience by adapting software behaviour based on user emotions.
* **Mental Health Monitoring:** Assists in tracking emotions over time, which can help identify emotional distress or disorders.
* **Customer Feedback Analysis:** In retail, analysing customer emotions helps companies tailor their offerings and improve satisfaction.
* **Education and Training:** Monitors students' reactions to adapt teaching methods in e-learning systems.
* **Security and Surveillance:** Detects unusual behaviour or distress in real-time surveillance applications.as a new feature.

## **Feasibility Study**

Prior to starting the project, a feasibility study is carried out to evaluate the system's practicality. This study is essential to determine if developing a new or improved system is financially viable, advantageous, functional, technologically feasible, and time-efficient. The details of the feasibility study are outlined below:

### **Technical Feasibility**

Technical feasibility is one of the initial assessments to be conducted once the project is identified. This analysis examines both the hardware and software requirements. For this project, tools such as Python, VS Code IDE, Jupyter Notebook, and Google Colaboratory are readily accessible.

### **Operational Feasibility**

Operational feasibility refers to an evaluation of how well a proposed system addresses the problem and leverages the opportunities identified during the scope definition phase. The factors listed below were taken into account when assessing the operational feasibility of the project:

* The system will be able to detect and capture facial images.
* The captured image is then categorized based on facial expressions or emotions.

### **Economic Feasibility**

Economic feasibility involves quantifying and identifying the positive economic benefits of a system. The system is considered economically feasible as it fulfills all requirements, including collecting data from:

* FER-2013
* Self-created dataset (Images downloaded from Google)

### **Schedule Feasibility**

The timeliness of a project is determined by its scheduling feasibility. Because the system is developed in such a way that it will complete within the stated time, it was considered schedule feasible.

## **Dissertation Structure**

The remainder of the dissertation is organized as follows: Chapter 1 contains introduction information, problem statements, intentions, and objectives. Chapter 2 presents a comprehensive analysis of stock price prediction strategies that employ various machine learning techniques. Chapter 3 details the data collecting and pre-processing techniques, as well as the research approach, which includes algorithm selection, model construction, and assessment measures. Chapter 4 summarizes the empirical findings from the comparison analysis. The fifth chapter addresses the study's findings, consequences, and limitations. Finally, Chapter 6 summarizes the major contributions and makes recommendations for future research.

1. **LITERATURE REVIEW**

Facial emotion detection has evolved significantly over the past decade, transitioning from basic image processing techniques to sophisticated deep learning methodologies. Early approaches primarily focused on geometric features, analysing the distances and angles between facial landmarks to infer emotions. Techniques such as Support Vector Machines (SVMs) and k-Nearest Neighbours (k-NN) were commonly employed for classification tasks.

## **Advances in Machine Learning Algorithms for FER**

Anagha S. Dhavalikar and Dr. R. K. Kulkarni proposed an Automatic Facial Expression Recognition system in their paper (Davalikar & Kulkarni, 2014) that first detects faces within images and then analyses them to identify emotional expressions, such as happiness, sadness, anger, and surprise. Their approach utilizes a combination of feature extraction and classification algorithms to achieve accurate recognition, which operates in three phases:

1. Face Detection
2. Feature Extraction
3. Expression Recognition

In the paper, face detection began with an analysis of the RGB color model, which included ISO illumination processing to identify the face and operations to preserve key facial features, such as the eyes and mouth. The proposed algorithm also utilized the Active Appearance Model (AAM) to extract facial features. This technology identified various facial features, known as Action Units (AUs), such as the eyes, eyebrows, mouth, and lips, and generated a file containing the parameters of the detected action points. The facial emotions were then input into the AAM model, which made adjustments based on the expression. The researchers used a simple Euclidean Distance method to compare the Euclidean distance between feature points in both training and query images. The recognition accuracy for this study ranged from 90% to 95%.

J. H. Immanuel, J. J. A. Arnold, J. M. MasillaRuban, M. Tamilarasan, and R. Saranya proposed a system named Emotion Based Music Recommendation System (Immanuel, et al., 2019) where they used FER-2013 dataset and model is trained using SVM. The features are extracted by processing the image and converting the facial expressions into a sequence of Action Units (AU). Here only 4 emotions Happy, Sad, Angry and Surprise are used.

The research by (Mellouk & Handouzi, 2020) emphasizes the growing importance of automatic facial emotion recognition (FER) and its applications in fields like safety, healthcare, and human-machine interfaces. The primary objective is to develop techniques that enable computers to interpret and decode facial expressions for more accurate emotion prediction. The paper reviewed recent advancements in Facial Expression Recognition (FER) using deep learning, emphasizing various methods and their effectiveness. It discussed the architectural approaches, the datasets utilized for training and validation, and the outcomes achieved by different methods. The study compared the performance of different FER methods, highlighting the progress made in accuracy and efficiency over time. It provided insights into which methods have been more successful and why. The paper aimed to guide researchers by offering a review of recent advancements, identifying gaps, and suggesting areas for improvement. It served as a resource for enhancing the effectiveness of FER systems in real-world applications.

The study by (Huang, et al., 2023) utilized a deep neural network (DNN) for facial emotion recognition (FER), specifically combining Convolutional Neural Networks (CNN), Squeeze-and-Excitation Networks (SENet), and Residual Neural Networks (ResNet) to enhance performance. The research highlighted that facial feature around the nose and mouth are crucial indicators for the DNN model in understanding emotions, underlining the importance of these areas in FER tasks. To train and validate the CNN models, the study used the AffectNet and Real-World Affective Faces Database (RAF-DB. These datasets are commonly used for emotion recognition and provide a variety of facial expression examples. When evaluated on RAF-DB, the AffectNet-trained model scored 77.37% accuracy, indicating the approach's generalizability across datasets. Further, transfer learning (pretraining on AffectNet and fine-tuning on RAF-DB) increased the accuracy to 83.37%, showcasing the effectiveness of transfer learning in improving FER performance. The study provides insights into how neural networks focus on critical facial features, contributing to a better understanding of model behavior in emotion recognition. The findings are expected to improve the accuracy of computer vision systems in real-world FER applications.

(Manalu & Rifai, 2024) investigated the issue of facial expression recognition (FER) using a dataset focused on emotion detection, identifying a total of 9 distinct emotions. They created hybrid models combining CNN and RNN techniques to address this challenge, employing both complete learning and transfer learning with MobileNetV2-RNN and InceptionV3-RNN. The traditional CNN-RNN approach achieved an accuracy rate of 63%, while the transfer learning models, MobileNetV2-RNN and InceptionV3-RNN, attained accuracy rates of 59% and 66%, respectively. The models developed demonstrate enhanced capability in recognizing subtle emotional cues, marking a noteworthy advancement in the area of facial expression recognition. This research carries important repercussions for cognitive science and practical uses, especially in enhancing interactive digital communication and emotional analysis.

The research paper published by (Kavitha & Vinodhini, 2024) used Convolutional Neural Networks (CNNs) and Linear Binary Patterns (LBPs) in facial recognition systems to identify emotions are popular methods in deep learning algorithms for facial feature extraction and emotion classification. The study dataset comprised 400 photos collected from 40 students (10 images per student) that were used to train the models, with 10 iterations (N=10) to improve algorithm accuracy. To enable model comparison, the photos were divided into two groups: Group 1 and Group 2. The Convolutional Neural Network (CNN) achieved an accuracy of 83.47% in emotion recognition, while Linear Binary Patterns (LBPs) had an accuracy rate of 82.15%. The accuracy difference suggests that CNNs may provide slightly better performance in emotion recognition.

## **Inputs of Multimodal Inputs**

Yong-Hwan Lee, Woori Han, and Young Kim Lee (Lee, et al., 2013) presented a method that uses Bezier curve fitting. This algorithm employed a multi-phase method for facial expressions and emotions. The first phase was to recognize and process the face region from the original input image. The second stage was to confirm the feeling in the location of interest. In the first step, face identification was performed using a colour still picture based on skin colour pixel with initialized spatial filtering that takes into account lighting circumstances. The Feature Map was then utilized to locate the position of the eyes and mouth, as well as to outline the general shape of the face. After extracting the regions of interest, this method extracts points from the feature map to apply the Bezier curve to the eye and mouth areas. It then analyses the difference between the Harsdorf distance and the Bezier curve, comparing the database image with the input face images. The size of the dataset was not limited. Even on enormous datasets, the technique remained highly efficient. The technique may also be used to 3D images, which can be utilized to extract information, but each point of study in the data has a global impact, and there are no outliers. If the dataset becomes skewed as a result of overfitting, efficiency suffers.

The study carried out by (Huang, et al., 2023) explored the application of Convolutional Neural Networks (CNN) in conjunction with a mix of Squeeze-and-Excitation Networks (SENet) and Residual Neural Networks (ResNet) for Facial Emotion Recognition (FER). For the purpose of training the CNN model, two primary facial expression databases were employed: AffectNet and the Real-World Affective Faces Database (RAF-DB). To boost performance, feature maps were extracted from the network's residual blocks and subjected to additional analysis. The research indicated that facial features, especially those located around the nose and mouth, serve as key indicators for neural networks when identifying emotions. These regions contain a wealth of information that aids the model in accurately discerning various emotional expressions. In terms of outcomes, the research demonstrated that the CNN model trained exclusively on AffectNet attained a validation accuracy of 77.37% when evaluated on the RAF-DB. Nevertheless, when the network model was initially pretrained on AffectNet and subsequently fine-tuned using the RAF-DB dataset, its accuracy saw a considerable improvement, reaching 83.37%. This observation underscores the advantages of transfer learning, showing that a model trained on one dataset (AffectNet) can be adapted to a different one (RAF-DB) to enhance performance. This finding emphasizes the significance of utilizing pretrained models and refining them with pertinent data to attain greater recognition accuracy in emotion detection tasks.

