A PRELIMINARY PROJECT REPORT ON

"Music Recommendation System Based on Face Recognition Technology"

SUBMITTED IN THE PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR THE AWARD OF THE DEGREE
OF

BACHELOR OF COMPUTER ENGINEERING

Submitted By

Aditya Deshmukh	Exam no: _ 01
Ganesh Adsul	Exam no: __02
Nikhil Hade	Exam no: __68
Kiran Talekar	Exam no: __62

Under the Guidance of
Prof. Iqra Shaikh

DEPARTMENT OF COMPUTER ENGINEERING

ALARD COLLEGE OF ENGINEERING AND MANAGEMENT,
Marunji, Pune, Maharashtra - 411057

SAVITRIBAI PHULE PUNE UNIVERSITY
2023-2024

CERTIFICATE

This is to certify that the project report entitles

"Music Recommendation System Based on Face Recognition Technology"

Submitted By:

Aditya Deshmukh	Exam no: _01
Ganesh Adsul	Exam no: __02
Nikhil Hade	Exam no: __68
Kiran Talekar	Exam no: __62

are Bonafede students at this institute and the work has been carried out by them under the supervision of Prof. Iqra Shaikh and it is approved for the partial fulfillment of the requirement of Savitribai Phule Pune University, for the award of the degree of **Bachelor of Engineering** (Computer Engineering).

Prof. Iqra Shaikh
Internal Guide
Dept. of Computer Engg.

Prof. Sayli Haldavanekar
Head of Dept.
Dept. of Computer Engg.

Prof._____
External Examiner,
SPPU

Dr. K.D Sapate
Principal,
Alard College of Engg. and Mngmt

Place: Pune.
Date:

ACKNOWLEDGEMENT

We take this opportunity to express our profound gratitude and deep regards to my guide **Prof. Iqra Shaikh** for her exemplary guidance, monitoring, and constant encouragement through- out the completion of this report. I am truly grateful to her efforts to improve my technical writing skills. The blessing, help and guidance given by her time to time shall carry me a long way in the journey of life on which I am about to embark. I take this privilege to express my sincere thanks to **Prof. Sayli Haldavanekar**, Head of Department of Computer Engineering, for her generous support.

Last but not the least I would also like to thank all those who have directly or indirectly helped me in completion of this report.

I would also thank my Institution and my faculty members without whom this project would have been a distant reality.

Adtiya Deshmukh
Ganesh Adsul
Nikhil Hade
Kiran Talekar
(B.E. Computer Engg.)

ABSTRACT

Music plays a significant role in improving and elevating one's mood as it is one of the important sources of entertainment and inspiration to move forward. Recent studies have shown that humans respond as well as react to music in a very positive manner and that music has a high impact on human's brain activity. Now a days, people often prefer to listen to music based on their moods and interests. This work focuses on a system that suggests songs to the users, based on their state of mind. In this system, computer vision components are used to determine the user's emotion through facial expressions.

Once the emotion is recognized, the system suggests a song for that emotion, saving a lot of time for a user over selecting and playing songs manually. Conventional method of playing music depending upon the mood of a person requires human interaction. Migrating to the computer vision technology will enable automation of such system. To achieve this goal, an algorithm is used to classify the human expressions and play a music track as according to the present emotion detected. It reduces the effort and time required in manually searching a song from the list based on the present state of mind of a person. The expressions of a person are detected by extracting the facial features using the Haar Cascade algorithm and CNN Algorithm. An inbuilt camera is used to capture the facial expressions of a person which reduces the designing cost of the system as compared to other methods.

Table of Contents

CHAPTER1 INTRODUCTION

1.1 Overview

Facial expressions are one of the natural means to communicate the emotions and these emotions can be used in entertainment and Human Machine Interface (HMI) fields In today's world, with the advancements in the areas of technology various music players are deployed with features like reversing the media, fast forwarding it, streaming playback with multicast streams. Although these features satisfy the basic requirements of the user, yet one has to manually surf for the song from a large set of songs, according to the current circumstance and mood. This is a time-consuming task that needs some effort and patience. The main objective of this work is to develop an intelligent system that can easily recognize the facial expression and accordingly play a music track based on that particular expression/emotion recognized.

The seven universally classified emotions are Happy, Sad, Anger, Disgust, Fear, Surprise and Neutral. The main objective of this work is to develop an intelligent system that can easily recognize the facial expression and accordingly play a music track based on that particular expression/emotion recognized. The seven universally classified emotions are Happy, Sad, Anger, Disgust, Fear, Surprise and Neutral. The algorithm that is used in developing the present system is Haar Cascade algorithm which utilizes eigenfaces to extract the facial features.

The designed algorithm is very efficient due to less computational time taken hereby increasing the performance of the system. This work finds its applications in various domains like Human Computer Interaction (HCI), therapeutic approach in health care etc. Most of the time the digital music is sorted and put together based on attributes such as artist, genre, albums, language, popularity and so on.

Many of the available online music streaming services recommend music based on user's preferences and his previous music listening history that employ content based and collaborative filtering recommendations. But these recommendations may not suite the current mood of the user. The manual classification of songs by learning user's preference of emotion is a time-consuming task. So, recommendations can also be achieved using the physiological and emotional status of the user which are mainly captured from the user's facial expression, ges2 tures, pulse rate, movement, speech/text interactions etc.

Several work is carried to detect emotions using facial landmarks to extract the features. Nguyen et al. detected three kinds of emotions namely positive, negative and blank using 68 facial landmarks. This system work proposes a CNN based approach to recommend music by analyzing the multimodal emotional information captured by facial movements and semantic analysis of the speech/text interactions of the user, thus, intensifying the decision of the system on recognized emotions in real-time. Machine learning has become very popular in recent years.

