**Facial Emotion Detection Using Machine Learning**

**ABSTRACT**

Facial emotion detection is a rapidly advancing field within artificial intelligence (AI) and computer vision, aimed at recognizing and classifying human emotions based on facial expressions. This paper presents a machine learning approach for detecting emotions from facial images using machine learning techniques which has emerged as a pivotal technology with wide-ranging applications. Our study introduces a novel approach to facial emotion detection, leveraging advanced machine learning techniques to interpret and analyse facial expressions. The primary objective of this research is to develop a robust system that accurately identifies human emotions through facial analysis, thereby providing valuable insights into the user's emotional state. Utilizing datasets like the FER2013, and self-created dataset, we explore different machine learning architectures, including Support Vector Machines (SVM), Random Forest (RF) and Logistic Regression. The proposed system captures images via a webcam, processes these images to extract key facial features, and then classifies the emotions using the most effective algorithms. The model achieves robust performance by capturing the subtle variations in facial expressions, thus enabling real-time emotion recognition in images and video streams. OpenCV is utilized for face detection and real-time emotion classification using live video feeds. The proposed solution can be applied to several domains, including human-computer interaction, mental health monitoring, and customer sentiment analysis. This research explores the challenges of emotion recognition, such as varying lighting conditions, occlusions, and the subjective nature of emotions.

**Keywords**

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

# **INTRODUCTION**

## **Background**

Facial emotion detection is an essential area in the intersection of artificial intelligence (AI) and human-computer interaction (HCI). Understanding human emotions through facial expressions enables machines to respond more intelligently and empathetically to user needs, enhancing the user experience across various applications (Zeng, et al., 2009). Emotions play a crucial role in communication, and facial expressions are often considered one of the most direct indicators of an individual’s emotional state. The mood of any human being can be detected by the facial emotion or expression expressed by him/her. Angry, disgust, fear, happy, neutral, sad, and surprise are all examples of human emotions. This umbrella of emotions can include a variety of additional emotions such as cheerful (a variation of joy) and contempt (a variation of disgust). These are very faint emotions. Facial muscle contortions are quite subtle, and recognizing them can be difficult because even minor variations result in distinct expressions (Kret, 2015). Also, because emotions are very context-dependent, different or even the same people's expressions of the same feeling may differ. While the focus may be on simply those portions of the face that express the most emotions, such as the mouth and eyes, how these gestures are extracted and classified remains a key question.

In the era of artificial intelligence, the ability of machines to understand and interpret human emotions has garnered significant attention. Facial emotion detection, a subset of affective computing, involves analysing facial expressions to identify underlying emotional states such as happiness, sadness, anger, fear, surprise, and disgust. This capability is crucial for enhancing human-computer interactions, improving mental health diagnostics, and tailoring user experiences across various industries. Traditional methods of emotion detection relied on manual observation and analysis, but with the advent of AI, particularly machine learning and deep learning, automated systems can now identify and interpret these emotions accurately and in real time (Gera & Balasubramanian, 2022). This capability opens up numerous opportunities in fields such as healthcare, education, security, and customer service. The potential applications of facial emotion detection are vast and varied. In the healthcare sector, emotion recognition systems can be used for mental health monitoring, detecting signs of depression or anxiety from facial cues. In customer service, businesses can use these systems to analyse customer satisfaction in real time based on facial reactions. Educational tools can also benefit, with systems that adapt content or teaching methods based on students' emotional engagement. For these purposes, neural networks and machine learning have been applied with good results.

Despite the remarkable progress in the field, facial emotion detection systems face significant challenges. Emotional expressions vary widely across cultural, gender, and age groups, introducing variability that complicates model accuracy. Real-world environments add further complexities, such as lighting variations, occlusions (e.g., glasses, masks), and dynamic backgrounds. Moreover, datasets used to train these systems are often imbalanced, with certain emotions underrepresented, leading to biases in detection outcomes. Addressing these issues requires comprehensive solutions, including the use of robust pre-processing techniques, balanced and diverse datasets, and state-of-the-art machine learning architectures.

This project aims to tackle these challenges by designing a robust system that detects emotions through facial expressions. Using a webcam interface, the system captures facial images and processes them through advanced image segmentation and feature extraction techniques. These features are then analyzed using machine learning models to accurately classify the emotions being expressed. By improving detection accuracy, the system can be applied in various contexts, from enhancing security systems to creating personalized user experiences. The integration of such technology into daily applications promises to revolutionize how humans and machines interact, 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 robust facial emotion detection system presents unique challenges. Variations in lighting conditions, facial orientations, and occlusions such as glasses, masks, or facial hair often hinder the system’s ability to detect emotions accurately. These real-world conditions must be accounted for during system development to ensure reliability. Additionally, emotions are conveyed through specific facial landmarks, such as the eyes, eyebrows, and mouth, which need to be detected and analysed precisely. The process typically involves several stages: locating the face within an image, extracting key features from facial landmarks, and applying machine learning algorithms to classify these features into predefined emotion categories. Models trained on extensive labelled datasets learn to recognize patterns in facial features, enabling them to distinguish between emotions like anger, joy, or surprise, even when differences are subtle. By employing advanced image processing techniques and machine learning, this project seeks to address these challenges and create a system capable of adapting to diverse environments and user profiles.

The proposed application will use state-of-the-art facial expression recognition algorithms to capture and analysed users' emotions effectively. The system begins by isolating the face from the input image or video and extracting key features associated with emotional expressions. These features are then fed into machine learning models, which classify the emotions into distinct categories. This technology has the potential to enhance various applications, including improving security systems by detecting suspicious behaviour, adapting educational content based on student engagement, and providing valuable feedback in therapeutic contexts. By accurately detecting emotions, the system can not only offer real-time insights but also enhance user experiences across industries, emphasizing the transformative role of emotional intelligence in artificial intelligence systems.

## **Research Aims/Objectives**

Facial emotion detection is a machine learning problem focused on recognizing and categorizing human emotions based on facial expressions. The primary objective is to develop a model capable of automatically identifying and classifying emotions such as happiness, sadness, anger, surprise, fear, disgust, and neutral expressions from images or video frames of human faces. This technology has significant applications in fields such as psychology, human-computer interaction, healthcare, and customer experience monitoring.