(Hameed, et al., 2024) introduced a novel human behaviour detection system that prioritizes privacy, leveraging Frequency-Modulated Continuous Wave (FMCW) radar alongside Machine Learning (ML) techniques to identify facial expressions. The research concentrated on five prevalent facial expressions: happy, sad, fearful, surprised, and neutral. Data was gathered in the form of Micro-Doppler signals, which were subsequently analysed to identify key features with the help of various advanced machine learning models, such as Super Learner, Linear Discriminant Analysis, Random Forest, K-Nearest Neighbours, Long Short-Term Memory, and Logistic Regression. After extracting the relevant features from the radar data, these were fed into different machine learning algorithms for classification purposes. The study's findings revealed a highly encouraging classification accuracy of 91%, underscoring the potential of FMCW radar and machine learning for facial expression recognition while preserving privacy. This method eliminates the requirement for conventional imaging techniques, providing a non-invasive and effective solution for recognizing human emotions.

## **Real World Applications and Challenges**

A. Lehtiniemi and J. Holm suggested a system for music recommendation based on an animated mood picture (Lehtiniemi & Holm, 2012). This program allows users to connect with a group of photographs to generate Pictorial music recommendations. This music suggestion program was produced by Nokia Research. This program employed text meta tags to determine how audio and genre signals are encoded. This method investigated real-time data for analysis. The data was retrieved from user reactions, resulting in great accuracy.

(Avata, et al., 2018) introduced a music recommendation system focused on emotions that determines the user's emotional state through signals gathered from wearable devices equipped with Galvanic Skin Response (GSR) and Photoplethysmography (PPG) sensors. Emotions play a crucial and central role in human life, affecting choices and behaviors. This study tackles the issue of recognizing emotions by forecasting two important emotional dimensions: arousal and valence, using multi-channel physiological data. The system evaluates these physiological signals to effectively identify emotional states, enabling personalized music suggestions that correspond to the user's emotional state.

The authors from (Guidel, et al., 2020) emphasized that facial expressions can swiftly convey an individual's emotional mood and mental state. The system created in the study centers on basic emotions such as happiness, sadness, anger, excitement, surprise, disgust, fear, and neutrality. Face detection for this project was accomplished through the use of a convolutional neural network. The authors introduce a system that analyzes users' facial expressions to ascertain their emotional state, which is subsequently utilized for personalized music suggestions. The article explores the combination of facial emotion recognition methods, likely employing machine learning algorithms to detect emotions from facial data collected via camera input. This technology aims to improve the user experience by suggesting music that aligns with the user's current emotional condition, thereby promoting a more interactive and customized experience with music streaming services.

(Sadhvika, et al., 2020) suggested that manually organizing a playlist and annotating songs based on a user's current emotional state is a labour-intensive and time-consuming task. While several methods have been developed to automate this process, existing algorithms are often slow, require additional hardware (such as EEG structures and sensors), and suffer from lower accuracy. The study presents an algorithm designed to automatically generate a music playlist based on a person’s facial expressions, thereby saving time and effort compared to manual methods. The proposed approach aims to reduce both the overall processing time and the cost of the system, providing a more efficient solution for music personalization.

A. Madhuri, M. Deepali, S. Upasana, and G. Megha (Athavle, et al., 2021) created a music recommendation system that utilizes facial emotion recognition, using the FER-2013 dataset and CNN for identifying emotions. Their goal was to reduce both the processing time needed to generate results and the overall cost of the proposed system. However, they utilized the FER-2013 dataset, which is highly unbalanced and contains misaligned images, variations in contrast, outliers, and incorrect labels.

Two researchers (Amjad & Aslam, 2024) described the application of FER in monitoring mental health, particularly among adolescents in schools. Machine learning tools assist in detecting bullying behaviours and emotional stress. Also, the study done by (Haque, 2024) discussed the role of FER in video surveillance systems. His work introduces fuzzy-based transformations to mitigate adversarial attacks on deep learning models, ensuring reliable emotion detection in security setups.

# **METHODOLOGY**

## **Project Methodology Overview**

In the initial phase of our study, significant effort was invested in planning and researching reliable and effective algorithms for facial emotion detection. This phase involved a thorough evaluation and discussion of various machine learning algorithms to determine which would be most appropriate for the project. Concurrently, we addressed the critical tasks of data collection and pre-processing, recognizing that fine and accurate results require high-quality and large-volume datasets. Since achieving better accuracy necessitates extensive data, we collected datasets by exploring various online resources. Given that we are new to this domain, we decided to utilize grayscale pixel values as a feature for the system. These values provide a simplified yet effective representation of facial data. For model training, we selected three machine learning algorithms: Support Vector Machine (SVM), Logistic Regression and Random Forest, as these are known for their strong performance in classification tasks. To identify the most suitable algorithm for our needs, we implemented and compared both models. The OpenCV framework was chosen for its robustness and utility in computer vision tasks, particularly in real-time facial recognition and processing applications.

The designed system focuses on delivering an engaging and user-friendly platform for recognizing facial emotions. Central to the system is a graphical user interface (GUI) that promotes smooth interaction for users. Users can effortlessly start the camera through the GUI, which captures their facial images in real-time. The system then focuses on the facial area by cropping the captured image, extracts pertinent features, and analyses these features using a machine learning algorithm to determine and predict the expressed emotion. The primary objective of the system is to accurately recognize and detect the user’s facial emotions, providing instantaneous feedback through the GUI. This setup renders the system straightforward and applicable for practical use in everyday scenarios.

Our approach utilizes a systematic workflow that starts with the gathering and pre-processing of facial image data. The datasets incorporated consist of FER-2013, which is a publicly accessible dataset focused on facial expressions, along with a custom dataset specifically developed for this research. These datasets feature labelled images that represent seven unique emotional categories: Angry, Disgust, Fear, Happy, Sad, Surprise, and Neutral. The pre-processing stage includes resizing all images to a standard dimension of 48x48 pixels, converting them to grayscale to streamline the data, and normalizing pixel values to ensure consistency. Furthermore, we employ data augmentation techniques such as rotating, flipping, and zooming images to enhance the diversity of the training data, thereby boosting the model's capability to generalize across various expressions and conditions.

Once pre-processed, the data is used to train the system to recognize patterns in facial expressions and associate them with the corresponding emotions. After training, the system is capable of processing real-time video streams captured via OpenCV. A pre-trained face detector is used to locate and isolate facial regions within the video frames. These regions are then passed through the trained machine learning model, which predicts the associated emotion. The system’s real-time processing capability makes it suitable for a wide range of practical applications.

For instance, in the healthcare sector, it can help in understanding patient emotions, leading to more accurate mental health assessments. In customer service, the system can gauge customer satisfaction by analyzing emotions during interactions. Furthermore, it enhances human-computer interaction by allowing systems to dynamically adjust to users' emotional states, ultimately improving the overall user experience and aiding better decision-making processes.

## **Implementation Tools**

### **Programming Language and Libraries**

* 1. **Python**

Python is a powerful and versatile programming language well-suited for addressing statistical challenges using machine learning techniques. It provides a wide range of utility functions that streamline the pre-processing phase. Python is known for its fast processing and cross-platform compatibility. It also offers easy integration with C++ and other image processing libraries, making it highly adaptable. With built-in methods and libraries for managing and manipulating various types of data, Python offers a robust environment for data-driven tasks. The pandas and NumPy frameworks, in particular, are invaluable for data manipulation. NumPy arrays, for example, allow for efficient handling of n-dimensional data, enabling the development of complex features for machine learning models.

* 1. **Scikit-learn**

Scikit-learn is a widely-used machine learning library for Python that works well with Matplotlib, NumPy, and various machine learning algorithms. Its interface is intuitive and easy to navigate, which makes it suitable for both newcomers and experienced users. The library offers a wide array of functions for data analysis and visualization, which are crucial for examining datasets. Scikit-learn also includes robust techniques for feature reduction, significance testing, and selection that can assist in crafting an optimal feature set for machine learning tasks. Its algorithms are adaptable and can be utilized for classification and regression challenges, along with their specific subclasses, making it an all-encompassing tool for diverse machine learning applications.

* 1. **Numpy**

NumPy is a widely-used Python package for managing large multidimensional arrays and matrices, offering a broad range of high-level mathematical operations (Bigelow, 2024). It is especially useful for performing essential scientific computations in Machine Learning.

* 1. **Pandas**

Pandas is a widely used Python package for data analysis, though it is not directly related to machine learning. Before training a machine learning model, the dataset must be prepared, and this is where Pandas proves particularly useful. It was specifically designed for data extraction, cleaning, and pre-processing, making it an essential tool for organizing and preparing data before applying machine learning algorithms (Codecademy Team, 2024).

* 1. **Matplotlib**

Matplotlib is a popular Python data visualization library. It is extremely handy when a developer wants to observe how data patterns are represented. It's a 2D charting tool that lets you create 2D graphs and charts.

* 1. **OpenCV**

OpenCV is an open-source library that supports a wide range of image processing functions and algorithm implementations, making it ideal for tasks like image transformations, such as grayscale conversion. It supports both C++ and Python, providing flexibility in development. OpenCV is a comprehensive package that can be integrated with other libraries to create a pipeline for image extraction or detection tasks. It offers a broad array of functions, including methods for extracting feature descriptors, making it a powerful tool for computer vision applications.

