iJRASET
International Journal For Research in Applied Science and Engineering Technology



INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 11 **Issue:** V **Month of publication:** May 2023

DOI: https://doi.org/10.22214/ijraset.2023.53028

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 11 Issue V May 2023- Available at www.ijraset.com

Musical Therapy through Facial Expression Recognizer using Machine Learning

Shruti Aswal[1], Sruthilaya Jyothidas[2], S Shankarakrishnan[3], Prof. Prabhu M[4]
[1, 2, 3, 4]*Computer Science and Engineering Department Adi Shankara Institute of Engineering and Technology Kochi, Kerala*

Abstract: ***Musical therapy is a type of treatment that uses music to help people improve their physical, emotional, cognitive, and social functioning.It is a non-invasive treatment that has been shown to have beneficial effects on mental health.Facial recognition technology has grown in prominence as a tool for analyzing human emotions and behaviors in recent years. We are investigating the possibility of facial recognition technology in musical therapy.In this project, we propose a novel approach to musical therapy that uses facial recognition technology to create personalized playlists for individuals based on their emotional state. We use a convolutional neural network (CNN) to analyze facial expressions and classify them into different emotional states, such as happiness, sadness, or anger. Based on the emotional state detected, we recommend music that is likely to have a positive effect on the individual's emotional well-being.A voice assistant is also integrated into the system to provide additional support and guidance to the users. The system is designed to be accessible to people of all ages and abilities, including those with disabilities.Our approach has the potential to improve the effectiveness of musical therapy by providing personalized recommendations based on the individual's emotional state. Moreover, our project provides a convenient and accessible platform for individuals to receive musical therapy without the need for in-person visits to a therapist.***

Keywords: Deep learning, Facial micro-expression recognition, CNN, Music Recommendation algorithm, Voice Assisted Therapy Bot

I. INTRODUCTION

With the advent of the information age, deep learning is widely used in image recognition, image processing, and especially facial expression recognition. Face recognition has become a research hotspot in the field of human-computer interaction, but it still has limitations on the application of image processing results. Image research often focuses on improving the accuracy of recognition, and the data in the image lacks the use of secondary processing, that is, in the actual production and life process, the image information has not been completely and efficiently used [1]. In this paper, a deep learning method is used to design and train a convolutional neural network expression recognition model. The results are combined with a music recommendation algorithm, and the music that adjusts the mood is recommended by judging the mood shown by the person.

With the advancement of technology, there is an increased interest in developing digital tools that can deliver music therapy in novel ways, particularly through the use of machine learning techniques.In this project, we present a novel approach to musical therapy that makes use of machine learning algorithms such as convolutional neural networks (CNNs) and Mediapipe to recognise face emotions and provide vocal help. Our software seeks to provide users with a personalised and immersive experience that will allow them to benefit from music therapy in a more engaging and effective way.The facial emotion recognizer at the heart of our programme analyses real-time video data from the user's camera to assess their emotional state using CNNs.We also highlight the advantages of using machine learning algorithms to personalize the therapy experience, and how it can lead to better outcomes for individuals with mental health conditions. We acknowledge the challenges of developing such an application, including the need for large and diverse datasets and the potential biases in the training data.

Overall, we believe that our approach to musical therapy using facial emotion recognition and voice assistance has great potential to improve the mental health and well-being of individuals. By leveraging the power of machine learning, we can create personalized and immersive experiences that can provide a more effective form of therapy than traditional methods. Our hope is that our application will inspire further research and innovation in the field of digital music therapy.

II. LITERATURE SURVEY

Works done by Shlok Gilda, Husain Zafar, Chintan Soni and Kshitija Waghurdekar [1], the music classification module divides songs into four different mood classes using audio features with an astounding 97.69% accuracy.

©IJRASET: All Rights are Reserved

The recommendation module suggests music to the user by connecting the user's emotions to the mood type of the song and accounting for their preferences. By offering music suggestions based on the user's mood, the music player may be able to keep the user more interested in the music and encourage longer listening sessions. By matching the music to the user's emotional condition, the music player may be able to improve their emotional wellbeing.

A new architecture network for CNN-based face emotion identification was introduced in this paper by Abir Fathallah, et al. [2]. Using the Visual Geometry Group methodology, they modified their architectural design to get better results (VGG). Several, primarily free databases (CK+, MUG, and RAFD) were used to evaluate their design. Their results show that the CNN method is particularly effective at extracting facial expressions from photographs stored in various public databases, leading to improvements in facial expression analysis.

