

TABLE OF CONTENTS

Abstract.....	I
List of Figures.....	II
List of Tables.....	III
List of Screens.....	IV
Symbols and abbreviations.....	V

S.No	Content		Page no.
1.	INTRODUCTION		1-4
	1.1 Introduction		1
	1.2 Aim of the project		1
	1.3 Project Domain CNN		1
	1.4 Scope of the Project		2
	1.5 Methodology		2
	1.5.1	Open CV	2
	1.5.2	Haar Cascade Classifier	3
	1.5.3	Convolutional Neural Network	3
2.	LITERATURE SURVEY		5-11
	2.1 Introduction		5
	2.2 Existing System		7
	2.3 Disadvantages of Existing system		7
	2.4 Proposed System		8
	2.5 Conclusion		11
3.	ANALYSIS		12-16
	3.1 Introduction		12
	3.2 Feasibility Study		12
	3.2.1	Economic Feasibility	12
	3.2.2	Technical Feasibility	13
	3.2.3	Social Feasibility	13

	3.3	Software Requirement Specification	13
	3.3.1	Hardware Specification	13
	3.3.2	Software Specification	14
	3.3.3	Standards and Policies	14
	3.4	General Architecture	16
4.	DESIGN		17-23
	4.1	Introduction	17
	4.2	Design Phase Diagrams	17
	4.2.1	Data Flow Diagram	17
	4.2.2	UML Diagram	18
	4.2.3	Class Diagram	19
	4.2.4	Sequence Diagram	20
	4.3	Module Description	21
	4.3.1	Audio Emotion Recognition	22
	4.4	Conclusion	23
5.	IMPLEMENTATION AND RESULTS		24-31
	5.1	Introduction	24
	5.2	How does the Feature Extractor work?	24
	5.3	Input and Output	25
	5.3.1	Input Design	25
	5.3.2	Output Design	26
	5.4	Code Implementation	27
	5.4.1	Sample Code	27
	5.4.2	Output Screens	29
	5.5	Conclusion	31

6.	TESTING AND VALIDATION	32-38
	6.1 Introduction	32
	6.2 Types of Testing	32
	6.2.1 Unit testing	32
	6.2.2 Integration testing	33
	6.2.3 Functional testing	34
	6.2.4 White Box Testing	36
	6.2.5 Black Box Testing	37
	6.2.6 Test Result	37
	6.3 Testing Strategy	38
	6.4 Conclusion	38
7.	CONCLUSION	39-40
	7.1 Project Conclusion	39
	7.2 Future Enhancements	40
8.	REFERENCES	41-42
	8.1 Referred Books	41
	8.2 Referred Sites	41
	8.3 References	41

6.	TESTING AND VALIDATION	32-38
	6.1 Introduction	32
	6.2 Types of Testing	32
	6.2.1 Unit testing	32
	6.2.2 Integration testing	33
	6.2.3 Functional testing	34
	6.2.4 White Box Testing	36
	6.2.5 Black Box Testing	37
	6.2.6 Test Result	37
	6.3 Testing Strategy	38
	6.4 Conclusion	38
7.	CONCLUSION	39-40
	7.1 Project Conclusion	39
	7.2 Future Enhancements	40
8.	REFERENCES	41-42
	8.1 Referred Books	41
	8.2 Referred Sites	41
	8.3 References	41

ABSTRACT

Human expression plays a vital role in determining the current state and mood of an individual, it helps in extracting and understanding the emotion that an individual has based on various features of the face such as eyes, cheeks, forehead or even through the curve of the smile. Music is basically an art form that soothes and calms human brain and body. Taking these two aspects and blending them together our project deals with detecting emotion of an individual through facial expression and playing music according to the mood detected that will alleviate the mood or simply calm the individual and can also get quicker song according to the mood, saving time from looking up different songs and parallel developing a software that can be used anywhere with the help of providing the functionality of playing music according to the emotion detected. By developing a recommendation system, it could assist a user to make a decision regarding which music one should listen to helping the user to reduce his/her stress levels. The image of the user is captured with the help of a webcam. The user's picture is taken and then as per the mood/emotion of the user an appropriate song from the playlist of the user is shown matching the user's requirement.

LIST OF FIGURES

S.No	Figure No	Description	Page No
1	1.5.3	Convolutional Neural Network	4
2	2.4	Emotion Classification	10
3	3.4	Architecture Diagram	16
4	4.2.1	Data Flow Diagram	17
5	4.2.2	UML Diagram	18
6	4.2.3	Class Diagram	19
7	4.2.4	Sequence Diagram	20

LIST OF TABLES

S.No	Table No	Description	Page No
1	6.2.6	Test Results	37

LIST OF SCREENS

S.No	Screen Shot No	Description	Page No
1	5.3.1	Input Design of Project	25
2	5.3.2	Output Design of Project	26
3	5.4.2.1	Output of Detected Camera	29
4	5.4.2.2	Output of Music	30
5	6.2.3	Functional Testing	36

LIST OF ACRONYMS AND ABBREVIATIONS

- CNN -- Convolutional Neural Network
- Open CV -- Open-Source Computer Vision Library
- PCA -- Principal Component Analysis
- HCI -- Human Computer Interface
- AER -- Audio Emotion Recognition
- MIR -- Music Information Retrieval
- PCM -- Pulse Code Modulation
- RNN -- Recurrent Neural Network
- KHz -- Kilo Hertz

1. INTRODUCTION

1.1 Introduction

Photo processing is a way where in the pixels are pre-processed and the capabilities are extracted out by using appearing diverse operations. There are literally styles of picture processing specifically, digital and analog photo processing. The analog pix are pictures which are in hard copies.

Example: scientific reports and pictures. Virtual Photo processing uses diverse algorithms to perform photograph manning on various digital pics. There are some strategies worried in virtual image processing. They are photograph enhancing, photo restoration, independent thing analysis linear filtration. Photograph processing also consists of exclusive steps like photograph enhancement restoration, and many others. Music is a melody that connects the soul and mind of the character together. It performs a critical function in human lifestyles. Feelings affect us physically as well as mentally. Our body reacts to one of a kind emotional state. The sturdy feelings are brought out by using listening to the track according to their scenario.

