EMOTION BASED MUSIC PLAYBACK SYSTEM

A PROJECT REPORT

Submitted by

SURYA V

(2019202057)

submitted to the Faculty of

INFORMATION AND SCIENCE TECHNOLOGY

in partial fulfillment for the award of the degree
of

MASTER OF COMPUTER APPLICATIONS

DEPARTMENT OF INFORMATION SCIENCE AND TECHNOLOGY
COLLEGE OF ENGINEERING, GUINDY
ANNA UNIVERSITY
CHENNAI 600 025

JUNE 2022

ANNA UNIVERSITY

CHENNAI - 600 025

BONAFIDE CERTIFICATE

Certified that this project report titled **”EMOTION BASED MUSIC PLAYBACK SYSTEM”** is the bonafide work of **SURYA V (2019202057)** who carried out project work under my supervision. Certified further that to the best of my knowledge and belief, the work reported herein does not form part of any other thesis or dissertation on the basis of which a degree or an award was conferred on an earlier occasion on this or any other candidate.

PLACE: CHENNAI
DATE: 13.06.22

MS. T. SINDHU
TEACHING FELLOW
INTERNAL GUIDE
DEPARTMENT OF IST, CEG
ANNA UNIVERSITY
CHENNAI 600025

COUNTERSIGNED

Dr.S. SRIDHAR
HEAD OF THE DEPARTMENT
DEPARTMENT OF INFORMATION SCIENCE AND TECHNOLOGY
COLLEGE OF ENGINEERING, GUINDY
ANNA UNIVERSITY
CHENNAI 600025

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 web application that can be used with the help of providing the functionality of playing music according to the emotion detected. The user would not have to waste any time in searching or to look up for songs and the best track matching the user's mood is detected, and songs would be shown to the user according to his/her mood. The image of the user is captured with the help of a webcam and then the image is processed to input into the CNN model that has been trained and exported. The songs are labelled using audio feature and then classified using machine learning algorithm according to moods. The user's picture is taken and then as per the mood/emotion of the user an appropriate list of songs is fetched, displayed and played.

Key Words : Emotion recognition, CNN, Music emotion recognition, Classification, Computer vision.

ACKNOWLEDGEMENT

Its my privilege to express my sincere thanks to my project guide **Ms. T. Sindhu**, Teaching Fellow, Department of Information Science and Technology, College of Engineering, Guindy, Anna University, Chennai for her keen interest, inspiring guidance, constant encouragement and support with my work during all the stages, to bring this thesis into fruition.

I deeply express my sincere thanks to **Dr. S. Sridhar**, Professor and Head of the Department, Department of Information Science and Technology, College of Engineering, Guindy, Anna University, Chennai for extending support.

I would like to express my sincere thanks to the project committee members, **Dr. Saswati Mukherjee**, Professor, **Dr. M. Vijayalakshmi**, Associate Professor, **Dr. E. Uma**, Assistant Professor, Department of Information Science and Technology, Anna University, Chennai for giving their valuable suggestions, encouragement and constant motivation throughout the duration of my project.

SURYA V

TABLE OF CONTENTS

LIST OF FIGURES

LIST SYMBOLS OF ABBREVIATIONS

CNN	Convolutional Neural Network
MER	Music Emotion Recognition
CSV	Comma Separated Value
JSON	Javascript Object Notation
API	Application Programming Interface
HTTP	Hyper Text Transfer Protocol
XML	Extensible Markup Language
FER	Facial Expression Recognition

CHAPTER 1

INTRODUCTION

This chapter discuss about overview, motivation, problem statement and organisation of the report.

1.1 OVERVIEW

Project Emoplayer [Emotion Based Music Player] is a novel approach which helps user to play the songs according to their emotion / mood. This proposed system implemented a web application that is made for web users only. Firstly user have to click lets start to access the camera application and then our trained CNN model will predict the user emotion and redirect towards the list of songs based on the emotion. User can access this application on desktop. To access in desktop user need windows7 or further version. The songs are collected from spotify and using audio features it is classified on emotions and saved to JSON file. The audio is played using in-build web player which fetches the song from spotify.

1.2 MOTIVATION AND PROBLEM STATEMENT

Music listeners have tough time creating and segregating the play - list manually when they have hundreds of songs. It is also difficult to keep track of all the songs sometimes songs that are added and never used , wasting a lot of device memory and forcing the user to find and delete songs manually. User's have to manually select songs every time based on interest and mood. User's also have difficulty to re-organize and playing music when play-style varies. Currently in existing application , music is organized using playlist and different types of moods. When user will click to their current type of mood it might happen that they don't like the play – list created by the developer because here developer is also playing the role of user so it's hard to define that's what kind of play -list user would like .As a music lover I thought to give a effort less music player like emotion recognition to play music

1.3 ORGANIZATION OF THE REPORT

The thesis is organized into 6 chapters, describing each part of the project with detailed illustration and system design diagrams. The chapters are as follows:

Chapter 1: consists of Introduction, Problem statement, Motivation and Objectives etc.

Chapter 2: consists of Literature survey details of the project alongside their detailed methodologies, advantages, disadvantages etc.

Chapter 3: : consist System design of the project with its preliminary design such as overall Architecture diagram and process flow diagram which tells about the modules integration in the project.

Chapter 4: consists of Detailed system design or module description with their input and algorithmic steps involved in each module to derive the output as per the user requirement.

Chapter 5: consists the details about the experiments that has been performed along with their outcomes. The detailed result of the project is also portrayed in this chapter.

Chapter 6: conclude the project report with all the results and implementation procedure that has been underwent during the project development. The future works and excellence of implemented project is detailed.

The above mentioned six modules are followed up with the references which deliberately explains and list all the reference documents used during the various phases of the project, which includes the journal papers, conference papers, white papers, articles and websites referred for tutorials.

CHAPTER 2

LITERATURE SURVEY

This chapter discusses the related works in the proposed system and their limitations and modules wise literature survey.

2.1 EMOTION BASED MUSIC RECOMMENDATION SYSTEM

Florence, S. Metilda, and M. Uma. [1] proposed a system that 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. 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 user would not have to waste any time in searching or to look up for songs. The proposed architecture contained three modules, namely, Emotion extraction module, Audio extraction module and Emotion-Audio extraction module. Although it had some limitations like the proposed system was not able to record all the emotions correctly due to the less availability of the images in the image dataset being used. The image that is fed into the classifier should be taken in a well-lit atmosphere for the classifier to give accurate results. The quality of the image should be at least higher than 320p for the classifier to predict the emotion of the user accurately. Handcrafted features often lack enough generalizability in the wild settings.