In a recent study, Ketki R. Kulkarni et al. [3] suggested a comparative examination of age-adjusted facial emotion recognition using the weighted least squares filter approach. The technique uses the Gabor and Log Gabor filters to extract facial information. The SVM classifier is trained using the known input images before classifying the unknown input images. According to the findings, utilizing the Log Gabor filter improves recognition accuracy but lengthens processing time when used in place of the Gabor filter.

In this work, Ahmet Elbir, et al. [4] reviewed studies on musical genre classification and suggestions. Digital signal processing is used to first extract the auditory properties of the music, after which music genres are identified and recommendations for further listening are made. Convolutional neural networks, another deep learning method, have also been applied to categorize musical genres, make song recommendations, and assess the effectiveness of the generated data. The SVM method performed the best overall on the GTZAN database, which was used in the study.

The study by S. L. Happy and A. Routray [5] suggested using a Gaussian Filter or mask to reduce noise before transferring a system image from the database to the facial landmark detection stage. For face detection in this instance, they used the Viola Jones approach, which combines Haar-like features with Adaboost learning. The Eyebrow corners detector, Eye detector, Noise detector, and Lip corner detector make up the feature detection stage. These active face patches are retrieved, and SVM is used to classify the attributes (Support Vector Machine).

It will retrieve test images from the database, classify them appropriately, and extract the attributes from hundreds of photographs. They used the CK+ (Cohn-Kanade) dataset and the JAFFE dataset to develop and test the database. The training database contains a total of 329 images. For social beings like humans, visual communication is a powerful form of engagement. A minor alteration in facial expression might indicate emotions such as happiness, grief, surprise, or anxiety. Each person's face may change depending on the surroundings, including the lighting, posture, and even the background. Recognizing facial expressions is still difficult due to a combination of these variables. Rahul Ravi, S.V. Yadhukrishna [6], and Rajalakshmi Prithviraj's research, aims to provide a fair comparison and insight into the accuracy of the two most widely used face expression recognition (FER) methods. The techniques employed here are local binary patterns and convolution neural networks.

This article [7] will take a look at various classification-based algorithms in order to propose a straightforward method for categorizing music into 4 mood groups and identifying users' moods from their facial expressions. You may put the two together to make a synchronized playlist.The accuracy of the second strategy, which uses SVM regression, is 81.6%, compared to the accuracy of the first strategy, which employs MLP, which is capped at 70%. When the HAAR classifier and the fisher face algorithm are combined, the face mood classifier has a 92% accuracy rate.

To identify emotions and play music in response, convolutional neural networks are used. CNNs employ multilayer perceptrons to achieve minimal processing.[8].CNNs have been shown to evaluate images more slowly than other image categorization methods. This demonstrates that CNN filters are superior to those employed in conventional algorithms. Direct visualization of features might not be as illuminating. Consequently, we use the back-propagation training method to activate the filters for improved visualization. The research suggests a Convolution Neural Network (CNN) for an autonomous face emotion categorization system that utilizes characteristics from the Speeded Up Robust Features (SURF)[9]. 91% accuracy is attained by the proposed approach, which makes it easier to identify human emotion through facial expressions.

In his work, H.Emmanuel James et al. presented a system focusing on human emotion recognition for the development of emotion-based music players, outlining the techniques employed by current music players to identify emotions[10].The webcam shows the individual's face.

The video recording is converted into a picture. The webcam image is preprocessed to remove facial expressions and turn them into a string of Action Units (AUs). combining 64 AU, (FACS) traits to identify each face. Facial emotions like joy, rage, sadness, and surprise are categorized after feature extraction. Web services are integrated with them.

©IJRASET: All Rights are Reserved | SJ Impact Factor 7.538 | ISRA Journal Impact Factor 7.894 |

TABLE I: RELATED WORKS

Paper	Method/Technology	Reference
SMART MUSIC PLAYER INTEGRATING FACIAL EMOTION RECOGNITION AND MUSIC MOOD RECOMMENDATION	CNN, ReLU, Max Pooling, Forward and Backward Propagation	[1]
FACIAL EXPRESSION RECOGNITION VIA DEEP LEARNING	Convolutional layer, pooling layer, fully connected layer, and fine tuning are all used in DNN	[2]
FACIAL EXPRESSION RECOGNITION	Gabor Filter, Log Gabor Filter, Support Vector Machine (SVM), Weighted Least Square (WLS) Filter (PCA).	[3]
MUSIC GENRE CLASSIFICATION AND RECOMMENDATION BY USING MACHINE LEARNING TECHNIQUES	KNN, Naive Bayes, Decision Tree, SVM, Random Forest	[4]
AUTOMATIC FACIAL EXPRESSION RECOGNITION USING FEATURES OF SALIENT FACIAL PATCHES	Gaussian filter and support vector machine	[5]
FACE EXPRESSION RECOGNITION USING CNN & LBP	LBP and CNN	[6]
MOOD BASED MUSIC PLAYER	Random Forest, and MLP, KNN, Support Vector Machines (SVM)	[7]
A MACHINE LEARNING BASED MUSIC PLAYER BY DETECTING EMOTIONS	CNN	[8]
AUTOMATIC HUMAN EMOTION RECOGNITION SYSTEM USING FACIAL EXPRESSIONS WITH CONVOLUTION NEURAL NETWORK	CNN , SURF	[9]
EMOTION BASED RECOMMENDATION SYSTEM	Hidden Markov Model, Image Pyramid, Histogram of Oriented Gradients, Linear Classifier	[10]