Depending on the type of application and the dataset available, certain types of machine learning techniques are more appropriate than others for different applications. The main types of learning algorithms include supervised learning, unsupervised learning, semi-supervised learning, and reinforcement learning. A neural network (NN) is a technique of machine learning that is generally effective at extracting critical features from complex datasets and deriving a function or model that expresses those features. The NN utilizes a training dataset to first train a model. After the model is trained, the NN can then be applied to new or previously unseen data-points and classify the data based on the previously trained model

1.2 MOTIVATION

- Our effort is an initial step to bridge the gap between the traditional music browsing system and the users' needs.
- This work proposes a CNN based approach to recommend music by analyzing the multimodal emotional information captured by facial movements and semantic analysis of the speech/text interactions of the user, thus, intensifying the decision of the system on recognized emotions in real-time.

1.3 PROBLEM STATEMENT

- To implement an application that provides evacuation route guidance and indoor navigation system by using Augmented Reality. This will help users to easily navigate through the buildings, companies, offices, college, etc. to reach 3 their respective location by using the path provided by the application and also in case of indoor fire environment it will provide the safe path to the users to get out from that place. the application should be able to fetch and play .mp3 and. Wav Files. The ability to support different type of service level music.

1.4 OBJECTIVE

- The main objective of this work is to develop an intelligent system that can easily recognize the facial expression and accordingly play a music track based on that particular expression/emotion recognized.
- The underlying objective of this paper is to design an accurate algorithm that would yield a list of songs from a user's playlist in conformance with a user's emotional state.
- The algorithm designed, requires less computational time, storage and reduces the cost incurred in employing additional hardware. It categorizes a facial image under 4 different facial expressions viz. Sad, Anger, Neutral and Happiness. The overall process is illustrated

CHAPTER2 LITERATURE SURVEY

2.1 STUDY OF RESEARCH PAPER

1.Paper Name: Research on Automatic Music Recommendation Algorithm Based on Facial Micro-expression Recognition Author:Ziyang Yu1 , Mengda Zhao1 , Yilin Wu1, Peizhuo Liu1 , Hexu Chen1 Abstract : In recent years, with the development and application of big data, deep learning has received more and more attention. As a deep learning neural network, convolutional neural network plays an extremely important role in face image recognition. In this paper, a combination of micro-expression recognition technology of convolutional neural network and automatic music recommendation algorithm is developed to identify a model that recognizes facial micro-expressions and recommends music according to corresponding mood. The facial micro-expression recognition model established in this paper uses FER2013 with a recognition rate of 62.1content-based music recommendation algorithm is used to extract the feature vector of the song and a cosine similarity algorithm is used to make the music recommendation. This research helps to improve the practicality of the music recommendation system, and the related results will also serve as a reference for the application of the music recommendation system in areas such as emotion regulation

3.**Paper Name:** EMOSIC- An Emotion Based Music Player for Android. Author: Karthik Subramanian Nathan , Manasi Arun and Megala S Kannan Abstract :Music plays a very important role in human0 s daily life and in the modern advanced technologies. Usually, the user has to face the task of manually browsing through the playlist of songs to select. Here we are proposing an efficient and accurate model, that would generate a playlist based on current emotional state and behavior of the user. Existing methods for automating the playlist generation process are computationally slow, less accurate and sometimes even require use of additional hardware like EEG or sensors. Speech is the most ancient and natural way of expressing feelings, emotions and mood and its and its processing requires high computational, time, and cost. This proposed system based on real-time extraction of facial expressions as well as extracting audio features from songs to classify into a specific emotion that will generate a playlist automatically such that the computation cost is relatively low. 7

3.Paper Name: Emotion aware Smart Music Recommender System using Two Level CNN. Author: Krupa K S, Ambara G Description : Music plays a significant role in improving and elevating one's mood as it is one of the important source of entertainment and inspiration to move forward. Recent studies have shown that humans respond as well as react to music in a very positive manner and that music has a high impact on human's brain activity. Nowadays, people often prefer to listen to music based on their moods and interests. This work focuses on a system that suggests songs to the users, based on their state of mind. In this system, computer vision components are used to determine the user's emotion through facial expressions and chatbot interactions. Once the emotion is recognized, saving a lot of time for a user over selecting and playing songs manual.

CHAPTER 3 REQUIREMENTS SPECIFICATIONS

3.1 Hardware Requirements

- Hardware: intel core
- Speed: 2.80 GHz
- RAM: 8GB
- Hard Disk: 500 GB
- Keyboard: Standard Windows Keyboard

3.2 Software Requirements

- Operating System: Windows 10(64 Bit)
- IDE: Spyder
- Programming Language: python version 3.7,3.8

3.3 FUNCTIONAL REQUIREMENTS

- Proposed system consists of 4 modules:

a) Feature point extraction: Feature points of each image gets detected.

b) Feature correspondence matching: Matching of selected feature points across various image frames.

c) Point estimation: Position estimation and vision system orientation during navigation.

d) Position refinement: Location estimate based, accurate location derivati

3.4 Non-Functional Requirements

3.4.1 Performance Requirements

- The performance of the functions and every module must be well. The overall performance of the software will enable the users to work recently. Performance of encryption of data should be fast. Performance of the providing virtual environment should be fast Safety Requirement
- The application is designed in modules where errors can be detected. This makes it easier to install and update new functionality if required.

3.4.2 Safety Requirements

- The application is designed in modules where errors can be detected and fixed easily. This makes it easier to install and update new functionality if required.

3.4.3 Software Quality Attributes

- Our software has many quality attribute that are given below: -
- Adaptability: This software is adaptable by all users.
- Availability: This software is freely available to all users. The availability of the software is easy for everyone.
- Maintainability: After the deployment of the project if any error occurs then it can be easily maintained by the software developer.
- Reliability: The performance of the software is better which will increase the reliability of the Software.
- User Friendliness: Since, the software is a GUI application; the output generated is much user friendly in its behavior.
- Integrity: Integrity refers to the extent to which access to software or data by unauthorized persons can be controlled.
- Security: Users are authenticated using many security phases so reliable security is provided.
- Testability: The software will be tested considering all the aspects.

3.5 SYSTEM REQUIREMENTS

- 100GB HDD (minimum)
- Intel 1.66 GHz Processor Pentium 4 (minimum)
- Internet Connectivity
- 2GB RAM (minimum).