2.2 REAL TIME EMOTION BASED MUSIC PLAYER

Muhammad, S., Ahmed, S.et al.[2] proposed a work focusing on detecting human being's emotions based on various facial cues and visual information. Then using the emotion detection model to associate these emotions with a music player that plays music mapped to the mood. The methods in this paper includes deep learning's CNN with pretrained transfer learning model, five layer model and global average pooling model. It uses FER dataset for training and testing purpose. They also implemented a web application using python flask framework to provide user interface experience. The limitations or research gap in this research manuscripts is that they have limited set of songs for each emotions and are not classified based on emotion.

2.3 FACE DETECTION USING HAAR CASCADE ALGORITHM

Arfi, Asif Mohammed, et al. [3] proposed a system to detect or recognise face of a person using six different Haar features by training handpicked database. They used sample images from various person with various pose and then trained them using numerous stages of classifying such as weak classifier and strong classifier to detect face if the image pass through stages of detector.The modules in this paper includes image capturing, generating dataset, converting image to gray scale, feature extraction and face detection process. The limitations of this work includes inaccurate results while rotation or shift of face and poor lightning conditions.

2.4 FACE DETECTION USING OPENCV

Khan, Maliha, et al. [4] proposed a paper which uses various technique for detecting face using different type of face detection algorithm that is available currently. The algorithm includes Haar cascade frontalface, Eigenfaces, Fisher-Face algorithm and local binary patterns histograms. This paper uses open CV, a deep learning library available in keras framework to capture real time video and then preprocess the video to convert into gray scale image to provide as input to the above mentioned algorithm to detect face and then compared their results.

2.5 ROBERT THAYERS MOOD MODEL

Bhat, Aathreya S., et al. [6]. In most existing methods of music mood classification, the moods of songs are divided according to psychologist Robert Thayer's traditional model of mood. The model divides songs along the lines of energy and stress, from happy to sad and calm to energetic, respectively. The eight categories created by Thayer's model include the extremes of the two lines as well as each of the possible intersections of the lines (e.g. happy-energetic or sad-calm).Faster tempos are associated with high-energy songs, and slower tempos with lower energy, sadder songs. Loudness, or intensity of a song can be connected with anger, while softer songs would suggest tenderness, sadness, or fear. Higher overall pitch can be an indicator of happiness, carefree and light moods within a song, while lower pitch implies a darker, sad, and serious tone. Timbre, the tonal component of a piece created by harmonics, is a curious indicator of mood. According to a group of researchers from the BNM Institute of Technology in Bangalore, India, "timbre stimulates human energy levels without regard to rhythmic or harmonic saturation. Sound sources that have simple harmonic profiles have 'darker' timbres and tend to soothe human emotions". This same group of researchers produced a correlation table of intensity, timbre, pitch, and rhythm in identifying various moods

2.6 MUSIC EMOTION RECOGNISTION USING CLUSTERING

Bakhshizadeh, Mahta, et al. [5] proposed a paper with one of the main goals is to generate personalized playlists automatically for each user. In this paper, a data scientific approach is provided to model the music moods which are created by clustering the tracks extracted from users' listening. Each Cluster consists of music tracks with similar audio features. Knowing which music track is currently being listened by users, their mood would be specified by determining the cluster of that music. It is presumed that playing the other music tracks contained in the same cluster as the next tracks will enhance their satisfaction. A suggestion for making the results visually interpretable which could help the corresponding music players with GUI design is provided as well. The module includes data collection and preprocessing from spotify REST api collecting audio feature of each song, model extraction and evaluation and playlist generation.

CHAPTER 3

SYSTEM DESIGN

This chapter explains about the overall architecture, flowchart of the proposed work. It also addresses the modules included in the proposed work.

3.1 SYSTEM ARCHITECTURE

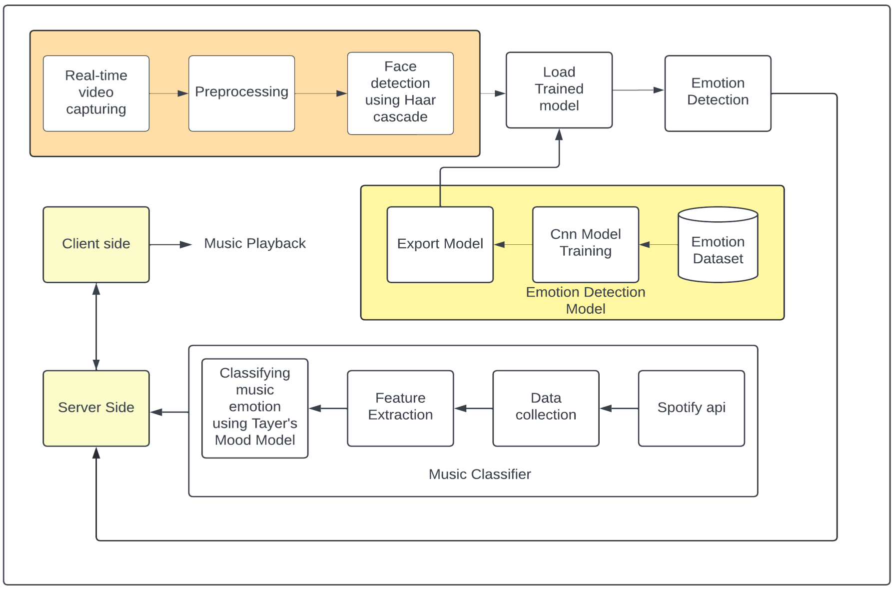


Figure 3.1: Overall Architecture of Proposed System.

Figure 3.1 depicts the overall system architecture. The proposed system works on Emotion based music playback. Emotion dataset are labelled and categorised with 5 labels such as happy, sad, neutral, angry, surprise. Then they are loaded to the CNN model for training the dataset and then export the model. For detecting face we use haar cascade classifier which detect face of a person and then the cropped face image is passed to trained model to predict the emotion of the user.To classify music, data is collected from spotify api and then clustered, labelled based on tayer mood model. Clustering model is created to predict the mood of the song. And then the song is used to map with the appropriate emotions and play music in the website.