©IJRASET: All Rights are Reserved

ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 11 Issue V May 2023- Available at www.ijraset.com

III. MUSIC AND EMOTION

According to pfizer news article, Listening to (or making) music increases blood flow to brain regions that generate and control emotions.2 The limbic system, which is involved in processing emotions and controlling memory, “lights” up when our ears perceive music. Emotional contagion refers to the tendency of individuals to mimic the emotions of those around them.

When a person listens to upbeat and cheerful music, the positive emotions conveyed by the music can be contagious and lead to an improvement in their own mood.Another possible explanation for the therapeutic effect of upbeat music is through the process of cognitive reappraisal. Cognitive reappraisal refers to the process of reinterpreting or reframing a situation in a more positive way, which can lead to a more positive emotional response.

A person who is sad can be suggested with songs that are motivating and uplifting.The researchers found that listening to music, especially uplifting and motivating music with inspiring lyrics, led to a significant reduction in anxiety and depression compared to rap or silence. Additionally, the researchers found that the effects of music on anxiety and depression were sustained over time, suggesting that music has a lasting therapeutic effect.

The researchers found that when a person is surprised with a song, their brain undergoes several changes that can affect their emotions, mood, and behavior. Hearing a familiar or pleasant song can activate the brain's reward centers, such as the ventral striatum and nucleus accumbens. These areas release dopamine, a neurotransmitter associated with pleasure and motivation, which can create a positive emotional response in the listener.

In addition, the surprise element of hearing a song unexpectedly can create a sense of novelty and excitement, which can increase the listener's attention and focus. This can lead to a heightened sense of awareness and a more profound emotional response to the music. Depending on the individual's personal associations and memories with the song, it can evoke a range of emotions, from joy and nostalgia to sadness and grief. Being surprised by a song can be an opportunity to explore new music from different languages around the world.

The impact of angry songs on people can be explained by the reason that listening to angry music did not increase aggression in participants but instead helped them regulate negative emotions. They concluded that listening to aggressive music could be an effective strategy for some individuals to regulate their emotions.. Emotional regulation refers to the ability to regulate or modulate one's own emotional responses in a way that is adaptive and effective in achieving one's goals.

IV. PROPOSED METHOD

People usually listen to music in accordance with their mood and interests, and music has long been known to affect how someone feels.The objective of our project is to uplift the user's spirits by playing music that fits their needs while simultaneously recording their facial expressions.

The ability to capture and recognize a person's emotion as well as to present music that is suited for that person's mood can help users to become more relaxed and have a more appealing overall result. The oldest and most natural method of expressing emotions, moods and feelings is through the use of facial expressions. The goal of this project is to classify facial expressions into several types of emotions to develop a low-cost music player that automatically generates sentiment-aware playlists based on the user's emotional state. User emotions are defined by the emotions module. To make song recommendations to the user, the music recommendation module combines the results of the emotion module and the music classification module. Compared to previous systems, this system offers much higher accuracy and performance. This is a new approach that allows users to automatically play songs based on their emotions. The user's facial expressions are recognized and music is played according to these expressions. The Convolution Neural Network algorithm is a machine learning technique used to identify emotions. This project is also integrated with voice assistance providing clear and concise instructions to guide the user. Overall, our proposed method for the Musical Therapy through Facial Emotion Recognizer with Voice Assistance project combines state-of-the-art machine learning and computer vision techniques to provide a personalized and effective music therapy experience. We believe that this approach has great potential to improve the mental health and well-being of individuals and we look forward to further research and innovation in this field.

V. METHODOLOGY

A. Data Pre-Processing

The dataset for the model must be gathered first. You should pick a dataset that the model can use to provide proper accuracy as well as one that can be applied to other specific areas.