CHAPTER 4 SYSTEM DESIGN

4.1 SYSTEM ARCHITECTURE

1. Image Acquisition In any of the image processing techniques, the first task is to acquire the image from the source. These images can be acquired either through camera or through standard datasets that are available online. The images should be in .jpg format. The images considered here are user dependent i.e., dynamic images. The number of sample training images considered here.

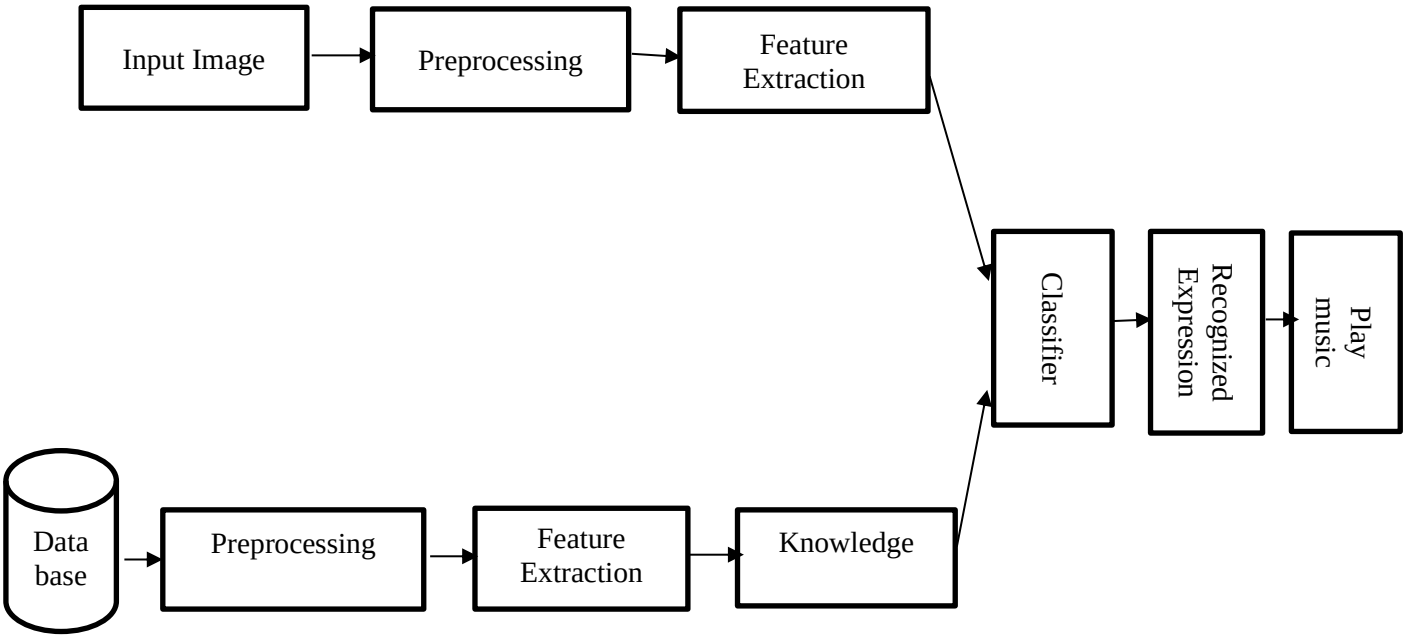

Figure 4.1: System Architecture

2. pre-processingc is mainly done to eliminate the unwanted information from the image acquired and fix some values for it, so that the value remains same throughout. In the pre-processing phase, the images are converted from RGB to Gray-scale and are resized to 256*256 pixels. The images considered are in .jpg format, any other formats will not be considered for further processing. During preprocessing, eyes, nose and mouth are considered to be the region of interest. It is detected by the cascade object detector which utilizes Jones-Viola algorithm.

3. Facial Feature Extraction After pre-processing, the next step is featuring extraction.
The extracted facial features are stored as the useful information in the form of vectors during training phase and testing phase. The following facial features can be considered "Mouth, forehead, eyes, complexion of skin, cheek and chin dimple, eyebrows, nose and wrinkles on the face". In this work, eyes, nose, mouth and forehead are considered for feature extraction purpose for the reason that these depict the most appealing expressions. With the wrinkles on the forehead or the mouth being opened one can easily recognize that the person is either surprised or is fearful. But with a person's complexion it can never be depicted. To extract the facial features PCA technique is used.

4. Expression Recognition: To recognize and classify the expressions of a person Euclidean distance classifier is used. It gets the nearest match for the test data from the training data set and hence gives a better match for the current expression detected. Euclidean distance is basically the distance between two points and is given by "(3.1)". It is calculated from the mean of the eigenfaces of the training dataset. The training images that correspond to various distances from the mean image are labeled with expressions like happy, sad, fear, surprise, anger, disgust and neutral. When the Euclidean distance between the eigenfaces of the test image and mean image matches the distance of the mean image and eigenfaces of the training dataset the expression is classified and named as per the labeled trained images. The smaller the distance value obtained; the closest match will be found. If the distance value is large enough for an image, then the system has to be trained for that individual.

5. Play Music: The last and the most important part of this system is the playing of music based on the current emotion detected of an individual. Once the facial expression of the user is classified, the user's corresponding emotional state is recognized. A number of songs from various domains pertaining to a number of emotions is collected and put up in the list. Each emotion category has a number of songs listed in it. When the user's expression is classified with the help of CNN algorithm, songs

4.2 DATA FLOW (DFD) DIAGRAMS

DFD which shows how the information moves through the system and how it is modified by a series of transformations. It is a graphical technique that shows information and the transformations that are applied as data moves from input to output.

