A MAJOR PROJECT II REPORT

ON

MUSIC RECOMMENDATION SYSTEM EMPLOYING FACIAL EMOTION RECOGNITION

Submitted in Partial Fulfillment of the Requirements

for the Award of the Degree of

MASTER OF TECHNOLOGY

IN

COMPUTER SCIENCE & ENGINEERING

Submitted By:

KAMAKSHA

Roll No-2K22/CSE/10

Under the Supervision of

Dr. VINOD KUMAR

(HOD, Department of Computer Science & Engineering)

DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING

DELHI TECHNOLOGICAL UNIVERSITY

(Formerly Delhi College of Engineering)

Shahbad Daulatpur, Main Bawana Road, Delhi-110042. India

MAY,2024

DELHI TECHNOLOGICAL UNIVERSITY

(Formerly Delhi College of Engineering)
Shahbad Daulatpur, Main Bawana Road, Delhi-42

CANDIDATE DECLARATION

I **KAMAKSHA** Roll number 2K22/CSE/10 student of M.Tech (Computer Science and Engineering), hereby declare that the work which is being presented in the major project - II entitled "**MUSIC RECOMMENDATION SYSTEM EMPLOYING FACIAL EMOTION RECOGNITION** " which is submitted by me to the Department of Computer Science and Engineering, Delhi Technological University, Delhi in partial fulfilment of the requirement for the award of the degree of Master of Technology, is original and not copied from any source without proper citation. This work has not previously formed the basis for the award of any Degree, Diploma Associateship, Fellowship or other similar title or recognition.

Place: Delhi **Kamaksha**

Date:

DELHI TECHNOLOGICAL UNIVERSITY

(Formerly Delhi College of Engineering)

Shahbad Daulatpur, Main Bawana Road, Delhi-42

CERTIFICATE

I hereby certify that the Major Project Report II titled **"MUSIC RECOMMENDATION SYSTEM EMPLOYING FACIAL EMOTION RECOGNITION "** which is submitted by **Kamaksha**, Roll No 2K22/CSE/10, Department of Computer Science and Engineering, Delhi Technological University, Delhi in partial fulfillment of the requirement for the award of the degree of Master of Technology, is a record of the project work carried out by the student under my supervision. To the best of my knowledge this work has not been submitted in part or full for any Degree or Diploma to this University or elsewhere.

Dr. Vinod Kumar
Professor
Department of CSE,
Delhi Technological University

Place: Delhi
Date: May, 2024

ACKNOWLEDGEMENT

I would like to express my deep appreciation to **Dr. Vinod Kumar**, Head of department at the Department of Computer Science and Engineering, Delhi Technological University, for his invaluable guidance and unwavering encouragement throughout the course of this research. His vast knowledge, motivation, expertise, and insightful feedback have been instrumental in every aspect of preparing this research plan. His expertise and scholarly guidance have significantly enhanced the quality of this thesis.

My heartfelt thanks go out to the esteemed faculty members of the Department of Computer Science & Engineering at Delhi Technological University. I extend my gratitude to my colleagues and friends for their unwavering support and encouragement during this challenging journey. Their intellectual exchanges, constructive critiques, and camaraderie have enriched my research experience and made it truly fulfilling.

While it is impossible to name everyone individually, I want to acknowledge the collective efforts and contributions of all those who have been part of this journey. Their constant love, encouragement, and support have been indispensable in completing this MTech report.

Kamaksha

(2K22/CSE/10)

ABSTRACT

Music is one of the most important forms of art created by humans. Several studies show that listening to music can fix various health issues like reducing mental stress, anxiety and helps in betters sleep quality mental awareness. It elevates a person's mood, and make them calm and relaxed. It also helps in experiencing all kind of emotions. Now a days people prefer songs based on their mood and our face acts as a mirror of it. Facial expression has become a gateway to a person's state of mind. In recent times recommending mu-sic through capturing emotions of the user has become an important and exciting field of study. This paper investigates some of the previous research that has gone in this field. It oversees the methodology used, algorithm implemented and performance measure of different techniques used in recommending music based on the user's state of mind. The paper proposes a music application, Mood Musica that recommend playlist based on facial expressions. Finally, the paper reveals the future scope of research and development in this field and provides insights into using facial expression as music recommendation factor.

LIST OF PUBLICATIONS

1. Kamaksha, "Facial Notes: Unlocking Music Recommendations through Expression (A Contextual Analysis)", Accepted at **"International Conference on Intelligent Computing and Communication Techniques (ICICCT)"**, at JNU New Delhi, India.

   Paper Id: 660

   Indexed by Scopus.

2. Kamaksha, "Mood Musica: Music System based on facial Expressions" Accepted at **"National Conference on Advanced Computer Science and Information Technology (NCACSI - 24)",** Udaipur, India.

   Paper Id: National Conference_7239168

   Indexed by Scopus.

TABLE OF CONTENTS

List of Tables

List of Figures

CHAPTER 1
INTRODUCTION

Since face acts as a bridge between user's mood and state of mind, several of the studies have been done in understanding this relationship and then it is used in various technologies. Music Recommendation is one of them. As music plays as a comforter during our low times and make a person cheerful or at least comfort him for the time being, it is imperative to use technology for the user's wellbeing and also as recreational purpose. Various research has been done in this field to put forward the needs of the user using this technology. With time various optimization has been implemented to enhance the user's experience and satisfaction. By recommending music that depends on the present nature of the user, understood from its facial expression, it has given a chance to increase the experience of the user by providing personalized music preferences. If we observe YouTube or any other entertainment platform, their algorithm recommends music or any other data based on the past searches or queries of the user. Although sometimes it does rightly predict what user really wants, but most of the time it fails to capture the current needs of the users. This happens because search platform uses static techniques and do not have the technology to know the ever changing of emotion of the user. They fail to understand that predicting the needs of the user must go beyond their past history and queries. Integrating techniques to capture current emotion of the user has become the essence of the time.

Facial expression plays an important part in our emotional understanding. We as a human are very easily understandable if been carefully examined facially. Our current feelings and emotions have startling effect on our face. We can easily differentiate a happy person and a person who is sad just by seeing their face. There are small nuances that can tell what a person's state of mind is and what it is currently feeling. This important relationship between our face and emotions plays a significant role understanding and developing systems that are dynamic in nature. This has led to integrating facial expression as emotion detector in recommending music to the user. This is dynamic in nature as users' mood or state of mind can also change while listening to music too. The systems that we are going to study are dynamic in nature i.e. they can observe the present state of mind of the user and can suggest music accordingly. There are various user preference-based models already being employed.

These models use different algorithms, data sets and different technical approaches. Despite the differences, emotion capturing is the most important features of such systems. Since the music is recommended based on facial expressions, these algorithm models must use efficient and fast algorithms to capture the data and process it in real time otherwise it will result in bad experience of the user. Also, they must have capable hardware for processing the information and mapping it to the recommendation. Now that we have established that capturing emotions is the most important operation in these systems. There are various studies that talks about it. Puri and Gupta [8] in their paper specified emotion capturing from images. Their work is based on computer vision and machine learning. In another work Oulid Djekoune al. [9] worked on emotion recognition based on deep learning and transfer learning. Also, Xavier Giro-i-Nieto et al. [10] uses convolutional Neural Netwok to recognize emotions in unrestricted environment.

1.1 Software Requirements

Software requirements are necessary for suggesting the basic minimum requirement of the system and system resources that the user must have so that it can run all the components related to the application smoothly. These are software that must be pre-installed before the installation of the actual software. They are generally not in included with the installation package of the software that we are going to install in the system.

1.1.1 Platform. A platform describes an environment on the system, either in hardware or software on which application can execute. Some of the daily used platforms include a pc, system or operating system, or programming languages and their executable libraries.

Operating system is one of the major aspects of running any software program. It provides the environment to the software to be installed and run properly. Since each operating system that exists in the markets have various differences between them, difference version of the same software is provided so that there is no problem in running the software across different platforms. Apart from different operating systems that exists in the market like windows, Linux, macOS, there is different version of the same operating system like windows have windows XP, windows 7, winnows 10. Most of the times difference version of the same operating system don’t support the same software. Hence along with the different software it is very important for different version of the same operating system to be compatible with the software.

1.1.2 APIs and drivers. API refers to application interface i.e. an interacting part of the software that acts as a mediator between the actual system and user. Through good API, is very necessary for the active and comfortable for user to participate in application usage. API includes hardware devices and high end displaces and also network devices.

1.1.3 Web browser. Most of the web applications and software that uses internet services or require internet technology uses the web browser that is present on the system of the user. Some of the examples of web browser includes, Google Chrome, Microsoft Internet Explorer, Safari etc. One of the important components in web technology is ActiveX component. It is very important for those application which uses internet services frequently. It is generally found in all the browsers. Even the web browsers present in mobile devices has ActiveX.

Software Requirements for this application is given as follows:

- Software: Anaconda, Visual Studio Code, OpenCV
- Primary Language: Python, C#
- Database: FER 2013 dataset, Kaggle songs dataset

1.2 Hardware Requirements

The typical set of requirements specified by any operating system or software application pertains to the physical computer resources, commonly referred to as hardware. Often, a list of hardware requirements is provided along with a hardware compatibility list (HCL), which is particularly important for operating systems. An HCL details hardware devices that have been tested for compatibility, and sometimes those that are incompatible, with a specific operating system or application. The sub-sections below explore the different facets of hardware requirements [14].

1.2.1 Architecture: Every OS has been tailor-made to work with a certain kind of computer hardware. Almost all programs have strict requirements about the OS and hardware architectures they can only work on. The majority of programs and operating systems need recompilation in order to function on different architectures, even if there are some that are architecture-independent. Check out this record of famous OSes and architectures that power them as well.

1.2.2 Processing power. Without enough processing capacity from the central processing unit (CPU), computer applications cannot operate. What most x86-based applications utilize to show computing capability is the kind and clock speed of the central processing unit (CPU). Despite having an impact on power and performance, a number of important CPU characteristics—like bus speed, cache, and MIPS— frequently ignored. This concept of capability is every so often inaccurate because AMD Athlon and Intel Pentium CPUs might have different throughput rates even though their clock speeds are the same. Since Intel Pentium CPUs are so prevalent, they are frequently included in this group.

1.2.3 Memory. All the software runs on the PC's random-access memory (RAM) during execution. All these operations must be taken into consideration when calculating the amount of memory that a program, together with the operating system, files, and auxiliary apps that it uses. The efficiency with which further, dissimilar software operates on a multitasking computer system is also measured while characterizing this demand. When preparing for secondary storage, considered the volume of hard drive space required for program installations, the temporary files are generated and kept when software installations and runs, and the possible utility of swap space.

1.2.5 Display adapter. Application or tools that needs an above-par PC graphics monitor, such as videos/image editors and massive-resolution games, frequently specifies the need for high-configuration display adapters in its system requirements [16].

1.2.6 Peripherals. The increased performance or functionality of certain peripherals is required by some software programs because of their rigorous and/or usage of certain devices. Drives for compact discs (CDs), keyboards, pointing devices, network adapters, etc., are also considered peripherals.

1.3 Feasibility Study

A feasibility study assesses the practicality of a project or system. Within this study, a thorough and objective analysis of a potential business or venture is undertaken to identify its strengths and weaknesses, potential opportunities and threats, resources needed for execution, and overall chances of success. When evaluating feasibility, two key factors to consider are the projected cost and anticipated performance [18].

Types of Feasibility Study

A feasibility study's credibility with potential funders and lenders depends on how objective it appears to be, as its goal is to assess the possibility that a project will succeed. The various fields that a study of this kind could cover, or the five different types of feasibility studies, are listed below.