©IJRASET: All Rights are Reserved

The pre-processed data is then fed into the CNN-based emotion recognition model and the Mediapipe-based face detection system, allowing us to accurately detect the user's emotional state and track their facial expressions throughout the process.

To preprocess data using Python, predefined libraries must be imported. These libraries are used to complete some particular tasks. The selected dataset must now be imported along with the model. The datasets' missing data must then be handled. Our machine learning model may run into serious issues if our dataset has some missing data. As a result, the dataset contains missing values, which must be handled.

The pre-processing module is an essential component of our system, ensuring that the input data is accurately and reliably processed to provide an effective music therapy experience.When it comes to face detection, we start by processing the image frames using Mediapipe. We convert the frames to grayscale to simplify the image and reduce processing time. Mediapipe then uses a machine learning model to detect the presence of faces. Once a face has been identified, we can extract various facial features like the eyes, nose, and mouth for further analysis. For emotion recognition, we use CNNs to identify emotions based on facial expressions. We represent the input data as a set of multidimensional arrays of pixel values, which are then processed through a series of convolutional layers, pooling layers, and fully connected layers.

We apply the Rectified Linear Unit (ReLU) activation function to introduce non-linearity and improve the model's ability to capture complex relationships between the input data and the output emotions. Both Mediapipe and CNNs rely heavily on NumPy arrays for manipulating and analyzing the data. NumPy provides a fast and efficient way to work with large arrays of data, allowing for complex mathematical operations to be performed quickly and accurately. The results of the face detection were automatically used to construct the face's location, breadth, and height. Photos were first cropped, and then they were further cropped in accordance with the output of the face detector before being used for training and testing.

B. Splitting the Dataset

In the field of machine learning, dividing a dataset into a training set and a validation set is a standard procedure. This split is necessary to enable us to assess how well our machine learning model performs on new data. The model is fitted (or trained) using the training set. During the training process, the model "learns" the patterns and relationships in the data, and uses this learning to make predictions on new data.

The validation set, on the other hand, is used to evaluate the model's performance on unseen data. By comparing the model's predictions on the validation set to the actual outcomes, we can assess the model's accuracy and identify any areas for improvement. If the model performs well on the validation set, this is an indication that it is not overfit to the training set, and that it is able to generalize its learning to new data. On the other hand, if the model performs poorly on the validation set, this may be an indication that it is overfit to the training set, and that it is not able to generalize its learning to new data.

Overall, the use of a validation set enables us to get a more accurate assessment of the model's performance and to identify any areas for improvement, ultimately helping us to build more robust and reliable machine learning models. In this project ,We divided the data into two sets—the training set and the testing set—for face detection. The training set is used to teach our machine learning model to recognise faces, while the testing set is used to gauge the model's effectiveness.

In a similar manner, we divided the data into training and testing sets for emotion recognition. We train our CNN model to identify emotions based on facial expressions using the training set, and we test our model's performance using the testing set.

C. Building the Model using CNN

A class of deep neural networks called convolutional neural networks (ConvNet/CNN) are frequently used to analyze visual imagery. Based on the sharedweight architecture of the convolution kernels that scan the hidden layers and translation in variance features, they are recognized as shift invariant or space invariant artificial neural networks. Heuristics are frequently used by engineers to develop the filters that are a common component of image processing algorithms. Convolutional neural networks (CNNs) differ significantly from conventional neural networks in that they use convolutions to handle the arithmetic in the background. A CNN uses convolutions to handle the underlying arithmetic, which makes it very different from a typical neural network. In at least one layer of the scheme, convolution is used in place of matrix multiplication.The organizational structure of CNN is depicted in the graph below. The input image goes through several layers of processing before being sent to CNN as preprocessed data. The first layer of feature retrieval from the input image uses convolution, a mathematical technique that uses an image filter and a kernel as inputs from the image. When the image is too large, the pooling layer will then assist in reducing the number of parameters; for this, we can use max pooling, average pooling, or sum pooling.

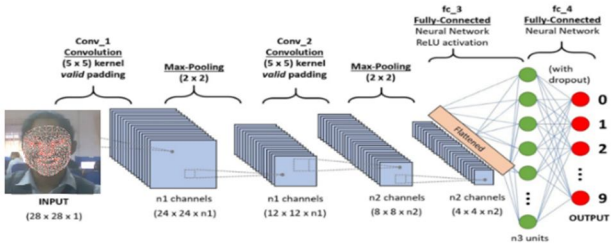


Fig. 5.1. Convolution Neural Network Architecture

Maximal pooling was applied to this model. The image convolution layer is applied a second time, and this time, the features are extracted using a variety of techniques. In the suggested model, ReLU is used. Flattening our matrix into a vector and feeding it into a fully connected layer that resembles a neural network is the last step in the CNN architecture/structure. The fully connected (FC) layer is referred to as this. The neural network's back-end is created using TensorFlow.