### **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 Data Mining 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) using machine learning lies in its potential to bridge the gap between human intuition and artificial intelligence, thereby enabling machines to interact with humans in a more intelligent, empathetic, and responsive manner. Facial expressions are a universal language of emotion, transcending linguistic and cultural barriers, and play a crucial role in human communication. By developing systems capable of understanding these expressions, we can improve the quality of interactions in various domains, such as healthcare, education, security, and customer service. Machine learning provides the computational foundation to analyse complex patterns in facial features, enabling automated systems to recognize emotions with precision 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.

Furthermore, the challenges associated with developing robust FER systems highlight the necessity of continued research. Variability in emotional expressions across cultures, genders, and age groups, coupled with real-world issues like lighting conditions, occlusions, and dynamic environments, complicates the accurate detection of emotions. Machine learning, with its ability to learn from large datasets and generalize patterns, offers a promising solution to these challenges. By advancing the development of FER systems, this research aims to address these limitations, making emotion recognition more reliable and applicable across diverse contexts. Ultimately, this study contributes to the broader goal of creating machines that are not only intelligent but also emotionally aware, fostering a new era of human-computer interaction.

## **Scope**

The scope of facial emotion detection using machine learning is vast and continues to expand as technology evolves. This area encompasses various applications across multiple domains, including healthcare, education, entertainment, and human-computer interaction. In healthcare, emotion detection can assist in monitoring patients' emotional states, enabling early intervention for mental health issues. In educational settings, it can provide real-time feedback to educators about student engagement and understanding, facilitating adaptive learning experiences. The entertainment industry benefits from this technology by personalizing content recommendations based on viewers' emotional responses, enhancing user engagement. Furthermore, facial emotion detection plays a critical role in the development of social robots and virtual assistants, enabling them to respond more naturally to human emotions, thus improving user experience. However, the scope also includes addressing challenges such as cultural differences in emotional expression, ethical considerations regarding privacy, and the need for robust algorithms that can perform accurately across diverse demographics and environments. As advancements in machine learning, particularly deep learning, continue to progress, the potential applications and effectiveness of facial emotion detection systems are likely to increase, paving the way for innovative solutions in both existing and emerging fields. 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**

Before starting the project, a feasibility study is carried out to measure the viability of the system. Feasibility study is necessary to determine if creating a new or improved system is friendly with the cost, benefits, operation, technology and time. Following feasibility study is given as below:

### **Technical Feasibility**

Technical feasibility is one of the first studies that must be conducted after the project has been identified. The technical feasibility study includes the hardware and software devices. The required technologies (Python, VS Code IDE, Jupyter Notebook, and Google colaboratory) existed.

### **Operational Feasibility**

Operational Feasibility is a metric that assesses how well a proposed system solves the problem and exploits the opportunities indicated during the scope definition. The following factors were taken into account when determining the project's technical feasibility:

* The system will recognize and capture a face image.
* The captured image is then (identified in which category).
* A music playlist is then provided based on the categorized emotion.

### **Economic Feasibility**

The purpose of economic feasibility is to determine the positive economic benefits that include quantification and identification. The system is economically feasible due to the availability of all requirements such as a collection of data from:

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

### **Schedule Feasibility**

The timeliness of a project is measured by its schedule feasibility. Because the system is constructed in such a way that it will complete within the specified time, it was deemed to be schedule viable.

## **Dissertation Structure**

The remainder of this dissertation is organized as follows: Chapter 1 provides the introductory information, problem statements, aims and objectives. Chapter 2 provides a comprehensive literature review of stock price prediction methodologies using different machine learning methodologies. Chapter 3 outlines the data collection and pre-processing procedures with detailed the research methodology, including algorithm selection, model development, and evaluation metrics. Chapter 4 presents the empirical results of the comparative analysis. Chapter 5 discusses the findings, implications, and limitations of the study. Finally, Chapter 6 summarizes the key contributions and suggests directions 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 first phase, face detection involves an RGB colour model analysis that includes ISO illumination processing to obtain a face and operations to maintain the required face details, such as the eyes and lips. The suggested algorithm also uses the Active Appearance Model (AAM) for extracting facial characteristics. This technique identifies several facial features, also known as Action Units (AUs), such as the eyes, eyebrows, mouth, and lips, and creates a file containing the properties of the observed action points. The facial emotions are then fed into the AAM model, which adjusts according to the expression. They used simple Euclidean Distance method where Euclidean distance between the feature points of the training images and that of the query image were compared. The true recognition rate for this study was 90% - 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 main goal is to develop techniques that can interpret and decode facial expressions for better emotion prediction by computers. The paper reviewed recent contributions to FER through deep learning, focusing on different methods and their effectiveness. It outlined the architectural approaches, the datasets used for training and validation, and the results achieved by various 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) employed a deep neural network (DNN) for facial emotion recognition (FER), specifically using a combination of Convolutional Neural Networks (CNN), Squeeze-and-Excitation Networks (SENet), and Residual Neural Networks (ResNet) for improved performance. The analysis revealed that facial features around the nose and mouth are key landmarks for the DNN model in recognizing emotions, highlighting the importance of these regions in FER tasks. The study utilized the AffectNet and Real-World Affective Faces Database (RAF-DB) for training and validating the CNN models. These datasets are widely used for emotion recognition and offer diverse facial expression samples. The model trained on AffectNet achieved 77.37% accuracy when validated on RAF-DB, demonstrating the generalizability of the model across different 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) studied facial expression (FER) problem on emotion recognition wearable dataset where they identified 9 emotions. The research made a hybrid CNN-RNN models for solving the problem which was developed through full learning and transfer learning using MobileNetV2-RNN and InceptionV3-RNN. The custom CNN-RNN model achieved an accuracy rate of 63%, while the MobileNetV2-RNN and InceptionV3-RNN transfer learning models yield 59% and 66%, respectively. The developed models demonstrate enhanced efficiency in distinguishing these nuanced emotions, a significant advancement in the field of facial expression recognition. This research holds substantial implications for cognitive science and real-world applications, particularly 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 consists of 400 images taken from 40 students (10 images per student) which were used to train the models, and the study involved iterations (N=10) to improve algorithm accuracy. The images were divided into two groups: Group 1 and Group 2 for model comparison. Convolutional Neural Network (CNN) achieved an accuracy of 83.47% for emotion recognition. Linear Binary Patterns (LBPs) achieved an accuracy of 82.15%. The accuracy difference suggests CNNs may offer slightly better performance in emotion recognition.