## **Development Environment**

* + - 1. **Jupyter Notebook**

Jupyter Notebook is an integrated development environment (IDE) that allows users to combine Python with various libraries essential for machine learning and data analysis. While some complex algorithms may take longer to process, the environment is interactive, enabling real-time output, such as plots and images. Jupyter Notebook serves as a comprehensive tool for all stages of a project, and it easily integrates with many popular libraries, including OpenCV and Scikit-learn, making it a versatile platform for developing and testing solutions.

* + - 1. **Google Colab**

Google Colab, also known as Collaboratory, is a Jupyter environment that works with CPUs, GPUs, and even TPUs, and is provided and supported by Google. It's just like any other Jupyter notebook, with the ability to code in Python and write explanations in Markdown, as well as all of the other Jupyter features. Google Colab is a product that emphasises collaboration. It is also hosted on Google's servers, so no download is required. The notebooks are also stored in your Google Drive account.

* + - 1. **Visual Studio Code**

Microsoft's Visual Studio Code (VS Code) is a source code editor that can be used on Windows, Linux, and macOS platforms. It includes a variety of powerful functionalities such as debugging features, built-in Git control, integration with GitHub, syntax highlighting, intelligent code suggestions, code snippets, and code refactoring tools. These capabilities make VS Code a flexible and effective editor for programming, offering a seamless coding experience along with an extensive selection of extensions and support for numerous programming languages.

## **System Design:**

System design depicts the overall design of the system. This section goes into detail about the system's design.

### **System Diagram**

The development of our system followed a structured and methodical approach, beginning with the planning phase. In this phase, we clearly identified and analyzed key requirements, including problem understanding, defining objectives, establishing the scope of the project, and determining the data requirements essential for building the system.

After grasping the requirements, the subsequent step involved gathering data, which was essential for training and assessing the system. Once the data was collected, it was divided into two groups: one for training and the other for testing. This division was crucial to guarantee that the system could be properly trained on one set of data and later evaluated on another to confirm its performance and ability to generalize.

The next step involved face detection, a crucial preprocessing stage that identifies and isolates facial regions within the dataset. This allows the system to focus on the relevant areas of the data, ensuring that only the essential facial features are processed. After detecting the faces, the system proceeded with feature extraction, a process aimed at capturing meaningful and distinctive features from the detected faces. These extracted features form the basis for classification, playing a key role in recognizing patterns in the data and enabling accurate emotion detection.

The training dataset, together with the extracted features, was utilized to train the model employing a range of machine learning classification algorithms. These algorithms were meticulously selected to enhance the model’s proficiency in accurately identifying and distinguishing between various emotional states. After training the model, it was tested on the testing dataset. The features from this testing dataset were assessed using the trained model to evaluate its accuracy and effectiveness.

Finally, the system classified the detected and tested features into specific emotion categories, such as angry, disgust, fear, happy, neutral, sad, and surprise. These emotion classes represent the system’s output, demonstrating its capability to analyze and classify facial expressions with accuracy.

The flowchart below provides a comprehensive illustration of the system's design, depicting the step-by-step process of training and testing the model.

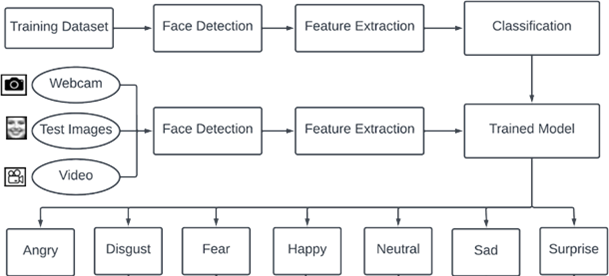
****

Figure 1: System Diagram

### **System Flowchart (Training and Testing of Model)**

The following figures illustrate the step-by-step process of training and testing a machine learning model. The process begins with data collection, where relevant data is gathered for training. This is followed by data preprocessing, where the data is cleaned and prepared for further analysis, ensuring consistency and quality. The next step is feature extraction, where the most relevant attributes of the dataset are identified and isolated, which will contribute to the learning process.

During the model training stage, these features are utilized to train the machine learning algorithm. Following the training, an evaluation of the model's performance is conducted. If the performance falls short of expectations, hyperparameter tuning is carried out to refine the model’s parameters. The model is subsequently re-evaluated, and this cycle is repeated continuously until the target performance is reached.

Once the model meets the performance criteria, it is saved for future use. This systematic workflow ensures the development of a robust and accurate machine learning model.

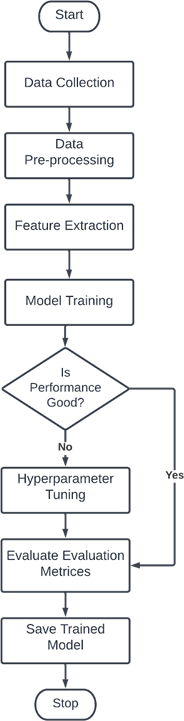


Figure 2: Model Training Flowchart

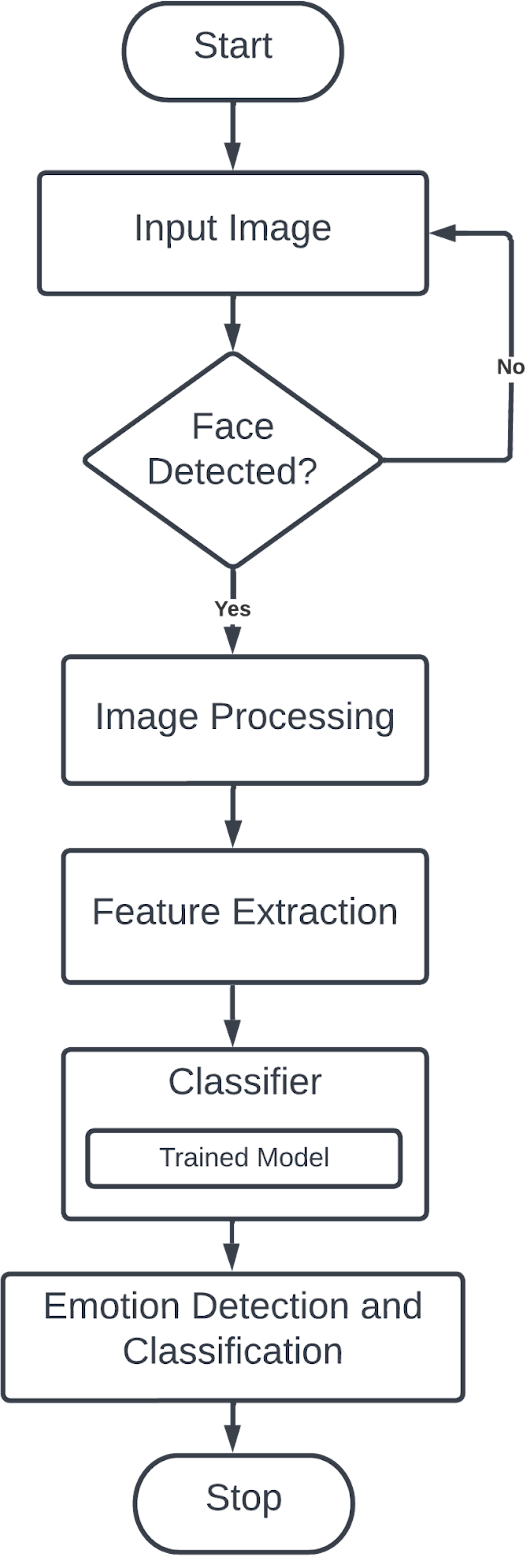


Figure 3: Model Testing Flowchart

## **Data collection**

For our research project, we collected different public datasets available on the internet including our own self-created dataset to train our model for facial emotion detection. We collected public image datasets from Kaggle and we downloaded images from Google, Facebook, etc., and requested our friends and requested different facial emotion images databases for access to the databases.

### **FER-2013 dataset**

This dataset, generated by Google in 2013, comprises of 48x48 pixel facial photos. The faces in the images have been automatically aligned, ensuring they are centered and occupy approximately the same amount of space in each image. This pre-processing step standardizes the dataset, making it more suitable for training the machine learning model. Below are some sample images from the FER-2013 dataset, showcasing different facial expressions. These images are representative of the dataset used for emotion recognition:



Figure 4: Image Example of FER-2013 Dataset

The FER-2013 dataset is a popular tool in the field of facial emotion recognition (FER), comprising 28,709 labelled images in the training set and 3,589 labelled images in the test set. Every image is classified into one of seven emotion categories: happy, sad, angry, fearful, surprise, disgust, or neutral, which aids in linking facial expressions to emotional states. The dataset features grayscale images, removing colour detail to concentrate on the facial characteristics that convey emotions, such as shapes and contours. It includes both posed and unposed headshots, with posed headshots showcasing individuals intentionally displaying facial expressions, while unposed images capture genuine, spontaneous expressions. This diversity enhances the dataset's realism for training models that need to accommodate various types of facial emotions. Furthermore, the "happy" category is the most prevalent, yielding a baseline accuracy of 24.4% through random guessing, indicating that a model that predicts "happy" for every image would reach this accuracy level. The dataset is divided into a training set (28,709 images) for developing the model and a test set (3,589 images) for assessing the model's performance on unseen data. In summary, FER-2013 serves as a crucial dataset for building emotion recognition systems by offering an extensive and varied collection of facial images representing different emotional states.