D. Feature Extraction

With the help of feature extraction, which starts with a collection of measured data, it is possible to learn more about machine learning, image processing, and pattern recognition. To be able to accomplish the required work, this reduced exemplification must be able to incorporate the relevant information from the input data into the chosen characteristics.Convolution, which multiplies matrices with a filter before moving on to a feature detector, is used to achieve this. We employ a sliding window in the feature detector that each channel moves over to summarize the features. Since neural networks are particularly sensitive to unnormalized data, results are standardized using batch normalization.

In social psychology, a microexpression is a facial expression that is simple to see and recognise as a means of communication. Facial expressions are crucial in human contact because they transmit our intents and aspirations as well as information about our emotions. Understanding and being able to read facial emotions naturally makes the desired conversation easier. The three processes in the classification of human facial expressions are face recognition, feature extraction, and facial expression classification.

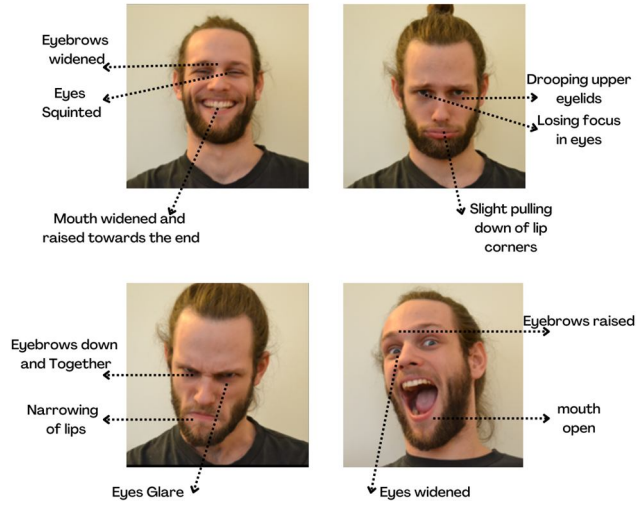


Fig. 5.2: Characteristic Features used to detect Facial Emotions

E. Generalization and Activation

Generalization is the main goal when training a machine learning model. This means that the model can perform well on new, unpublished data and not just on training data presented to it during the training phase. This is important because in real-world applications the model will need to make predictions about data it has never seen before.

If a model has good generalizability, it should be able to make accurate predictions about this new data. A commonly used technique to help improve the generalizability of a model is skipping.Dropout is a regularization technique that involves randomly deleting part of the input of a neural network during training.

This has the effect of reducing overfitting, which is when a model performs well on training data but poorly on new data. By randomly removing inputs, the model is forced to learn several different combinations of input features, which helps the model generalize better to new data. The activation function is another important component in the architecture of a neural network.

This is a mathematical function that is applied to the output of each neuron in the network and determines whether that neuron is activated or not. There are different types of activation functions, including ReLU (Edited Linear Unit), leaky ReLU, and ELU (Exponential Linear Unit).The type of trigger function selected can have a significant impact on the model's performance. The hidden layer activation function can be used to monitor the training of the training dataset from the network model.

By visualizing the activation of hidden units, it is possible to see which features are modeled and which are ignored.This can be useful to identify problems with the model and to improve its performance. In the output layer, the selected activation function can affect the type of prediction the model will generate. The ReLU activation function (Rectified Linear Unit) is a commonly used activation function in neural networks, especially in deep learning models.

It is a simple function that returns the maximum between 0 and the input value.we need to first define the input layer, hidden layers, and output layer of the neural network. Each layer will have its own set of weights and biases that are learned during training.Next, we can use the ReLU function as the activation function for each of the hidden layers. The output layer can use a different activation function, depending on the specific task being performed.

F. Model Training and Evaluation

In this section, we focus on model training and evaluation.As we know,The convolution of x and y is represented by the symbol x y. It is the integral of the product of one of the functions and the product of the other function, which is a pair, after inversion and translation. As a function of the degree of translation, the number of features obtained following the cumulative operation is displayed in the param column.

$$(\mathrm{x} * \mathrm{y})(\mathrm{t}) \stackrel{def}{=} \int_{R^n} x(\tau) y(t-\tau)\, d\tau$$

The fully connected layer comes after the flattening layer, which expands the feature data generated from the convolution and clustering layer, and performs dimensional transformation for the incoming data. At this point, the AI model can predict and classify. To produce accurate classification prediction results, the layer is fully connected between the features generated by the convolution layer and the clustering layer in this layer and modifies the weights and biases.