Technical Feasibility

The technological resources of the firm are the primary subject of this investigation. This aids companies in determining whether their methodological staff can carry out their ideas and whether their technological resources are sufficient. Determining the projected system's technical viability also entails assessing its hardware, software, and other requirements. As an extreme example, a corporation now cannot attempt to build Star Trek transporters in their facilities due to logistical issues.

Economic Feasibility

In order to assist businesses in assessing a project's feasibility, costs, and advantages prior to allocating financial resources, this assessment usually includes a cost-profit analysis [19]. It serves as an unbiased evaluation of the project and enhances its reputation in addition to helping decision-makers identify the benefits the project will have economically for the company [21].

Legal Feasibility

This analysis's goal is to ascertain whether any laws, including those governing social media, data security, and partitioning, are broken by the proposed project. Think about the following situation: A business has selected a certain location for a new office complex. The feasibility study may come in rather handy if the organization's selected location turns out to be unzone. For that organization, realizing early on that their concept was unrealistic has just saved a great deal of time and work

Operational Feasibility

This evaluation entails conducting a study to analyze and ascertain if the organization's needs can be adequately met through project completion[22]. Operational feasibility studies also scrutinize how well a project plan fulfills the requirements identified during the requirements analysis phase of system development [22].

Scheduling Feasibility

The accomplishment of the project rely upon this analysis since no project can succeed if it is not handed over duly complete on schedule. A firm estimates the amount of time needed to finish the project to establish the scheduling feasibility. Following the consideration of these variables, the feasibility study may identify any potential project constraints, including:

- Internal Project Constraints, encompassing resource, financial, technological and technical constraints, among others.
- Limitations from inside the company, such as export, marketing, financial, etc.
- Regulatory, environmental, legal, and logistical considerations are examples of external constraints.

1.4 UNDERLYING ALGORITHMS

CNN ALGORITHM

The Convolutional Neural Network stands as a primary category for conducting image classification and recognition tasks within neural networks. Its applications span various domains including scene labeling, object detection, and facial recognition, among others.

A Convolutional Neural Network (CNN) receives an image as input and classifies it into specific categories like car, bird, or even sun [10]. The computer can only input an image as a vector of pixels, with the perception depending on the vector specification. The representation is typically denoted as h * w * d, where h represents height, w represents width, and d represents the number of color channels [21].

A RGB image is represented as a 6 * 6 * 3 array of matrices, while a grayscale image is represented as a 4 * 4 * 1 array of matrices.

In a CNN, an input image processed under a series of layers such as convolutional layers, pooling, fully connected layers, and filters [7]. At last, the Softmax function is applied to objects that are needed to be classified, giving them a value between 0 and 1.

Convolution Layer

The convolutional layer is the first of the layer and used for extracting features from the input vector of the input image as the initial stage in extracting features from an input image [19]. By leveraging a small square of input data to learn image features, this layer preserves the pixel relationships [19]. It uses a matrix operation involving two inputs: the input vector and a kernel [19].

- The dimension of the image matrix is $h \times w \times d$.
- The dimension of the filter is $fh \times fw \times d$.
- Consequently, the dimension of the output is $(h-fh+1) \times (w-fw+1) \times 1$.

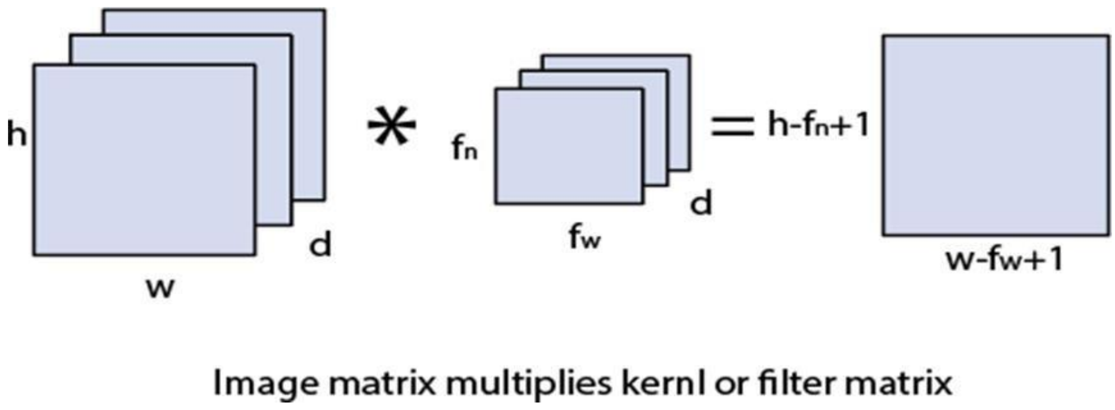


Fig. 1.1 Matrix Kernel Multiplication

Let's begin with a 5*5 image with pixel values ranging from 0 to 1, and a filter matrix of size 3*3 as follows:

$$\begin{bmatrix} 1 & 1 & 1 & 0 & 0 \\ 0 & 1 & 1 & 1 & 0 \\ 0 & 0 & 1 & 1 & 1 \\ 0 & 0 & 1 & 1 & 0 \\ 0 & 1 & 1 & 0 & 0 \end{bmatrix} \times \begin{bmatrix} 1 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 1 \end{bmatrix}$$

5×5 − Image Matrix 3×3 − Filter Matrix

Fig. 1.2 Image and Filter Matrix

The result of multiplying the 5*5 image matrix by the 3*3 filter matrix is referred to as the "Feature Map," which is presented as the output.

$$\begin{bmatrix} 1 & 1 & 1 & 0 & 0 \\ 0 & 1 & 1 & 1 & 0 \\ 0 & 0 & 1 & 1 & 1 \\ 0 & 0 & 1 & 1 & 0 \\ 0 & 1 & 1 & 0 & 0 \end{bmatrix} \times \begin{bmatrix} 1 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 1 \end{bmatrix} = \begin{bmatrix} 4 & 3 & 4 \\ 2 & 4 & 3 \\ 2 & 3 & 4 \end{bmatrix}$$

Convolved Feature

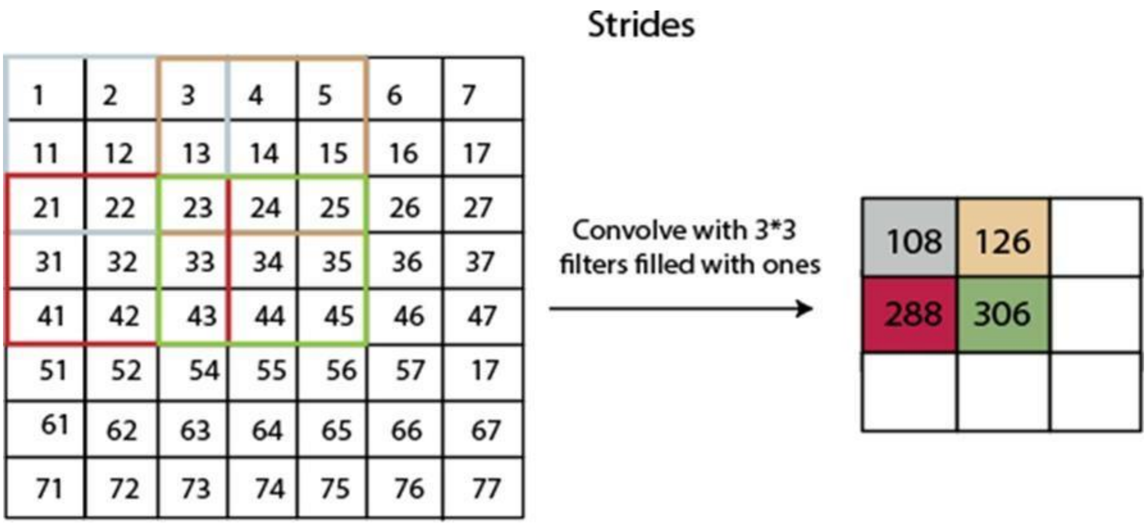
Fig. 1.3 Feature Map

Applying various filters to a picture through convolution allows for the applications of effects like edge detection, sharpening and blur.

Strides

The stride means how much the filters moves over the input image vector. If the stride becomes 1 the filter moves 1 pixel over the input image. Hence if the stride is set to 2, the filter shifts by 2 pixels over the input image at one time. The illustration below demonstrates the convolution process with a stride of 2.

Fig. 1.4 Strides



Padding

In order to construct the Convolutional Neural Network, padding is essential. If the input vector is shrunk, if a neural network with hundreds of layers is applied to the that input vector the resultant filtered image will be small.

What happens if we convolve a greyscale image with a three-by-three filter on top of it.

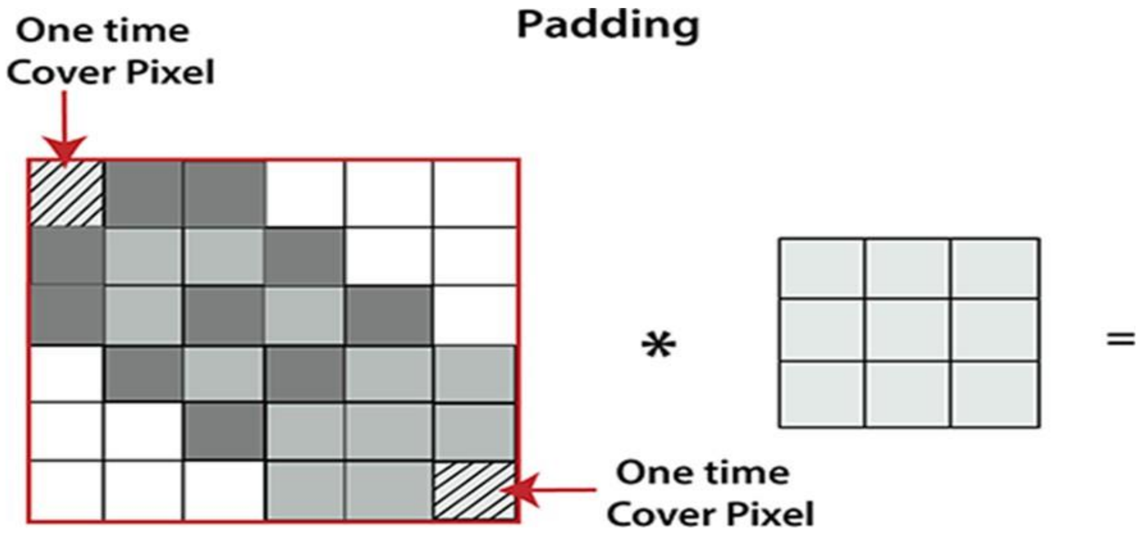


Fig. 1.5 Padding

As depicted in the image above, the corner pixels are covered only once, whereas the middle pixels are covered multiple times, providing more information. Consequently, two downsides arise:

- Shrinking outputs: With a larger stride, the output size reduces more rapidly.
- Loss of information at the corners of the image: Since corner pixels are covered only once, there's a risk of losing valuable information.

To address these issues, padding is introduced to the image. Padding involves adding an extra layer around the border of the image.

Pooling Layer

The pooling layer is crucial in the pre-processing of an image, as it helps reduce the number of parameters, particularly when dealing with large images. Pooling involves "downscaling" the image obtained from previous layers, i.e. to decrease the density of the pixel, the image is shrinked.

Spatial pooling, also known as down sampling or subsampling, diminishes the dimensionality of each map while preserving essential information [15]. There are several types of spatial pooling, including:

Max Pooling

Max pooling serves as a process with discretization on a sample with the primary goal of downscaling an input representation, thereby reducing its dimensionality.

This process enables assumptions to be made about features within the regions of a particular field. Max pooling involves applying a filter over the input vector and choosing the max of the overlapped pixels.