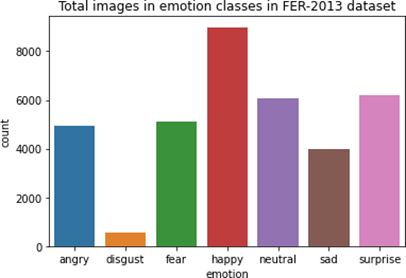


Figure 5: Image Distribution in FER-2013 Dataset

### **Self-Created dataset**

For the project, we created our own dataset by collecting images from our friends, downloading images from google, by requesting some databases like KDEF databases, etc. We have collected about 4336 images including all 7 emotions where all emotions contain about an equal number of images. The collected images are cropped in the face area using Haar-Cascade algorithm and removed most of the background part which is noise or outliers for the feature extraction. In our self-created dataset, there are RGB images that will be converted to grayscale for feature extraction and model training. This is the balanced dataset that contained almost equal data image for each of the emotion class.

Some of the images of the self-created dataset are shown below:

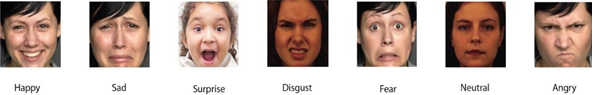


Figure 6: Images in Our Self-Created Dataset

Below chart shows the image distribution present in our self-created dataset which almost contains almost equal images for every emotion class.

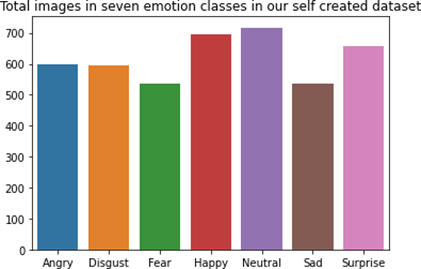


Figure 7: Images Distribution in Self-Created Dataset

Table 1: Dataset Collection and Description

|  |  |  |
| --- | --- | --- |
| Datasets | Sample Details | Available Description |
| FER-2013 | * Training set and Test set having about 28,709 and 3,589 images in each set * 48x48 pixels images * Contains 7 emotions images * Imbalanced dataset | “Google created this dataset in 2013 by collecting the results of an image search for each emotion and their synonyms.” |
| Self-Created Dataset | * Total of 4336 images * Contains only the face area of humans | “The dataset is created by ourselves that contains images from our friends, images from google and different facial databases.” |

## **Data Pre-processing:**

For this study, the model was trained and evaluated using two datasets: the FER-2013 dataset and a custom-created dataset. During the training process, 80% of the images were used for training the model, while the remaining 20% were reserved for testing. This split allowed for effective training of the model on the majority of the data, while the testing set provided an opportunity to assess the model's performance on unseen images, ensuring its ability to generalize to new data.

### **FER-2013 Dataset Challenges and Modifications**

The FER-2013 dataset is widely recognized for its imbalanced class distribution and various quality issues. For example, in the training set, the "happy" emotion class contains approximately 7,215 images, whereas the "disgust" emotion class has only 436 images. Additionally, the dataset exhibits significant challenges such as intra-class variation, occlusions, variations in contrast, and the presence of eyeglasses. There are also outliers, misaligned images, and incorrectly labelled samples. Some images lack faces entirely or are misaligned, complicating the training process.

To address these issues, the dataset was augmented by duplicating it and incorporating images from the testing set into the training set. To maintain balance, an equal number of images from other emotion classes were randomly selected and added to the training set. This process helped mitigate the imbalance between classes and provided the model with a more diverse set of training samples.

* + 1. **Custom Dataset and Image Normalization**

Both the FER-2013 dataset and the custom-created dataset underwent pre-processing to standardize the facial images. In this step, all images were resized to a uniform resolution of 48x48 pixels to maintain consistency across both datasets. Facial regions were then detected and cropped using the Haar Cascade Classifier, specifically the haar\_cascade\_frontalface\_default.xml file, which is a commonly used algorithm for frontal face detection. This pre-processing ensured that the model would focus on the relevant facial features and enable efficient learning from the data.

The face detection algorithm automatically identified facial regions, generating information about the position, width, and height of the detected faces. Based on these results, the images were cropped to focus solely on the facial area. These cropped and normalized images were then used for both training and testing the model.

Figure 8: Original Image Figure 9: Cropped Image

The below code shows the code to crop the image in required position.

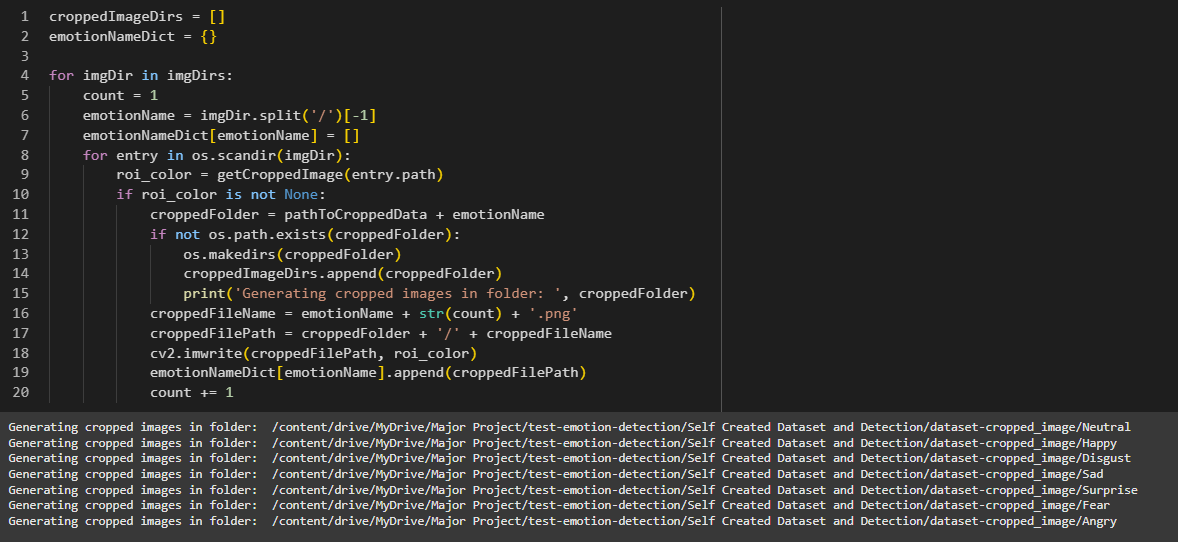


Figure 10: Code Snippet to Crop the Image to get Face Data

By addressing dataset imbalances and applying consistent pre-processing techniques, the study aimed to enhance the robustness and accuracy of the emotion recognition model.

## **Modules Used in Our System**

Our study is focused on detecting and classifying the facial emotion seen in user’s face. So, we have divided our project into three modules: Face Detection module, Feature Extraction Module and Emotion Classification module.

### **Face Detection Module**

Detecting and recognizing facial emotions is a complicated task aimed at identifying human feelings through facial expressions. The system utilizes a supervised learning method, examining images that display various emotional expressions, including anger, disgust, fear, happiness, neutral, sadness, and surprise. This process consists of several essential phases, including training and testing, along with important steps such as image acquisition, face detection, image preprocessing, feature extraction, and emotion classification. In the face detection phase, the system pinpoints and isolates the facial areas within the images, while the feature extraction phase concentrates on capturing significant facial characteristics. The features obtained are then utilized to categorize the emotions into seven separate types. The images may be either still photographs or dynamic video sequences, captured using devices like cameras or webcams, depending on the specific use case.

**Face Detection**

Face detection is the critical first step in the emotion recognition process. As part of computer vision, it involves creating algorithms that enable computers to interpret and analyze visual data. The primary task of face detection is to accurately locate facial areas within images or video frames while minimizing distractions from unnecessary elements such as backgrounds or external noise. This accuracy is essential for ensuring the effectiveness of subsequent processes, such as emotion classification.

One of the significant strengths of face detection systems is their ability to function in real-time. This capability allows them to process both static images and video frames dynamically, making them suitable for various applications, including live video monitoring, surveillance, and interactive systems. The effectiveness of face detection depends on classifiers—specialized algorithms that determine whether a specific part of an image contains a face. These classifiers are trained on large datasets of facial images to improve accuracy and reliability, ensuring consistent performance across different conditions.

**Haar Cascade Classifier for Face Detection**

In this project, the Haar Cascade Classifier is employed for face detection. This classifier is pre-trained with an extensive set of facial data, enabling it to precisely identify facial regions. Using a machine learning approach, the Haar Cascade Classifier is trained with a cascade function and multiple input files to refine its detection capabilities. It isolates facial features while filtering out irrelevant elements, ensuring the detection process is both efficient and accurate.

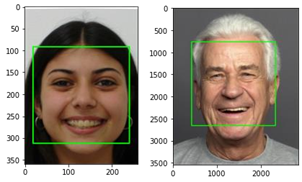


Figure 11: Face Detection Using Haar-Cascade Classifier Algorithm

The figure below shows the code to detect the face which is going to be cropped and processed for feature extraction.