4.2.1 DFD Level 0 Diagram

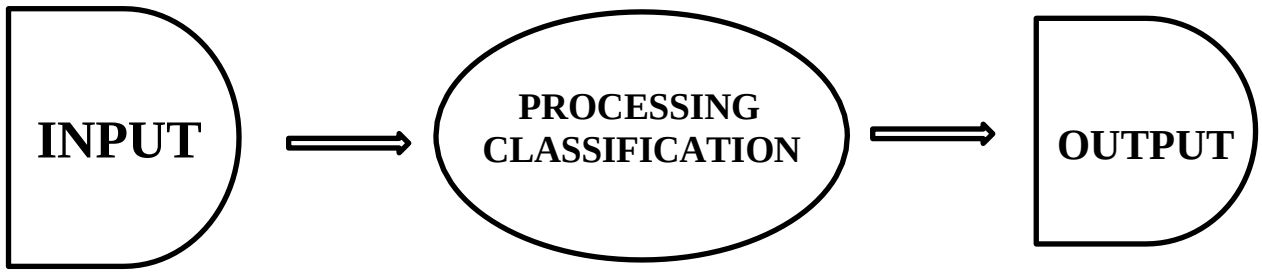

Figure 4.2: DFD Level 0 Diagram

4.2.2 DFD Multi Level Diagram

The data flow diagram (DFD) is one of the most important modelling tools. It is used to model the system components. These components are the system process, the data used by the process, an external entity that interacts with the system and the information flows in the system.

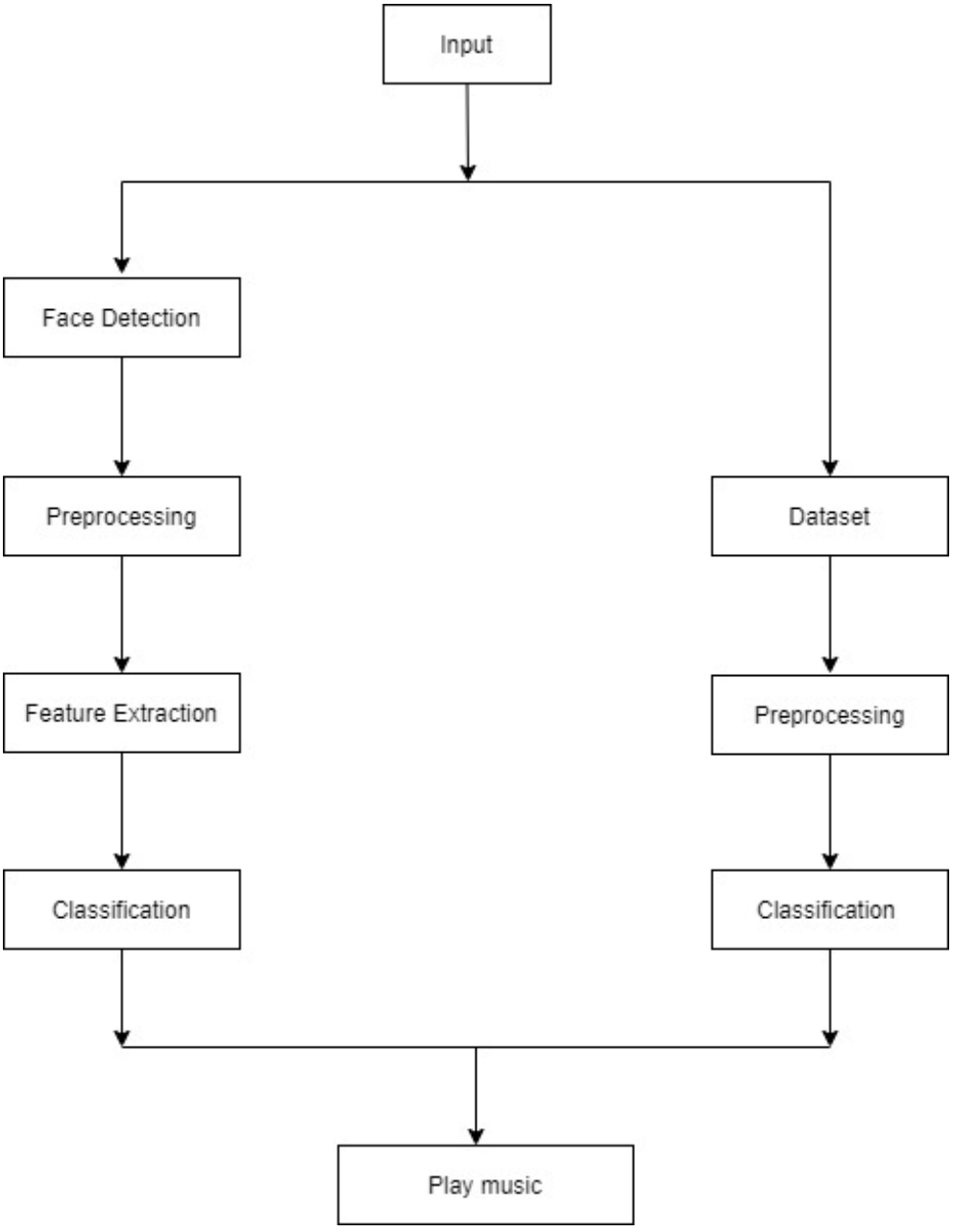

Figure 4.3: DFD Multi Level Diagram

4.3 UML DIAGRAMS

4.3.1 Activity diagram

Activity diagrams are graphical representations of work flows of step wise activities and actions with support for choice, iteration and concurrency. In the Unified Modeling Language, activity diagrams are intended to model both computational and organizational processes. Activity diagrams show the overall ow of control.