3.2 DETAILS OF DATASET

For Emotion recognition, emotion image datasets are collected from the Kaggle named FER. The dataset consists of 48x48 pixel gray scale images of faces. The faces have been automatically registered so that the face is more or less centered and occupies about the same amount of space in each image. The facial expression is categorized into one of seven categories (0=Angry, 1=Disgust, 2=Fear, 3=Happy, 4=Sad, 5=Surprise, 6=Neutral). This work ignores disgust, fear due to less samples compared to others to provide better result and mapping of disgust, fear songs is inappropriate.

3.3 DATA PRE-PROCESSING

Dataset preprocessing includes the image augmentation of the dataset. Image augmentation is a process of creating new training examples from the existing ones. To make a new sample, you slightly change the original image. For instance, you could make a new image a little brighter; you could cut a piece from the original image; rotate, scale, shift etc.. you could make a new image by mirroring the original one, etc

3.4 FACE DETECTION

The main objective of this session is to capture images so here we are using the common device i.e, webcam. for that purpose this system is using the computer vision library.This makes it easier to integrate it with other libraries which can also use NumPy and it is mainly used as a real time computer vision.In the initial process when execution starts it starts to access the camera stream and captures images and converting the image into gray scale and cropping for further process. This system use an haar cascade algorithm which is a pretrained face detection model that could take the images and process and display face detected if it detect face. Then the image is cropped and made input to the CNN model as shown in the Figure 3.2

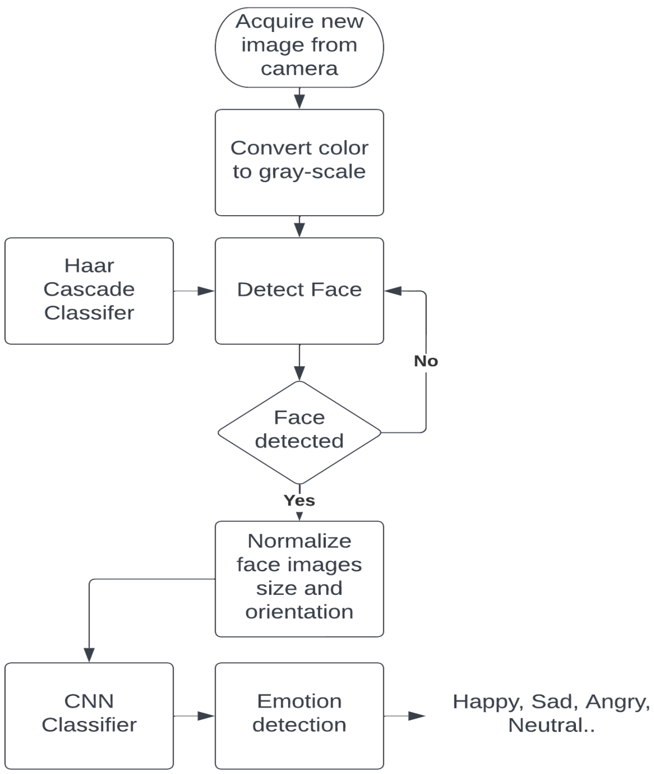


Figure 3.2: Flow Diagram of Face detection

3.5 EMOTION DETECTION

The proposed system builds a CNN model to extract feature from image layer by layer. For this purpose we have used four blocks of convolutional layers with size doubling every layer from 32. Each layer has pooling layer of size of 3*3 with max-pooling. Activation Function used is ‘ELU’. It has two dense or fully-connected layer with 2304 nodes. Finally we have final dense layer or output layer with 5 nodes classifying happy, sad, angry etc..

3.6 WEB DEVELOPMENT

The system builds a web interface for the user to access via graphical user interface. Initially the user will have to initiate Web cam to capture the video of the user and the video is pre-processed to detect face and predict the emotion using model file stored in the server running on flask. And then the page is redirected to the playlist page based on emotion and it retrieves list of songs from the classified song data and render the content on the html page.

3.7 MUSIC EMOTION RECOGNITION

to classify song based on emotion we to acquire the data from spotify. Firstly establish a api connection with spotify api and after successful establishment the track and audio feature of each song is retrieved. Then the system combines all the collected dataframe into a single dataframe. Pre-process the data to remove unwanted and null entries. Label data using Tayer’s mood model and group songs using clustering. Split the data to train and test the model using clustering like linear regression or Support vector machine (SVM) etc.. and then the model is evaluated and the output is stored in JSON format as show in the Figure 3.3

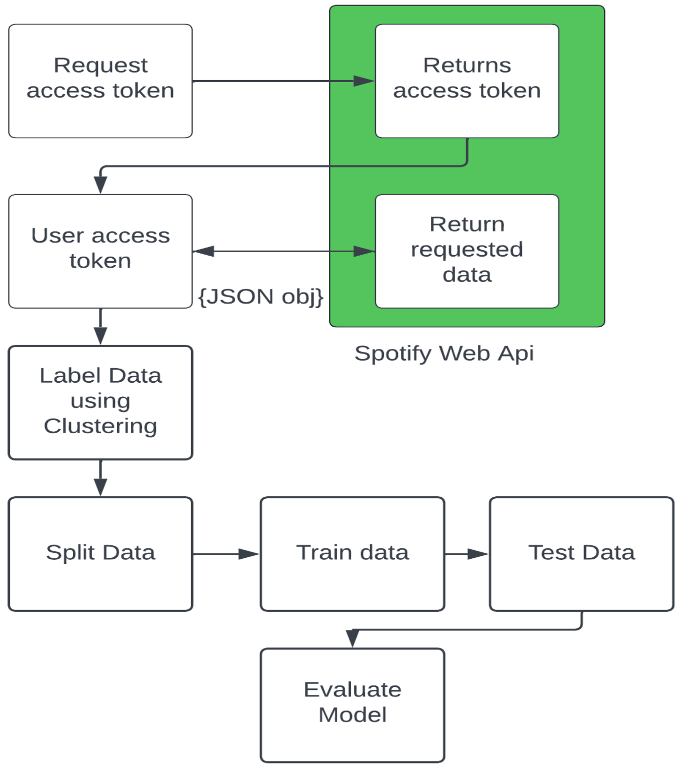


Figure 3.3: Flow Diagram of Music Emotion Recognition

3.8 FLOW-CHART DESIGN

Figure 3.4 shows the predicted workflow and design of the proposed system.

