

Emotion based Music Recommendation System

Mihir Brahmane , Suraj Jawalkar , Amit Narute , Mayuri Jagtap
Guide : Prof. Vishal Nayakwadi
Department of AI&DS Engineering ,
Zeal College of Engineering & Research ,
Narhe , Pune

Abstract : *The human face assumes a significant part in knowing a person's state of mind. The expected information is extricated from the human face clearly utilizing a camera. One of the uses of this information can be for extricate the data to reason the mind-set of a person. This information can then be utilized to get a rundown of tunes that follow the "state of mind" you got from the information given before. This kills the tedious and repetitive undertaking of physically isolating or gathering melodies into various records and helps in creating a fitting playlist in light of a person's close-to-home elements. The goal of the Facial Expression-Based Music Player is to fetch the data and interpret it before making a playlist with the provided properties. Consequently, our proposed framework canters around identifying human feelings for creating a feeling-based music player, which are the methodologies utilized by accessible music players to distinguish feelings, the approach our music player follows to recognize human feelings, and the way things are smarter to involve our framework for feeling identification. A concise thought regarding our frameworks' working, playlist age, and feeling order is given. Watchwords: Tensor-Flow & Keras , MediaPipe , Convolutional Neural Networks , Deep Learning , StreamLit RTC , Holistic .*

1. INTRODUCTION

The new possibility in the field of music information retrieval is for computers to automatically analyze and comprehend music . Numerous research topics in this field are pursued by researchers, including computer science, digital signal processing, mathematics, and statistics applied to musicology. This is due to the high diversity and richness of music content. Late improvement in music data recovery incorporates programmed sound kind/temperament arrangement, music closeness calculation, sound craftsman recognizable proof, sound to-score arrangement, question by-singing/murmuring, etc. Content-based music suggestions are one of the most achievable applications that can be given. From the setting data, we can accomplish more intelligent context-based music suggestions. Multidisciplinary

endeavors like feeling portrayal, feeling Identification / acknowledgment , highlight based arrangement, and satisfying-based suggestion are required for the accomplishment in the satisfied-based music proposal framework. Music scientific classification has been depicted really utilizing a feeling descriptor . A suspicion for feeling portrayal is that feeling can be viewed as a bunch of nonstop amounts and planned into a bunch of genuine numbers. A circumflex model in which each effect is displayed over two bipolar dimensions was proposed by researchers as a pioneering effort to describe human emotions. Those two aspects are charming, unsavory, and arousing. As a result, it is possible to define each affect word as a particular combination of pleasure and arousal. Afterward, another specialist adjusted Russel's model to music. "Excitement" and "valence" are the two primary aspects in Thayer's model. Feeling terms were portrayed as quiet to enthusiastic along the excitement aspect. Feeling terms were named as quiet to vivacious and pessimistic to good along the valence aspect. With Thayer's model, the two-layered feeling plane can be partitioned into four quadrants with eleven feeling modifiers set over them. Then again, Xiang et al. proposed a "psychological state change organization" for depicting feelings of change in people. Test data is used to calculate the probability of each transition between two states. In any case, different feelings, for example, apprehensive and energized, aren't thought of. Automatic emotion detection and recognition in music is expanding rapidly thanks to digital signal processing advancements and a variety of efficient feature extraction techniques. Many other potential applications, such as music entertainment and systems for human-computer interaction, may benefit greatly from emotion detection and recognition . Feng introduced the primary exploration of feeling recognition in music. There are Several Deep Learning Models which are used to build this Mood Based Song Recommendation System .