During classification, overfitting is avoided thanks to the dropout layer in the fully linked layer. Throughout training, the built-in model will undergo regular corrections to lessen the loss trend. The loss curve of the model shows a tendency to decline, which also suggests that on-going training can improve prediction accuracy.

G. Testing the Model

We used a method known as "train test split" in order to divide the dataset into two distinct ways, one for training and the other for testing. Now that we have tested the testing dataset, we can obtain the results.

$$\text{Accuracy} = \frac{\mathrm{TN} + \mathrm{TP}}{\mathrm{TP} + \mathrm{TN} + \mathrm{FP} + \mathrm{FN}}$$

TP: Samples were accurately classified as positive.
TN: The samples were correctly identified as negative.
FP: Mistakenly classified as positive.
FN: A false negative sample was detected.

Table 5.1. Positive and negative samples confuse predicted outcomes with each other

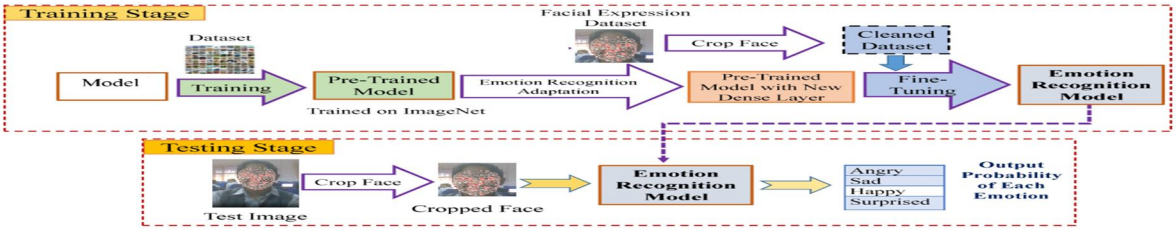


Fig. 5.3. Training Stage and Testing Stage

H. Emotion Detection

A component or feature that is intended to identify and categorize a person's emotions based on their facial expressions is known as an emotion module for a facial emotion recognition system. There are several approaches to building an emotion module, but most systems rely on machine learning algorithms to analyze images or video frames of the person's face and identify specific features that are associated with different emotions.

Some common techniques used in emotion modules include:

1) *Feature extraction*: This involves identifying and extracting specific features from the face that are relevant to emotion recognition, such as the shape and movement of the eyebrows, the size and position of the pupils, and the curvature of the lips.
2) *Classification algorithms*: Once the relevant features have been extracted, a classification algorithm is used to analyze the data and assign a label to the emotion being expressed. We will be using the Convolutional neural networks(CNN) Algorithm.
3) *Training data*: It is frequently necessary to train the system on a sizable dataset of images or videos that have been correctly labeled with the appropriate emotion in order to increase the emotion module's accuracy. This allows the system to learn the characteristics of different emotions and improve its ability to recognize them in new images. Emotion modules can help improve the user experience by allowing the system to respond appropriately to the user's emotional state.

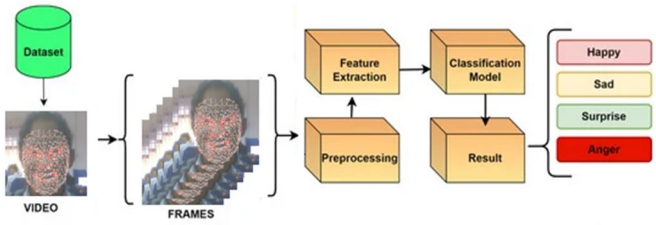


Fig. 5.4. Emotion Detection Process

I. Music Recommendation Module

There is a database of songs that has between 8 and 10 tracks for each emotion. As is common knowledge, music may undoubtedly improve our mood. By displaying subtitles in accordance with detected emotions, the music

The player's GUI recommends a playlist to users. As a result, if a user is depressed, the algorithm will suggest a music playlist that will uplift him or her. As a result, the user's mood will automatically improve. A more individualized and engaging music listening experience can be delivered by a music recommendation module that analyzes facial expressions to determine the user's mood. Such a system could work by using a camera to capture images or video of the user's face and feeding this data into an emotion recognition module. The emotion recognition module would analyze the user's facial expressions and identify their current mood, which could then be used to generate music recommendations that are tailored to that mood. For example, if the emotion recognition module detects that the user is feeling happy or excited, the music recommendation module could recommend upbeat, energetic tracks to match their mood. On the other hand, if the user is feeling sad or relaxed, the recommendation module could suggest more mellow or soothing tracks. There are several approaches that could be used to build a music recommendation module that takes into account the user's mood.