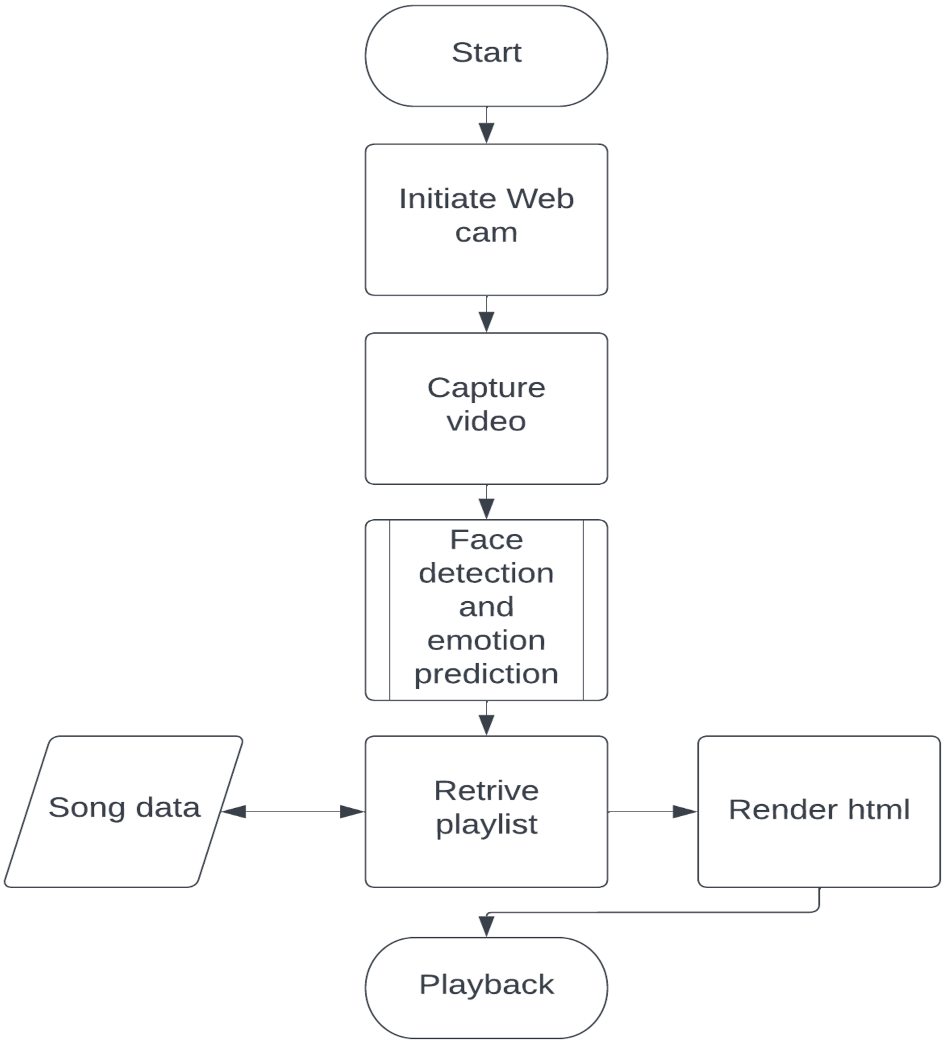


Figure 3.4: Workflow of the proposed system

CHAPTER 4

ALGORITHM IMPLEMENTATION

This chapter explains about the algorithmic implementation of the proposed work and details of the each modules.

4.1 FACE DETECTION USING HAAR CASCADE

A Haar cascade is defined as: a sequence of ”square shaped” functions which together form a family of wavelets or a base. It is also focused on ”Haar Wavelets,” which organize pixels on the picture into squares, based on a hair wavelet approach which was suggested in the paper by Paul Viola and Michael Jones Figure 4.1 shows the Haar features used to detect face. It is a learning approach based on computers, where many positive and negative representations are used to construct a cascade function. This is then used to detect objects Using ”integral image” principles in order to compute ”features” identified by the Haar cascades.

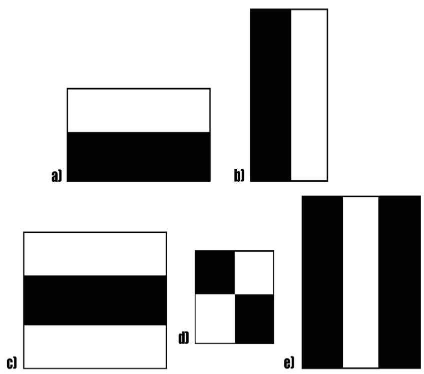


Figure 4.1: A sample of Haar features used by viola and jones

4.1.1 Pre-Processing

It is the process of converting the video into number of frames. This is done by CV2 from OpenCV Library. Processing a video means, performing operations on the video frame by frame. Frames are nothing but just the particular instance of the video in a single point of time. Frames can be treated as similar to an image. So, whatever operations we can perform on images can be performed on frames as well. The collected frames are to be converted into grayscale images and the images are resized into the sub window size (48×48 pixel).

4.1.2 Face Detection

The main objective of this process has identified where the facial image has located. If facial features are found, it will be indicated by a bounding box around the region of interest starts to read the frame. Before extracting feature, the system has converted every frame captured per second to a gray scale and extracted the entire feature one by one over it. Next, if any face found over the frame, then it starts to track the frame simultaneously till the frame available. However, when face features have not detected, the system skips the frame and continues to extract features over the next frame. The face detection recognition system presented here is constructed by using Haar cascades frontal face classifier.

4.2 CNN MODEL DETAILS

Figure 4.2 depicts the CNN model and overall architecture of the proposed system to extract feature from image layer by layer.