## **Inputs of Multimodal Inputs**

Yong-Hwan Lee, Woori Han, and Young Kim Lee (Lee, et al., 2013) proposed a system based on Bezier curve fitting. This algorithm uses a multiple-phase process for facial expression and emotions. The first phase involves detecting and processing the facial area from the original input image. The second step is verifying the emotion in the region of interest. Face detection in the first phase uses a colour still image based on skin colour pixel with initialized spatial filtering, accounting for lighting conditions. The Feature Map is then used to measure the locations of the eyes and lips, as well as the overall facial shape. After extracting the regions of interest, this method extracts points from the feature map to apply the Bezier curve to the eye and mouth techniques, studying the difference between the Hausdorff distance and the Bezier curve between the database image and the input face images. The size of the dataset isn't restricted. The algorithm shows high efficiency even on large datasets. The algorithm can also be used for 3D pictures; thus, a 3D picture can also be exploited for information but each point of study in the data has a global influence, and there are no outliers. This can cause overfitting, decreasing efficiency if the dataset is distorted due to overfitting.

The research done by (Huang, et al., 2023) used CNN, the combination of squeeze-and-excitation network and the residual neural network, for the task of FER. The research used AffectNet and the Real-World Affective Faces Database (RAF-DB) as the facial expression databases that provide learning samples for the CNN. The feature maps were extracted from the residual blocks for further analysis. The research concluded that the features around the nose and mouth are critical facial landmarks for the neural networks. Accorgin to paper, the network model trained on AffectNet achieved 77.37% accuracy when validated on the RAF-DB, while the network model pretrained on AffectNet and then transfer learned on the RAF-DB results in validation accuracy of 83.37%.

(Hameed, et al., 2024) proposed a novel and privacy-preserving human behaviour recognition system that utilizes Frequency Modulated Continuous Wave (FMCW) radar combined with Machine Learning (ML) techniques for classifying facial expressions. The study focused on five common facial expressions: Happy, Sad, Fear, Surprise, and Neutral. The recorded data was obtained in the form of a Micro-Doppler signal, and state-of-the-art ML models such as Super Learner, Linear Discriminant Analysis, Random Forest, K-Nearest Neighbour, Long Short-Term Memory, and Logistic Regression are employed to extract relevant features. These extracted features from the radar data were then fed into ML models for classification. The results showed a highly promising classification accuracy of 91%.

## **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. To define the encoding of audio and genre signals, this application uses text meta tags.This algorithm studied real-time data for analysis. The data was extracted from the reactions of the users, and hence, provides high accuracy.

(Avata, et al., 2018) proposed an emotion-based music recommendation system that learns the user's emotion from signals obtained through wearable computing devices that are integrated with galvanic skin response (GSR) and photoplethysmography (PPG) physiological sensors in their paper. Emotions are a basic part of human nature. They play a vital role throughout life. In this paper, the emotion recognition problem is taken into account as arousal and valence prediction from multi-channel physiological signals. The authors from (Guidel, et al., 2020) stated that a human's state of mind and current emotional mood can be easily observed through their facial expressions. This system was developed by considering basic emotions (happy, sad, anger, excitement, surprise, disgust, fear, and neutral). Face detection in this project was implemented using a convolutional neural network. The authors propose a framework that analyses users' facial expressions to determine their emotional state, which in turn informs personalized music recommendations. The paper discusses the integration of facial emotion recognition techniques, likely utilizing machine learning algorithms to classify emotions based on facial features captured via camera input. The system aims to enhance user experience by suggesting music that aligns with their current mood, thus fostering a more engaging and tailored interaction with music streaming platforms.

(Sadhvika, et al., 2020) advised that manual segregation of a playlist and annotation of songs, following the current emotional state of a user, is a labour-intensive and time-consuming task. Numerous algorithms have been proposed to automate this process. However, existing algorithms are often slow, increase the overall system cost by using additional hardware (e.g., EEG structures and sensors), and have less accuracy. The paper presents an algorithm that automatically generates a playlist of audio based on a person's facial expressions, saving time and labour in performing this task manually. The proposed algorithm aims to reduce the overall computational time and the cost of the designed system. A. Madhuri, M. Deepali, S. Upasana, and G. Megha (Athavle, et al., 2021) proposed a system for music recommendation based on facial emotion recognition, using the FER-2013 dataset and CNN for emotion detection. They aimed to reduce the computational time involved in obtaining results and the overall cost of the designed system. However, they used the FER-2013 dataset, which is highly unbalanced, with images that are not aligned, have contrast variations, 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 proposed system focuses on providing an interactive and user-friendly platform for detecting facial emotions. At its core is a graphical user interface (GUI) designed to enhance user interaction. This interface allows users to engage with the emotion detection system seamlessly. Using the GUI, users can activate the camera, which captures their facial images in real time. The system then isolates the facial region by cropping the captured image, extracts relevant features, and processes these features through a machine learning algorithm to classify and predict the displayed emotion. The primary objective of the system is to identify and detect the user’s facial emotions accurately, offering real-time feedback through the GUI. This approach makes the system intuitive and practical for real-world applications.

Our system operates through a structured workflow, starting with the acquisition and pre-processing of facial images. The datasets used include FER-2013, a publicly available dataset of facial expressions, along with a self-created dataset tailored to the study. These datasets contain labelled images corresponding to seven distinct emotional categories: Angry, Disgust, Fear, Happy, Sad, Surprise, and Neutral. Pre-processing involves resizing all images to a uniform dimension of 48x48 pixels, converting them to grayscale to simplify the data, and normalizing pixel values for consistency. Additionally, data augmentation techniques such as image rotation, flipping, and zooming are applied to increase the variability of the training data and improve the model's generalization capabilities.

Once pre-processed, the data is used to train the system to recognize patterns in facial expressions and map them to the corresponding emotions. After training, the system is equipped to process real-time video streams captured via OpenCV. A pre-trained face detector is employed to locate and isolate facial regions within the video frames. These regions are then classified by the trained machine learning model, which predicts the associated emotion. The system’s real-time processing capability makes it suitable for diverse practical applications. For example, in the healthcare industry, it can aid in understanding patient emotions for improved mental health assessments. In customer service, the system can evaluate customer satisfaction by analysing emotions during interactions. Additionally, it enhances human-computer interaction by enabling systems to adapt dynamically to users' emotional states, thereby improving overall user experience and decision-making processes.