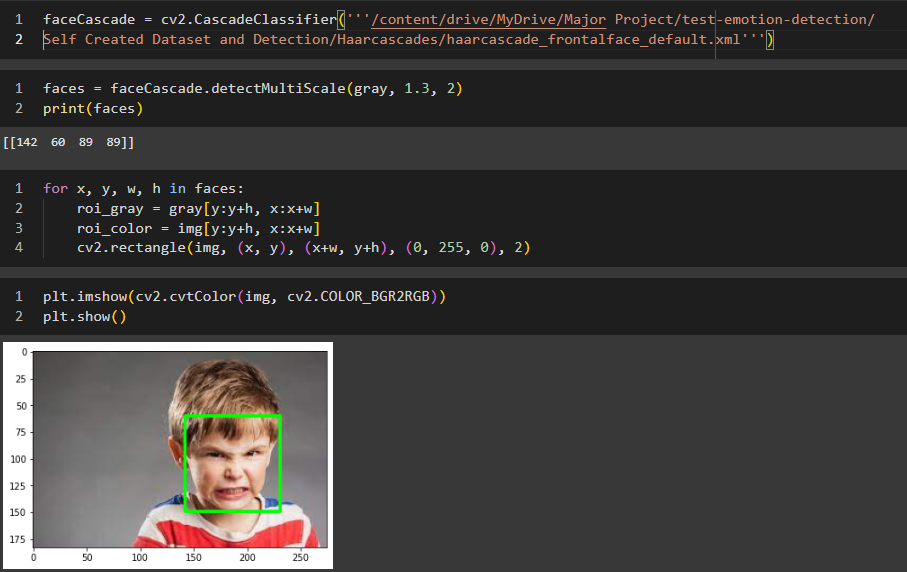


Figure 12: Code Snippet Showing Face Detection

As previously noted, the Haar Cascade Classifier is employed to identify faces within an image, after which the recognized face is extracted for additional analysis. Pre-processing of images is vital for enhancing data quality by eliminating noise and standardizing differences in pixel placement or brightness. This involves methods such as color normalization to maintain uniformity across images, which is crucial for precise emotion classification.

The overall system integrates face detection and emotion classification in a seamless manner. Once the face is detected, the relevant facial features are extracted from the region of interest and analyzed to classify the emotion into one of the seven predefined categories: anger, disgust, fear, happiness, neutral, sadness, and surprise. The system’s versatility is enhanced by its ability to process both static images and dynamic video data, making it adaptable to real-world scenarios where emotions need to be detected in various contexts. By combining face detection, feature extraction, and classification techniques, this system effectively bridges the gap between visual data and emotional interpretation, offering valuable insights in applications such as human-computer interaction, healthcare, and customer service.

In summary, facial emotion detection and recognition rely heavily on accurate face detection as a foundational step. The use of technologies like the Haar Cascade Classifier ensures precision and efficiency, making this system an essential application of computer vision in understanding human emotions.

### **Feature Extraction Module**

In a pattern classification problem, one of the most crucial steps is feature vector selection. After preprocessing the image of the face, the relevant features must be extracted to build an effective classification model. Feature extraction plays a significant role in ensuring the system can accurately recognize patterns in the image data. Several challenges must be overcome during this process, including variations in scale, pose, translation, and changes in illumination levels. These factors can significantly affect the quality and accuracy of the features extracted from the images.

In our project, we focus on using grayscale pixel values as the primary feature extractor. This approach is both simple and effective, particularly in scenarios where computational efficiency is important.

* + 1. **Grayscale Pixel Value as Feature**

Grayscale pixel values are used as the feature extractor in our system. Unlike an RGB or BGR image, which contains three color channels (Red, Green, and Blue), a grayscale image consists of just a single channel representing intensity levels. This simplicity makes grayscale images easier to process and analyze, as they reduce the complexity involved with multiple color channels.

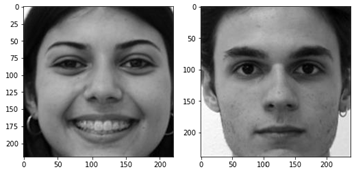


Figure 13: Grayscale Image

Once the image is converted to grayscale, it becomes a 2D matrix where each element represents the intensity of a pixel. The pixel values themselves serve as features for the classification process. Using grayscale values as features is particularly beneficial because it directly leverages raw pixel data, making it one of the simplest and most efficient methods for feature extraction.

If the image is 48x48 pixels in size, the resulting grayscale image will have a total of 4608 features (since 48 x 48 = 2304 pixels, and each pixel has a single grayscale value). These features are then used by the classification algorithm to identify patterns and make predictions about the emotion displayed in the image.

To convert an RGB or BGR image into grayscale, we use a function from OpenCV, a widely used computer vision library. The conversion is performed using the following command:



Figure 14: Code to Change BGR image to Grayscale

Below figure shows the code setup example used in our project:



Figure 15: Code Snippet to Convert RGB Image to Grayscale

### **Emotion Classification Module (Model Building)**

The emotion classification module is a critical component of the system, tasked with identifying and categorizing emotions based on facial features. Once the face detection process identifies a face in an image, the facial region is cropped and pre-processed to improve image quality. This pre-processing step may involve noise removal, normalization, and other techniques to enhance the accuracy of subsequent analysis. Afterward, the system extracts relevant facial features, such as the position of the eyes, mouth, and other key landmarks, which are crucial for distinguishing between different emotions.

The features that have been extracted are then utilized by machine learning algorithms that are designed to identify and categorize a range of emotions, including anger, happiness, sadness, fear, surprise, disgust, and neutrality. The algorithm is developed by training it on a substantial collection of labeled facial images, which allows it to grasp the intricate patterns and variations associated with each emotion. After the model has undergone training, it is capable of accurately classifying the emotions displayed in new, previously unseen facial images, enabling the system to effectively detect and interpret human emotions based on visual information.

In this project, grayscale pixel values are used as the primary feature for classification. These pixel values provide a straightforward yet effective representation of the facial data, which helps in distinguishing between different emotions.

To classify emotions, we implemented three widely used machine learning algorithms:

* 1. **SVM**

SVM is a widely used pattern recognition algorithm. Based on statistical learning theory, SVM is a state-of-the-art machine learning method. Near-optimal class separation may be achieved with SVM. Based on the characteristics that are displayed, SVMs are trained to categorize facial expressions. Strong generalization performance is provided by SVMs, the maximum hyperplane classification technique that is based on the findings of statistical learning theory.

In order to effectively move non-linearly separable input data to a high-dimensional feature space where linear techniques can be used, kernel functions are utilized. Because SVMs exhibit strong classification accuracy even with little amounts of training data, they are especially well-suited to a dynamic, interactive approach to expression identification.

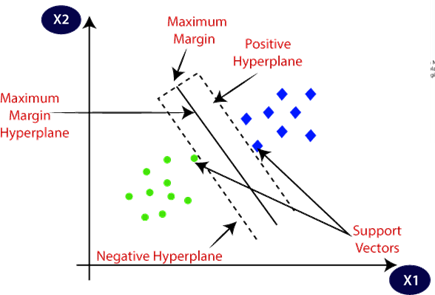


Figure 16: Graph for SVM (Sumbatilinda, 2024)

In a classification issue, an optimal division occurs when the hyperplane that separates the classes maximizes the margin. This indicates that the space between the hyperplane and the closest data points from either class is as extensive as can be. This hyperplane, referred to as the decision surface, acts as the boundary that segments the data points into various categories.

The Support Vector Machine (SVM) algorithm is commonly used to solve various classification problems, including face recognition, genetic analysis, and text categorization. In SVM, a training set of labeled samples is provided, and the algorithm identifies the optimal hyperplane that separates the classes in a high-dimensional space. By maximizing the margin, SVM ensures better generalization and robustness in classifying new, unseen data. This makes SVM a powerful tool in machine learning, particularly for tasks that require precise classification with clear class boundaries. An SVM (Support Vector Machine) tries to find the optimal hyperplane that best separates the different classes of data while minimizing classification errors. In an SVM (Support Vector Machine), for an input vector xi, the classification is determined by calculating the distance between the input vector and the hyperplane. The hyperplane is the decision boundary that separates the different classes. The distance is often referred to as the decision function, and it helps to determine on which side of the hyperplane the input vector lies, thereby classifying the input as belonging to one of the two classes.

Support Vector Machines (SVMs) are primarily designed as binary classifiers, which means they categorize data into two distinct classes. They operate by identifying the optimal hyperplane that maximizes the separation margin between the two classes, thereby ensuring effective classification. For any new input vector (xi), the sign of the decision function indicates the class assignment for that input vector. A positive distance means the input is classified into one class, whereas a negative distance signifies that it belongs to the other class.

For example, if the decision function is computed as:

Where:

* 𝑤 is the weight vector (normal to the hyperplane),
* xi is the input vector,
* b is the bias term

Then:

* If f(xi) > 0, the input vector xi belongs to class +1.
* If f(xi) < 0, the input vector xi belongs to class -1.
  + 1. **Random Forest Classifier**

Random Forest is a robust algorithm for supervised learning that integrates multiple decision trees to enhance performance. This technique is typically trained using a method known as bagging (Bootstrap Aggregating), which consists of training each decision tree on a random portion of the dataset. The fundamental concept of bagging is that merging numerous learning models (in this context, decision trees) can produce more precise predictions than relying on a single model. Each decision tree is trained independently on various random subsets of the data, and the overall prediction is obtained by either averaging the outcomes (in regression tasks) or determining the majority vote (in classification tasks) from all the trees.

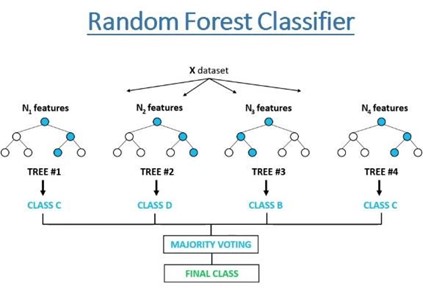


Figure 17: Graph for Random Forest Classifier (Khushaktov, 2023)

The Random Forest algorithm utilizes the bagging (Bootstrap Aggregating) technique to improve the performance of individual decision trees by combining their outputs. In bagging, the training data is randomly divided into multiple subsets, and each subset is used to train a separate decision tree. This random selection helps reduce overfitting and increases the model’s generalization ability. When constructing the trees in the random forest, each node is split based on a randomly selected subset of features rather than the most significant feature, which introduces variability and reduces correlation among the individual trees. This process enhances the overall performance of the random forest.