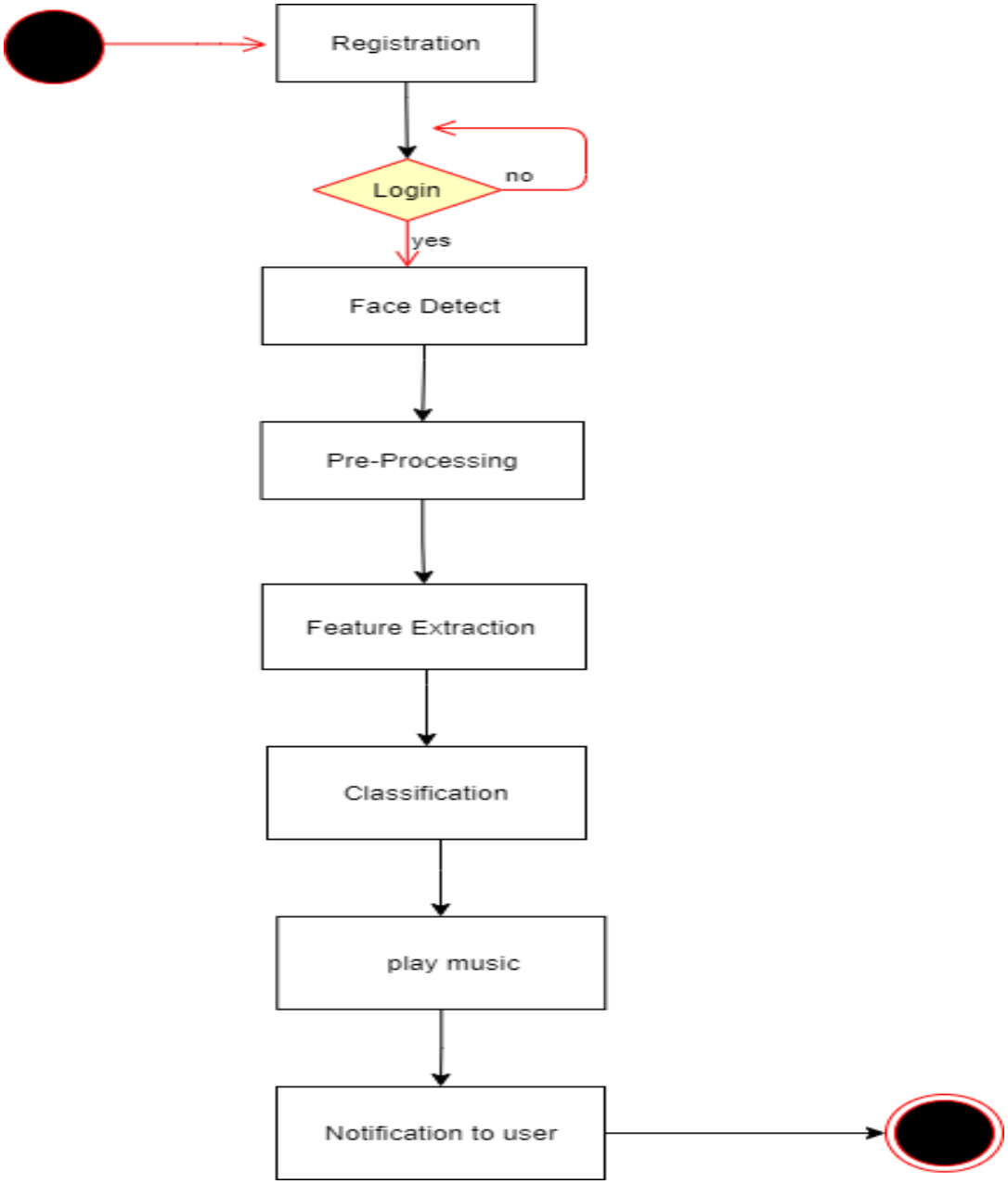

Figure 4.4: Activity Diagram

4.3.2 Sequence Diagram

A sequence diagram is a kind of interaction diagram that shows how processes operate with one another and in what order. It is a construct of a message sequence chart. A sequence diagram shows object interactions arranged in time sequence. It depicts the objects and classes involved in the scenario and the sequence of messages exchanged between the objects needed to carry out the functionality of the scenario.

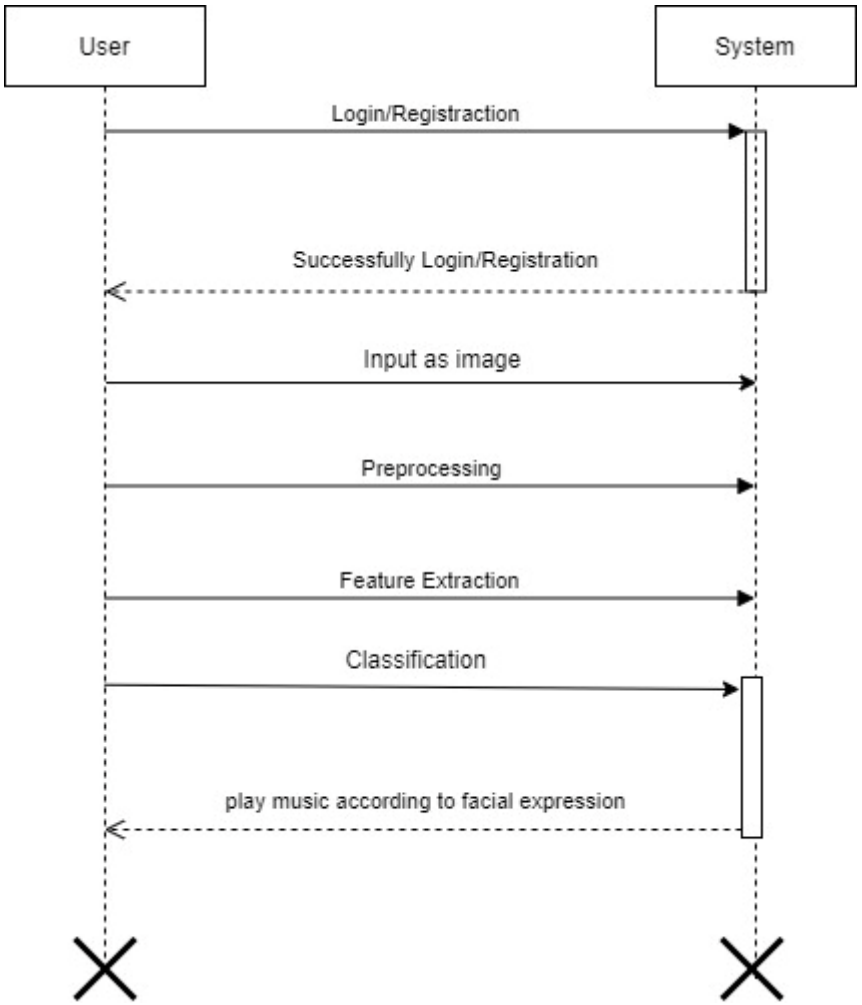

Figure 4.5: Sequence Diagram

4.3.3 Use case Diagram

Generate data from various resources. Provides data from Agri-food transaction module and forward to database. Perform all transactions and generate the final results using block chain.