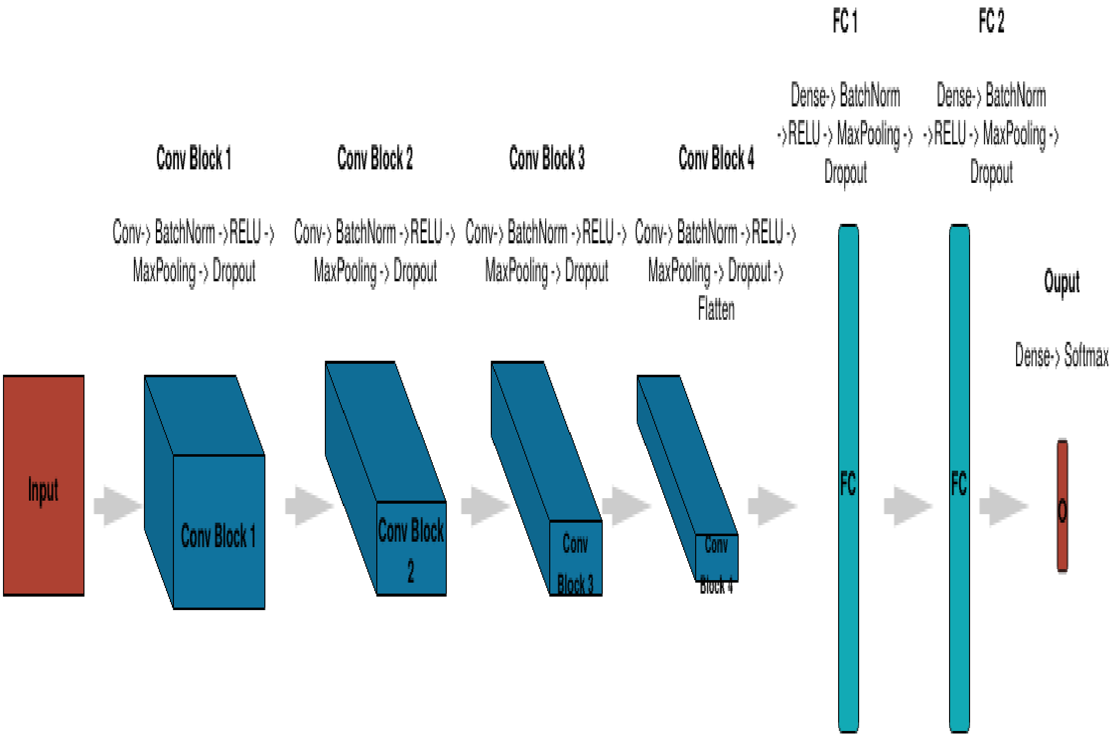


Figure 4.2: CNN Model Architecture

4.2.1 Sequential Layer

Sequential is a container of Modules that can be stacked together and run at the same time. Sequential layer is used to store feature vector returned by the ResNext model in a ordered way. So that it can be passed to the LSTM sequentially.

4.2.2 Pooling Layer

The pooling operation involves sliding a two-dimensional filter over each channel of feature map and summarising the features lying within the region covered by the filter. For a feature map having dimensions nh x nw x nc, the dimensions of output obtained after a pooling layer is

$$(nh - f + 1)/sx(nw - f + 1)/sxnc$$

4.2.3 Relu Activation Function

A Rectified Linear Unit is activation function that has output 0 if the input is less than 0, and raw output otherwise. That is, if the input is greater than 0, the output is equal to the input. The operation of ReLU is closer to the way our biological neurons work. ReLU is non-linear and has the advantage of not having any backpropagation errors unlike the sigmoid function, also for larger Neural Networks, the speed of building models based off on ReLU is very fast. Figure 4.3 shows the Relu activation function.

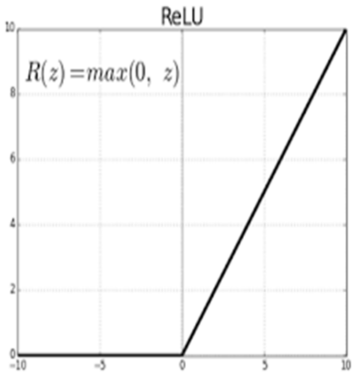


Figure 4.3: Relu Activation Function

4.2.4 Flatten Layer

After implementing convolutional operations and pooling, the resultant outputs with multiple features and dimensions. To flatten out the dimensions and feature map into a single column. This single column is a long vector of input data that can be passed onto the ANN.

4.2.5 Fully Connected Layer

Fully Connected layers in a neural networks are those layers where all the inputs from one layer are connected to every activation unit of the next layer. In most popular machine learning models, the last few layers are full connected layers which compiles the data extracted by previous layers to form the final output. It is the second most time consuming layer second to Convolution Layer

4.2.6 Dropout Layer

Dropout layer with the value of 0.4 is used to avoid overfitting in the model and it can help a model generalize by randomly setting the output for a given neuron to 0. In setting the output to 0, the cost function becomes more sensitive to neighbouring neurons changing the way the weights will be updated during the process of back propagation. Fig. 4.4 depicts the overview of dropout layer.

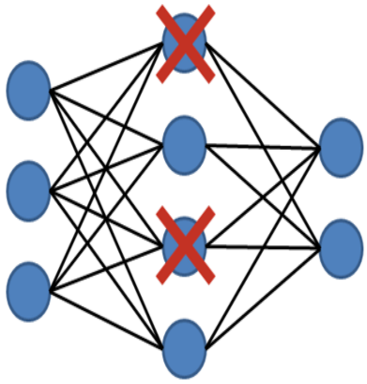


Figure 4.4: Dropout Layer

4.3 MODEL TRAINING

We build our own CNN model to extract feature from image layer 8 by layer. For this purpose we have used four blocks of convolutional layers with size doubling every layer from 32. Each layer has pooling layer of size of 3*3 with max-pooling. Activation Function used is 'ELU'. It has two dense or fully-connected layer with 2304 nodes. Finally we have final dense layer or output layer with 5 nodes classifying happy, sad, angry etc

4.3.1 Train Test Split

The dataset is split into train and test dataset with a ratio of 70percent 24282 train samples and 30percent of 5937 test samples.

4.3.2 Training

The training is done for 30 epochs with a learning rate of 1e-3 (0.0010),weight decay of 1e-3 (0.0010) using the Adam optimizer.

4.3.3 Adam Optimizer

Adam is an optimization solver for the Neural Network algorithm that is computationally efficient, requires little memory, and is well suited for problems that are large in terms of data or parameters or both. Adam is a popular extension to stochastic gradient descent. To enable the adaptive learning rate Adam optimizer with the model parameters is used.