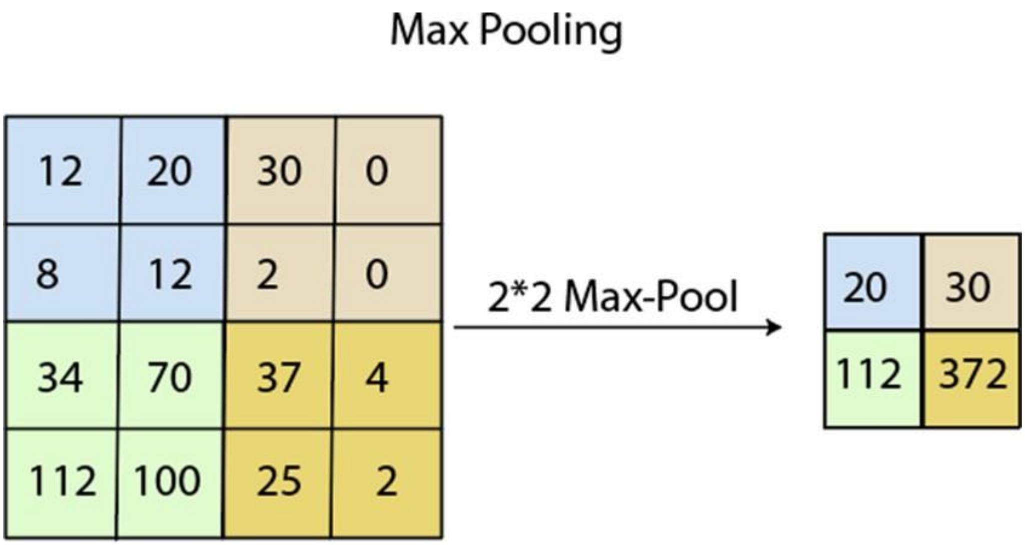


Fig. 1.6 Fig. Depicting Max Pooling

Down scaling is maintained through calculating averaging pooing among the input vector and divided into pooing regions and mean values are calculated for each region.

Syntax

These are function calls for creating an average pooling layer:

layer = averagePooling2dLayer (poolSize);

layer = averagePooling2dLayer (poolSize, Name, Value);

In the first call, a basic average pooling layer is created with the specified pool size.

In the second call, additional options can be provided using name-value pairs to customize the layer.

Sum Pooling

In this the respected field or sub fields are taken over the input vector just was in case of max pooling. But here instead of going for maximum, sum or addition of pixels on the respected subfields are taken and outputted to feature vector

Fully Connected Layer

It is an important component where inputs from preceding layers are flattened into a vector and processed. It facilitates the transformation of output into the number of classes desired by the network[17].

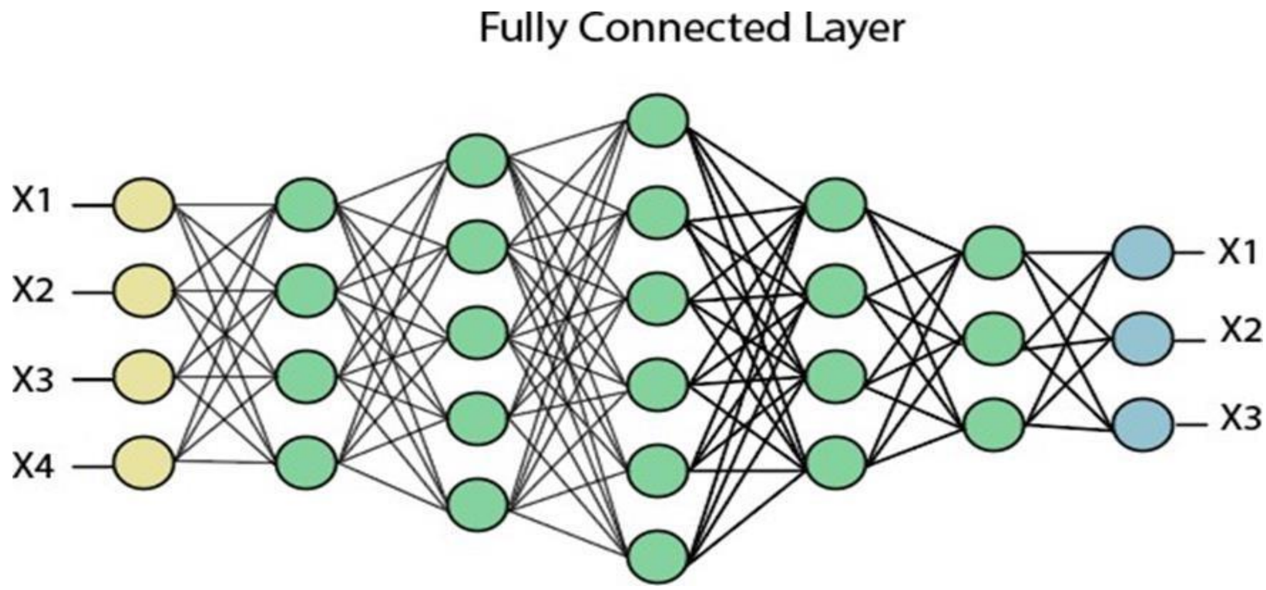


Fig. 1.7 Fully connected Layer

In the diagram, the feature map matrix undergoes transformation into a vector, represented as x1, x2, x3... xn, through fully connected layers[17]. These features are amalgamated to construct a model, and an activation function like softmax or sigmoid is employed to classify the outputs into categories like car, dog, truck, etc.

CHAPTER 2
LITERATURE SURVEY

2.1 Survey Papers

The goal of the topic is to evaluate some of the research that has already been done in the field of recommending music based on user's current emotion. Related topics have been studied in the past, such as the fusion of recommendation systems and facial expression detection software. These studies have looked into a variety of techniques to analyze facial expressions and infer emotional states, including computer vision, machine learning, and deep learning. Although there are multiple already analyzed papers in the field, less talks about the various algorithm and implementation techniques that this field requires. Moreover, various of them talk about the practical overview of the algorithm implemented also analyzing various performance measures. Here our aim is to provide an overview of the modern algorithms and technologies and their use cases

Recent breakthroughs in computer vision, machine learning, and artificial intelligence (AI) have led to considerable improvements in music recommendation systems. This review of the literature attempts to give a thorough overview of the work that has been done in facial expression-based music providing systems, emphasizing important techniques, discoveries, and developments.

As seen from the table in the next page various works has been done in capturing the emotions. Some of the works were closely related to the said field where generation of music playlist were done while some of the work was done in developing technology that would later become important in the field of study. While implementing the technology of any project there are several working components that are required in order to get a technology to work. We will also study various other projects that helped in developing this project

One of the earlier works in this category has been done by Anand et. Al. [1]. In his paper they proposed a music providing systems that uses deep learning algorithms. They extracted features from a dataset containing a large amount audio signal and based on that they created a personalized playlist. Some of the features include rhythm. melody, pace of the songs etc. These features as we can see helps in determining the essence of a song which in terms helps in determining the mood of the user. While training the models a neural network or any other deep neural network is used for feature extraction. These models are trained using various deep learning techniques. As the user interacts with the model, the model analyzes their listening history and preference and using those patterns recommend tailorized playlist for the user to listen. Based on the quality or accuracy of the playlist use rate the playlist. As they want to continuously improve the model, the system includes a feedback mechanism that use user's feedback to garner the preferences of the user. After succession of training and feedback, the model improves its accuracy significantly. It is one of the good studies that shows a model implementing this model.

Table 1: Papers showing research in the field

Author(s)	**Title**	Remarks
Anand R, Sabeenian RS, Deepika Gurang, Kirthika R, Shaik Rubeena [1]	AI-based Music Recommendation system using Deep Learning Algorithms	Proposed a music recommendation system based on deep learning algorithms, leveraging features extracted from audio signals for personalized recommendations.
Shantha Shalini K, Jaichandran R, Leelavathy SA, Raviraghul R, Ranjitha J, Saravanakumar [2]	Facial Emotion Based Music Recommendation System using computer vision and machine learning techniques	Developed a system using computer vision techniques to analyze facial expressions and ML algorithms to recommend music based on detected emotions.
Ankita Mahadik et al. [3]	Mood-based music recommendation system	Proposed a system that recommends music based on the user's mood, utilizing sentiment analysis techniques to infer emotional states.
Aman Tamboli et al. [4]	Music Recommendation System	Developed a collaborative filtering-based recommendation system, leveraging user-item interactions to generate personalized music recommendations.
Xinglin Wen [5]	Using deep learning approach and IoT architecture to build the intelligent music recommendation system	Proposed an IoT-enabled music recommendation system using deep learning techniques for feature extraction and recommendation generation.
Xiao Han, Fuyang Chen, Junrong Ban [6]	Music Emotion Recognition Based on a Neural Network with an Inception-GRU Residual Structure	Proposed a neural network architecture for emotion recognition in music, achieving high accuracy in emotion classification tasks.
Ferdos Fessahaye et al. [7]	T-RECSYS: A Novel Music Recommendation System Using Deep Learning	Introduced T-RECSYS, a deep learning-based recommendation system that incorporates temporal dynamics and user preferences for improved recommendations.

This model analyzes user’s photo or video stream to get the current facial expression of the user. Various computer vision algorithms are used for identifying the different parts of the face. Once the emotion is detected, features are extracted using machine learning algorithm. Once the training is done user tests the model and a tailored playlist is generated based on the preference of the user. Once again based on the generated errors, model is trained and its accuracy is improved. This system also interacts with the user in real time continuously tracking user’s facial expressions.

Similarly, Ankita Mahadik et ai.[3] also proposed a system similar to this. They developed system that gives music based on current mood of the user using sentiment analysis techniques to interpret or map the emotional states or expressions. Sentiment analysis is most popularly used in Natural Language Processing (NLP). It takes the features extracted from the input image and rate them a sentiment score. It indicates the emotional element in the sentiment analysis system. Various hints like user’s entered text and data are used and forwarded to the model to get the current mood of the user. It is very important as it can process a wide variety of text information correctly and accurately. It can correctly predict the tone of the text. This results in more accurate generation of the personalized playlist for the user. Hence the user’s music experience gets elevated to another level.

In another paper Aman Tamboli et al. [4] developed a collaborative filtering-based recommendation system, leveraging user- item interaction to generate personalized music recommendation. Collaborative filtering is a technique that can filter out items that a user might like or dislike based on the reactions of similar users[4]. It first search among a very big group of people and then identifying the patters among them and clustering the according to similar likes and dislikes. Generally deep learning models are used in such systems. Similarity is measured using various parameters like cosine similarities. Recommended systems frequently use this algorithm to generate tailor made playlist for a particular user. Music data of all the people in the group is taken and their preference is categorized. These systems are then used to filter the music that are liked or disliked by users. Then the system creates user tailored music by analyzing these patterns. By analyzing the previous user’s likes and dislikes this model improves user’s overall expectation and experience.

Another model that efficiently performs user's recommended music experience is by Xinglin Wen et al. [5]. IoT (Internet of Things) is the technology used by them to recommend by them. They proposed an IoT enabled music providing system using deep learning techniques for feature generation and music recommendation. IoT devices are non-computing hardware devices like network devices, switches or plugs that can connect to a network and can transfer data. Wen presented IoT enabled music recommender that Makes use of IoT architecture and deep learning architecture. In this user's data is collected in real time using IoT devices like wearable sensors or smart watches or speaker like Alexa or Google pod. These devices collect the user data and forwards it to deep neural architecture. Deep neural networks are then used to process this data along with extra features like user preferences.

Based on this data model training is done. After model is trained, testing is done by the user. Model generate the recommended playlist depending on the user's current mood and current situations. If any error is there it is fed back to the models and model's accuracy is improved based on back propagation and hyper parameter tuning.