## **Implementation Tools**

### **Programming Language and Libraries**

* 1. **Python**

Python is a powerful scripting language and is very useful for solving statistical problems involving machine learning algorithms. It has various utility functions which help in pre- processing. Processing is fast and it is supported on almost all platforms. Integration with C++ and other image libraries is very easy, and it has in-built functions and libraries to store and manipulate data of all types. It provides the pandas and NumPy framework which helps in the manipulation of data as per our needs. A good feature set can be created using the NumPy arrays which can have n-dimensional data.

* 1. **Scikit-learn**

Scikit-learn is the machine learning library in python. It includes matplotlib, NumPy, and a variety of machine learning algorithms. The API is very easy to use and understand. It has many functions to analyse and plot the data. A good feature set can be formed using many of its feature reduction, feature importance, and feature selection functions. The algorithm it provides can be used for classification and regression problems and their subtypes.

* 1. **Numpy**

NumPy is a popular Python library for processing large multi-dimensional arrays and matrices using a large collection of high-level mathematical functions (Bigelow, 2024). It is extremely useful for basic scientific computations in Machine Learning.

* 1. **Pandas**

Pandas is a well-known Python data analysis toolkit. It has nothing to do with Machine Learning. The dataset must be prepared before the training, as we all know. Pandas come in helpful in his scenario because it was designed expressly for data extraction and preparation (Codecademy Team, 2024).

* 1. **Matplotlib**

Matplotlib is a well-known Python data visualisation package. It's very useful when a coder needs to see how data patterns are represented. It's a 2D charting package that allows you to make 2D graphs and charts.

* 1. **OpenCV**

It is the library we will be using for image transformation functions such as converting the image to grayscale. It is an open-source library that can be used for many image functions and has a wide variety of algorithm implementations. C++ and Python are the languages supported by OpenCV. It is a complete package that can be used with other libraries to form a pipeline for any image extraction or detection framework. The range of functions it supports is enormous, and it also includes algorithms to extract feature descriptors.

## **Development Environment**

* + - 1. **Jupyter Notebook**

Jupyter Notebook is the integrated development environment (IDE) for combining Python with all of the libraries we'll be using in our solution. Although certain complex computations take longer to execute, it is interactive. Plots and photos appear in real-time. It may be utilised as a one-stop-shop for all of our needs, and most libraries, such as OpenCV and Scikit-learn, can be simply incorporated.

* + - 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's also hosted on Google's servers, so there's no need to download anything. The notebooks are also preserved in your Google Drive account.

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

Microsoft's Visual Studio Code is a source code editor for Windows, Linux, and macOS. Debugging, embedded Git control, GitHub, syntax highlighting, intelligent code completion, snippets, and code refactoring are all supported.

## **System Design:**

System design shows the overall design of the system. In this section, we discuss in detail the design aspects of the system.

### **System Diagram**

Our system design started with planning phase where the requirements like problem understanding, objectives, scope, data requirements were understood. After the data acquisition, the data was partitioned into two parts: training dataset and testing dataset which is followed by face detection method. After the face detection process, the features were extracted and trained the model using different machine learning classification algorithms and test set features were tested using the trained model which is then classified into different emotion class. Our system followed this design path. The figure below shows the flowchart of training and testing of the model**.**

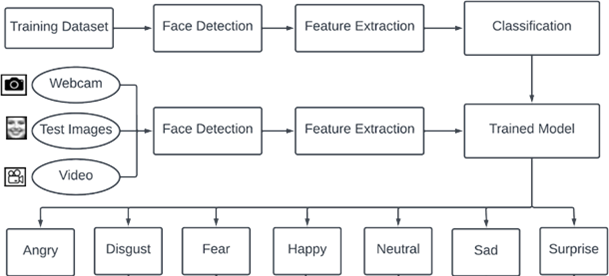
****

Figure 1: System Diagram

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

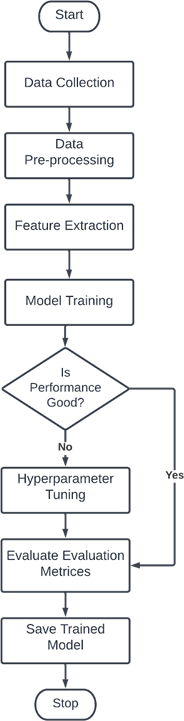


Figure 2: Model Training Flowchart

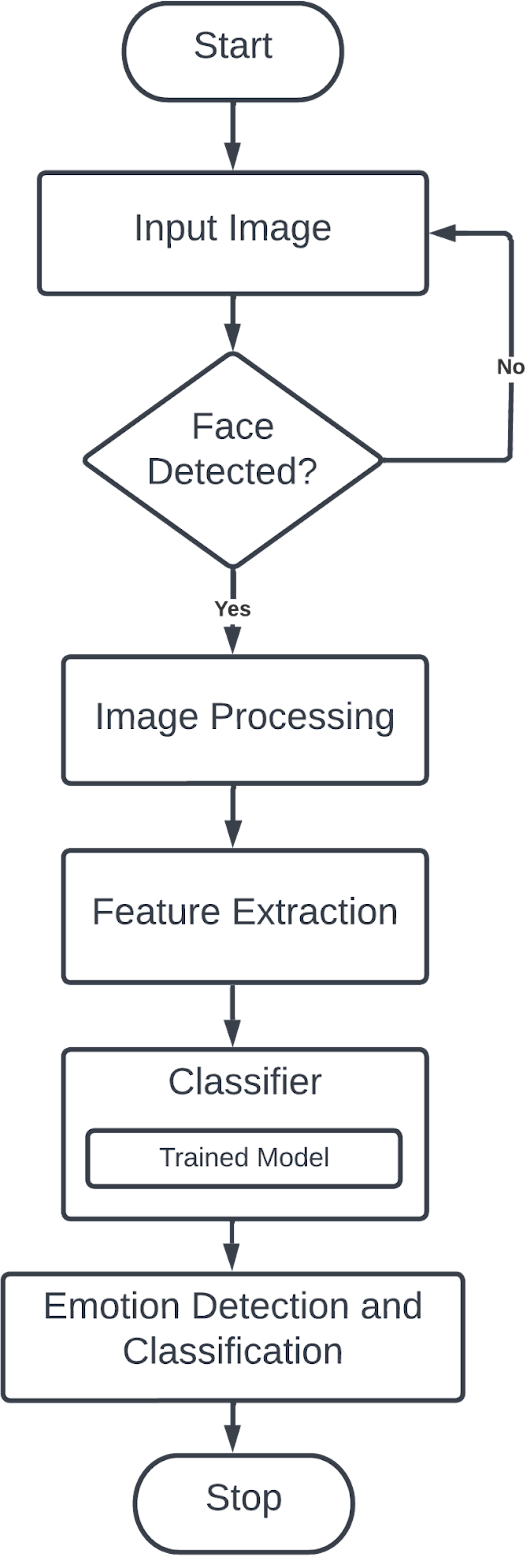


Figure 3: Model Testing Flowchart

## **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 was prepared by Google in 2013 and consists of face images of 48x48 pixels. The faces have been automatically registered so that the face is more or less centred and occupies about the same amount of space in each image. Some images of the FER-2013 dataset are presented below:



Figure 4: Image Example of FER-2013 Dataset

The FER-2013 dataset consists of 28,709 labelled images in the training set, and 3,589 labelled images in the test set. Each image in FER-2013 is labelled as one of seven emotions: happy, sad, angry, afraid, surprise, disgust, and neutral with happy being the most prevalent emotion, providing a baseline for random guessing of 24.4%. The images in FER-2013 consist of both posed and unposed headshots, which are in grayscale as shown in above figure.

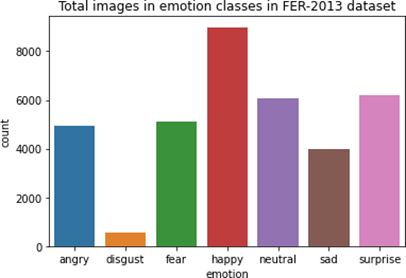


Figure 5: Image Distribution in FER-2013 Dataset

### **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:

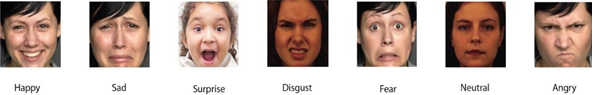


Figure 6: Images in Our Self-Created Dataset

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

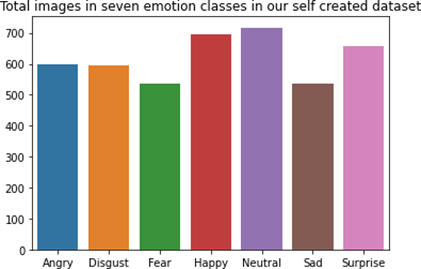


Figure 7: Images Distribution in Self-Created Dataset

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 | “This dataset was made by Google in 2013 by gathering the results of a Google \image search of every emotion and synonyms of the emotions.” |
| 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 our study, the model was trained and tested using two datasets named- FER-2013, and our own self-created dataset. Our model is trained on 80% of train images and tested on 20% of test images.

The FER-2013 dataset is very imbalanced because in the training set, happy emotion class has about 7215 images but disgust emotion class has only 436 images. It also has intra-class variation problems, occlusion, contrast variation, eye-glasses images, outliers, etc. Some samples do not contain faces and some images are not aligned and some of them are incorrectly labelled. We copied the same dataset and doubled them and added the testing set images to the training set and a same number of images of other emotion classes were randomly selected and made the train image set.

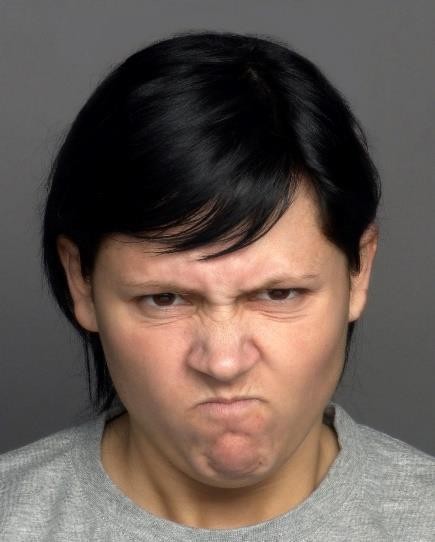
For our self-created dataset and the other datasets that is FER-2013, we normalized the faces to 48x48 pixels. Based on the structure of the face, facial images of different pixels were cropped by detecting the frontal face using the face detector algorithm which is Haar Cascade Classifier (haar\_cascade\_frontalface\_default.xml). The results of face detection including face location, face width, and face height were automatically created. Finally, images were cropped in accordance with the result given by the face detector, and further cropped images were used for training and testing.

Figure 8: Original Image Figure 9: Cropped Image

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

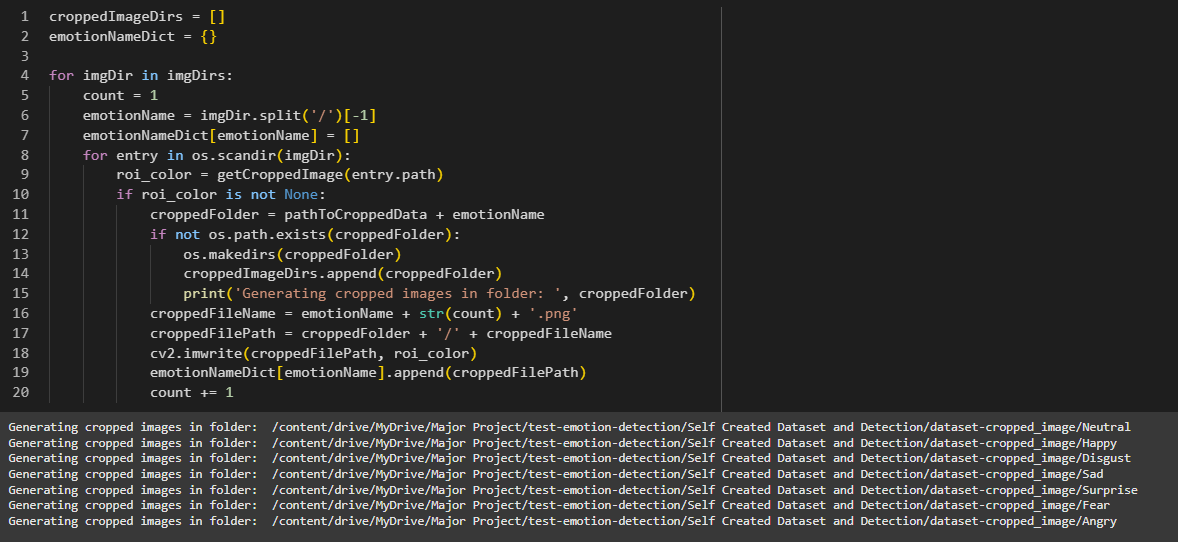


Figure 10: Code Snippet to Crop the Image to get Face Data

The detailed data preprocessing is described in the ‘Face Detection Module’ section below:

## **Modules Used in Our Proposed System**

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

### **Face Detection Module**

The facial emotion detection and recognition module is trained using a supervised learning approach in which it takes images of different facial expressions like anger, disgust, fear, happiness, neutral, sad, and surprise. The system includes the training and testing phase followed by image acquisition, face detection, image preprocessing, feature extraction, and emotion classification. Face detection and feature extraction are carried out from face images and classified into seven emotion classes.Images used for facial expression recognition are static images or image sequences. Images of faces can be captured using a camera or webcam.