2. LITERATURE SURVEY

Renuka R. Londhe et al. proposed a paper which zeroed in on the investigation of changes in the ebbs and flows of the face and the powers of the relating pixels. The author classified the emotions using Artificial Neural Networks (ANN). The creator likewise proposed different methodologies for the playlist. Zheng et al. proposed two critical classifications for facial element extraction, which included appearance-based highlight extraction and mathematical-based include extraction, which included extraction of a few fundamental marks of the face like mouth, eyes, and eyebrows. It determines the attitude of the client by utilizing look. People frequently express their inclination by their demeanors, hand motions, and by raising the voice of tone, yet generally people express their sentiments by their face. A feeling-based music player diminishes the time intricacy of the client. For the most part, individuals have an enormous number of melodies on their playlist. Playing melodies haphazardly doesn't fulfill the mind-set of the client. This framework assists clients with playing melodies naturally as indicated by their state of mind. The picture of the client is caught by the Web camera, and the pictures are saved. The pictures are first changed over from RGB to paired design. This course of addressing the information is known as an element point location strategy. This cycle should likewise be possible by utilizing the Haar Fountain innovation given by OpenCV. The music player is created by utilizing a Java program. It deals with the data set and plays the tune as indicated by the mindset of the user. Zeng et al. explored different advances in human influence acknowledgment. He concentrated on various strategies for dealing with recordings of affective states in either audio or visual form. The paper gives a point-by-point survey of general media-figuring techniques. The impact is portrayed as a model of feeling classes which incorporate joy, misery, dread, outrage, repugnance, and shock. The issues surrounding the development of an automatic, spontaneous affect recognizer that assisted in emotion detection were the primary focus of this paper. It likewise recognized a few issues that have been missed or stayed away from in uni-modular presented feeling acknowledgment. Parul Tambe et al. proposed a thought that computerized the communications between the clients and the music player, realized every one of the inclinations, feelings, and exercises of a client, and gave melody choice thus. The different looks of clients were recorded by the gadget to decide the feeling of the client and anticipate the class of the music. Jayshree Jha et al. proposed a feeling-based music player utilizing picture handling. This showed how different calculations and methods that were proposed by various creators in their examination could be utilized for associating the music player alongside human feelings.

It has in this manner helped in decreasing the endeavors of clients in making and overseeing playlists and giving a superb encounter to the music audience members by presenting to them the most appropriate melody as per the client's ongoing articulation. Anukritine et al. concocted a calculation that gives a rundown of melodies from the client's playlist as per the client's inclination. The calculation that was planned was centered around having less computational time and, furthermore, decreasing the expense remembered for utilizing different equipment. The fundamental thought was to isolate the feelings into five classifications, i.e., euphoria, miserable, outrage, shock, and dread, likewise given a profoundly exact sound data recovery approach that removed pertinent data from a sound sign quicker than expected. Aditya et al. fostered an android application that goes about as a tweaked music player for a client, utilizing picture handling to dissect and introduce tunes to clients as indicated by the client's state of mind. The application was created utilizing Shroud and furthermore OpenCV to carry out facial acknowledgment calculations. This paper likewise showed examination between different calculations utilized in facial discovery. The pictures of the client were caught using the front camera of the cell phone. By capturing their emotions, it aimed to satisfy music enthusiasts. A. Habibzad et al. proposed another calculation to perceive the facial inclination, which included three phases: pre-handling, highlight extraction, and grouping. The initial segment portrays different stages in picture handling that incorporate preprocessing and separating used to extricate different facial highlights. The subsequent part upgraded the eye and lip oval attributes, and in the third part, the eye and lip ideal boundaries were utilized to arrange the feelings. The acquired results showed that the speed of facial acknowledgment was obviously better than other regular methodologies. Prof. Nutan Deshmukh et al. zeroed in on making a framework that brings the feeling of the client utilizing a camera and afterward robotizes the outcome utilizing the feeling location calculation. This calculation catches the mind-set of the client after each concluded time frame, as the client's temperament may not be similar after some time; it might change. An emotion-based music system can be created using the proposed algorithm in an average of 0.95 to 1.05 seconds, which is faster than previous algorithms and lowers design costs. A system that makes use of Brain-Computer Interfaces, or BCIs, was described by Chang Liuet al. . BCI utilizes gadgets to convey messages to its handling frameworks. EEG equipment is utilized to screen the individual's mental perspective. The disadvantage of the plan is that they require the contribution of the client's mind consistently to play out the characterization. A calculation in view of MID is utilized to ceaselessly screen and cycle the signs got from the cerebrum of the client and utilize these signs to screen