A GRU (Gated Recurrent Unit) based structure was presented by Han et. Al [6]. They proposed recurrent neural network-based architecture for emotion recognition and music recommendation achieving high accuracy in emotion classification tasks. A recurrent neural network is a type of artificial neural networks that specializes in temporal and sequential data. They process sequential data as they process input data without any predetermined size limit. These models recognize patterns and relationship by remembering data as current output becomes the input to next time step. A GRU is same as LSTM which is a special type of RNN used for remembering long term dependencies. With GRU an inception module is used by this model which is used in image recognition tasks. It is modified to extract features of higher dimension from music audio data hence identifying patterns in both spatial and temporal dimensions. After feature extraction this data is inputted to GRU structure where mapping is done. After successful mapping recommended music playlist is generated for the user. For backpropagation, error correction and accuracy, same operations are performed as other already mentioned models.

T-RECSYS, a music providing system that adds user preferences and temporal dynamics with deep learning approaches, was introduced by Fessahaye et al. [7]. Through the capture of temporal patterns in music listening behavior, this technology addresses the dynamic nature of user preferences across time. Adaptive and customized music suggestions are made possible by T-RECSYS, which models and predicts changes in user preferences using deep learning algorithms. The deep learning model can forecast future user preferences based on historical behavior since it learns to identify patterns and correlations in the data during training. Crucially, the model is updated and modified on a regular basis in response to new data, so it stays up to date and takes into account the user's changing preferences.

There are various studies that talks about it. Puri and Gupta [8] in their paper specified emotion capturing from images. Their work is based on computer vision and machine learning. In another work Oulid Djekoune al. [9] worked on emotion recognition based on deep learning and transfer learning. Also, Xavier Giro-i-Nieto et al. [10] uses convolutional Neural Netwok to recognize emotions in unrestricted environment. There are other research works that talks about more advance form of applications of the said project. A Facial Expression-Based Music Recommender System Using Deep Learning by A. Rahim, R. S. Rahman, S. F. Ahmed, and M.
I. Ahmed [11]. This paper proposes a music providing system that incorporates facial expression analysis using deep learning techniques. Emotion-Based Music Recommendation Using Fusion of Content-Based and Collaborative Filtering Techniques by S. Kumar and S. Rajput [12]. This paper presents a music providing system that adds facial emotion analysis with content-based and collaborative filtering techniques. Music Emotion Recognition: A State-of-the-Art Review by Y. Song, L. Li, and T. P. Martin. [13] although not specifically about music recommendation, paper, provides a comprehensive review of music emotion recognition techniques, which could be relevant for developing emotion-aware recommendation systems

2.3 Algorithms already in use

Deep Learning-based Facial Expression Recognition with Keras: It enables real-time detection of up to five distinct facial emotions. It operates atop a Convolutional Neural Network (CNN) constructed using Keras with TensorFlow as its backend in Python. The system can classify facial emotions. Utilizing OpenCV for image processing tasks, the algorithm identifies faces from a live webcam feed, processes them, and feeds them into neural network for emotion detection. By facilitating one to one learning directly from input images.

Hybrid Music Recommendation Approach: It completely relies on collaborative filtering for music recommendation has various drawbacks, the most prominent being the "Cold Start" issue. This arises from the fact that music albums are tagged as frequently as they are discovered or listened to, leading to limited or absent tags for new or undiscovered music. Furthermore, listeners are more inclined to tag songs they strongly enjoy compared to those they only mildly appreciate or dislike.

Viola–Jones Object Detection Framework: It stands as a widely adopted algorithm for object detection renowned for its fast detection speed despite slow training. This algorithm operates using Haar basis feature filters, eliminating the need for multiplications. Efficiency is significantly enhanced by first generating the integral image.

CHAPTER 3

METHODOLOGY

In the coming sections we are going to study about our model which is based on Machine Learning and CNN. Machine learning enables machines to learn and improve without being programmed by an external source. It involves following steps.

1. Data Gathering
2. Preparing the data
3. Model Choosing
4. Training The dataset
5. Evaluation of the dataset and system
6. Prediction done

Data Collection: Collecting data is a vital phase in machine learning, where the accuracy of a model heavily depends on the data collected [17]. The information is organized in a tabular format and employed to train the model to get the patterns and correlations among features [19]. After preprocessing, the data is divided into distinct sets for model training and assessment, specifically the set that is used as training and testing.

Data Preparation: Preparing the data involves loading it into an appropriate environment and organizing it to ensure that the learning process remains unaffected. It is crucial to identify relationships among various variables, and the data is divided into two parts: one set will be used for training the model (the training set) and the other for checking the performance of the model (the testing set). Additional manipulations, such as normalization and error correction, are also carried out.

Model Selection: The next step is selecting the appropriate model, which is essential for accomplishing the task effectively.

Training the Model: After selecting the model, the data is used to progressively enhance its predictive capability. Values are assigned to parameters A and B in the model, and the output is predicted and compared with the actual values. These parameters are then adjusted to better align with the predictions, and this iterative process continues.

Dataset Evaluation: After completing the training, it is crucial to evaluate the dataset to determine the model's effectiveness. Testing is performed on data that was not used during training and has been reserved specifically for this purpose.

Prediction: The ultimate goal of machine learning is to leverage data to answer various questions, which underscores its value. The model developed through the preceding process can be used to predict the desired outcomes.

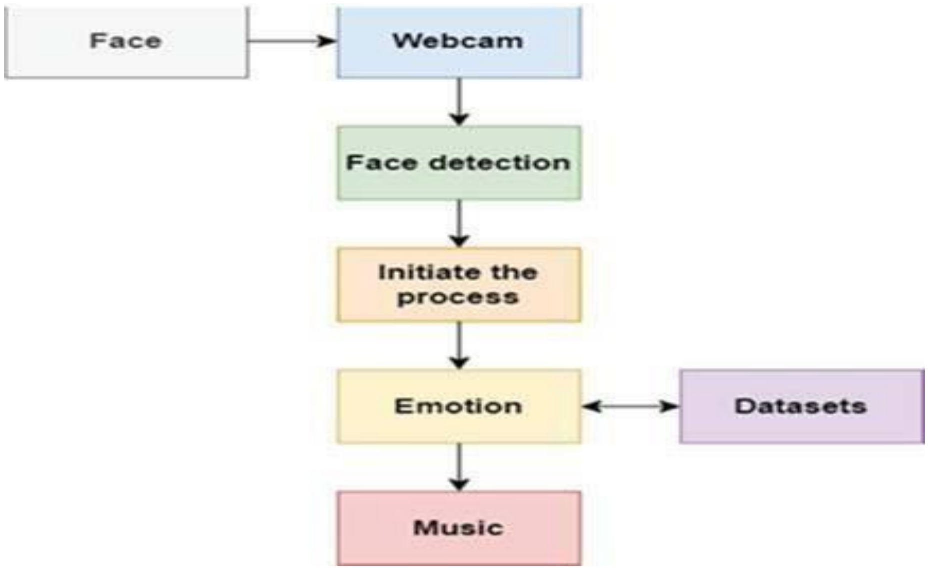


Fig. 3.1 Flowchart of Implementation

3.1 Input Design

It is that link that bridges a connection between user and the application interface (API). It generally includes having some information about converting raw data or transaction data in data which can be processed by the system.

This conversion is accomplished in a many ways, including direct user data entry or computer scanning of written or printed materials. The main aim is to restrict data that can be entered into the system, errors can be minimized, get rid of faulty process and keep the UI simple for the user. It is set up to guarantee privacy preservation, usability, and security. Determining the required input data, appropriately structuring or coding the data, guiding the user through conversation, and putting techniques for input validation and error handling procedures into practice are all aspects of input design.

3.2 Output Design

An output that satisfies end user expectations and efficiently conveys information is considered high quality. Any system uses outputs to communicate its processing results to users and other systems. The way information is presented in hard copy format and to meet urgent demands is determined by the output design. It is the main and most direct information source for users. The system's ability to assist user decision-making is improved by well-designed output.

3.3 Data Flow Diagrams

It is set of Diagrams which are simply graphs that shows input, different operations that can be performed on that input and output that is generated as a result of these operations. DFDs provide a visual representation of the changes and paths taken by information as it moves through a system.

Use Case Diagrams

It is a behavioral diagram created using UML and developed by carefully going through all the cases or situation a user can go through. The important function of USD is to provide a visual picturization of the functioning of the all the process of a system and relationship between all those situations or cases. A use case diagram shows system operation for each player in the process and also their roles.

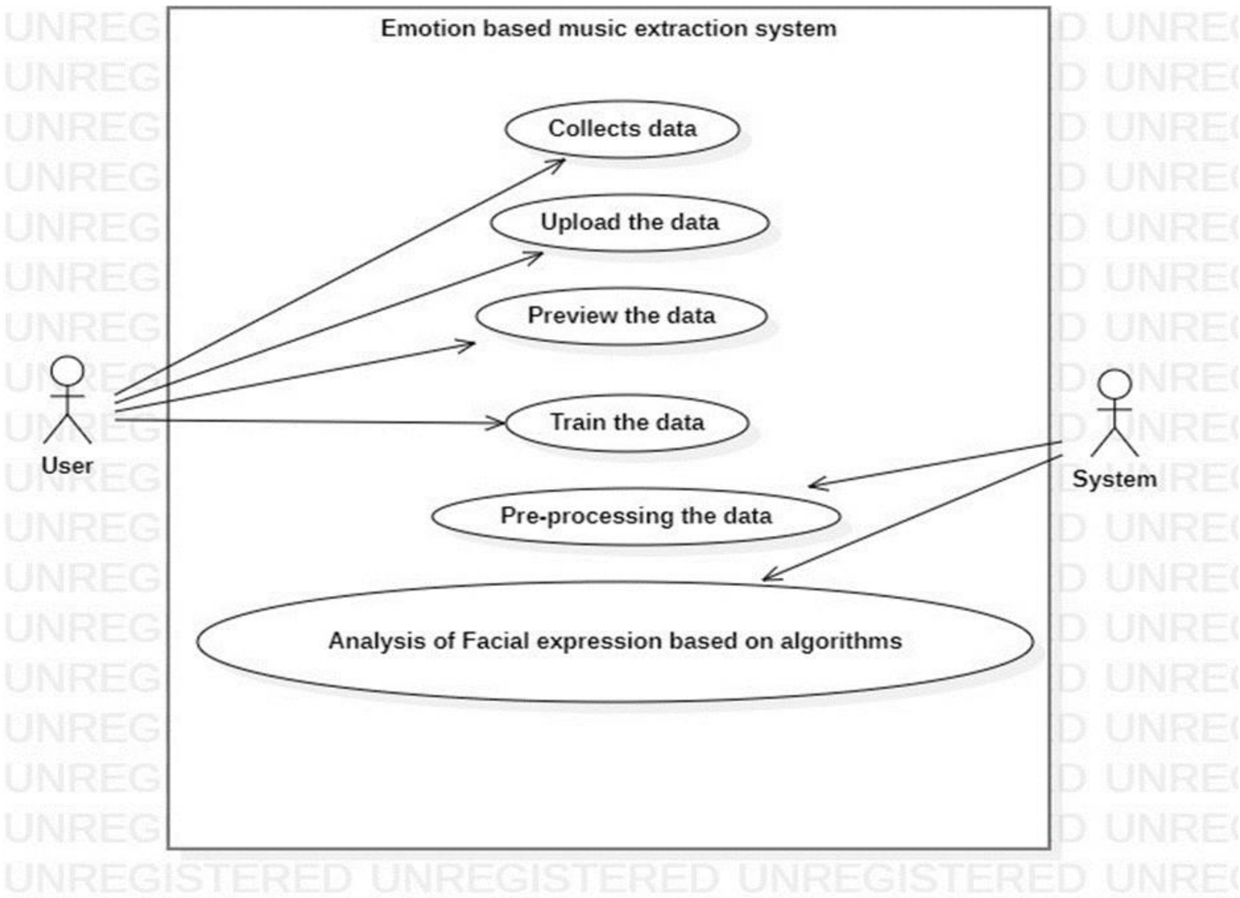


Fig. 3.2 diagram showing Use Case

Sequence Diagram

It is pictorial representation of sequence and style of interaction between process. Its organization follows the following guidelines of a message chart. It is also known as event diagrams or timing diagrams.