* + - 1. **Face Detection and Cropping**

It is one of the applications that fall under the purview of computer vision technology. This is the process of developing and training algorithms to correctly locate faces or objects in object detection or related systems in images. Real-time detection from a video frame or photos is possible. Face detection employs classifiers, which are algorithms that determine if an image contains a face or not. To improve accuracy, classifiers are trained to recognize faces using a large number of images. In our project, we use the Haar Cascade Classifier which is trained with pre-defined varying face data, allowing it to accurately detect different faces. Face detection’s major goal is to reduce external noise and other elements in order to spot the face within the frame. The cascade function is trained with a group of input files using a machine learning approach.

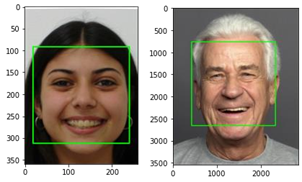


Figure 11: Face Detection Using Haar-Cascade Classifier Algorithm

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

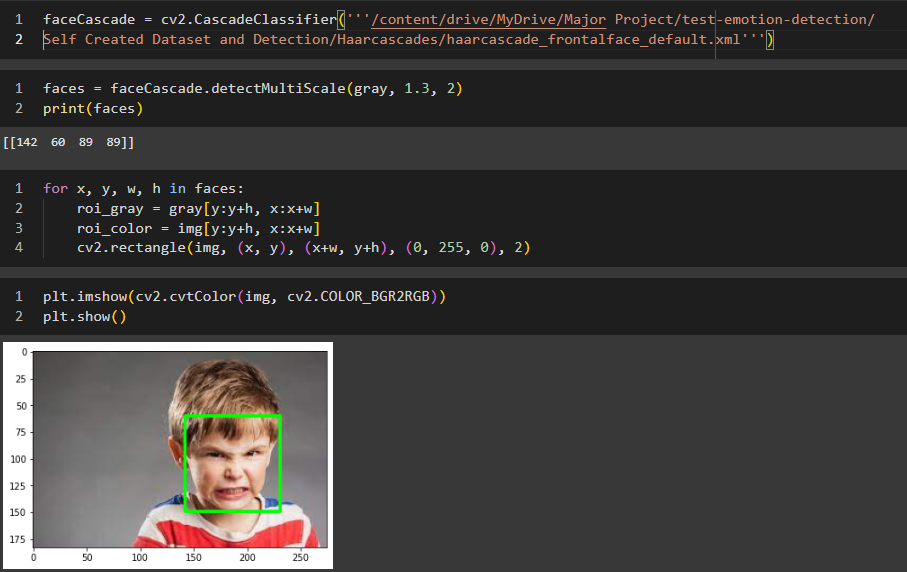


Figure 12: Code Snippet Showing Face Detection

As mentioned above, to detect the face from any image, Haar Cascade Classifier is used after which the face is cropped for further processing. Image pre-processing includes the removal of noise and normalization against the variation of pixel position or brightness by color normalization, etc.

* + - 1. **Feature Extraction**

In a pattern classification problem, feature vector selection is the most important step. After preprocessing the image of the face, the relevant features are extracted. Scale, pose, translation, and fluctuations in illumination level are all fundamental challenges in image classification. In our project, we use grayscale pixel value as a feature extractor.

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

We used grayscale pixel values as a feature. An RGB or BGR image has 3 channels but a grayscale image has a single channel. An RGB or BGR image can be converted to grayscale by using OpenCV (gray = cv2.cvtColor(image, cv2.COLOR\_RGB2GRAY)). That’s why it is the simplest way to create features from an image. The main focus is to use the raw pixel values as separate features. If the image is 48x48 pixels then it will have a total of 4608 features in grayscale.

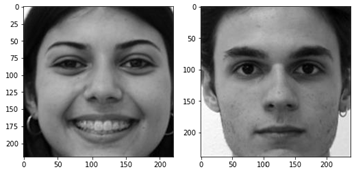


Figure 13: Grayscale Image

The figure below shows the code that changes the colorful image to grayscale image.



Figure 14: Code Snippet to Convert RGB Image to Grayscale

* + - 1. **Image Resizing and Normalization**

The extracted face images had varying dimensions, requiring them to be resized to a standardized size of 64x64 pixels. This resizing ensured consistency in image dimensions, making them compatible with the input requirements of a machine learning algorithms. Additionally, resizing reduced computational demands and memory usage, which is critical for efficient model training. After resizing, the pixel values were normalized to a range between 0 and 1. This normalization process accelerated the training phase by ensuring that all pixel values were on a uniform scale, minimizing bias toward higher-intensity pixels.

Standardizing the image dimensions was essential for training machine learning models, as they require fixed-size inputs. Similarly, normalization enhanced the training process by improving model convergence, as models tend to perform better when input features share a consistent scale.

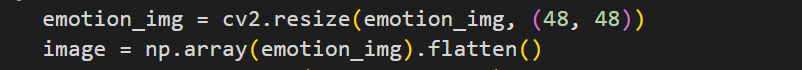


Figure 15: image Resizing and Normalization

### **Emotion Classification Module**

After the face is detected and the facial part is cropped and pre-processed then the facial feature is extracted. Then the extracted feature is fed to the machine learning algorithm to obtain the trained model to detect, recognize and classify the emotions that are present in the facial image. The classification is based on the feature that is given by grayscale pixel value as a feature. For emotion detection, recognition, and classification, we used three machine-learning algorithms: Support Vector Machine, Logistic Regression and Random Forest Classifier. We trained all three algorithms and took the model that had the best accuracy.