One approach might be to use a machine learning algorithm to analyze the user's listening history and identify patterns that are associated with different moods. This could involve training the system on a dataset of songs labeled with the moods they are most commonly associated with, such as happy, sad, energetic, etc. The system could then use this information to recommend tracks that match the user's current mood. The music recommender according to mood is designed to provide a personalized and engaging music listening experience that can enhance the user's emotional state and improve their overall mood. The module can be customized to fit the user's preferences and can be integrated with various music streaming services to provide seamless access to recommended songs and playlists. As we continue to develop our music recommendation system, we've decided to include a diverse range of languages like English, Punjabi, Tamil, Malayalam, and Hindi in our song dataset. We believe that listening to songs in our regional languages can have a positive impact on our mental health and wellbeing. When we listen to songs in our native language, it can evoke a sense of nostalgia and familiarity, which can be incredibly comforting and soothing.

Additionally, it can help us feel more connected to our culture and community, which is an important factor in maintaining good mental health. By including regional language songs in our music recommendation system, we hope to provide a more personalized and inclusive listening experience for our users.

©IJRASET: All Rights are Reserved | SJ Impact Factor 7.538 | ISRA Journal Impact Factor 7.894 |

J. Voice Assisted Therapy Module

This module can provide a range of therapeutic activities, including mindfulness exercises, cognitive-behavioral therapy techniques, and guided meditation sessions. The voice assistant can also offer personalized feedback, insights, and suggestions based on the user's interaction with the module and their specific needs.

In addition, the assistant can provide information on mental health resources, such as nearby therapists, support groups, and crisis hotlines. It can also help users track their mental health progress and provide reminders for therapy appointments and medication schedules. The voice assistant for therapy is designed to offer an accessible, convenient, and cost-effective way for individuals to access mental health resources and support. The module can be customized to fit the user's preferences and needs, and can provide a confidential and non-judgmental environment for individuals to work on their mental health.