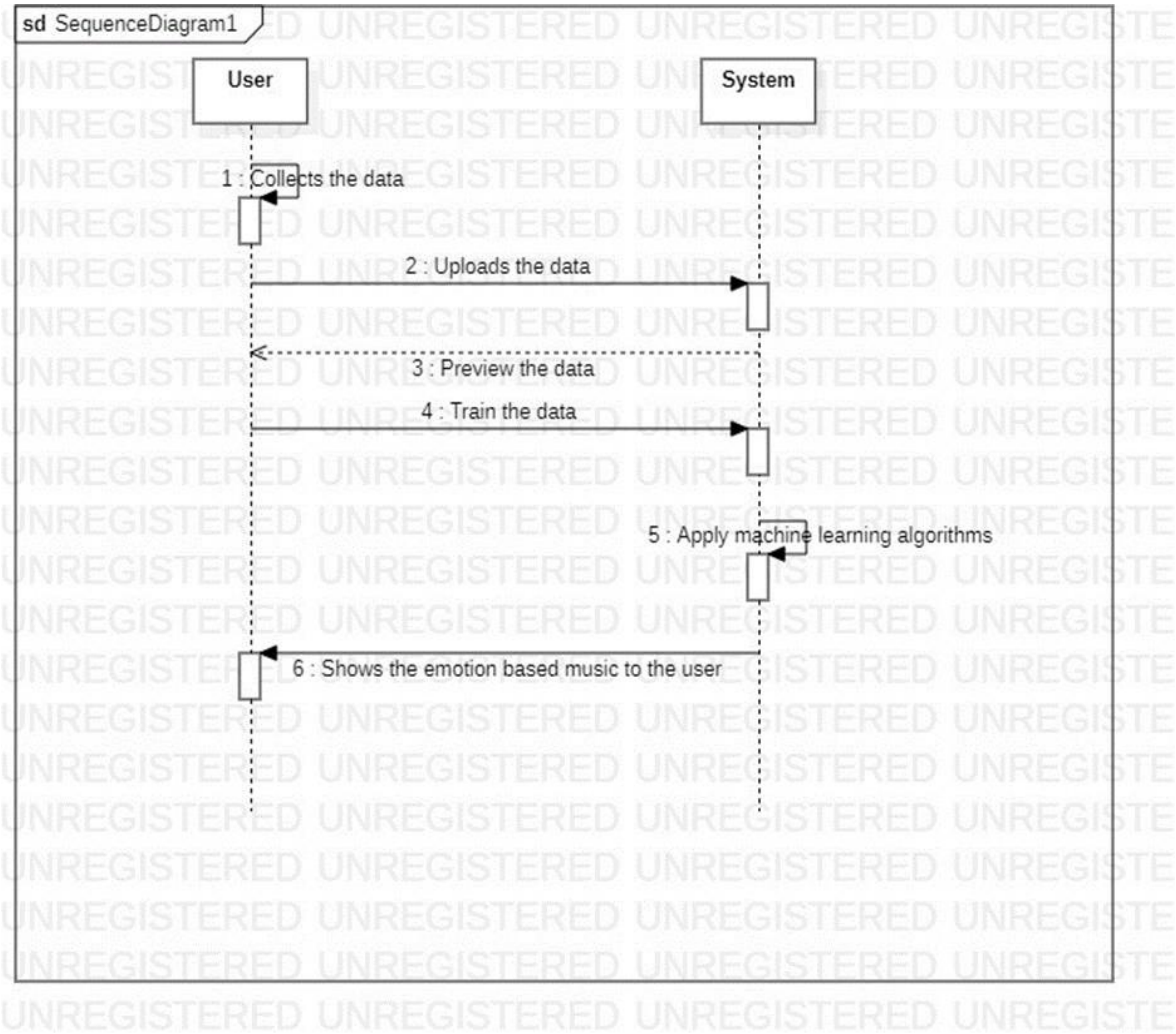


Fig. 3.3 Diagram Showing sequence

Activity Diagrams

Activity diagrams are graphic designs representing a set of sequential operations that include decisions, repetition, and concurrent execution of tasks and actions. Activity diagrams are a technique used in the Unified Modeling Language (UML) to represent the sequential processes of business and operational components of a system [23]. An overview of the system's control Flow is given by these diagrams.

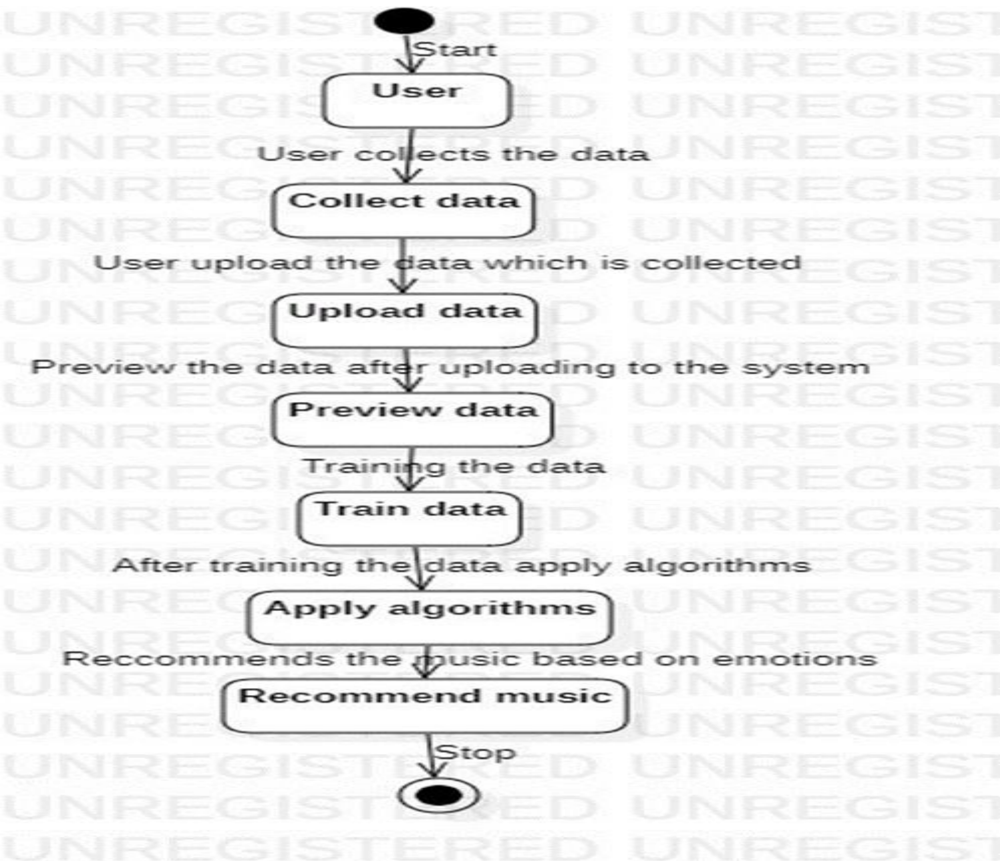


Fig. 3.4 Activity Diagram

3.4 Modules

Data collection Module

In the survey user were asked three simple questions: 1. What type of music they generally listen when they are feeling happy? 2. What kind of music they would like to hear in a depressing situation? 3. What type of music they listen when they are upset?

Emotion Extraction Module

An image of the feeder is taken with the help of camera or webcam. To improve the effectiveness of the classifier—which is in charge of recognizing the user's face in the image—the frame from the camera feed is transformed into a grayscale image after it is taken.

The image is then forwarded to the classification section after the conversion procedure. The user's face is extracted from the webcam stream frame by the algorithm using feature extraction techniques. Then, specific features are fetched from the input face and sent to the section of the trained network so that it can identify the emotion that the user has displayed.

The classifier uses these photos as training data. As a result, based on the information learned during the training phase, the classifier can determine the locations of facial landmarks when it is given a completely new and unknown collection of photos. The coordinates of the recently discovered face landmarks are then returned by the classifier.

To accurately identify the user's indicated emotion, the network is trained on a large dataset.

Audio Extraction Module

The user is then shown music or other sounds that matches the expressed feeling after the user's emotion has been identified. The user is presented with a list of songs that correspond with the selected emotion, from which they can select their favorite song. The songs that are shown are sorted according to how the user usually listens.

Emotion – Audio Integration Module

The songs that are relevant to the emotions that were extracted are preserved and played on the internet. For example, music from the cheery database will play if the user labels the mood or facial expression as "happy." The proposed method uses OpenCV to capture images with an inbuilt camera and perform preprocessing.

The model, trained using Fisher face, is saved in an XML file. The user selects their preferred language for music and their favorite singer either by selection or typing. The system then captures current facial expressions using the inbuilt app and stores the images in a specific format.

In our proposed system, Mood Musica, we integrated computer vision and machine learning techniques to integrate and detect facial expressions. To get the results for the training, we used Visual Studio Code for coding. We have used a CNN model along with Emgu and OpenCV library for facial recognition and feature extraction and classification. We have used C# for window application development and user interface. We can divide our model in two parts – mood detection module and music recommendation module.

A. Mood detection

This module is divided into three parts:

Face capturing - Frame extraction using inbuilt camera is a process in which the user is able to capture images using the camera that is built into their system.
The camera captures frames, and these frames are then saved as images in a particular size.

Several applications, including image processing, object detection, and facial recognition, benefit from this method. It is very easy to see that this process needs a camera with good resolution and clarity to capture high-quality frames.

Additionally, the images captured by the camera need to be processed and manipulated to achieve the desired results.

Face Detection – It is ability of a system to detect facial features in any image or frame. After face detection the coordinates of a bound box is given. We have used the python library OpenCV for face detection and facial features extraction. We have used Emgu library to integrate OpenCV to windows application. It is cross platform .NET wrapper for OpenCV. Its main purpose is to allow let developer access OpenCV functions and algorithms, making it easier to develop computer vision application using C# or any other .NET language.

Mood Detection – Mood detection is based on the inferring extracted facial features. Based on given facial features, it will be classified in the existing mood emotion like happy, sad, surprise, fear and angry. For this task, traditional keras mod-ule which is generally preferred for such operations could be used but it would have used a lot of time and would be machine heavy. Hence, we have used a pre-trained CNN model, FER (facial expression Recognition) which is trained on the Facial Expression Recognition Challenge (FER2013) dataset. There are various other models too but FER provides a set of images with annotated facial expressions making it easier to compare. Also, it consists of 35000 labelled images of 48×48 pixels which contains different facial expressions. It allows robust model training and evaluation. We trained our models for 50 epochs and achieved an ac-curacy of 82%.

Image Processing

A. Converting to blob object

The first stage of image processing begins with the conversion of the input image into a blob object. This conversion process entails resizing the image to match the network's input specifications, standardizing pixel values, and reordering channels as necessary. The resultant blob object serves as a structured representation of the image, apt for subsequent input into deep learning models.

B. Defining the class and Declaring Bounding Box

After converting the blob, classes are designated to recognize objects of interest. Bounding boxes are defined around these classes, delineating the spatial limits of each object. This stage lays the groundwork for subsequent object detection, furnishing essential details for model training and assessment.

C. Convert the array to a NumPy Array

In order to streamline data manipulation, the blob object is transformed into a NumPy array. NumPy arrays provide flexibility and efficiency, facilitating smooth compatibility with deep learning frameworks.

This conversion simplifies the management and manipulation of image data throughout subsequent processing phases.