When a new data point is presented, it is processed through every tree in the ensemble. Each tree produces its own prediction, and the overall output is calculated by combining the predictions from all trees. In the case of regression tasks, this is usually achieved by taking the average of all tree predictions. For classification tasks, the final class label is decided by a majority vote, based on the most frequently predicted class from the individual trees.

The random selection of data samples and features at each node enables Random Forest to create a strong and precise model capable of managing complicated tasks like facial expression classification. By utilizing the capabilities of numerous decision trees and combining their results, Random Forests offer a viable approach for challenges involving high-dimensional and noisy data.

* + 1. **Logistic Regression**

Machine learning also employs logistic regression, a method from the field of statistics. Logistic regression is a technique that analyzes past data to forecast a binary outcome, such as yes or no, or to classify into different groups. The model predicts a dependent variable by assessing the relationship between one or more independent variables. It is the preferred method for solving binary classification problems.

The logistic function, the method's central component, is the source of the name logistic regression. Also known as the sigmoid function, the logistic function is provided by:

f (x) = 1/(1+e-x)

f(x) is the output of the logistic regression model and has the values in the range [0,1] where e is the base of the natural logarithms, x is the feature of the datasets. The graph for logistic regression is shown below:

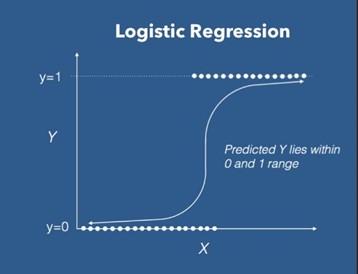


Figure 18: Graph for Logistic Regression

We used a dataset of face picture labels with accompanying emotions to train all three algorithms. The accuracy of each model's emotion classification was assessed. We chose the model with the greatest accuracy for emotion detection, identification, and classification after evaluating the models' performances.

By utilizing the advantages of machine learning methods, this stage guarantees that the system can reliably recognize emotions with accuracy and efficiency. The emotion classification module is an essential part of the entire emotion detection system, utilizing a mix of preprocessing, feature extraction, and model validation.

## **Hyperparameter Tuning**

Hyperparameter tuning is a parameter that is supplied as an argument to the estimator classes constructor (Anon., 2017). Hyperparameters direct the model's parameter selection.

### **Hyperparameter for SVM**

* **C:** inverse of regularization strength. Regularization is the process that constrains the size of the model coefficients. C is a floating-point number; it's 1.0 by default and we increase the regularization by making the number smaller.
* **Kernel:** Specifies the kernel type to be used in the algorithm. If none is given, ‘rbf’ will be used. Other kernels are {‘linear’, ‘poly’, ‘rbf’, ‘sigmoid’, ‘precomputed’}.
* **Gamma:** Kernel coefficient for ‘rbf’, ‘poly’ and ‘sigmoid’. Parameters are {‘scale’, ‘auto’}.
* **Probability:** Whether probability estimations should be enabled. This must be enabled before using fit; it will slow down the process because it utilizes 5-fold cross-validation internally, and predict\_proba might not match predict.

### **Hyperparameter for Random Forest**

* **n\_estimators:** The number of trees in the forest. Default value is 100.

### **Hyperparameter for Logistic Regression**

* **C:** inverse of regularization strength. Regularization is the process that constrains the size of the model coefficients. C is a floating-point number; it's 1.0 by default and we increase the regularization by making the number smaller.
* **Solver:** Algorithm to use in the optimization problem. Default is ‘lbfgs’. Other values are {‘newton-cg’, ‘lbfgs’, ‘liblinear’, ‘sag’, ‘saga’}.

## **System Evaluation**

Evaluation of the system is done using the following performance evaluation metrics:

* 1. **Precision**

Precision evaluates an algorithm's accuracy in making predictions by determining how reliable a given label, whether positive or negative, is in relation to the target class (Huilgol, 2024). It is defined as the ratio of correctly identified instances to the total number of predicted instances.

**precision = tp / (tp+fp)**

**Where:**

**tp = true positive**

**fp = false positive**

* 1. **Recall**

Recall depends on the number of correctly identified samples (true positives) as well as those that were incorrectly identified (false positives) and those that were missed (false negatives) (Huilgol, 2024). The ratio of accurately assigned expressions to the overall number of expressions is referred to as recall.

**recall = tp / (tp+fn)**

**Where:**

**tp = true positive**

**fp = false negative**

* 1. **F-score:**

The F-score is a combined metric that balances sensitivity and specificity, rewarding algorithms with higher sensitivity while challenging those with greater specificity. When β equals 1, the F-score treats precision and recall equally. A β value greater than 1 prioritizes precision, while a value less than 1 places more emphasis on recall.

**F-measure = ((2+1) \*precision\*recall)/(2\*precision\*recall)**

Precision, recall, and F-measure are essential metrics for assessing classification models, particularly when differentiating correct label assignments across various classes. These metrics concentrate on positive examples, each providing a distinct viewpoint on model effectiveness. Precision evaluates the correctness of positive predictions by determining the ratio of true positives to predicted positives, whereas recall measures the model's effectiveness in identifying all actual positive instances. Since these metrics emphasize different facets of performance, the F-measure (F1-score) is frequently utilized as an integrated metric. It calculates the harmonic mean of precision and recall, reconciling their trade-offs for a more thorough evaluation. It is crucial to understand that "recall" is occasionally incorrectly referred to as "accuracy," but accuracy usually describes the proportion of all correctly classified instances, including both positive and negative cases. The F-measure offers a more detailed perspective by factoring in both precision and recall when choosing the best model for expression category mappings.

# **Results, Analysis and Discussion**

## **Experiments and Results**

The experiments that were carried out, their outcomes, and the project's overall outputs are described in this section. The Face Detection Module and the Emotion Detection Module are the two primary components that make up the suggested system. The model is trained and tested by the Emotion Detection Module, which also determines which model is best for real-time emotion detection.

### **Experiments Based on Emotion Detection Module**

For recognizing facial expressions, we employed two datasets: FER-2013 and a tailored dataset created exclusively for this project. Our objective was to build a reliable system that can correctly identify facial emotions across different situations. To train and choose the best model, we used machine learning methods, particularly Support Vector Machine (SVM) and Random Forest, both of which are well-suited for classification challenges in computer vision.

Our approach followed supervised learning, where the system was trained on labeled data from both datasets. A key part of the process was the random sampling of data, ensuring the model learns to generalize well and avoids overfitting to specific examples. For training, 80% of the data from each dataset was used, while the remaining 20% was set aside for testing and validation to assess the model’s performance on unseen data.

To enhance the performance of both SVM and Random Forest models, we fine-tuned the hyperparameters of each algorithm. Hyperparameter tuning plays a critical role in improving the model's accuracy by adjusting the learning process to better suit the data. For instance, in SVM, the choice of kernel function, regularization parameter, and other settings can significantly influence the model’s ability to classify facial expressions. Likewise, for Random Forest, parameters such as the number of trees, tree depth, and splitting criteria were optimized to boost the model’s predictive power.

The ultimate goal of this module is to create a comprehensive, accurate system for recognizing facial emotions in real time. By leveraging, these machine learning techniques and optimizing the model, we aim to achieve high performance in emotion detection, making it effective across a variety of real-world applications.

* + - 1. **Experiment on FER-2013 Dataset**

As we already mentioned that this dataset is very imbalanced where happy emotion class have lots of data as compared to fear emotion class which made our model biased towards happy emotion class. So, we randomly took 500 images from each emotion class to make it balanced. We trained 80% data to make the machine learning model and used 20% data to test the model. The images in FER-2013 dataset are of 48x48 pixels and grayscale. So, for our study, we are using grayscale pixel values as the features.

After training the model and tuning the hyperparameters, we got the following accuracy and best hyperparameter using grayscale pixel values which are presented below:

Table 2: Model, Accuracy score and Hyperparameters (FER-2013)

|  |  |  |  |
| --- | --- | --- | --- |
|  | **Model** | **Best Score** | **Best Params** |
| 1 | SVM | 0.4281 | {‘C’:100, ‘kernel’: ‘rbf’} |
| 2 | Random Forest | 0.3260 | {‘n\_estimators’:10} |
| 3 | Logistic Regression | 0.3025 | {‘C’: 1} |

As from the above table, SVM has the best accuracy score. So, we have taken the SVM. Here, the best parameters for SVM are C = 100, kernel = rbf.

Table 3: Performance Evaluation Metrics of SVM on FER-2013 dataset

|  |  |
| --- | --- |
| **Evaluation Types** | **Result Percentages** |
| Precision | 42% |
| Recall | 43% |
| F1-score | 42% |
| Accuracy | 43% |

For the SVM model on FER-2013 dataset, we have precision of 42%, recall 43%, F1-score 42% and accuracy of 43%. The figure below shown is the confusion matrix heatmap that shows the true positive, true negative, false positive and false negative relationship between different emotion classes.

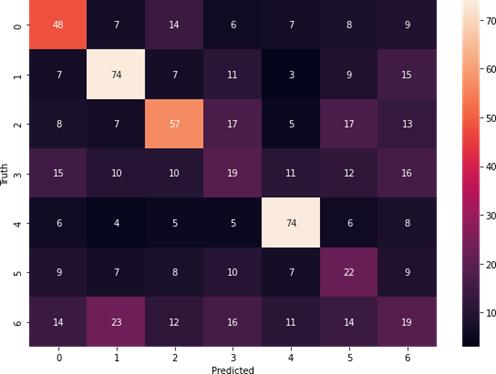


Figure 19: Confusion Matrix Heatmap using SVM model

Here, 0 = Sad, 1 = Surprise, 2 = Happy, 3 = Angry, 4 = Disgust, 5 = Neutral, 6 = Fear

* + - 1. **Experiment on Our Self-Created Dataset**

This dataset is created by ourselves where we have downloaded images from Google, Facebook, etc., requested images from our friends and requested database access to many facial databases. The images are cropped in the facial part using Haar-Cascade Classifiers and resized, gray scaled and feature extracted. In this dataset also, we used 80% of the images for training and 20% of the images for testing.