and create feelings that the client is at present encountering effectively. Swati Vaid et al. investigated EEG. Electroencephalography (EEG) is a type of clinical science that records the electrical movement from the neurons of synapses. From within the brain's cells, the electrical activity of the neurons is recorded. In view of the recorded action of the neurons, a guess is made, and the feeling of the individual is assessed from that examination. This strategy is referenced above, in spite of the fact that it effectively gets the action of synapses yet neglects to fill the need for convey ability and financial matters.

3. METHODOLOGY

A) Dataset :- The 48x48 grayscale portraits of faces that make up the Emotional Datasets which were used to create the model. The seven emotions that are assigned to each image are Anger, Joy, Fear, Happiness, Sadness, Surprise and Neutral. The public test set consists of 1568 examples, while the training set consists of 7321 samples. Music Mood Collection where dataset that is used is a labelled dataset of size 264. It has 14 sections, including Name, Collection, Music Artist, User Id, Release Date, Prevalence, Danceability, Energy, Liveness, Valence, Beat, Key, Song Language and State of Mind. Dataset has been linked Directly from the Music Platforms like Spotify or YouTube through Stream-Lit RTC libraries. Therefore , no need to collect and Store different musical datasets in a Personal Data Storage Disk for training and application purpose which saves time and processing with memory. Data is Directly Processed on Music Platform through the Facial Inputs of User and Personal Parameters. This Leads in fetching and Sorting all the Music Recommendation Process through Facial and Emotional Data Processing and Integrating.

B) Feeling Identification :- Face detection is one application of computer vision technology. Calculations are made and prepared in this method to accurately find faces or items in object identification or related frameworks in photographs. It is possible to detect something in real time from an image or frame of a video. Face detection primarily aims to identify the face that is contained within the frame by minimizing external noises and other elements. This strategy depends on AI, and an assortment of information documents is utilized to prepare the outpouring function. This utilizes AI methods to extract preparing information with a serious level of exactness. We utilize the pre-prepared network, which is a consecutive model, as an erratic element extractor while performing highlight extraction, permitting the information picture to advance to the following layer, halting there, and involving that layer's results as our highlights. Utilize a couple of channels on the

grounds that the underlying convolutional network layers recover the most significant level qualities from the caught picture. As we add further levels, we duplicate the quantity of channels by a few, contingent upon how enormous the channel was in the first layer. The picture's class will either be parallel or multi-class to recognize various kinds of dress or to distinguish digits. The learned properties in a brain network can't be perceived in light of the fact that brain networks are like a "black box." Thus, the CNN model is used in Image Classification and Face Detection. CNN essentially returns the discoveries subsequent to getting an information picture. The model that was advanced by loads utilizing CNN is stacked to distinguish feelings. At the point when a client takes a continuous picture, the picture is submitted to a CNN model that has proactively been prepared. The CNN model then, at that point, predicts the inclination and adds a mark to the picture. CNN Models are integrated with Deep Neural Networks and Tensor-Flow through a python Programming language and Libraries for Efficient Facial Recognition and Emotion Detection Tasks. Further , For an additional basis , Media-Pipe Libraries are used for Hand gesture Identification and Classification which will Recognise the patterns of Hand gesture through a emotion Body Language methods and Classification. It is also used with Holistic Function which Captures the functions of Hand Gestures and Recognise the Patterns and Emotions attached to a particular function (ex:- Fist Closed for anger , Palm Open for Happiness or Joy).