Loading the Pre- Trained Model

A. Reading the Network Layer

Loading a pre-trained model entails comprehending its architecture through examination of the network layers. This process ensures alignment and understanding of the model's structure, thereby enabling subsequent fine-tuning or feature extraction tailored to specific tasks

B. Extracting the Output Layers

Upon loading the model, the output layers are retrieved. These layers encompass feature maps and class scores generated during the forward pass. Extracting the output layers is essential for acquiring predictions and gaining insights into the model's interpretation of the input image.

C. Appending Image Annotation files and Images

During this stage, image annotation files, which offer ground truth information, are associated with their corresponding images. This pairing establishes a comprehensive dataset essential for both model training and evaluation, allowing the algorithm to glean insights from annotated examples.

D. Converting BGR to RGB

Differences in color representation across libraries require converting the image from BGR to RGB. This adjustment ensures uniformity in color interpretation across various platforms, preparing the image for further processing and visualization.

E. Creating the Mask and Resizing the Image

A mask is created to emphasize areas of interest within the image, assisting in subsequent feature extraction. At the same time, the image is processed and molded into a dimension which is easily processable by the next layers, guaranteeing consistent input sizes for the model. This process is crucial for preserving uniformity and reliability across different datasets and scenarios.

F. Noise Removal

Noise Removal: Techniques are employed to eradicate errors or unwanted artifacts from the image, which could lead to unreliable outcomes. Gaussian blur, median blur, and bilateral filtering are commonly utilized methods for noise removal. These techniques aid in refining the image, diminishing distortions or irregularities in pixel values that might have arisen during image capture or processing. By eliminating these discrepancies, the image becomes more suitable for accurate analysis and classification.

Data Augmentation

A. Randomizing the Image

Data augmentation is essential for bolstering the resilience and diversity of training datasets utilized in machine learning models. An essential approach involves introducing variability into images through randomization, which encompasses applying random transformations. These transformations include adjustments to brightness, contrast, and color intensity, thereby exposing the model to a wider spectrum of visual scenarios.

Through randomization, the risk of overfitting is mitigated, as the model encounters diverse representations of the same object, enhancing its ability to generalize to unseen data.

B. Rotating the Image

Rotation serves as a crucial data augmentation technique, enriching the dataset's understanding of object orientations. Through the application of random rotation angles to images, the model gains insights into recognizing objects from diverse perspectives. This process enhances the model's capability to address real-world scenarios, where objects may manifest at different orientations. By implementing this augmentation technique, the model avoids over-dependence on specific object orientations inherent in the original dataset, fostering improved adaptability to new instances.

C. Transforming the Image

Transformation, which includes scaling, shearing, and flipping, introduces geometric variations to images during augmentation. This technique enriches the dataset by simulating diverse spatial relationships between objects. Scaling adjusts the size, shearing distorts shapes, and flipping horizontally or vertically creates mirrored versions. The model gains from exposure to these transformed instances, becoming more adaptable to variations in scale, shape, and orientation.

Overall, data augmentation, employing randomization, rotation, and transformation, strengthens machine learning models, enabling effective generalization to unseen data and enhancing their performance in real-world applications.

D. Feature Extraction

CNN, a machine learning technique, comprises two layers: an input layer and an output layer. Trained on a dataset of pre-labeled images, it can classify distinctive facial characteristics in the input data.

Notable visual features extracted in this project encompass the nose, lips, and eyes. These features manifest as points, exemplified by the distance between the eyebrows, the mouth's width, and the upper eyelid's height. Through analyzing these features, CNN can discern the emotions conveyed by the user in the image, encompassing happiness, sadness, or neutrality.

Functional Calculation

The positioning of an individual's nose, mouth, and eyes on their face is determined through the calculation of all extracted features in this stage. Face motion detection relies on this calculation.

Music Recommendation

The dataset of songs chosen for the recommendation of the music was found on Kaggle. Then we started looking for good cloud storage platform to store and retrieve the songs. Initially various famous cloud storage like Google cloud, Dropbox, AWS, etc. were considered but they quickly turned down as they were costly as well as providing limited storage options.

We also looked for several cheap options like Ampache and Mega were considered but their service was mainly web based.

Finally, we used Cloudflare Cloud service as it mainly is a backend server. It also provides a handsome free cloud storage.

Integration and Working

We integrated the camera feed with facial features extraction using expressions using Emgu.CV library. Following steps were taken to integrate the features:

Video Capture: we used 'VideoCapture' class from the Emgu.CV to initialize the camera and initialize capturing video frames. This function allows smooth capturing of frames from the camera.

Frame Processing: we used ‘ProcessFrame’ to process each captured frame. We performed various operation like face detection and facial features extraction.

Face Detection: we used ‘CascadeClassifier’ class of the Emgu.CV to perform face detection on each frame. It uses pre-trained Haar cascade classifier to enable face detection.

Face Region Extraction: We extracted a region of interest corresponding to detected face

Facial expression/ emotion recognition: we used pre trained facial expression recognition model to classify the facial expression in the extracted region. We used DnnInvoke.ReadNetFromTensorflow’ method from Emgu.CV.

Output: The output is the predicted facial expression which is displayed in the form of a printed text. This allows real time processing of facial emotion classification.

For the integration of the mood detection and music recommendation module, we used above steps. Once the integration is done, a project on Cloudflare is created which contains the songs loaded from Kaggle dataset. Then the Cloudflare database is linked to the windows application. We created a window application with intuitive user interface.

The Mean Absolute Error (MAE) is used as a statistical metric for assessing precision of prediction models [23]. It quantifies the average absolute difference between predicted and actual values within a dataset [23]. In the context of song recommendation systems, MAE offers a means to gauge the accuracy of the system's predictions. Essentially, it reflects the average absolute difference between the predicted and actual song recommendations across all test samples [24]. By providing a single numerical value, MAE encapsulates the overall performance of the model. A smaller MAE value indicates superior model performance.

3.5 The test

We proceed in the testing stage in the same manner as in the training step; the only thing that is different is that "test set" data is used as an example for input for testing i.e. to test the difference between actual and predicted value. It is known as Loss. The loss function used is MAE. The Mean Absolute Error (MAE) serves as a statistical metric for assessing the precision of prediction models. It quantifies the average absolute disparity between predicted and actual values within a dataset [23]. In the context of song recommendation systems, MAE offers a means to gauge the accuracy of the system's predictions. Essentially, it reflects the average absolute difference between the predicted and actual song recommendations across all test samples [24]. By providing a single numerical value, MAE encapsulates the overall performance of the model. A smaller MAE value indicates superior model performance.

We report the results, concentrating only on calculating the test loss for user-rated films in the test dataset. This guarantees that the basis of our evaluation is not previously seen data, resulting in a more precise appraisal of the system's performance with novel inputs. By using data that the system hasn't seen during training, the procedure makes sure that its capabilities are evaluated, providing information about its generalization and predictive capabilities.

CHAPTER 4
RESULTS AND ANALYSIS

It is very hard to predict actual human emotion and mood as each face has unique facial expressions and features. But since some of the facial features are same, with proper facial expression algorithm and training it can be guessed up to a certain extent. In our case we have used the algorithm and used the dataset having quality labelled data. With poor quality of the images, most of the time model fails to work if provided with bad quality images. One of the things that user can do is use a high-resolution camera so that result is not negatively impacted by the low resolution. We have used the window application and laptop cameras are generally subpar as compared to their mobile counterparts. Sufficient lighting is used to increase the srtructure of the input image of the windows webcam. We have successfully developed and runs the applications and captured following screenshots.

Expression	Description	Primary Emotion
Raised Eyebrows	Brows are raised, eyes are wide open	Surprise
Smiling	Mouth is curved upward, eyes may wrinkle	Happiness
Frowning	Frowning	Sadness
Rolled Eyes	Eyes are rolled upwards	Disgust

Table 4.1 Facial Feature

Features	Description
Tempo	Speed of the songs
Rhythm	The pattern of the beats in the song
Melody	The actual tuning of the song
Harmony	combination of musical notes played at same time
Lyrics	The words of a song
Instrumentation	Actual instrument played in the song
Loudness	volume of the song
Energy	The perceived intensity or activity level of the song

Table 4.2 Song Features

Algorithm		Accuracy	Precision	Recall	F1 Score
CNN (Proposed Model)		0.835	0.862	0.827	0.844
SVM		0.884	0.876	0.865	0.871
KNN		0.78540	0.79	0.78	0.77

Table 4.3. Accuracy of the Algorithms

This research presents a music recommendation system that works on the feelings captured to use the user's face using a camera. The system uses OpenCV for capturing and preprocessing images, Haar Cascade Classifier for face detection, and CNN for extracting the features and emotion [16]. Based on the detected emotion, the system gives a playlist of songs to listen to [16]. The goal of this system is to improve user interaction with the system and provide a form of never had experience for the user.

Fig 2 depicts the home page of our application Mood Musica. Users are asked to register for the application and login if already the member. Fig 3 shows the user in happy mood and ‘happy face detected’ msg printed. Once the user clicks on the music, a playlist is shown having happy emotion. Fig 3 shows same for the sad emotion.