1.2 Aim of the project

Our project aims to recommend and play the appropriate music, based on user's current emotional state with the help of image processing techniques through user's facial expressions.

1.3 Project Domain CNN

A standout amongst the most prevalent types of deep learning techniques encompasses CNN. It makes use of a framework much like a multilayer perceptron that has been meant for reduced processing requirements. As it extracts capabilities without delay from photos, the need for guide function extraction is removed. The detection of various functions of a photograph is learnt by using CNN lets to loads of hidden layers every of which increases the complexity of the learnt image capabilities. This computerized function extraction makes models of deep studying notably correct for laptop vision duties together with item class. This software has capabilities enables to efficiently capture the spatial and temporal dependencies in a photo.

1.4 Scope of the Project

Facial expressions are an incredible indicator of the country of a mind for a person. Indeed, the maximum natural way to specific emotions is thru facial expressions. Humans have a tendency to hyperlink the track they listen to, to the emotion they're feeling. The song playlists even though are, at times too massive to sort out robotically. It'd be beneficial if the track player changed into "smart sufficient" to sort out the music based totally on the modern-day state of emotion the person is feeling. The project units out to use diverse techniques for an emotion popularity gadget, analyzing the impacts of different techniques used. By means of the use of Emo participant we will effortlessly play the songs in step with the emotion of the user.

1.5 Methodology

1.5.1 Open CV

Open CV (Open supply laptop imaginative and prescient Library) is a library of programming capabilities specifically targeted on actual- time computer vision. It smiles a C++ implementation library. There is a java cv library that's derived from Open CV using this we can put into effect viola and Jones face detection algorithm. Face detection is vital as it will classify most effective if face is present. Expression reputation is also achieved using graphical based totally type approach.

Audio documents might be scanned and capabilities could be extracted from them and in step with the mood we get the playlist. The library is go- platform. It especially geared toward actual-time picture processing. The performance of library can be multiplied if Intel's included performance Primitives is set up at the system. It occurs due to the presence of proprietary optimized workouts.

Open CV gives multiple numbers of capabilities for face recognition and facial detection. Open CV comes with a teacher in addition to detector. if you need to teach your personal classifier for any object like mobile, pen etc. you can use Open CV to create one.

1.5.2 Haar Cascade Classifier

Object Detection the usage of Haar function-based totally cascade classifiers is an effective item detection approach proposed by using Paul Viola and Michael Jones of their paper, " rapid object Detection the use of a Boosted Cascade of simple capabilities" in 2001. It's far a gadget learning based totally technique where a cascade feature is trained from numerous positive and terrible pix. It's far then used to discover items in other pictures. For every feature, it reveals the satisfactory threshold that allows you to classify the faces to superb and bad. Glaringly, there may be errors or mis classifications. We choose the functions with minimum mistakes fee, because of this they are the features that most as it should be classify the face and non-face photos.

For this they brought the idea of Cascade of Classifiers. In preference to making use of all 6000 features on a window, the functions are grouped into extraordinary tiers of classifiers and carried out one-by means of-one. If a window fails the first degree, discard it. We don't forget the continue to being capabilities on it. If it passes, practice the second degree of capabilities and maintain the system. The window which passes all tiers is a face region. The writer detector had 6000+ capabilities with 38 tiers with 1, 10, 25, 25 and 50 functions in the first five levels. An average 10 capabilities out of 6000+ are evaluated in line with sub-window.

1.5.3 Convolutional Neural Network

CNN is a neural community that has one or more convolutional layers and are used particularly for image processing, classification, segmentation and also for different automobile correlated facts. A convolution is largely sliding a clear out over the input.

It is processed below deep studying version; for processing facts that has a grid sample. Designed to mechanically and adaptive study spatial hierarchies of capabilities, from low- to high-level styles. It smiles a mathematical construct this is usually composed of three varieties of layers i.e., Convolution, Pooling, and absolutely linked Layers. The Convolution and Pooling, perform feature extraction. The fully related Layer, maps the extracted features into final output, such as classification. A convolution layer plays a key role in CNN, which consists a specialized sort of linear operation.

CNNs have been used for understanding in NLP and speech popularity, despite the fact that regularly for RNN are used. Pooling and Padding are the critical capabilities enables CNN to achieve in photo processing.