C) Music Recommendation :- Every feature in our dataset has a magnitude that indicates its intensity, and these features are also regarded as acoustic features of that particular song. Greatness might be measured on various scales. Along these lines, there are 4-5 principal highlights among 10-14 that contribute more. Along these lines, melodies are characterized by various classifications like cheerful, miserable, nonpartisan, and vivacious. For this, we utilized the Convolution Brain Organization's powerful, discriminative highlights. Recognition of a particular emotion is done by using 264 neurons . To determine which features were most essential for classifying the image, load the input image for which you wish to view the feature map. Involving the ReLu actuation capability in convolution brain network engineering, channels or element locators are applied to the info picture to create highlight guides or enactment maps. Edges, vertical and horizontal lines, bends, and other characteristics already present in the image can be identified using feature detectors or filters. It is feasible to use pooling from the base, greatest, or normal. In any case, when contrasted with min or normal pooling, max pooling gives better execution. Categorical Cross Entropy is the Loss Function used to reduce the error rate and RMS Prop Optimizer is used for Optimizing the Working of the Model .

D) User Interface :- Using deep neural networks, the method teaches the best feature abstraction. Profound Brain Organizations are an effective methodology for facial emotion recognition, personalized music recommendation and some more. Convolutional brain organizations have been demonstrated exceptionally compelling in regions like picture acknowledgment and order. The proposed framework can identify the looks of the client utilizing a convolutional brain organization model. In this venture, a primary site page is planned utilizing the StreamLit structure, where a picture of the client is captured. The picture caught is then shipped off the model to anticipate the feeling of the client. When the inclination is identified, the Spotify programming interface is called by the Python module Spotify to demand music tracks, which are then shown in the UI. The detect emotion capability deals with feeling identification. There are four CNN layers in the model, and it is trained for 50 epochs. UI Execution The UI is worked with the stream-lit system. When the page is stacked, a following Musical Platform Website through changing Interface is opened to catch the picture of the client. Keras Backend Library with 264 Neuron classifiers is used to determine whether a face is present in the captured image. CV2 Module is used to provide an easy-to-use interface for working with Real-Time Image and Video processing. There is also a selection options of textbox with checklist for personalization of the Language of the song and the artist of a Playlist for Privacy and Individual Requirement Management. The Spotipy module uses the emotion that has been detected to display the emotion on the screen. In addition, the Spotipy module searches the Spotify library for songs that correspond to the user's mood and then displays those songs on the screen. The tracks are implanted so that the client could pay attention to the tune in the web application itself or explore the spotify application by tapping on the specific track.

4. PROPOSED SYSTEM

The proposed framework benefits us by introducing associations between the client and the music player. The motivation behind the system is to catch the face appropriately with the camera. Caught pictures are taken care of by the Convolutional Brain Organization, which predicts the inclination. Then, at that point, the feeling got from the caught picture is utilized to get a playlist of tunes. The primary point of our proposed framework is to give a music playlist, consequently changing the client's temperaments, which can be cheerful, miserable, normal, or

shocked. The proposed framework distinguishes the feelings, and on the off chance that the point includes a gloomy inclination, a chosen playlist will be introduced that contains the principal reasonable kinds of music that will upgrade the mind-set of the individual decidedly. There are four modules in music recommendation based on facial emotion recognition.

- Real-Time Capture :- The system is responsible for Capturing the accurate user's face in Real Time.
- Face Recognition :- User's face will be used as input. The convolutional brain network is modified to assess the highlights of the client picture.
- Emotion Detection :- The system extracts features from the user image to determine the user's emotions and captions are generated based on the user's feelings.
- Music Recommendation :- The song is proposed by the suggestion module to the client by planning their feelings according to the temperament sort of the melody.

CNN is used for Image Processing and Face Detection. Tensor-Flow is used to Simplify the Complex tasks and Keras is use to handle the complex tasks. RMS Prop optimizer is used for optimizing the Model performance and Categorical cross entropy is used as a Loss Function with ReLU activation Function because it never let values became negative while training. MediaPipe and Holistic is used to capture Hand Gestures and Identify them. StreamLit RTC is used to deploy apps in various platforms and control the user Interfaces. CV2 is used for providing east-to-use Interface for Image Processing. Various Functional Options are provided for Personal Requirements through Text Box and Check-List Options.