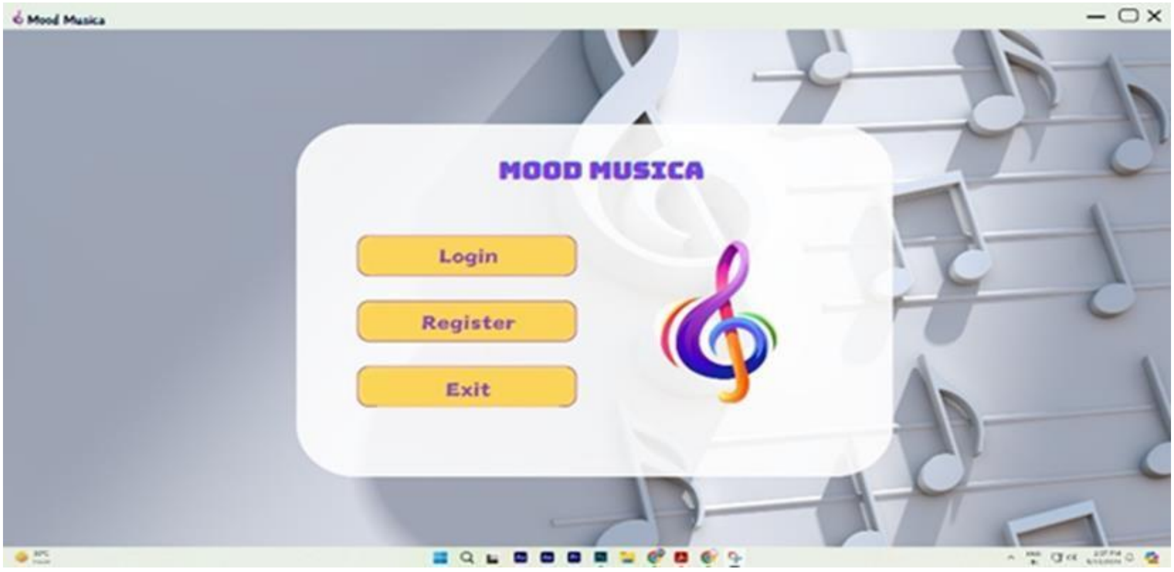


Fig. 4.1 Mood Musica Home Page

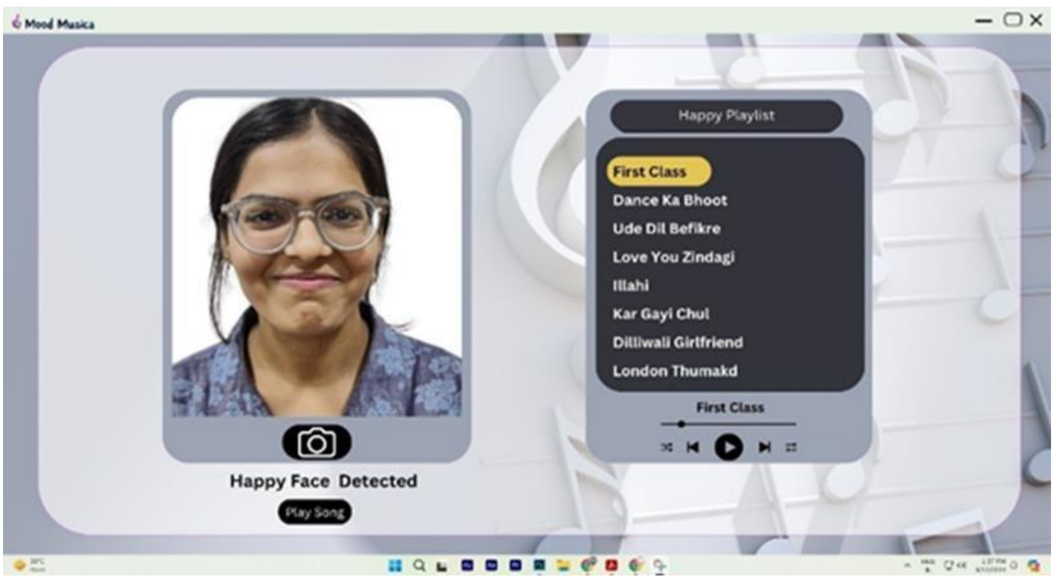


Fig 4.2 Happy Face Detected

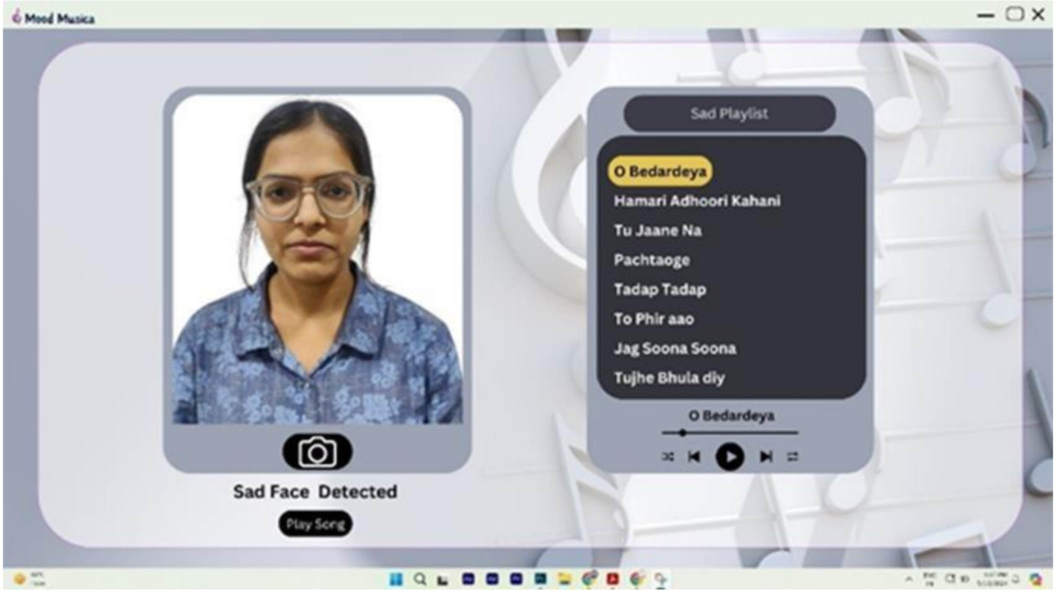


Fig 4.3 Sad Face Detected

When the user opens the app for the first time, the app asks for Registration/ Login Credentials. Once user confirms, an interface showing the camera snap button which is integrated to the system camera app.

Once the user clicks on the snap button, user is taken to camera app to take the picture. The picture is then forwarded to the face detection program. If there is no face then error is printed to the screen to tell the user that no face is there for consideration. If face is detected it is provided to mood detection algorithm. Algorithm classify the expression and then after paly music button is clicked, music playlist is displayed to the user.

Overall, the proposed system aims to deliver an automatic music playlist designed to positively enhance a person's mood. If the system detects negative emotions, it will offer a curated playlist featuring music types most likely to improve the user's experience. By analyzing the facial expressions and emotions, the system can provide personalized music recommendations tailored to their specific needs.

CHAPTER 5
CONCLUSION & FUTURE SCOPE

We have seen in this paper that various study has already been done successfully in music recommendation based on user's emotion captured through facial expression. We have seen different technologies like computer vision, and many more technologies been deployed in order to capture the emotion and the successfully recommending the music that relays on the current personal mood of the user. Different algorithms are deployed and we saw how different algorithm integrate different facial data and capture such data to capture the true emotion of the user. We can totally create a whole new application or can integrate such technologies not the existing platform like YouTube or google etc. Also, many more data collection technologies must be inferred to further advance the user's satisfaction. Also, user interface design must be user friendly and process of capturing of the facial emotion must be minimal i.e. user won't want to carry a separate device for emotion capturing. This transitioning of emotion capturing should be smooth. Ethical concerns such as privacy must also be adhered to such that no data must be collected or used without the consent of the user as privacy infringement and data misuse are not new things in today's time. There is still a load of work not just in this field but facial expression can have a wide range of application on other fields too. With continuous research and advancement in the field, we can create an idol system that recommend music based on the dynamic emotion of the user. We have developed a music recommended system, Mood Musica, which recommends music based on the emotion of the user. It captures the emotion of the user by analyzing the facial expression which are like the gateway to one's mood and emotion. Once the emotion is detected, underlying algorithm is able to return a playlist based on the detected emotion. Although capturing emotion is very hard it can be done if model is trained accurately to determine a set of emotions. Our model was able to give an accuracy of the 82% on the validation set.

We were able to detect four emotions related to the extracted features of the facial emotion. These include – happy, sad, neutral, and angry. For the further work, we can look to integrate more emotions like surprise and fear. Also, it is very hard to predict the current emotion of the user just by seeing the facial expression. We need more features like live heart rate and body temperatures can be used. These along with facial expression can also leads to better detection of the mood of the user leading to accurate results.

REFERENCES

[1] Anand R, Sabeenian RS, Deepika Gurang, Kirthika R, Shaik Rubeena. "AI-based Music Recommendation system using Deep Learning Algorithms." International Journal of Innovative Research in Computer and Communication Engineering 2019.

[2] Shantha Shalini K, Jaichandran R, Leelavathy SA, Raviraghul R, Ranjitha J, Saravanakumar. "Facial Emotion Based Music Recommendation System using computer vision and machine learning techniques." International Journal of Engineering Research & Technology 2020.

[3] Ankita Mahadik et al. "Mood-based music recommendation system." International Journal of Computer Applications 2018.

[4] Aman Tamboli et al. "Music Recommendation System." International Journal of Innovative Research in Computer Science & Technology 2017.

[5] Xinglin Wen. "Using deep learning approach and IoT architecture to build the intelligent music recommendation system." IEEE International Conference on Big Data and Smart Computing 2020.

[6] Xiao Han, Fuyang Chen, Junrong Ban. "Music Emotion Recognition Based on a Neural Network with an Inception-GRU Residual Structure." IEEE Access 2021.

[7] Ferdos Fessahaye et al. "T-RECSYS: A Novel Music Recommendation System Using Deep Learning." IEEE Transactions on Neural Networks and Learning Systems 2022.

[8] Puri, Raghav & Gupta, Archit & Sikri, Manas& Tiwari, Mohit& Pathak, Nitesh & Goel, Shivendra. (2020). Emotion Detection using Image Processing in Python.

[9] Djekoune, O., Aiadi, O., & Koudil, M. (2020). Deep Learning for Emotion Recognition on Small Datasets using Transfer Learning. 2020 5th International Conference on Advanced Technologies for Signal and Image Processing (ATSIP), 1-6.

[10] Giro-i-Nieto, X., Burghardt, T., & Solà, J. (2017). Emotion Recognition in the Wild via Convolutional Neural Networks and Mapped Binary Patterns. 2017 IEEE International Conference on Computer Vision Workshops (ICCVW), 2736-2744.

[11] Rahim, A., Rahman, R. S., Ahmed, S. F., & Ahmed, M. I. (2020). A Facial Expression-Based Music Recommender System Using Deep Learning. IEEE Access, 8, 66031-66039.

[12] Kumar, S., & Rajput, S. (2020). Emotion-Based Music Recommendation Using Fusion of Content- Based and Collaborative Filtering Techniques. Journal of Information Processing Systems, 16(5), 1199- 1212.

[13] Song, Y., Li, L., & Martin, T. P. (2020). Music Emotion Recognition: A State-of-the-Art Review. IEEE Transactions on Affective Computing, 11(3), 587-601.

[14] Ramanathan R; Kumaran R; Rohan R Gupta; R; Prabhu V (2017) "An Intelligent Music Player Based on Emotion Recognition", 2nd International Conference on Computational Systems and Information Technology for Sustainable Solution (CSITSS).

[15] Chankuptarat K; Sriwatanaworachai R; Chotipant S,(2019) "Emotion- Based Music Player",5th International Conference on Engineering, Applied Sciences and Technology (ICEAST).

[16] Gilda S, Zafar H, Soni C and Waghurdekar K (2017) "Smart Music Player Integrating FacialEmotion Recognition and MusicMood Recommendation", IEEE.

[17] Song Y, Dixon S, Pearce M, (2012) "Evaluation Of Musical Features For Emotion Classification", University of London, ISMR.

[18] Abdat F; Maaoui C; Pruski A, (2011) "Human-Computer Interaction Using Emotion Recognition from Facial Expression", UKSim 5th European Symposium on Computer Modeling and Simulation.

[19] Bhardwaj A; Gupta A; Jain P; Rani A; Yadav J (2015) "Classification of human emotions from EEG , 2nd International Conference on Signal Processing and Integrated Networks (SPIN).

[20] Laurier C and Herrera P, Cloud M: A Real-Time Music Mood Visualization Tool, Music Technology Group Universitat Pompeu Fabra Ocata 1, 08003 Barcelona, Spain.

[21] Hirve1 R, Jagdale2 S, Banthia3 R, Kalal4 H & Pathak5 KR, (2016)"EmoPlayer: An Emotion Based Music Player", Imperial Journal of Interdisciplinary Research (IJIR) Vol-2, Issue-5.

[22] Pujari, R. K., Acharya, T., & Ramachandra, K. (2020). Emotion-based music recommendation using facial expression recognition. In 2020 International Conference on Artificial Intelligence in Information and Communication (ICAIIC) (pp. 176-181). IEEE.