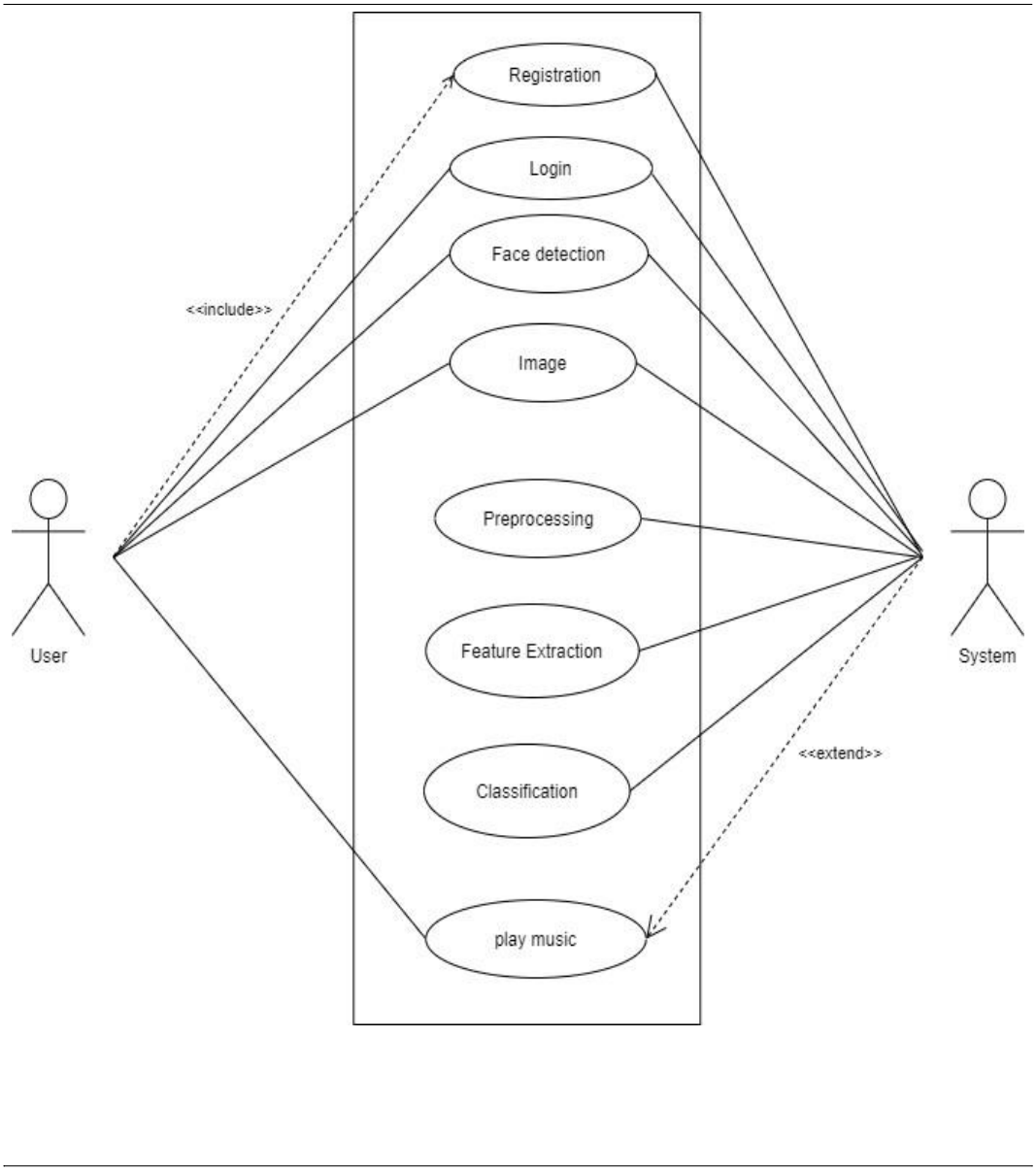

Figure 4.6: Use Case Diagram

4.3.4 Class Diagram

In a class diagram, the classes are arranged in groups that share common characteristics. A class diagram resembles a flowchart in which classes are portrayed as boxes, each box having three rectangles inside. The top rectangle contains the name of the class; the middle rectangle contains the attributes of the class; the lower rectangle contains the methods, also called operations, of the class. Lines, which may have arrows at one or both ends, connect the boxes. These lines define the relationships, also called associations, between the classes.