RAM :- 4 GB or higher

ROM :- 100 GB or higher

Programming Framework :- Windows 10 or higher

Processor :- I3 Processing System or higher

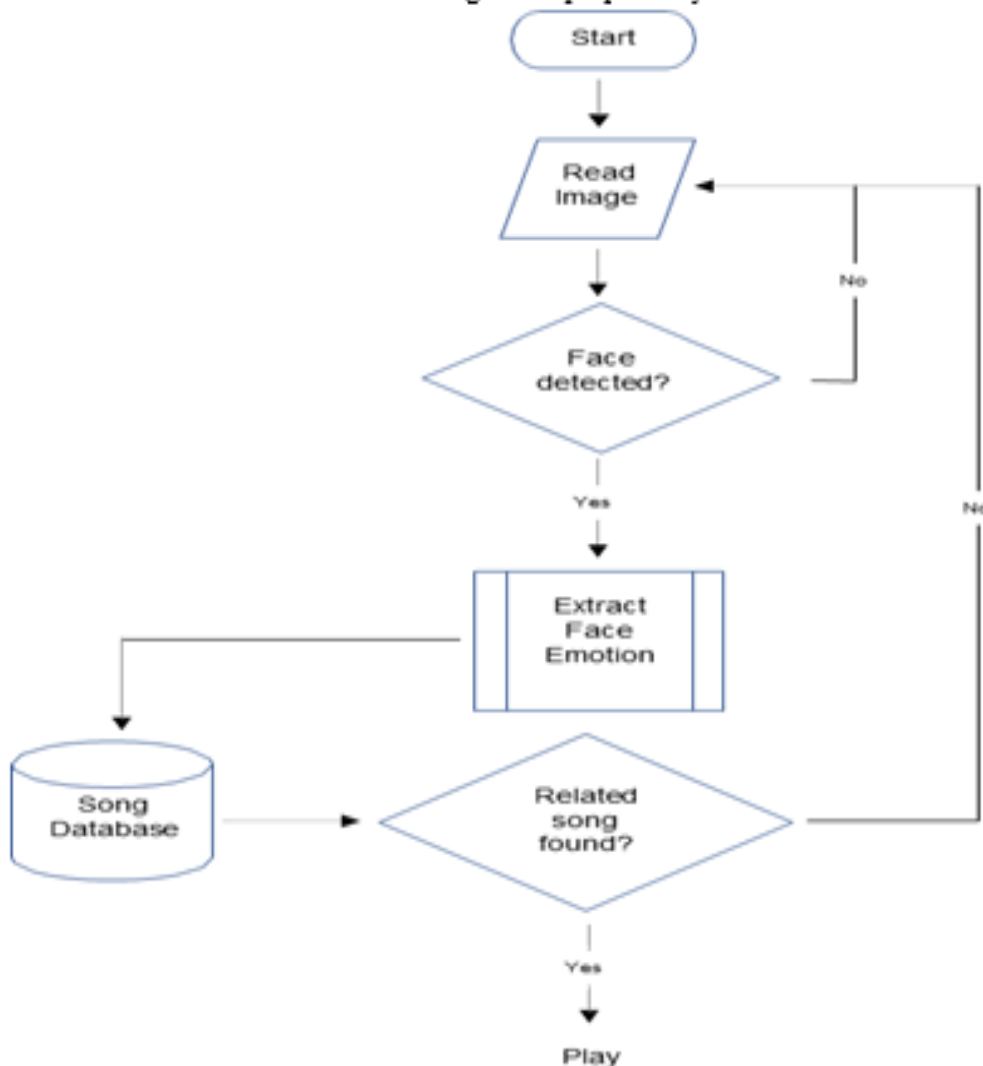
Coding Language :- Python 3.5 or higher version

Programming Platform :- Jupyter Notebook

Coding Libraries :- Tensorflow , StreamLit

A quality result is one that meets the necessities of the end client and presents the data plainly. In any framework consequences of handling are conveyed to the clients and to other framework through yields.