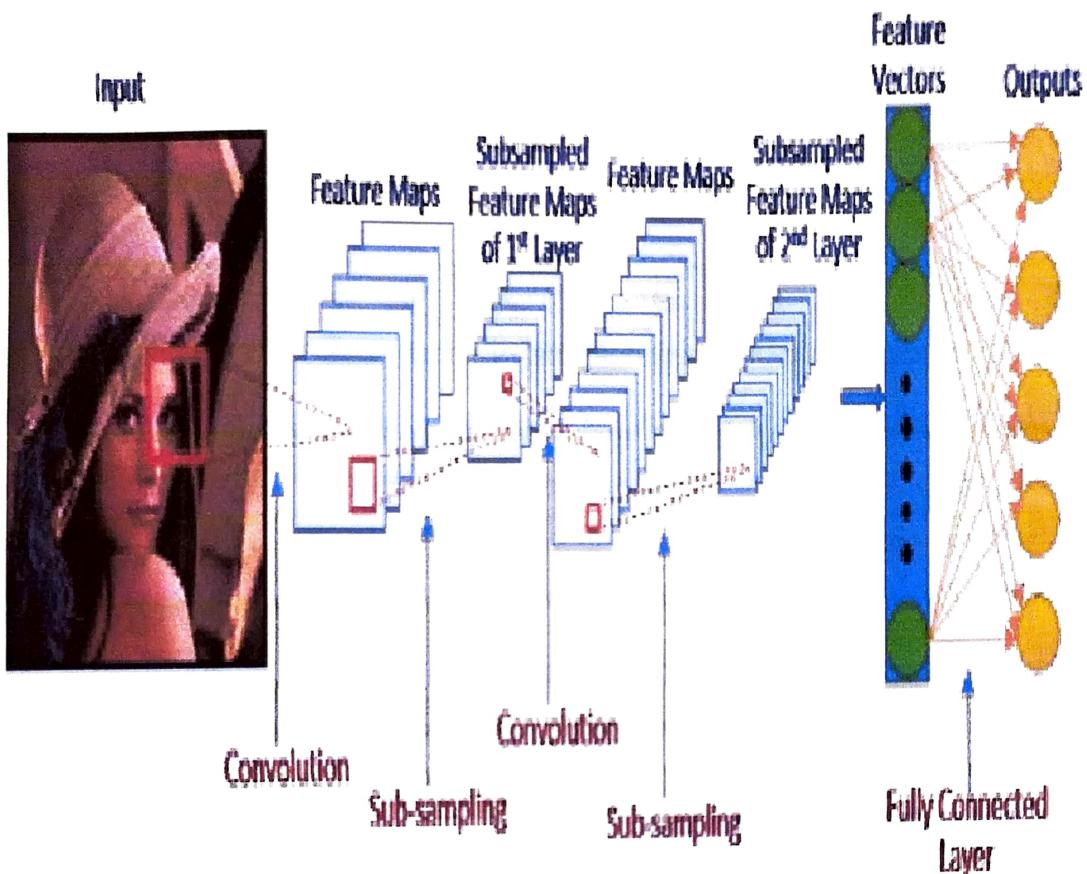


Figure 1.5.3: **Convolutional Neural Network**

2. LITERATURE SURVEY

2.1 Introduction

A literature survey is a piece of discursive prose, not a list describing or summarizing one piece of literature after another. It is an iterative process, assessing and distilling information. One of the key purposes of the literature survey is to investigate a problem that no one else has addressed.

S Metilda Florence and M Uma (2020),"Emotional Detection and Music Recommendation System based on User Facial Expression" Research paper (2020)

A classifier that is used to detect or obtain the facial landmarks from the face of the user is trained on HELEN dataset. HELEN dataset contains more than 2000 images. The network is trained with the help of CK extensive data set. The system still is not able to record all the emotions correctly due to the less availability of the images in the image dataset being used.

Krittrin Chankuptarat; Raphatsak Sriwatanaworachai; Supannada Chotipant Emotion-Based Music Player Research paper (2019 5th International Conference on Engineering, Applied Sciences and Technology (ICEAST))

This paper proposes an emotion-based music player, which is able to suggest songs based on the user's emotions; sad, happy, neutral and angry. The application receives either the user's heart rate or facial image from a smart band or mobile camera. It then uses the classification method to identify the user's emotion. This paper presents 2 kinds of the classification method; the heart rate-based and the facial image-based methods. Then, the application returns songs which have the same mood as the user's emotion.

K. S. Nathan, M. Arun and M. S. Kannan," EMOSIC — An emotion-based music player for Android," 2018 IEEE International Symposium on Signal Processing and Information Technology (ISSPIT), Bilbao, 2018, pp. 371-276.

Nathan proposed a very useful two-dimensional model (stress energy v / s), plotted on two axes and whose emotions are represented by a two-dimensional coordinate system based on two axes or on the four quadrants, which is represented by the two-dimensional diagram be formed.

The musical mood names and AV values of a total of 20 subjects were tested and analyzed in the work of Kannan. Based on the results of the analysis, the aircraft AV was divided into 8 regions (clusters), which illustrate the mood using an efficient data mining algorithm for k-means clusters.

Abat, F. Maaoui, C., and A. Prusk. Human-computer interaction using emotion recognition from facial expression. 2017 UK- Sim 5th European Symposium on Computer Modelling and Simulation (2017).

This paper describes emotion recognition system based on facial expression. A fully automatic facial expression recognition system is based on three steps: face detection, facial characteristic extraction and facial expression classification. We have developed an anthropometric model to detect facial feature points combined to Shi & Thomasi method. The variations of 21 distances which describe the facial features deformations from the neutral face, were used to coding the facial expression. Classification step is based on SVM method (Support Vector Machine). Experimental results demonstrate that the proposed approach is an effective method to recognize emotions through facial expression with an emotion recognition rate more than 90% in real time. This approach is used to control music player based on the variation of the emotional state of the computer user.

Wang, A., Chen, G., Yang, J., Zhao, S., Chang, C. Y. (2016). A comparative study on human activity recognition using inertial sensors in a smartphone, IEEE Sensors Journal, 16(11), 4566- 5 4578.

The aim of the history is to help situate the scientific community in the state of the art for the HAR area in the context of smartphones and to present a motivation for the planning and execution of the next steps that will help define the future evolutionary milestones of the area. This article presented a detailed description of each step of the methodology commonly used to recognize human activities with smart phones equipped with inertial sensors. The steps his paper works from the literature, along with tips for the best practices, are presented. In particular, issues related to the features used in classification models were highlighted. Based on his own theory proposed, the approach to extraction of features is done by manually.

2.2 Existing System

- Mood Player

This app makes use of face detection and temper reputation to work- out the person's temper and supported this; it offers a personalized play listing. The face detection rule is predicated on Open CV library and additionally the mood detection half of is going to be based on pattern matching. If we all understand he know-how that is required, we generally tend to use the remaining. Information that joins every song with tags that describe it. These implementations are rectangular degree designed so one can get a list in keeping with the person temper's and provide these functionalities:

1. Set your mood manually i.e., Happy or Sad
2. Analysing your mood periodically by capturing camera
3. Set tempo of the music from Calm to Energetic.

- Stereo Mood

Stereo temper can be a cellular or pill utility. With the press of a button, we'll have a ready-made play list for every time in our life. We'll choose our temper from our tags, concentrate, find out new music, share and tag our emotions in tune. This application gives the subsequent functionalities:

1. Sharing your mood through social network like WhatsApp, Snapchat.
2. Labelling the particular song with different user defined names.
3. Upgrading the User's profile with its unique feature algorithm.