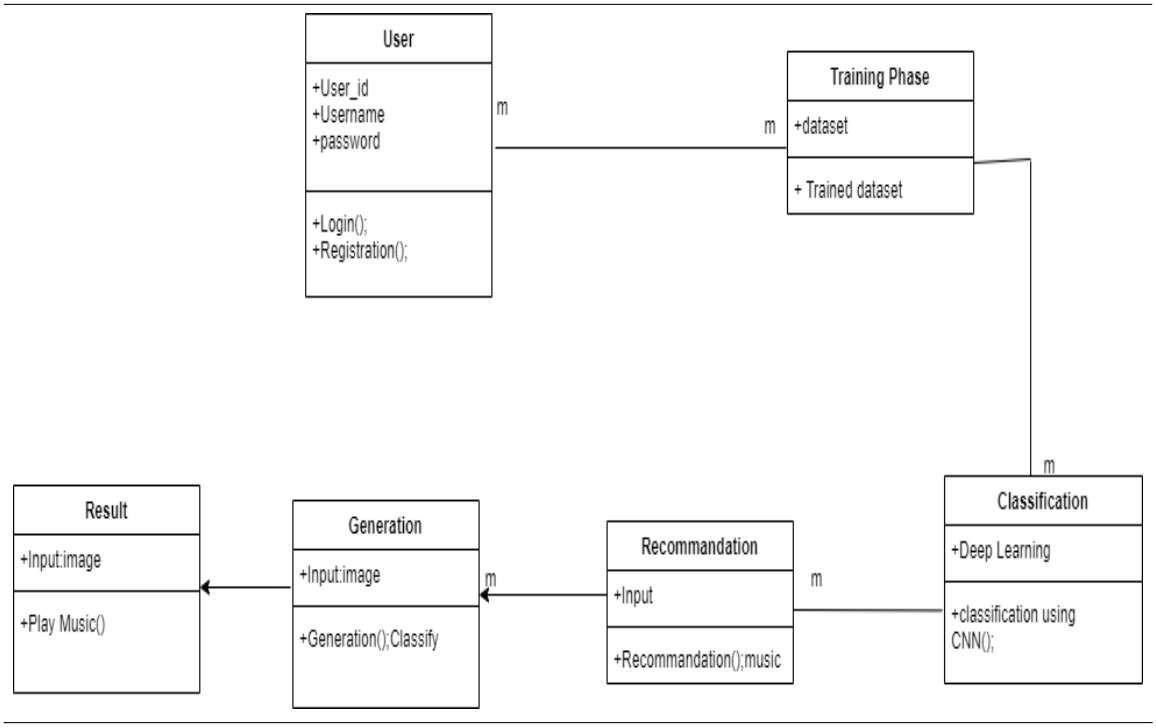

Figure 4.7: Class Diagram

4.3.5 ER Diagram

An Entity Relationship (ER) Diagram is a type of flowchart that illustrates how "entities" such as people, objects or concepts relate to each other within a system.

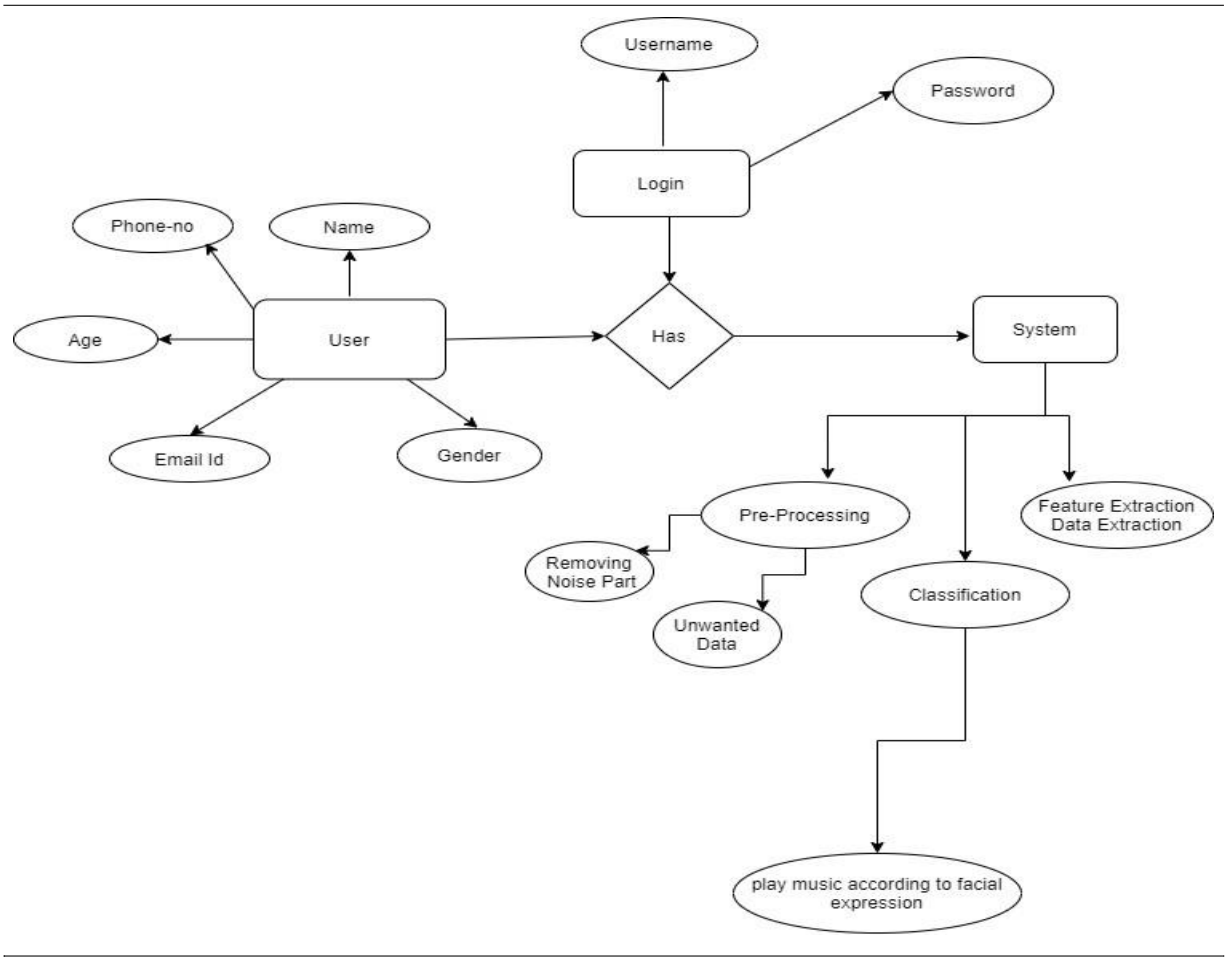

Figure 4.8: Entity Relationship Diagram.

CHAPTER 5 SYSTEM IMPLEMENTAION PLAN

5.1 Description of Tools

Anaconda Navigator

Anaconda Navigator is a desktop graphical user interface (GUI) included in Ana- conda distribution that allows you to launch applications and easily manage conda packages, environments, and channels without using command-line commands. Nav- igator can search for packages on Anaconda.org or in a local Anaconda Repository.It is available for Windows, macOS, and Linux.

Spyder

Spyder is short for" Scientific Python Development Environment." It's intended foruse as a workbench for scientific computing with Python, and that's reflected in thefeature set, the packaging, and the overall behavior of the IDE. Spyder has useful features for general Python development, but unless you work mainly with I Pythonand scientific computing packages, you're probably better off with a different IDE.