The model, accuracy and best hyperparameter using grayscale pixel values as features are presented below:

Table 4: Model, Accuracy Score and Hyperparameters (Self-Created Dataset)

|  |  |  |  |
| --- | --- | --- | --- |
|  | **Model** | **Best Score** | **Best Params** |
| 1 | SVM | 0.7769 | {‘C’:10, ‘kernel’: ‘rbf’} |
| 2 | Random Forest | 0.5965 | {‘n\_estimators’:10} |
| 3 | Logistic Regression | 0.6196 | {‘C’: 1} |

As from the above table, SVM has the best accuracy score. So, we have taken the SVM. Here, the best parameters for SVM are C=10, kernel = rbf.

Table 5: Performance Evaluation Metrix using SVM in Self-Created Dataset

|  |  |
| --- | --- |
| **Evaluation Types** | **Result Percentages** |
| Precision | 77% |
| Recall | 77% |
| F1-score | 77% |
| Accuracy | 77% |

For the SVM model on FER-2013 dataset, we have precision of 77%, recall, 77%, F1-score 77% and accuracy of 77%.

The figure below shown is the confusion matrix heatmap that shows the true positive, true negative, false positive and false negative relationship between different emotion classes.



Figure 20: Confusion Matrix Heatmap using SVM in Self-Created Dataset

Here, 0 = Fear, 1 = Sad, 2 = Angry, 3 = Happy, 4 = Surprise, 5 = Neutral, 6 = Disgust

## **Emotion Detection Module Output**

For our study, we proposed to build a Facial Emotion Detection System using Machine Learning, so for the system to run, we built a graphical user interface (GUI). The GUI built for the project is shown below:

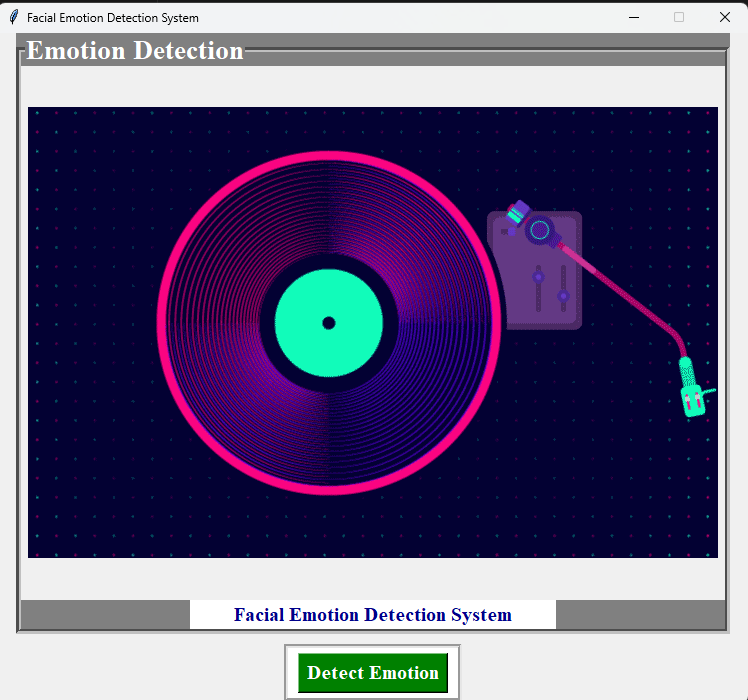


Figure 21: GUI for our Study

### **Emotion Detection Module Output**

When clicked on the “Detect Emotion” button present in the Button Section, the webcam is opened entitled as “Face Emotion Recognition”. The output of the emotion detection module is shown below:

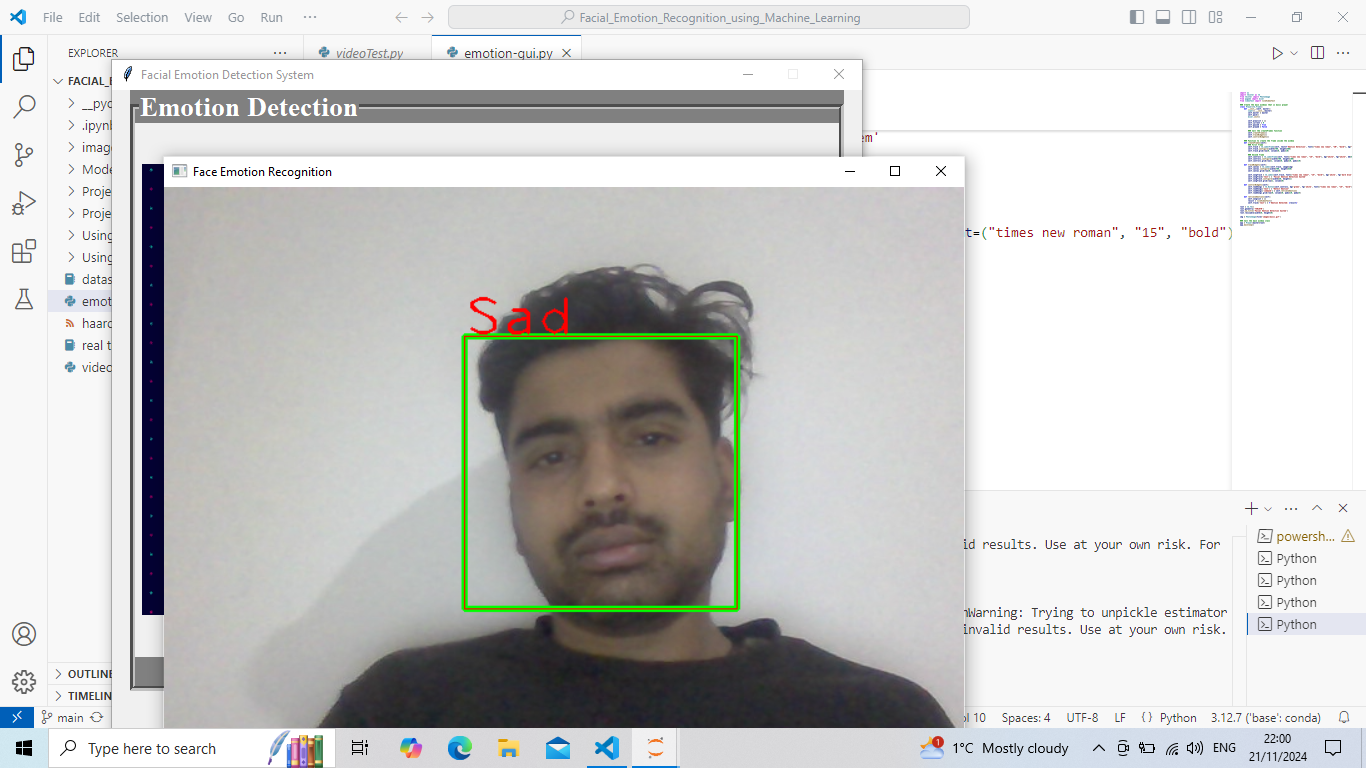


Figure 22: Emotion Recognition Using Webcam

## **Analysis of Experiment Results**

The objective of our research was to create a system capable of identifying and categorizing facial emotions. To accomplish this, we introduced and implemented two fundamental machine learning algorithms: SVM and Random Forest, utilizing two datasets—FER-2013 and a dataset we developed ourselves. The SVM model achieved an overall accuracy of 77% on our custom test dataset, surpassing the Random Forest model, which achieved an accuracy of 59%. The effectiveness of the model was further demonstrated through precision, recall, and F1-score metrics, with an average F1-score of 0.77 across all emotion categories. Nevertheless, both algorithms faced challenges with the FER-2013 dataset, which was imbalanced and contained a significant number of outliers and errors.

The confusion matrix highlighted high accuracy in detecting dominant emotions such as happiness and sadness, while recognition rates for emotions like fear and disgust were slightly lower. Data augmentation and transfer learning played a crucial role in reducing overfitting and improving the model’s ability to generalize across different facial expressions and conditions.

Although the outcomes are encouraging, applying this in real-world settings will necessitate tackling issues like inconsistent lighting and obstructions. Upcoming efforts will aim to enhance the model and broaden the dataset to reflect a wider range of demographic diversity, promoting fairness and reliability in practical applications. This organized strategy offers an in-depth evaluation of the model's effectiveness and highlights areas that require enhancement.

# **Discussion and Conclusion**

## **Discussion and Future Enhancement**

The findings underscored the capacity of machine learning techniques to effectively predict and categorize facial emotions. The significant accuracy obtained reflects the model's proficiency in learning and generalizing from intricate facial characteristics. Nevertheless, the diminished performance for specific emotion categories, such as fear and disgust, highlights the necessity for enhanced feature extraction and potentially the inclusion of contextual data to bolster classification in these domains.

Challenges related to dataset bias and real-world variability were partially mitigated through comprehensive preprocessing and data augmentation techniques. Nevertheless, future work could explore the integration of multimodal data, such as combining facial expressions with vocal tone or physiological signals, to further improve emotion recognition accuracy.

Facial Emotion Detection (FED) using machine learning represents a significant advancement in human-computer interaction, offering numerous applications across diverse fields. In this discussion, we delve into the implications, challenges, and future directions of this technology.