2.3 Disadvantages of Existing system

1. It requires the user to manually select the songs.
2. Randomly played songs may not match to the mood of the user.
3. User has to classify the songs into various emotions and then for playing the songs user has to manually select a particular emotion proposed System.

The proposed system tries to provide an interactive way for the user to carry out the task of creating a playlist. the working is based on different mechanisms carrying out their function in a predefined order to get the desired output. The working can be stated as follows:

1. Proposed system upgrades with a unique UI which enables the user to get files of audio.
2. Files are being detected, they are scanned for audio features which it's extracted via Internet or Author's gallery.
3. The extracted feature values are subjected to label the classified feature value with its parameters provided.
4. These parameters include a limited set of genre types based on which the audio feature values will be processed.
5. After this, the songs are segregated into different playlists based on the feature extraction process. Hence lists of similar sounding songs or songs belonging to similar genres are generated.

2.4 Proposed System

The proposed system can detect the facial expressions of the user and based on his/her facial expressions extract the facial landmarks, which would then be classified to get a particular emotion of the user. Once the emotion has been classified the songs matching the user's emotions would be shown to the user.

Thus, our proposed system focuses on detecting human emotions for developing emotion-based music player. A brief idea about our systems working, playlist generation and classification.

In this system live face emotion is detected using webcam where in the facial expressions are detected, captured and classified into their respective type accordingly the system plays music from the collected facial expressions. This system makes use of Convolutional Neural Network (CNN) model for image classification. The model trains itself according to the results thus improving the efficiency and effectiveness of the system.

Emotion Based Music Player is a useful application for music listeners with a smart phone and an internet connection. The Application is accessible by anyone who creates a Profile on the system. The Application is designed to meet the following needs of the users as described below;

1. Creating an account or signing up, signing in

2. Adding songs

3. Removing songs

4. Updating songs

5. Personalized playlist & Recommendations.

6. Capturing Emotions using a camera.

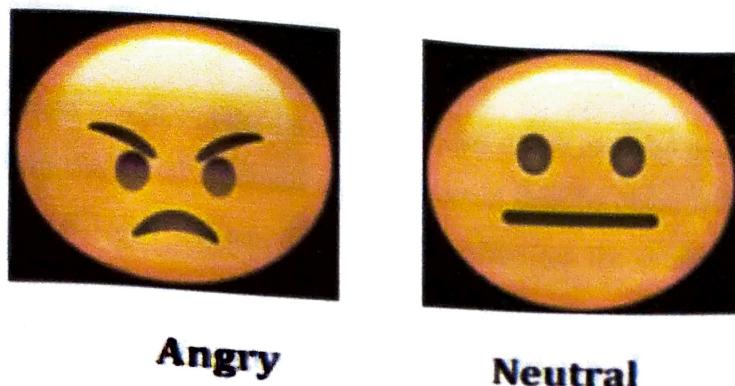
• **Detected emotions:**



Happy



Sad



Accordingly, to which we can classify emotion directory for playing song we have chosen this 4 Emotions.

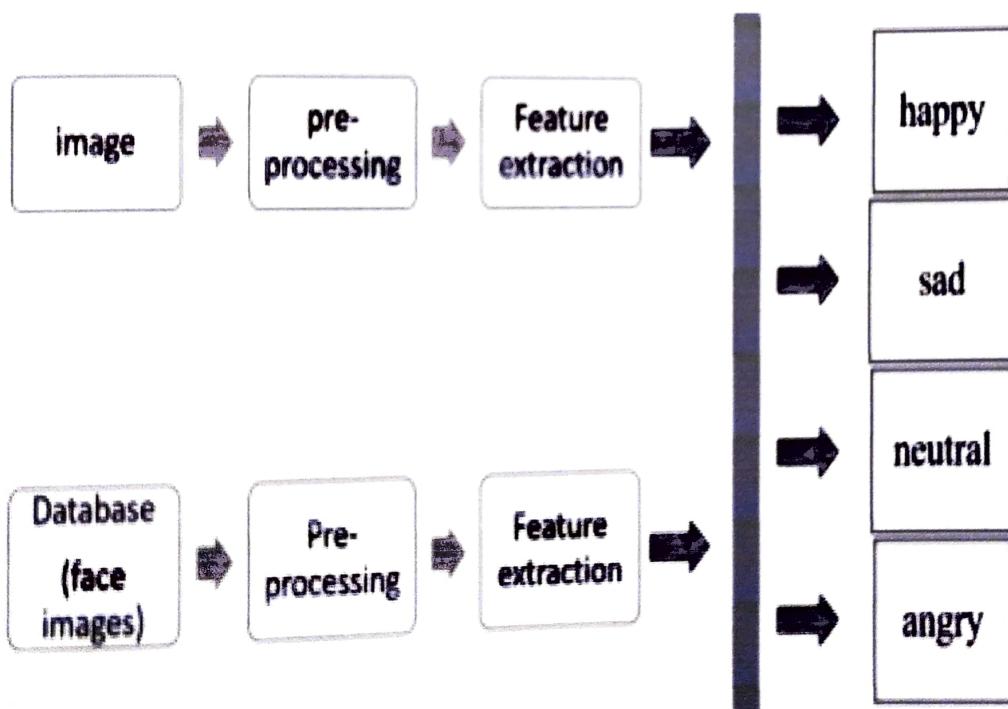


Figure 2.4: Emotion Classification

2.5 Conclusion

The proposed system can detect the facial expressions of the user and based on his/her facial expressions extract the facial landmarks, which would then be classified to get a particular emotion of the user. Once the emotion has been classified the songs matching the user's emotions would be shown to the user. Throughout the years, the results from the studies proved that different music style can actually influence individuals in different ways. Our proposed system simply calm the individual and can also get quicker song according to the mood, saving time from looking up different songs.

3. ANALYSIS

3.1 Introduction

The most common set of requirements defined by any operating system or software application is the physical computer resources, also known as hardware. Software Requirements deal with defining software resource requirements and pre-requisites that need to be installed on a computer to provide optimal functioning of an application. These requirements or pre-requisites are generally not included in the software installation package and need to be installed separately before the software is installed.