Flow diagram of proposed system



5. RESULT & ANALYSIS

We assessed some of the examinations, in which Convolutional brain networks are used for detecting the Accurate emotions and Recommending the Optimal Songs. Correlation of elated calculations and Exactness values are given for each review. The effectiveness of emotion detection is enhanced by using a convolutional neural network. Hyperparameters for the prepared CNN network are working at higher precision. The weight update at the end of each batch is controlled by the learning rate. A few ages of the cycles of the whole preparation dataset are given to the organization during preparation. Group size is the quantity of examples displayed in the organization before the loads are refreshed. The model can learn nonlinear prediction boundaries thanks to activation functions. Adam may be an alternative to stochastic

gradient descent as an optimization method for training deep learning models. The misfortune capability clear cut cross-entropy is utilized to measure profound learning model errors, commonly in single-name, multi-class order issues. The little edge in the distinction between the consequences of preparation and approval shows the model isn't overfitting. This is because of the accessibility of additional information for preparing the model. Calculations of 5 emotion recognition was experimented for testing the model accurate rate and efficient performance metrics. A test accuracy of Happy , Sad , Anger , Neutral , Joy were calculated and the results of each Accuracy Rate were 94% , 92% , 95% , 90% , 90% respectively and Loss Function Calculated for each emotion was 1.766 , 1.344 , 1.456 , 1.099 , 1.267 . The Overall Test Accuracy was 92.72%. In training Process overall Epochs iterations were 50 and observations states that accuracy rates was drastically increasing per epoch while loss functions was

highly decreasing each epoch. Even when the User trails have changed the characteristics as gender, wearables like glasses or earrings have not affected the system's performance that much and also testing some facial changes like beards or changing the hair styles have not that much significance of affecting the performance metrics and loss/error functions of the model. This demonstrates that the model's capacity to generalize is affected by the kind of noise that is added. The model anyway returns great outcomes for every one of the examinations completed, with F1 scores of more noteworthy than 70% for every one of the tests and exactness of around 95% approximate calculated by training processes.

6. CONCLUSION & FUTURE WORK

All in all, our proposed feeling-based music suggestion framework, utilizing facial pictures and real time video captures for overflow calculations, accomplished a precision of around 70%. This shows that it is feasible to involve looks as a dependable contribution to foresee the feelings of a client and suggest fitting music in a similar manner. The framework gives a customized music experience to clients, which is a significant calculation of the present reality where individuals are continuously searching for redone encounters. The proposal framework recommends melodies in light of the feelings distinguished, which upgrades the client's mind-set and gives a superior encounter. In any case, there is still an open door for development in the framework's precision. One elective choice is to research different AI models that might create improved results. Furthermore, stretching out the dataset used to prepare the model might support working on the framework's precision. Generally, our framework gives a promising way to deal with customized music suggestions and can be stretched out to different regions where feeling acknowledgment assumes a significant part, for example, medical care and client support. The future extent of this examination could include the investigation and consolidation of further developed facial acknowledgment and feeling discovery calculations, like profound learning and brain organizations, to additionally work on the exactness of the feeling-based music proposal framework. Furthermore, the framework could be extended to incorporate more music classes and customized proposals in view of client listening history and inclinations. The incorporation of user feedback to enhance the recommendation algorithm and the user experience as a whole is one more potential area of future research. Moreover, the framework could be applied to other domains other than music, like movies or television show proposals, to give a more customized and connecting experience for clients. It will definitely improve the user interface in musical applications and also provide a high technological advancement not just in Computer vision industries but all over every fields of science and technology. It can also be helpful for recommending songs

for disabled people who can just express their emotions to activate a customize playlist for them to listen what they want to listen as emotions are the part of every individual or living being even if they are disabled. It can be established as an Application or may be integrated with well known music platforms like Spotify.