**SVM**

SVM is a pattern recognition algorithm that is commonly used. SVM is a cutting-edge machine learning technique based on statistical learning theory. SVM is capable of achieving a near-optimal separation of classes. SVMs are trained to classify face expressions based on the features presented. SVMs are the maximal hyperplane classification method that relies on statistical learning theory results to provide strong generalization performance.

Kernel functions are used to efficiently transfer non-linearly separable input data to a high- dimensional feature space where linear methods can be applied. SVMs are particularly well suited to a dynamic, interactive approach to expression recognition since they display good classification accuracy even when just a little quantity of training data is given.

When the hyperplane and the training data of any class are the largest, an optimal separation is attained. The decision surface is this dividing hyperplane. SVM has been used to solve a variety of classification problems, including text categorization, genetic analysis, and face detection. Given a training set of labelled samples:

D = {(xi , yi | xi ϵ Rn, yi ϵ {-1,1}}p

A SVM tries to find a hyperplane to distinguish the samples with the smallest errors.

w.x – b = 0

For an input vector xi, the classification is achieved by computing the distance from the input vector to the hyperplane. The original SVM is a binary classifier.

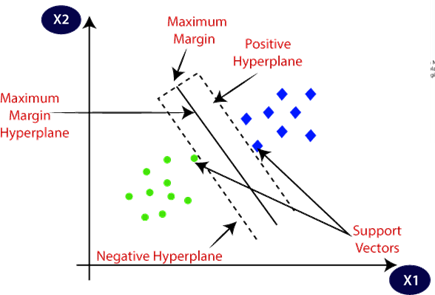


Figure 16: Graph for SVM (Sumbatilinda, 2024)

* + 1. **Random Forest Classifier**

Random forest is a supervised learning algorithm. It produces a "forest" out of a collection of decision trees that are often trained using the "bagging" method. The main idea of the bagging method is that combining many learning models enhances total output. The random forest has the advantage of being able to address classification and regression problems, which are common in machine learning systems nowadays.

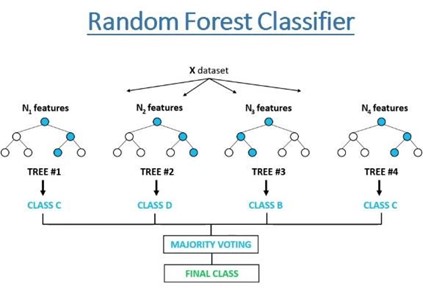


Figure 17: Graph for Random Forest Classifier (Khushaktov, 2023)

The bagging approach is used to train Random Forests. Bagging, also known as Bootstrap Aggregating, entails picking subsets of the training data at random, fitting a model to these smaller data sets, then aggregating the results. Random forest is trained to perform facial expression classification using the features proposed. The random forest adds more randomness to the model as it grows the trees. Instead of looking for the most important feature when dividing a node, it seeks the best feature from a random group of features. As a result, there is a great deal of variability, resulting in a better model.

Each new data point in the Random Forests algorithm goes through the same procedure as before, except this time it visits all of the trees in the ensemble, which was constructed using random samples of both training data and features. The aggregation functions utilized will vary depending on the task at hand. It employs the mode or most frequent class predicted by individual trees (also known as a majority vote) for classification issues, and the average prediction of each tree for regression tasks.

* + 1. **Logistic Regression**

Logistic regression is another technique borrowed by machine learning from the field of machine learning. Logistic regression is a statistical analysis method to predict a binary outcome, such as yes or no, or classify different classes, based on prior observations of a data set. A logistic regression model analyzes the relationship between one or more existing independent variables to predict a dependent data variable. It's the method of choice for binary classification issues.

Logistic regression is named for the function used at the core of the method, the logistic function. The logistic function, also called the sigmoid function is given 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:

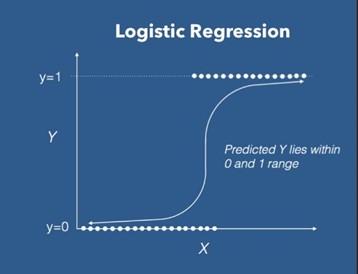


Figure 18: Graph for Logistic Regression

## **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 to enable probability estimates. This must be enabled prior to calling fit, will slow down that method as it internally uses 5-fold cross- validation, and predict\_proba may be inconsistent with 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 the algorithm's predictive capacity by estimating the predictive value of a label, which can be positive or negative depending on the class for which it is derived (Huilgol, 2024). The percentage of correctly assigned expressions in relation to the total number of aspects is known as precision.

precision = tp / (tp+fp)

Where,

tp = true positive

fp = false positive

* 1. **Recall**

Recall is a function of both successfully classified (true positives) and misclassified (false positives) samples (false negatives) (Huilgol, 2024). The fraction of correctly assigned expressions to the total number of expressions is known as recall.

recall = tp / (tp+fn)

Where,

tp = true positive

fp = false negative

* 1. **F-score:**

The F-score is a composite metric that rewards algorithms with better sensitivity while challenging those with higher specificity. When β = 1, the F-score is evenly balanced. When β > 1, it favors precision; otherwise, it favors recall.

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

All three measures distinguish between correct label classification within various classes. They concentrate on a single subject (positive examples). As a result, accuracy and recall assess different properties, and we require a combined quality measure to figure out which aspect of expression category mappings is the best match. The F-measure (fm) calculates the harmonic mean of precision and recall and allows both attributes to be considered at the same time. It's worth noting that the total recall is also referred to as accuracy.

# **Results, Analysis and Discussion**

## **Experiments and Results**

This section gives the details of the experiments performed, their results and the obtained output of the whole project. Our proposed system consists of two modules: Face Detection Module and Emotion Detection Module. Emotion detection module focuses on training and testing the model and selecting the best model for real time emotion detection.

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

For facial emotion detection, we used two datasets FER-2013 and a self- created dataset. We used three machine learning algorithms SVM, Logistic Regression and Random Forest to train our model and select the best model for our project. The aim of this module is to develop a complete facial emotion recognition system. First of all, the system was trained using different random samples in each dataset by supervised learning. In each dataset, the data were partitioned into two parts for training and testing where 80% for training and 20% for test. For betterment and increasing the accuracy of the model, we have tuned hyperparameters for each algorithm.

* + - 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 and 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.