3.2 Feasibility Study

Feasibility observes pursuits to objectively and rationally uncover the strengths and weaknesses of an existing or proposed system, opportunities and threats gift inside the environment the sources required to carry thru, and ultimately the prospects for achievement. In its only phrases, the two criteria to judge feasibility are cost required and fee to be attained. A properly designed feasibility has a look at should offer a historical background of the task; typically, feasibility research precedes technical development and assignment implementation. A feasibility looks at evaluates the challenge capability for success.

3 key issues involved within the feasibility evaluation are: -

- Economic Feasibility
- Technical Feasibility
- Social Feasibility

3.2.1 Economic Feasibility

This examine is achieved to check the economic effect that the system could have on the corporation. The amount of fund that the organization can pour into the work and improvement of the gadget is confined. The costs must be justified. hence the advanced system as well in the budget and this turned into completed due to the fact most of the technologies used are freely to be had.

3.2.2 Technical Feasibility

This exam is finished to test the specialized achievable, that is, the specialized requirements of the framework. The created bodywork has to have an unassuming necessity; because it was negligible or in-valid adjustments are required for actualizing this framework.

3.2.3 Social Feasibility

The aspect of observe is to check the extent of popularity of the system by using the person. This includes the process of training the consumer to use the gadget effectively. The person should no longer feel threatened through the system, as an alternative should be given it as a need. the extent of attractiveness by means of the customers completely relies upon at the strategies which might be employed to teach the consumer approximately the gadget and to make him acquainted with it. His level of confidence should be raised so that he's also able to make a few constructive complaints, that is welcomed, as he is the final user of the gadget.

3.3 Software Requirement Specification

A Software requirements specification document describes the intended purpose, requirements and nature of a software to be developed. The SRS is a specification for a specific software product, program, or set of applications that perform particular functions in a specific environment.

3.3.1 Hardware Specification

- Input Device: - Webcam, Keyboard, Mouse
- Output Device: - Standard Monitor 9
- Processors: Intel Atom® processor or Intel® Core™ i5 processor 8th generation
- Disk space: 10 GB
- Operating systems: Windows* 8 or later, macOS, and Linux
- Python* versions: 3.6.X
- Included development tools: conda*, conda-env, Jupyter Note- book* (IPython), heroku clc, GIT cmd.
- Compatible tools: PyCharm*, Commandprompt

- Included Python packages: NumPy, PIP*, gunicorn.
- Graphics: Nvidia GeForce 940M or later, Intel Graphics 15.0 or later

3.3.2 Software Specification

Python:

- Windows: Python 3.9.5
- PIP and NumPy: Installed with PIP, Ubuntu*, Python 3.9.5, NumPy 1.20.3
- IOS/Mac: Python 3.9.5

Open CV:

- Windows: Open CV 4.5.2
- Open CV: Open CV 4.5.2 with installed HaarCascade Library of OpenCL1
- IOS/Mac: Open CV 4.5.2

Browser:

- Chrome 91 or above
- Edge 91 or above

3.3.3 Standards and Policies

Python:

Python is an interpreted, item-orientated language with dynamic semantics. Python helps modules and programs, which evokes seasoned- gram modularity and code reuse. The Python interpreter consequently the widespread popular library is to be had in supply or binary shape without charge for all most important platforms.

Standard Used: ISO8601

Operating System:

A 64-bit processor is able to storing more computational values, including memory cope with, because of this its capable of get right of entry to over four billion instances the bodily reminiscence of a 32-bit processor.

NumPy:

NEP 29 is recommended in Python and NumPy version. They assist as a community policy fashionable. This NEP recommends that everyone projects across the medical Python atmosphere adopt a not unusual “time window- based totally” policy for guide of Python and NumPy versions. well knowing a recommendation for task assist of minimal Python and NumPy versions will enhance downstream mission planning. Standard Used: NEP 29

Open CV:

OpenCV (Open-source computer vision Library) is an open-source computer vision and machine learning software program library.

OpenCV changed into constructed to provide a commonplace infrastructure for system vision applications. The library has greater than 2500 optimized algorithms, which incorporates a complete set cutting- edge each classic and computer vision and system state-of-the-art algorithms. those algorithms can be used to detect and apprehend faces, be- come aware of objects, classify human moves in movies, track digital camera actions, tune transferring objects, ex- tract 3-D fashions latest objects, produce 3-d factor clouds from stereo cam- eras, stitch photographs collectively to supply a 11 high decision photograph brand new a whole scene, find comparable photos from an photograph database, state-of-the-art crimson eyes from photographs taken the use of flash, follow eye movements, apprehend surroundings and establish markers to overlay it with augmented fact, and so etc...