4.3.4 Cross Entropy

To calculate the loss function Cross Entropy approach is used because we are training a classification problem.

4.3.5 Softmax Layer

A Softmax function is a type of squashing function. Squashing functions limit the output of the function into the range 0 to 1. This allows the output to be interpreted directly as a probability. Similarly, softmax functions are multi-class sigmoids, meaning they are used in determining probability of multiple classes at once. Since the outputs of a softmax function can be interpreted as a probability (i.e.they must sum to 1), a softmax layer is typically the final layer used in neural network functions. It is important to note that a softmax layer must have the same number of nodes as the output later. In our case softmax layer has 5 output nodes i.e Happy, Angry, Sad, Neutral, Surprise, also Softmax layer provide us the confidence(probability) of prediction.

4.3.6 Export Model

After the model is trained, we have exported the model to the flask server running on the local machine. So that it can be used for prediction on real time data.

4.3.7 Emotion prediction

The model is loaded in the application. The new video for prediction is preprocessed and passed to the loaded model for prediction. The trained model performs the prediction and return if the image is a happy,sad,angry,neutral,surprise.

4.4 MUSIC EMOTION RECOGNITION

This module consists of 3 principal sections containing data collection and preprocessing, model extraction, playlist generation.

4.4.1 Data Collection

The first thing needed in this module is a dataset containing list of songs and their audio features. For this purpose the paper uses Spotify api to get the audio features of every music track in the playlist, then their Spotify IDs are given to Spotify Web API. Based on simple REST principles, the Spotify Web API endpoints return JSON metadata about music artists, albums, and tracks, directly from the Spotify Data Catalogue. Web API also provides access to user's related data, like playlists and music that the user saves in the Your Music library. Such access is enabled through selective authorization, by the user. ”Get Audio Features for a Track” is one of the various services provided by Spotify API which gives audio feature information for a single track identified by its unique Spotify ID. The mentioned audio features given by the Spotify API involve the values in Table 1 for each music track.

4.4.2 Model extraction

To cluster the music tracks in the dataset, a set of these values must be chosen to form a vector for each track. Since the goal is to match the moods of the users to the music mood, some basic features such as duration, key, time-signature, tempo, and loudness seem to be irrelevant in comparison to superior values such as valance, arousal and energy. Also, since the presence of an audience in the recording does not appear to be relevant enough to the emotional moods, the value of liveness is better not to be involved in the mentioned problem. Rule base method is a method created for the detection of emotions using arousal prediction data, valence area and thayer labels.There are 4 main rules that are used to detect Thayer based emotions on the arousal and valence area. If arousal is high and valence is high it fits in first cluster and arousal high valence low it belongs to second cluster and so on.

4.4.3 Playlist Generation

Once the model is created, it is used to predict the mood of the song by passing the track id of the song and generate a set of playlist based on the emotion and then export the list of songs as JSON format.

CHAPTER 5

IMPLEMENTATION AND RESULTS

This chapter showcases the results and final output of the proposed work.

5.1 WEB APPLICATION DASHBOARD MODULE

Figure 5.1 depicts the Dashboard screen provided. It is running on the localhost with flask server.

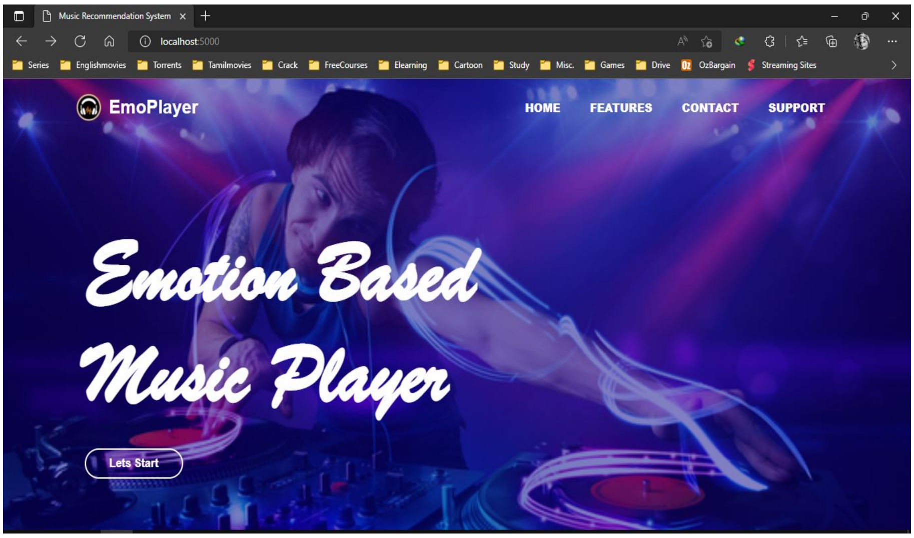
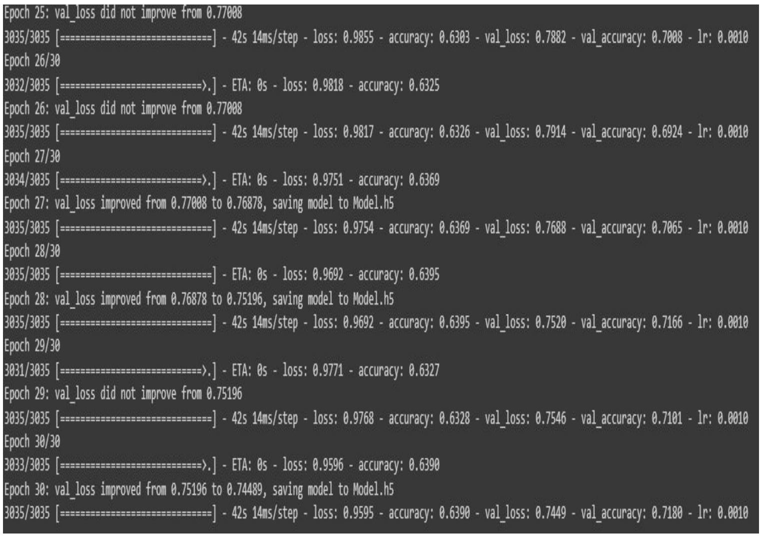