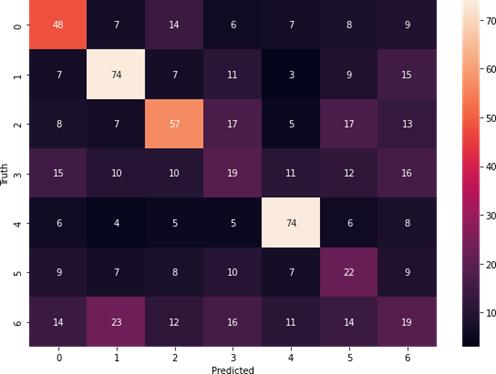


Figure 19: Confusion Matrix Heatmap using SVM model

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:

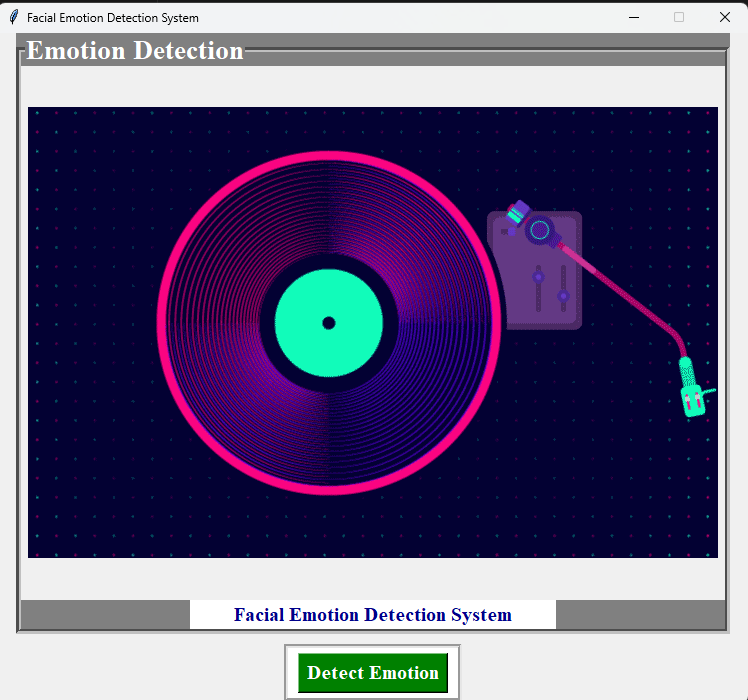


Figure 21: GUI for our Study

### **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:

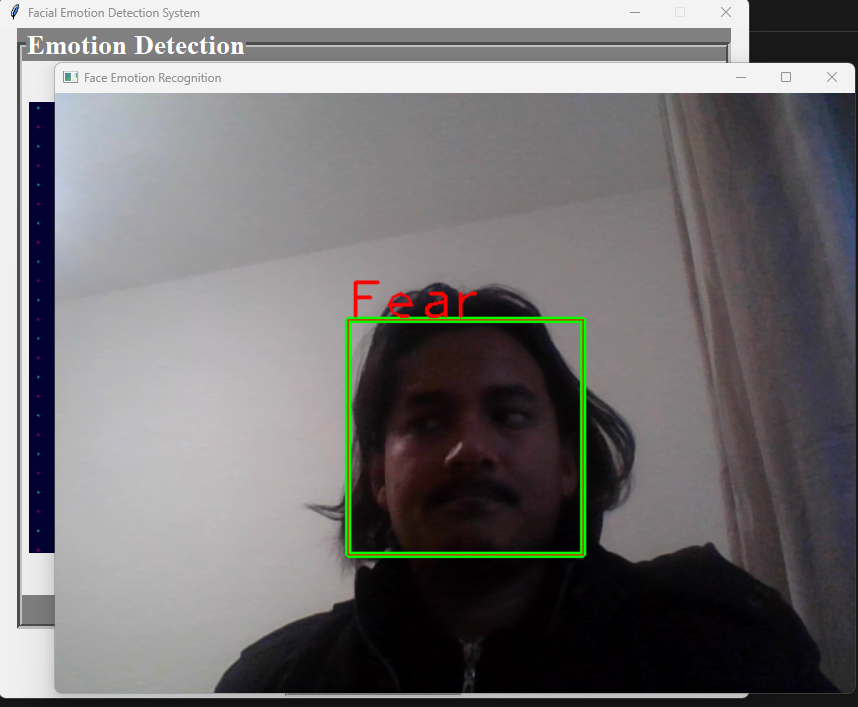


Figure 22: Emotion Recognition Using Webcam

## **Analysis of Experiment Results**

As the objective of our study is to build a system that detect the facial emotion and classify it. So, for the study we proposed and used two basic machine learning algorithms: SVM and Random Forest in our two datasets: FER-2013 and our own self-created dataset. The proposed SVM model achieved an overall accuracy of 77% on our self-created test dataset, outperforming random forest model, which recorded accuracies of 59%. Precision, recall, and F1-score metrics further corroborated the model's effectiveness, with an average F1-score of 0.77 across all emotion classes. But both the SVM and Random Forest algorithm could not perform well on FER-2013 dataset which was imbalanced, had lots of outliers and faults.

The confusion matrix revealed high accuracy in detecting predominant emotions like happiness and sadness, while emotions such as fear and disgust exhibited slightly lower recognition rates. Data augmentation and transfer learning contributed significantly to mitigating overfitting and enhancing the model's generalizability across diverse facial expressions and conditions.

While the results are promising, real-world deployment will require addressing challenges such as varied lighting conditions and occlusions. Future work will focus on refining the model and expanding the dataset to include a broader demographic representation, ensuring fairness and robustness in diverse real-world applications." This structured approach ensures a comprehensive and insightful analysis, providing a clear understanding of the model's performance and areas for improvement.

# **Discussion and Conclusion**

## **Discussion and Future Enhancement**

The results demonstrated the prediction and classifying capability of machine learning approaches in facial emotion detection tasks. The high accuracy achieved underscores the model's capability to effectively learn and generalize from complex facial features. However, the lower performance in certain emotion classes like fear and disgust highlights the need for more nuanced feature extraction and possibly the incorporation of contextual information to improve classification in these categories.

The challenges of dataset bias and real-world variability were partially addressed through comprehensive preprocessing and data augmentation techniques. Nonetheless, future work could explore the integration of multi-modal data, such as combining facial expressions with vocal tone or physiological signals, to enhance emotion recognition accuracy further.

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 does not perform well in extremely bad light conditions and poor camera resolution thereby provides an opportunity to add some functionality as a solution in the future. For a better user experience, we can transform this GUI into a mobile application in the future.

## **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.