7. REFERENCES

- [1] Londhe RR and Pawar DV 2012 Analysis of facial expression and recognition based on statistical approach *International Journal of Soft Computing and Engineering* **2**
- [2] Kabani H, Khan S, Khan O and Tadvil S 2015 Emotion based music player *International Journal of Engineering Research and General Science* **3** 750-6
- [3] Gupte A, Naganarayanan A and Krishnan M Emotion Based Music Player-XBeats *International Journal of Advanced Engineering Research and Science* **3** 236854
- [4] Hadid A, Pietikäinen M and Li SZ 2007 Learning personal specific facial dynamics for face recognition from videos *International Workshop on Analysis and Modeling of Faces and Gestures* pp1-15 Springer Berlin Heidelberg
- [5] Zeng Z, Pantic M, Roisman GI and Huang TS 2008 A survey of affect recognition methods Audio, visual, and spontaneous expressions *IEEE transactions on pattern analysis and machine intelligence* **31** 39-58
- [6] Patel AR, Vollal A, Kadam PB, Yadav S and Samant RM 2016 MoodyPlayer a mood based music player *Int. J. Comput. Appl.* **141** 0975-8887
- [7] ParulTambe, YashBagadia, Taher Khalil and Noor UIAin Shaikh 2015 Advanced Music Player **5**
- [8] Lucey P, Cohn JF, Kanade T, Saraghi J, Ambadar Z and Matthews I 2010 The extended cohn-kanade dataset (ck+) A complete dataset for action unit and emotion-specified expression In 2010 *IEEE computer society conference on computer vision and pattern recognition-workshops* 94-101 IEEE
- [9] Kanade T, Cohn JF and Tian Y 2000 Comprehensive database for facial expression analysis In *Proceedings Fourth IEEE International Conference on Automatic Face and Gesture Recognition* 46-53 IEEE
- [10] Luoh L, Huang CC and Liu HY 2010 Image processing based emotion recognition In 2010 *International Conference on System Science and Engineering* 491-494 IEEE
- [11] Vivek JD, Gokilavani A, Kavitha S, Lakshmanan S and Karthik S 2017 A novel emotion recognition based mind and soul-relaxing system In 2017 *International Conference on Innovations in Information, Embedded and Communication Systems* 1-5 IEEE
- [12] Jyoti Rani and Kanwal Garg 2014 Emotion Detection Using Facial Expressions A Review *International Journal of Advanced Research in Computer Science and Software Engineering* **4**

- [13] Joshi A and Kaur R 2013 A Study of speech emotion recognition methods *Int. J. Comput. Sci. Mob. Comput.* **2** 28-31
- [14] Shoaib M, Hussain I, Mirza HT and Tayyab M 2017 The role of information and innovative technology for rehabilitation of children with autism a systematic literature review In2017 *17th International Conference on Computational Science and Its Applications* 1-10 IEEE
- [15] Dubey M and Singh L 2016Automatic emotion recognition using facial expression a review *International Research Journal of Engineering and Technology (IRJET)* **3** 488-92
- [16] Anwar S, Milanova M, Bigazzi A, Bocchi L and Guazzini A 2016 Real time intention recognition In *IECON 2016-42nd Annual Conference of the IEEE Industrial Electronics Society* 1021-1024 IEEE
- [17] Rázuri JG, Sundgren D, Rahmani R, Moran A, Bonet I and Larsson A 2015 Speech emotion recognition in emotional feedbackfor human-robot interaction *International Journal of Advanced Research in Artificial Intelligence* **4** 20-7
- [18] Dureha A 2014 An accurate algorithm for generating a music playlist based on facial expressions *International Journal of Computer Applications* **100**