3.4 General Architecture

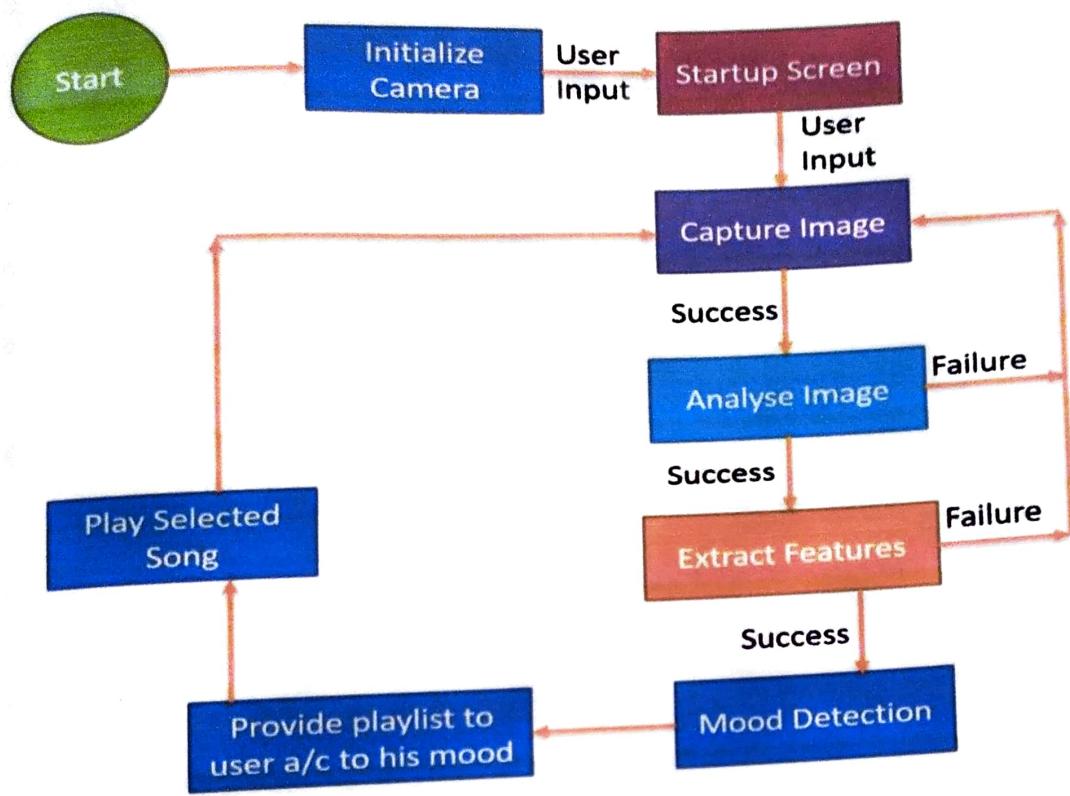


Figure 3.4: Architecture Diagram

Fig 3.4 shows The Environment of the system design and how the connectivity is done to each and every thing in this system.

4. DESIGN

4.1 Introduction

The Design Phase is where you look at the many potential solutions and narrow down the choices to determine the most effective and efficient way to construct the solution. The Design Phase answers the questions about "how" you will build the best solution. The aim is to develop one or more designs that can be used to achieve the desired project goals.

4.2 Design Phase Diagrams

4.2.1 Data Flow Diagram

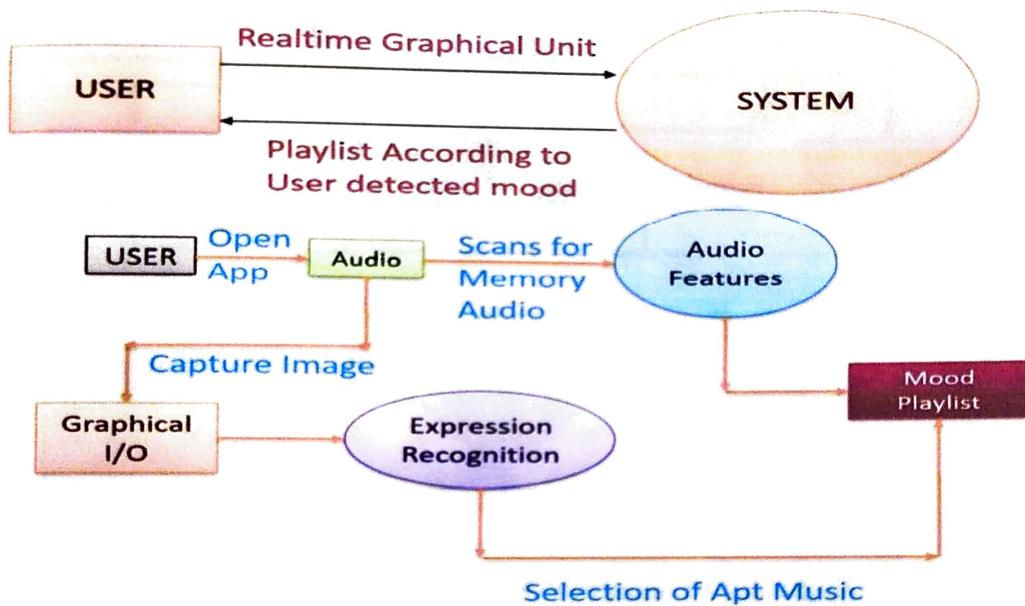


Figure 4.2.1: Data Flow Diagram

Description:

Step-1: The data is collected from capturing image.

Step-2: These data will be pre-processed under Haar Casade Algorithm.

Step-3: The meta data will extract the value of Lips.

Step-4: According to the value, the music is played with the expression extracted at capturing.