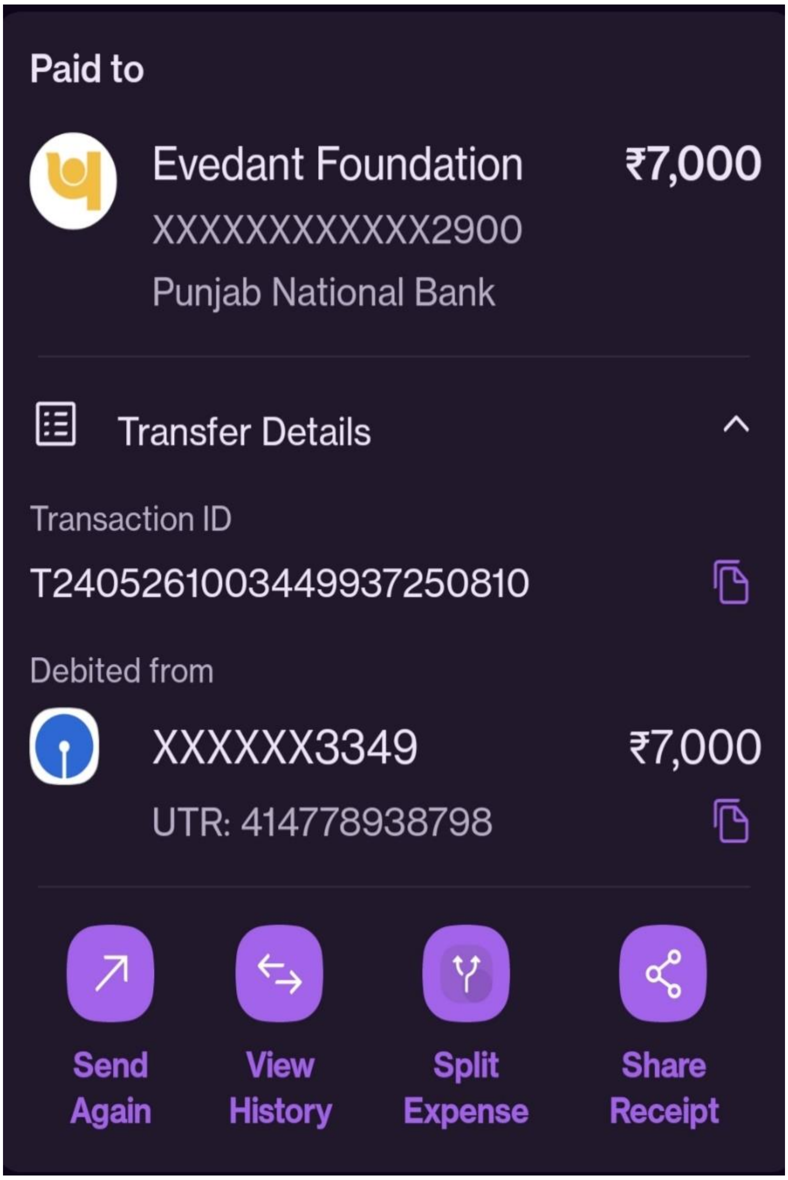
[23] Sajjan, V., & Nayak, J. (2019). A hybrid emotion-based music recommendation system. In 2019 International Conference on Vision Towards Emerging Trends in Communication and Networking (ViTECoN) (pp. 1-6). IEEE.

[24] Kumar, S., Sagar, V. & Punetha, D. A comparative study on facial expression recognition using local binary patterns, convolutional neural network and frequency neural network. Multimed Tools Appl (2023). https://doi.org/10.1007/s11042-023-14753-y.

<u>LIST OF PUBLICATION AND THEIR PROOF</u>

1. Kamaksha, "Facial Notes: Unlocking Music Recommendations through Expression (A Contextual Analysis)", Accepted at **"International Conference on Intelligent Computing and Communication Techniques (ICICCT)"**, at JNU New Delhi, India.

   Paper Id: 660



Gmail

kamaksha <kamaksha.d31@gmail.com>

Notification of acceptance of paper id 660

1 message

Microsoft CMT <email@msr-cmt.org> Fri, May 10, 2024 at 1:31 PM
Reply-To: ICICCT Conference <contactus@icicct.in>
To: "kamaksha ." <kamaksha.d31@gmail.com>

Dear Dr./ Prof. kamaksha .,

Congratulations...

Your paper / article paper id 660: Facial Notes: Unlocking Music Recommendations through Expression (A Contextual Analysis) has been accepted for publication in International Conference on Intelligent Computing and Communication Techniques at JNU New Delhi, India.

Kindly save your paper by given paper id only (eg. 346.docx, 346.pdf, 346_copyright.pdf)

Registration Link:

https://forms.gle/mSsHa8GMLtMkWuaq8

Please ensure the following before registration and uploading camera ready paper.
1. Paper must be in Taylor and Frances Format.
Template and copyright with author instruction are given in below link: https://icicct.in/author_inst.html

2. Minimum 12 references should be cited in the paper and all references must be cited in the body. Please follow the template.
3. The typographical and grammatical errors must be carefully looked at your end.
4. Complete the copyright form (available at template folder).
5. The regular fee (Available in registration section) will be charged up to 6 pages and after that additional Rs.1000 for Indian authors / 10 USD for foreign authors per additional page will be charged.
6. Reduce the Plagiarism below 10% excluding references and AI Plagiarism 0%.
Each Illustration must include a caption and an alternative text description to assist print impaired readers ('Alt Text').
(Alt Text is mandatory for each Illustrations)
7. Certificate will be issued by the name of registered author (Single author only).
8. Certificates may be issued to all other authors on the extra payment of 1000/- INR per author.
9. Last Date of registration and uploading copyright and camera-ready copy: 31/05/2024.
10. Make a single payment which includes registration fee + Extra certificates fee + Extra page fees.
11. Permissions: Kindly make sure the permissions for each copyrighted artwork file have been cleared ahead of the submission, with the details listed in the Permission Verification form (attached). All permission grants must be submitted along with your final manuscript.
Figures: Please make sure no figures are missing, and all figures are high resolution
Tables: Please ensure that there are no missing tables, and the tables in your manuscript are not pasted as figures
Citation: Kindly ensure there are no missing citations in your manuscript

Registration Link: https://forms.gle/mSsHa8GMLtMkWuaq8

Registration Fee to be deposited in below account

Bank Account Details :

Indian Account Details:

Account Holder Name: EVEDANT Foundation

Account Number: 0674002190422900

IFSC Code: PUNB0067400
SWIFT Code: PUNBINBBGNM

Branch: Punjab National Bank, Navyug Market, Ghaziabad

Account Type: Current Account

International Conference on Intelligent Computing and Communication Techniques
28 & 29 June 2024.

Thanks and Regards
Convener
International Conference on Intelligent Computing and Communication Techniques
contact details : icicctcon@gmail.com, https://icicct.in/index.html

1. Kamaksha, “Mood Musica: Music System based on facial Expressions” Accepted at **“National Conference on Advanced Computer Science and Information Technology (NCACSI - 24)”,** Udaipur, India.

Paper Id: National Conference_7239168

Dear kamaksha,

We are pleased to inform you that your online payment transaction has been successfully processed. The details of your payment are as follows:

Order ID: 6652bcb7bee87

Reference ID: 113301952992

Amount: 4875.00 INR

Name: kamaksha

Email ID:
kamaksha.d31@gmail.com

Phone Number: 9971208406

Thank you for choosing **National Conference**
. We appreciate your prompt payment.

For assistance or inquiries, feel free to reach out to us at info@nationalconference.in or **+91 9677007228**.

Best Regards,
National Conference
info@nationalconference.in
+91 9677007228

Gmail

kamaksha <kamaksha.d31@gmail.com>

Acceptance Letter for Your Research Article - National Conference on Advanced Computer Science and Information Technology (NCACSI - 24)

1 message

National Conference <info@nationalconference.in> Tue, May 21, 2024 at 3:34 PM
To: kamaksha.d31@gmail.com

Dear **Kamaksha** ,

Greetings!

We are delighted to inform you that your research article titled **"Mood Musica: Music System based on facial Expressions"** has been accepted for presentation at the upcoming **"National Conference on Advanced Computer Science and Information Technology (NCACSI - 24)"** scheduled for **15th June 2024 in Udaipur, India**.

To progress further, we kindly ask you to move on to the conference registration phase through the provided link:

Conference Registration Link: https://www.nationalconference.in/event/registration.php?id=2482541

For an immediate registration experience, our team is excited to offer you a special coupon code "**INSTANT250**" providing an **250 INR** concession on your registration fee.

Act promptly, as this offer remains valid until the **25th May 2024**. Ensure your registration is completed, and the coupon code is applied before the deadline for maximum savings.

Here's a concise guide on redeeming your coupon code:

Step 1: Click "Register Now," select the relevant fee, and complete the form with your details.

Step 2: Enter the online gateway, furnish the necessary information, confirm, and proceed to the payment checkout.

Step 3: Locate the coupon code section on the CC Avenue checkout page. Enter the coupon code given above and click "Apply" to access the discount.

Step 4: Input your card details and finalize the authentication process.

Payment Confirmation: After completing the registration, share the payment confirmation with us.

For any questions or assistance, feel free to contact us via email. You can also reach us at **+91 9677007228.**

--

Thanks & Regards,

Raj Kumar

Coordinator

National Conference

info@nationalconference.in

www.nationalconference.in

Contact: +919677007228

NATIONAL CONFERENCE

National Conference_7239168.pdf
1441K

Simi

PAPER NAME

kamaksha_2k22cse10.pdf

WORD COUNT **9760 Words**	CHARACTER COUNT **52170 Characters**
PAGE COUNT **42 Pages**	FILE SIZE **1.1MB**
SUBMISSION DATE **May 23, 2024 8:54 PM GMT+5:30**	REPORT DATE **May 23, 2024 8:55 PM GMT+5:30**

6% Overall Similarity

The combined total of all matches, including overlapping sources, for each database.

- 3% Internet database
- 0% Publications database
- 5% Submitted Works database

Excluded from Similarity Report

- Crossref database
- Crossref Posted Content database
- Bibliographic material
- Cited material
- Small Matches (Less then 8 words)

kamaksha_2k22cse10.pdf

Delhi Technological University

Document Details

Submission ID
trn:oid:::27535:59850342

Submission Date
May 23, 2024, 8:54 PM GMT+5:30

Download Date
May 23, 2024, 8:57 PM GMT+5:30

File Name
kamaksha_2k22cse10.pdf

File Size
1.1 MB

42 Pages

9,760 Words

52,170 Characters

How much of this submission has been generated by AI?

0%

of qualifying text in this submission has been determined to be generated by AI.

Caution: Percentage may not indicate academic misconduct. Review required.

It is essential to understand the limitations of AI detection before making decisions about a student's work. We encourage you to learn more about Turnitin's AI detection capabilities before using the tool.

Frequently Asked Questions

What does the percentage mean?
The percentage shown in the AI writing detection indicator and in the AI writing report is the amount of qualifying text within the submission that Turnitin's AI writing detection model determines was generated by AI.

Our testing has found that there is a higher incidence of false positives when the percentage is less than 20. In order to reduce the likelihood of misinterpretation, the AI indicator will display an asterisk for percentages less than 20 to call attention to the fact that the score is less reliable.

However, the final decision on whether any misconduct has occurred rests with the reviewer/instructor. They should use the percentage as a means to start a formative conversation with their student and/or use it to examine the submitted assignment in greater detail according to their school's policies.

How does Turnitin's indicator address false positives?
Our model only processes qualifying text in the form of long-form writing. Long-form writing means individual sentences contained in paragraphs that make up a longer piece of written work, such as an essay, a dissertation, or an article, etc. Qualifying text that has been determined to be AI-generated will be highlighted blue on the submission text.

Non-qualifying text, such as bullet points, annotated bibliographies, etc., will not be processed and can create disparity between the submission highlights and the percentage shown.

What does 'qualifying text' mean?
Sometimes false positives (incorrectly flagging human-written text as AI-generated), can include lists without a lot of structural variation, text that literally repeats

DELHI TECHNOLOGICAL UNIVERSITY

(Formerly Delhi College of Engineering)
Shahbad Daulatpur, Main Bawana Road, Delhi-42

PLAGIARISM VERIFICATION

Title of the Thesis **MUSIC RECOMMENDATION SYSTEM EMPLOYING FACIAL EMOTION RECOGNITION**

Total Pages **42** Name of the Scholar **Kamaksha**

Supervisor **Dr Vinod Kumar**

Department **Computer Science and Engineering**

This is to report that the above thesis was scanned for similarity detection. Process and outcome is given below:

Software used: **Turnitin** Similarity Index: **6%** Total Word Count: **9760**

Date

Candidate's Signature **Signature of Supervisor**