The biggest reason not to use Spyder as a general-purpose Python develop- ment environment isn't the feature set, but the setup process. Spyder is not deliveredas a standalone executable in the manner of a product like Visual Studio or PyCharm. Instead, it's installed as a Python package.

DBSqlite3

DB Browser for SQLite (DB4S) is a high quality, visual, open-source tool to create, design, and edit database files compatible with SQLite.

DB4S is for users and developers who want to create, search, and edit databases. DB4S uses a familiar spreadsheet-like interface, and complicated SQL commands do not have to be learned.

Controls and wizards are available for users to:

- Create and compact database files
- Create, define, modify and delete tables
- Create, define, and delete indexes
- Browse, edit, add, and delete records
- Search records
- Import and export records as text
- Import and export tables from/to CSV files
- Import and export databases from/to SQL dump files
- Issue SQL queries and inspect the results
- Examine a log of all SQL commands issued by the application
- Plot simple graphs based on table or query data

5.2 Programming Language Description

Python

Python is an interpreted, object-oriented, high-level programming language with dynamic semantics. Its high-level built in data structures, combined with dynamic typing and dynamic binding, make it very attractive for Rapid Application Devel- opment, as well as for use as a scripting or glue language to connect existing com- ponents together. Python's simple, easy to learn syntax emphasizes readability and therefore reduces the cost of program maintenance. Python supports modules and packages, which encourages program modularity and code reuse. The Python in- terpreter and the extensive standard library are available in source or binary form without charge for all major platforms, and can be freely distributed.

5.3 Algorithm Details

5.2.1 CNN Algorithm

A CNN is a kind of network architecture for deep learning algorithms and is specifically used for image recognition and tasks that involve the processing of pixel data. There are other types of neural networks in deep learning, but for identifying and recognizing objects, CNNs are the network architecture of choice.

CNNs are used for image classification and recognition because of its high accuracy.It was proposed by computer scientist Yann LeCun in the late 90s, when he was inspired from the human visual perception of recognizing things

The widely used algorithms in this context include denoising, region growing, edge detection, etc. The contrast equalization is often performed in image-processing and contrast limited adaptive histogram equalization
steps:

- Convolutional Layer.
- Pooling Layer.
- Fully Connected Layer.
- Dropout.
- Activation Functions

CHAPTER 6 CONCLUSION AND FUTURE SCOPE

6.1 Advantages

- It can capture the user's emotional state in real-time by analyzing facial expressions.
- This enables the system to recommend music that aligns with the user's current mood, making the listening experience more emotionally engaging.
- Facial expression analysis can help create highly personalized music recommendations.
- Users don't need to manually input their emotional state or preferences, as the system can infer this information from their facial expressions.
- Music has the power to influence and regulate emotions.

6.2 Limitations

- The accuracy of facial expression recognition technology can vary.
- Facial expressions do not always provide sufficient context for music recommen-dations.
- A person may smile because they find a song's lyrics ironic or humorous, which may not align with their actual emotional state.

6.3 Applications

- Music Streaming Services.
- Mood Regulation.
- Healthcare and Therapy.
- User Experience Enhancement

6.4 Conclusion and Future Scope

The proposed work presents facial expression recognition system to play a song according to the expression detected and also classify music Type. It uses CNN approach to extract features, and Euclidean distance classifier classifies these ex- pressions. In this work, real images i.e. user dependent images are captured utilizing the in-built camera.

There are continuously new ways and methods being developed that make use of image processing algorithms to extract the emotion of the user and make use of the extracted emotion to treat the user. Emotion recognition has gained a lot of importance in all aspects of life and if a robust algorithm implemented which can accurately classify the emotions of the person, then a great deal of advancement in the industry can be achieved with the help of this. The system has successfully been able to capture the emotion of a user. It has been tested in a real-time environment for this predicate. However, it has to be tested in different lighting conditions to determine the robustness of the developed system

Future Scope

- Reduce the time required to train the classifier
- Use of EEG signals to make the software even more optimized and to detect the exact mood /emotion of the us

Appendices

1. Appendix A: Model Architecture

- Provide detailed diagrams, charts, and descriptions of the neural network or machine learning model used for face recognition and music recommendation.

2. Appendix B: Hyperparameter

- List the hyperparameters used during training, along with any optimization algorithms and strategies applied to fine-tune the models.

3. Appendix C: Evaluation Metrics

- Explain the evaluation metrics used to assess the system's performance and provide the calculations and results for these metrics.

4. Appendix D: Model Training Details

- Include information about the hardware and software setup used for training, training time, and any issues or challenges encountered during the training process.

5. Appendix E: Results and Performance

Present the results of experiments and evaluations, including tables, charts, and figures that showcase performance metrics like accuracy, precision, recall, F1 score, and others.

References

1. S L Happy and Aurobinda Routray, “Automatic Facial Expression Recognition using Features of salient Facial Patches,” in IEEE Trans. On Affective Computing, January- March 2015, pp. 1-12.
2. Hafeez Kabani, Sharik Khan, Omar Khan and Shabana Tadvi, “Emotion based Music Player,” Int. J. of Eng. Research and General Sci., Vol. 3, Issue 1, pp. 750-756, January- February 2015.
3. Li Siquan, Zhang Xuanxiong. Research on Facial Expression Recognition Based on Convolutional Neural Networks [J]. Journal of Software, 2018, v.17; No.183 (01): 32-35.
4. Hou Yuqingyang, Quan Jicheng, Wang Hongwei. Overview of the development of deep learning [J]. Ship Electronic Engineering, 2017, 4: 5-9.
5. Liu Sijia, Chen Zhikun, Wang Fubin, et al. multi-angle face recognition based on convolutional neural network [J]. Journal of North China University of Technology (Natural Science Edition), 2019, 41 (4): 103-108.
6. Li Huihui. Research on facial expression recognition based on cognitive machine learning [D]. Guangzhou: South China University of Technology, 2019.
7. Li Yong, Lin Xiaozhu, Jiang Mengying. Facial expression recognition based on cross- connection LeNet-5 network [J]. Journal of Automation, 2018,44 (1): 176-182.