### **Implications and Applications**

FED systems have transformative potential in several domains. In healthcare, they can assist in diagnosing and monitoring mental health conditions by analyzing patients' emotional states. In education, these systems can provide real-time feedback on student engagement and comprehension, allowing educators to tailor their teaching methods. Additionally, in customer service and marketing, understanding customer emotions can enhance user experience and improve service delivery. Entertainment platforms can leverage FED to personalize content recommendations based on viewers’ emotional reactions, thereby increasing user engagement and satisfaction.

### **Challenges and Limitations**

Despite its potential, FED faces several challenges. One major issue is the variability in emotional expression across individuals and cultures. Emotions are often expressed differently based on cultural norms, and a system trained on a limited dataset may not generalize well across diverse populations. Moreover, real-world conditions such as varying lighting, occlusions (e.g., glasses, masks), and facial hair can impact the accuracy of emotion detection. Privacy concerns are also significant, as FED systems require capturing and analyzing facial data. Transparent policies regarding data privacy, consent, and usage should be established. Users must be informed about how their data is being used and given the option to opt out. Researchers and developers should also be mindful of biases in their models and strive to create fair and equitable systems., raising ethical questions about data collection, consent, and usage.

Also, it’s very difficult to develop a project without any limitations. Since our project being developed for academic purpose under certain budget and time frame, there are limitations. Sometimes, some of the emotions like sad and fear can’t be detected by our system. Also, the system can’t perform well in extremely bad light conditions and poor camera resolutions.

### **Future Enhancements**

To overcome current limitations, future research should focus on developing more robust and generalizable models. This includes collecting diverse, representative datasets that encompass a wide range of facial expressions, ethnicities, and age groups. Advances in transfer learning and data augmentation can help mitigate the challenges posed by limited data availability. Additionally, integrating multi-modal data, such as combining facial expressions with physiological signals (e.g., heart rate, voice tone), can enhance the accuracy and reliability of emotion detection.

Privacy and ethical considerations will become increasingly important as FED technology becomes more pervasive. Future systems must adopt privacy-preserving techniques like federated learning to protect user data while maintaining functionality. Additionally, ethical AI frameworks will be essential to ensure that FED technologies are deployed responsibly, minimizing biases and ensuring fairness in their applications.

The current system faces limitations when operating in challenging environments, such as extremely poor lighting conditions or with low-quality camera resolutions. These factors can significantly impact the accuracy and performance of the system, highlighting a potential area for improvement. In the future, additional features or algorithms could be integrated to address these issues, such as better image preprocessing techniques or adaptive lighting adjustments, which could help the system perform more reliably under these conditions.

Furthermore, to enhance user accessibility and convenience, transforming the existing graphical user interface (GUI) into a mobile application could provide a more seamless and flexible user experience. A mobile version would allow users to access the system on their smartphones, making it easier to use in real-time scenarios. This could broaden the system's potential use cases, enabling it to reach a larger audience and offering greater portability and ease of access compared to the current desktop-based interface.

## **Conclusion**

In conclusion, facial emotion recognition (FER) using machine learning represents a transformative technology with a wide array of applications across healthcare, security, education, marketing, and human-computer interaction. Basic machine learning algorithms are foundational in facial emotion recognition (FER) and often serve as starting points before transitioning to more complex deep learning models. Advances in deep learning, particularly with convolutional neural networks (CNNs), recurrent neural networks (RNNs), and transformer-based models, have substantially improved the accuracy and efficiency of FER systems. These models can now analyze subtle facial cues and recognize nuanced emotional expressions with impressive precision.

However, significant challenges remain. Environmental factors such as lighting variations, face occlusions, and diverse facial orientations introduce complexities that can degrade model accuracy. Additionally, cultural differences in emotional expression and mixed-emotion states make it difficult to categorize emotions universally. Ethical concerns, including privacy risks and the potential for emotional profiling, add further layers of complexity, necessitating careful consideration around the deployment of FER systems in real-world scenarios.

Looking forward, the field of FER could benefit from models that integrate multimodal data sources, such as vocal tone, physiological signals, and contextual data, to enhance emotional understanding. Research on generalizable and domain-adaptive models is also critical to improve FER performance across diverse demographics and settings. Moreover, as FER technologies continue to improve, establishing ethical frameworks and regulations for responsible use will be paramount in ensuring privacy and fairness. Ultimately, with ongoing advancements, FER can contribute meaningfully to more empathetic, responsive, and effective human-machine interactions, fostering a future where technology better understands and adapts to human emotions.

# **References**

Amjad, A. I. & Aslam, S., 2024. Beyond Borders: Examining Bullying, Social Networks, and Adolescent Mental Health in Developing Regions. *Sec. Mental Health and Wellbeing in Education,* 9(2024).

Anon., 2017. *Hyperparameter tuning for machine learning models..* [Online]   
Available at: https://www.jeremyjordan.me/hyperparameter-tuning/  
[Accessed 20 October 2024].

Athavle, M., Mudale, D., Srivastava, U. & Gupta, M., 2021. Music Recommendation Based on Face Emotion Recognition. *Journal of Informatics Electrical and Electronics Engineering,* 02(02), pp. 1-11.

Avata, D., Yaslan, Y. & Kamasak, M. E., 2018. Emotion Based Music Recommendation System USing Wearable Physiological Sensors. *IEEE Transactions on Consumer Electronics,* 14(8).

Bigelow, S. J., 2024. *What is NumPy? Explaining how it works in Python.* [Online]   
Available at: 15

Codecademy Team, 2024. *Introduction to Pandas and NumPy.* [Online]   
Available at: https://www.codecademy.com/article/introduction-to-numpy-and-pandas  
[Accessed 15 October 2024].

Davalikar, A. S. & Kulkarni, R. K., 2014. Face Detection and Facial Expression Recognition System. *International Conference on Electronics, Circuits and Systems (ICECS),* pp. 1-7.

Gera, D. & Balasubramanian, V., 2022. Emotion recognition using multimodal data and machine learning techniques: A survey. Information Fusion. Volume 77, pp. 45-64.

Guidel, A., Sapkota, B. & Sapkota, K., 2020. *Music Recommendation by Facial Analysis; Engineering Sarokar.* [Online]   
Available at: https://engineeringsarokar.com/music-recommendation-by-facial-analysis/  
[Accessed 25 July 2024].

Hameed, H. et al., 2024. RF sensing enabled tracking of human facial expressions using machine learning algorithms. *Scientific Reports 14,* 27800(2024).

Haque, S. B. U., 2024. A fuzzy-based frame transformation to mitigate the impact of adversarial attacks in deep learning-based real-time video surveillance systems. *Applied Soft Computing,,* 167(Part C).

Huang, Z.-Y.et al., 2023. A study on computer vision for facial emotion recognition. *SciRep,* 23 May, 8425(2023), pp. 1-13.

Huang, Z.-Y.et al., 2023. A study on computer vision for facial emotion recognition. *Sci Rep 14,* 8423(2023).

Huilgol, P., 2024. *Precision and Recall in Machine Learning.* [Online]   
Available at: https://www.analyticsvidhya.com/blog/2020/09/precision-recall-machine-learning/  
[Accessed 20 October 2024].

Immanuel, J. H. et al., 2019. Emotion Based Music Recommendation System. *International Research Journal of Engineering and Technology (IRJET),* 06(03), pp. 2096-2100.

Kavitha, R. & Vinodhini, G. A. F., 2024. Emotion recognition system using linear binary pattern algorithms and compare the accuracy with convolutional neural network. *2ND INTERNATIONAL INTERDISCIPLINARY SCIENTIFIC CONFERENCE ON GREEN ENERGY, ENVIRONMENTAL AND RENEWABLE ENERGY, ADVANCED MATERIALS, AND SUSTAINABLE DEVELOPMENT: ICGRMSD24,* 3193(1).

Khushaktov, M. F., 2023. *Introduction Random Forest Classification By Example.* [Online]   
Available at: https://medium.com/@mrmaster907/introduction-random-forest-classification-by-example-6983d95c7b91  
[Accessed 20 October 2024].

Kret, M. E., 2015. Emotional expressions beyond facial muscle actions. A call for studying autonomic signals and their impact on social perception. *Frontiers in Psychology,* 6(711).

Lee, Y.-H., Han, W. & Kim, Y., 2013. Emotional Recognition from Facial Expression Analysis Using Bezier Curve Fitting. *16th International Conference on Network-Based INformation Systems,* pp. 250-254.

Lehtiniemi, A. & Holm, J., 2012. Using Animated Mood Pictures in Music Recommendation. *16th Conference on Information Visualization,* pp. 143-150.

Manalu, H. V. & Rifai, A. P., 2024. Detection of human emotions through facial expressions using hybrid convolutional neural network-recurrent neural network algorithm. *Intelligent Systems with Applications,* 21(2024), pp. 1-18.

Mellouk, W. & Handouzi, W., 2020. Facial emotion recognition using deep learning: review and insights. *The 2nd International Workshop on the Future of Internet of Everything (FIoE),* 175(2020), pp. 689-694.

Sadhvika, C. H., Gupta, A. & Reddy, P. S., 2020. Emotion Based Music Recommendation System. *Journal of Emerging Technologies and InnoVative Research (JETIR),* 7(4).

Sumbatilinda, 2024. *Support Vector Machine (SVM) Algorithm..* [Online]   
Available at: https://medium.com/@sumbatilinda/support-vector-machine-svm-algorithm-064566b5d411  
[Accessed 20 October 2024].

Zeng, Z. et al., 2009. A survey of affect recognition methods: Audio, visual, and spontaneous expressions. *IEEE Transactions on Pattern Analysis and Machine Intelligence,* 31(1), pp. 39-58.

**Appendix**