4.2.2 UML Diagram

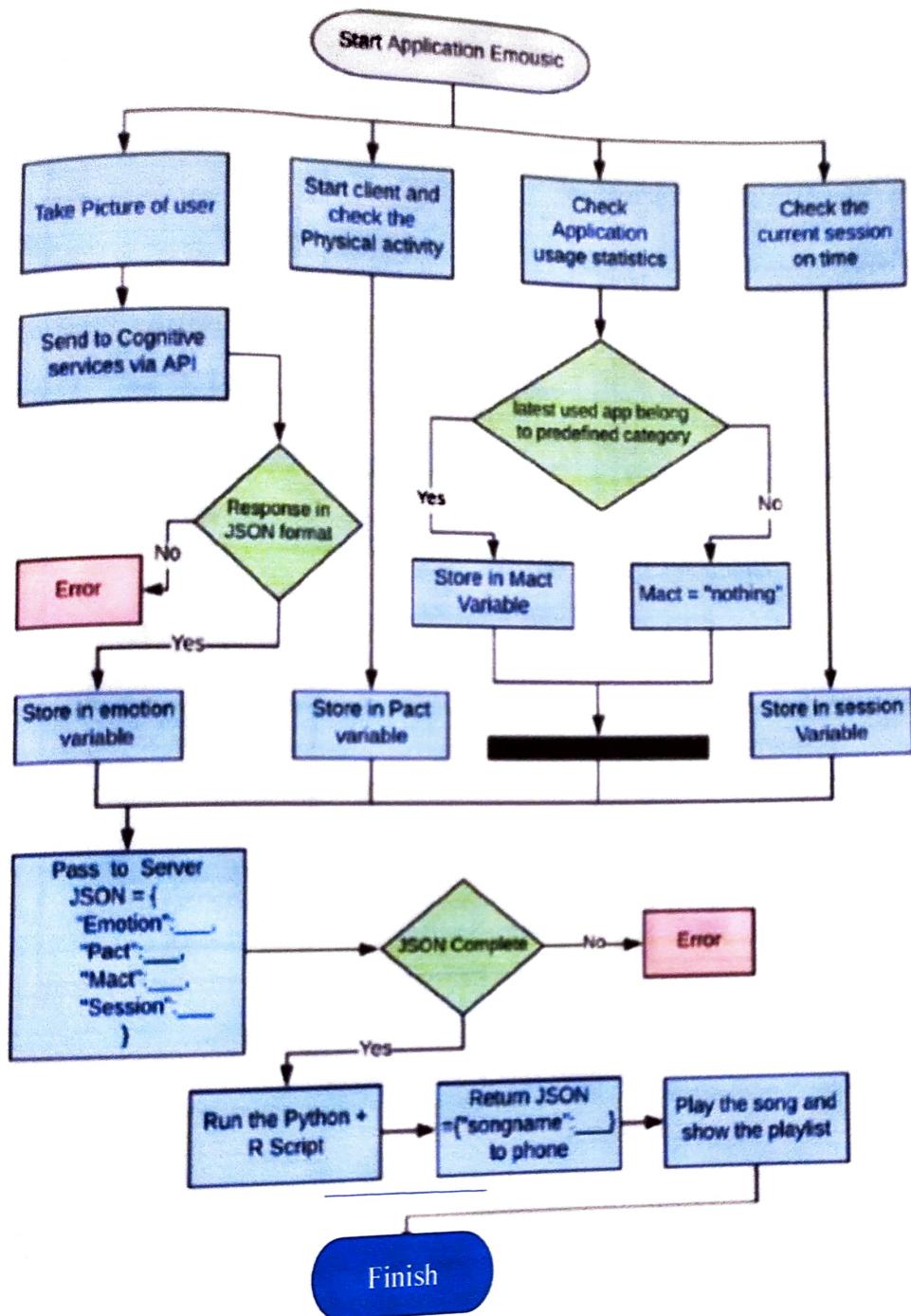


Figure 4.2.2: UML Diagram

4.2.3 Class Diagram

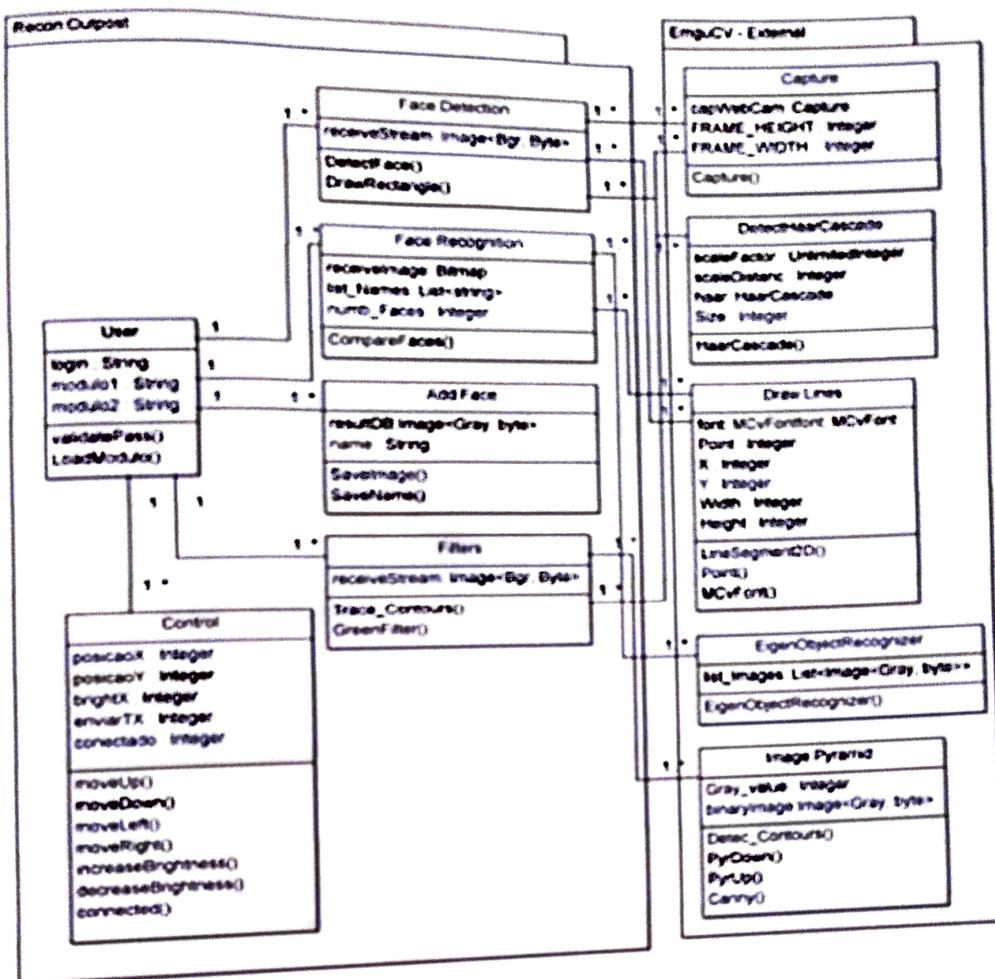


Figure 4.2.3: Class Diagram

Description

Step-1: Here the entities are Face Detection, Face Recognition, Capture, User, Detect Haar Cascade.

Step-2: The attributes of each entity are represented in the oval shape.

Step-3: The relation between entities is represented in the diamond shape.

Step-4: The relations between the entities are represented as (1...1), (0...n), (1...n).

Step-4: The relations between the entities are represented as (1...1), (0...n), (1...n).