Figure 5.1: Web Application Dashboard Screen

5.2 TRAINING AND TESTING ANALYSIS

The whole dataset contains 29000 labelled image. The Dataset is divided as 80percent for training and the remaining consider as the testing dataset. Figure 5.2 shows the validation accuracy and loss with 30 epoch is 71.8percent and 74percent.



```
Epoch 25: val_loss did not improve from 0.77008
3035/3035 [==============================] - 42s 14ms/step - loss: 0.9855 - accuracy: 0.6303 - val_loss: 0.7882 - val_accuracy: 0.7008 - lr: 0.0010
Epoch 26/30
3032/3035 [============================>.] - ETA: 0s - loss: 0.9818 - accuracy: 0.6325
Epoch 26: val_loss did not improve from 0.77008
3035/3035 [==============================] - 42s 14ms/step - loss: 0.9817 - accuracy: 0.6326 - val_loss: 0.7914 - val_accuracy: 0.6924 - lr: 0.0010
Epoch 27/30
3034/3035 [============================>.] - ETA: 0s - loss: 0.9751 - accuracy: 0.6369
Epoch 27: val_loss improved from 0.77008 to 0.76878, saving model to Model.h5
3035/3035 [==============================] - 42s 14ms/step - loss: 0.9754 - accuracy: 0.6369 - val_loss: 0.7688 - val_accuracy: 0.7065 - lr: 0.0010
Epoch 28/30
3035/3035 [==============================] - ETA: 0s - loss: 0.9692 - accuracy: 0.6395
Epoch 28: val_loss improved from 0.76878 to 0.75196, saving model to Model.h5
3035/3035 [==============================] - 42s 14ms/step - loss: 0.9692 - accuracy: 0.6395 - val_loss: 0.7520 - val_accuracy: 0.7166 - lr: 0.0010
Epoch 29/30
3031/3035 [============================>.] - ETA: 0s - loss: 0.9771 - accuracy: 0.6327
Epoch 29: val_loss did not improve from 0.75196
3035/3035 [==============================] - 42s 14ms/step - loss: 0.9768 - accuracy: 0.6328 - val_loss: 0.7546 - val_accuracy: 0.7101 - lr: 0.0010
Epoch 30/30
3033/3035 [============================>.] - ETA: 0s - loss: 0.9596 - accuracy: 0.6390
Epoch 30: val_loss improved from 0.75196 to 0.74489, saving model to Model.h5
3035/3035 [==============================] - 42s 14ms/step - loss: 0.9595 - accuracy: 0.6390 - val_loss: 0.7449 - val_accuracy: 0.7180 - lr: 0.0010
```

Figure 5.2: Training and Testing

5.3 FACE DETECTION

Figure 5.3 shows the detected face using HAAR cascade classifier and bounding box is drawn in the region of interest.

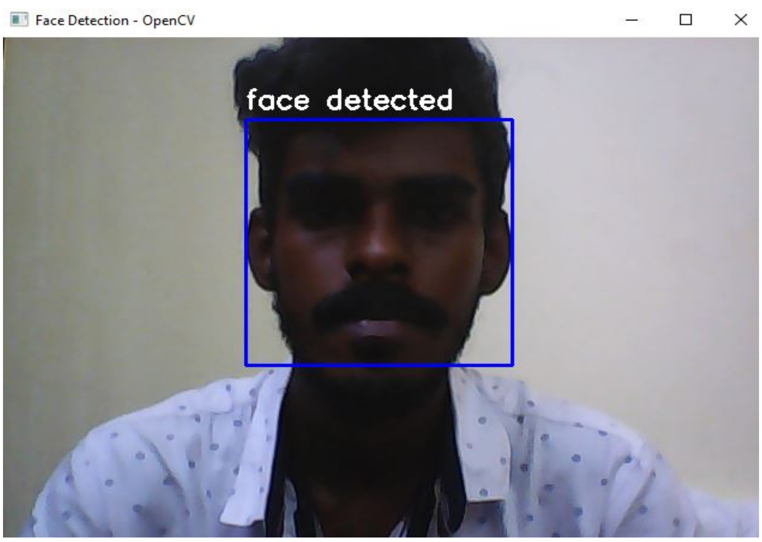


Figure 5.3: Face detection

5.4 EMOTION DETECTION

Emotion of the person is detected and displayed in real time by making bounding box around the face. Figure 5.4, Figure 5.5, Figure 5.6 shows the neutral, happy and angry face of the user.