4.2.4 Sequence Diagram

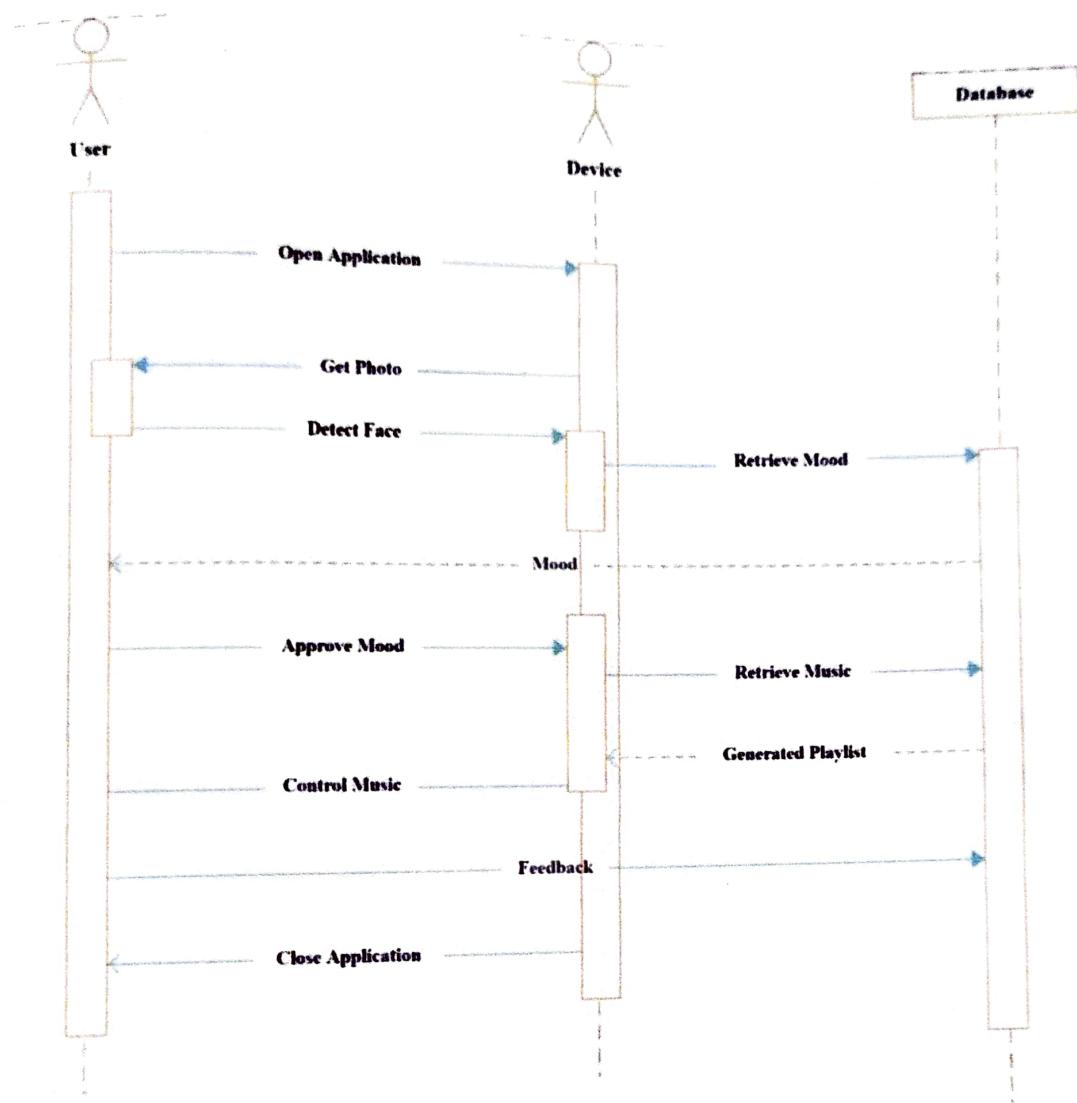


Figure 4.2.4: Sequence Diagram

Description

Sequence diagram essentially delineates collaboration between objects in a successive request for example the request where these communications happen. We can likewise utilize the terms occasion graphs or occasion situations to allude to a grouping chart. Arrangement out-lines depict how and in what request the items in a framework work.

4.3 Module Description

Module 1:-Emotion Extraction Feature

The photograph of the person is captured with the assist of a digital camera / web-cam. once the image captured, the frame of the captured picture from webcam feed is converted to a grayscale image to enhance the performance of the classifier, that's used to perceive the face gift within the image. as soon as the conversion is complete, the photograph is dispatched to the classifier set of rules which, with the assist of characteristic extraction strategies can extract the face from the body of the net digicam feed.

From the extracted face, individual functions are obtained and are sent to the trained network to come across the emotion expressed by way of the person. those photos may be used to train the classifier so that when a completely new and unknown set of snap shots is offered to the classifier, it can extract the location of facial landmarks from the one's photos based totally on the know-how that it had already ac- quired from the training set and return the coordinates of the new facial landmarks that it detected. The network is educated with the assist of CK significant records set. this is used to perceive the emotion being voiced through the person. It's achieved by using OpenCV.

Module 2:-Audio Extraction Feature

In this module, a listing of songs paperwork the input. As songs are audio documents, they require a positive amount of pre-processing Stereo indicators obtained from the internet are transformed to 16-bit PCM mono sign around a variable sampling rate of 48.6 kHz. The conversion system is finished the usage of Audacity method. The pre- processed signal undergoes an audio characteristic extraction, wherein features like rhythm firming is obtained using MIR 15 Toolbox. pitch is extracted using Chroma Toolbox and different capabilities like centroid, spectral flux, spectral roll off, kurtosis, 15 MFCC coefficients are extracted using Auditory Toolbox. Audio alerts are categorized into 7 kinds viz unhappy, marvel, excitement, pleasure, anger, satisfied, lonely. The extractions of converted files are carried out by using three device container units. MIR 15 Toolbox, Chroma Toolbox, and Auditory Toolbox which might be inbuilt in Haar Cascade Classifier.