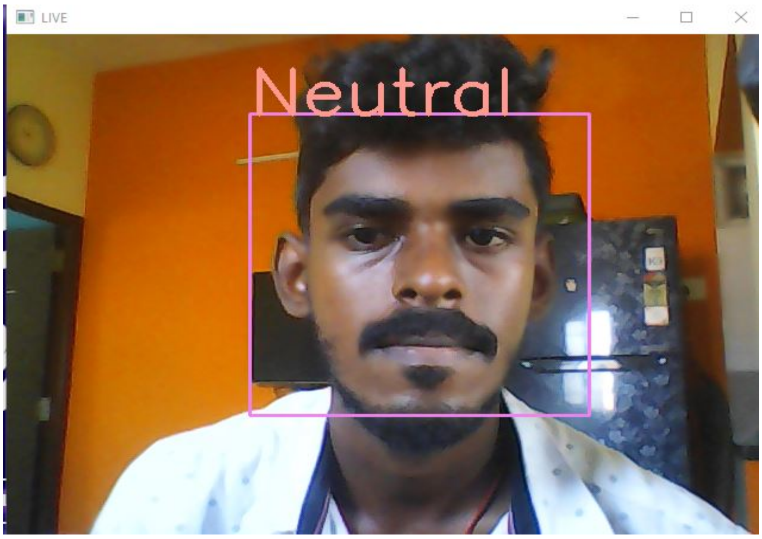


Figure 5.4: Neutral Emotion Detected

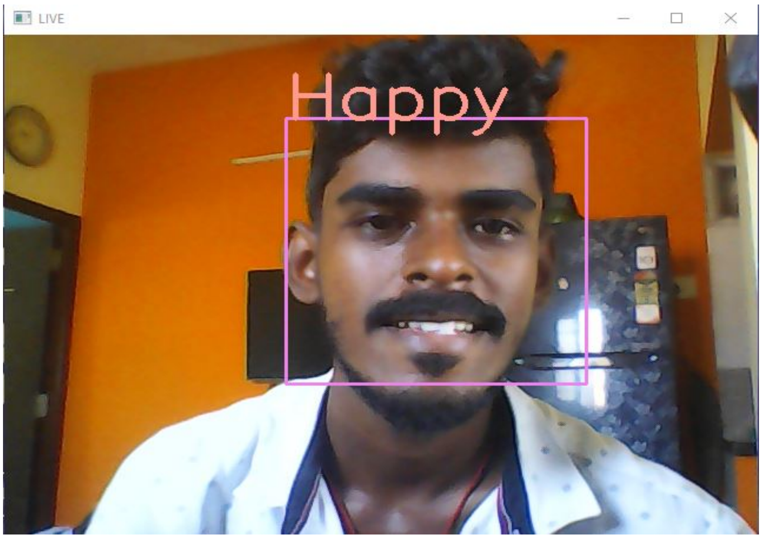


Figure 5.5: Happy Emotion Detected

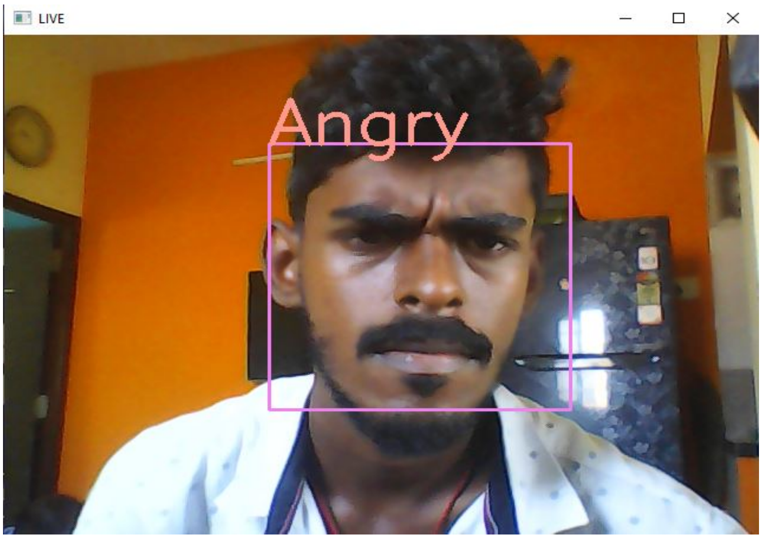


Figure 5.6: Angry Emotion Detected

5.5 HAPPY SONG PAGE IN WEB APPLICATION

Figure 5.7 Displays the list of happy songs based on user happy emotion by fetching the song list from JSON file.

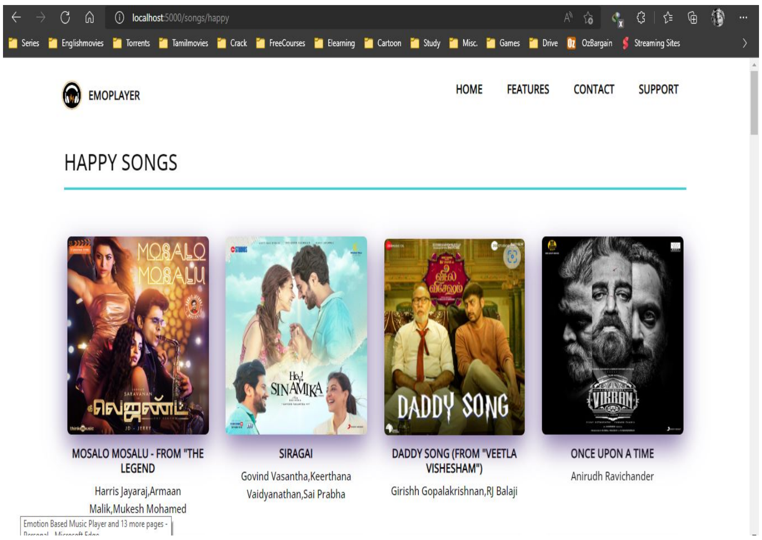


Figure 5.7: Happy Song Playlist

CHAPTER 6

CONCLUSION AND FUTURE WORK

6.1 CONCLUSION

In this paper, an emotion detection model is proposed to recommend music based on one's mood. Our work aims to achieve the highest possible accuracy 71.8% and managed to predict the mood of the song pretty acurately. This was a breakthrough as such a lightweight model is easily mountable on small devices, which adds to real-world scenarios' applicability.

6.2 FUTURE WORK

As part of future work, our models' accuracy could be further improved by applying landmark detection techniques that cancel out irrelevant facial features from the image during training. The application can be extended to recommend and classify songs in real-time. It can also be improved in web camera loading time and prediction time of emotion. The songs are only classified into four clusters and can be extended to many clusters in the future.

REFERENCES

[1] S Metilda Florence and M Uma. Emotional detection and music recommendation system based on user facial expression. In *IOP Conference Series: Materials Science and Engineering*, volume 912, page 062007. IOP Publishing, 2020.

[2] Sulaiman Muhammad, Safeer Ahmed, and Dinesh Naik. Real time emotion based music player using cnn architectures. In *2021 6th International Conference for Convergence in Technology (I2CT)*, pages 1–5. IEEE, 2021.

[3] Asif Mohammed Arfi, Debasish Bal, Mohammad Anisul Hasan, Naeemul Islam, and Yasir Arafat. Real time human face detection and recognition based on haar features. In *2020 IEEE Region 10 Symposium (TENSYMP)*, pages 517–521. IEEE, 2020.

[4] Maliha Khan, Sudeshna Chakraborty, Rani Astya, and Shaveta Khepra. Face detection and recognition using opencv. In *2019 International Conference on Computing, Communication, and Intelligent Systems (ICCCIS)*, pages 116–119. IEEE, 2019.

[5] Mahta Bakhshizadeh, Ali Moeini, Mina Latifi, and Maryam Tayefeh Mahmoudi. Automated mood based music playlist generation by clustering the audio features. In *2019 9th International Conference on Computer and Knowledge Engineering (ICCKE)*, pages 231–237. IEEE, 2019.

[6] Aathreya S Bhat, VS Amith, Namrata S Prasad, and D Murali Mohan. An efficient classification algorithm for music mood detection in western and hindi music using audio feature extraction. In *2014 fifth international conference on signal and image processing*, pages 359–364. IEEE, 2014.