VI. RESULTS

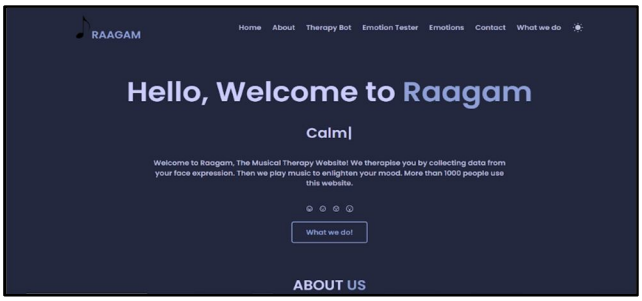


Fig. 6.1 RAAGAM - Website

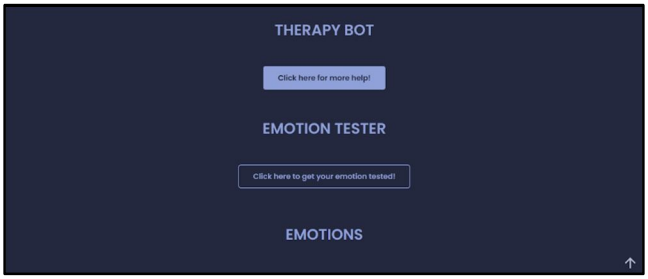


Fig. 6.2. Other Features of the Website

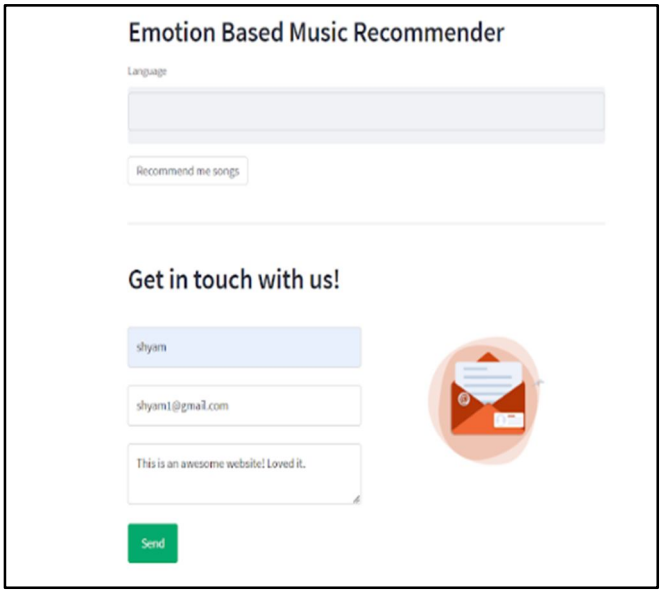


Fig. 6.3. Emotion Based Music Recommender

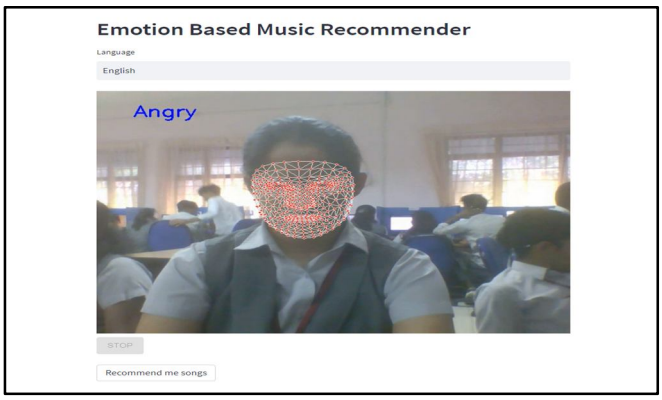


Fig. 6.4. Detection Of the "Angry" Emotion

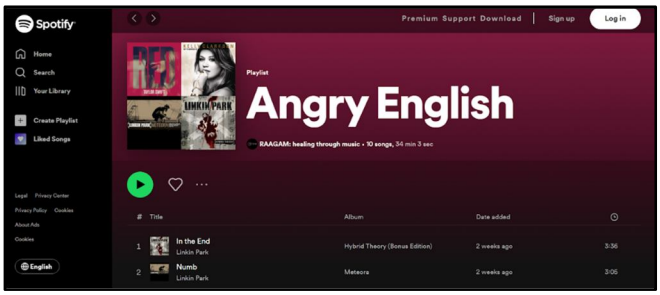


Fig. 6.5. Playlist Recommending Songs Based on Emotion

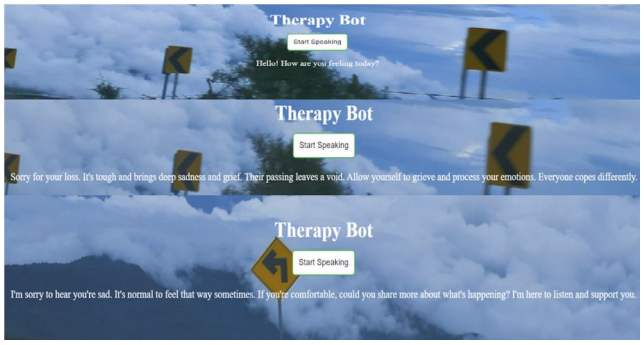


Fig. 6.6. Voice Assisted Therapy Bot

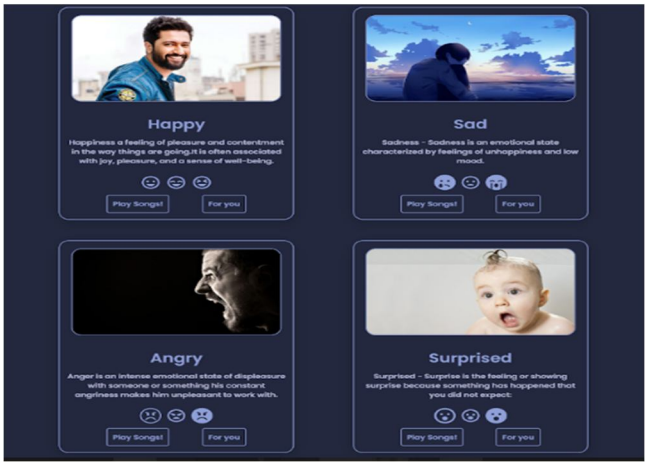


Fig. 6.7. Emotions and Other Informations

VII. FUTURE ENHANCEMENT

We believe that there are some unique future enhancements that could take the system to the next level. One potential enhancement is to incorporate social media analysis into the recommendation algorithm. By analyzing a user's social media activity and posts, the system could gain a deeper understanding of their interests and preferences, leading to more personalized music recommendations.Another alternative based on additional emotions which are excluded in our system are disgust and fear.
This emotion included supporting the playing of music automatically. The future scope also within the system would be to style a mechanism that might be helpful in music therapy treatment and help the music therapist to treat the patients suffering from mental stress, anxiety, acute depression, and trauma.Finally, incorporating machine learning and artificial intelligence techniques could lead to even more accurate emotion recognition and personalized music recommendations. With these advancements, the system could become a truly advanced and powerful tool for music lovers seeking a more personalized listening experience.

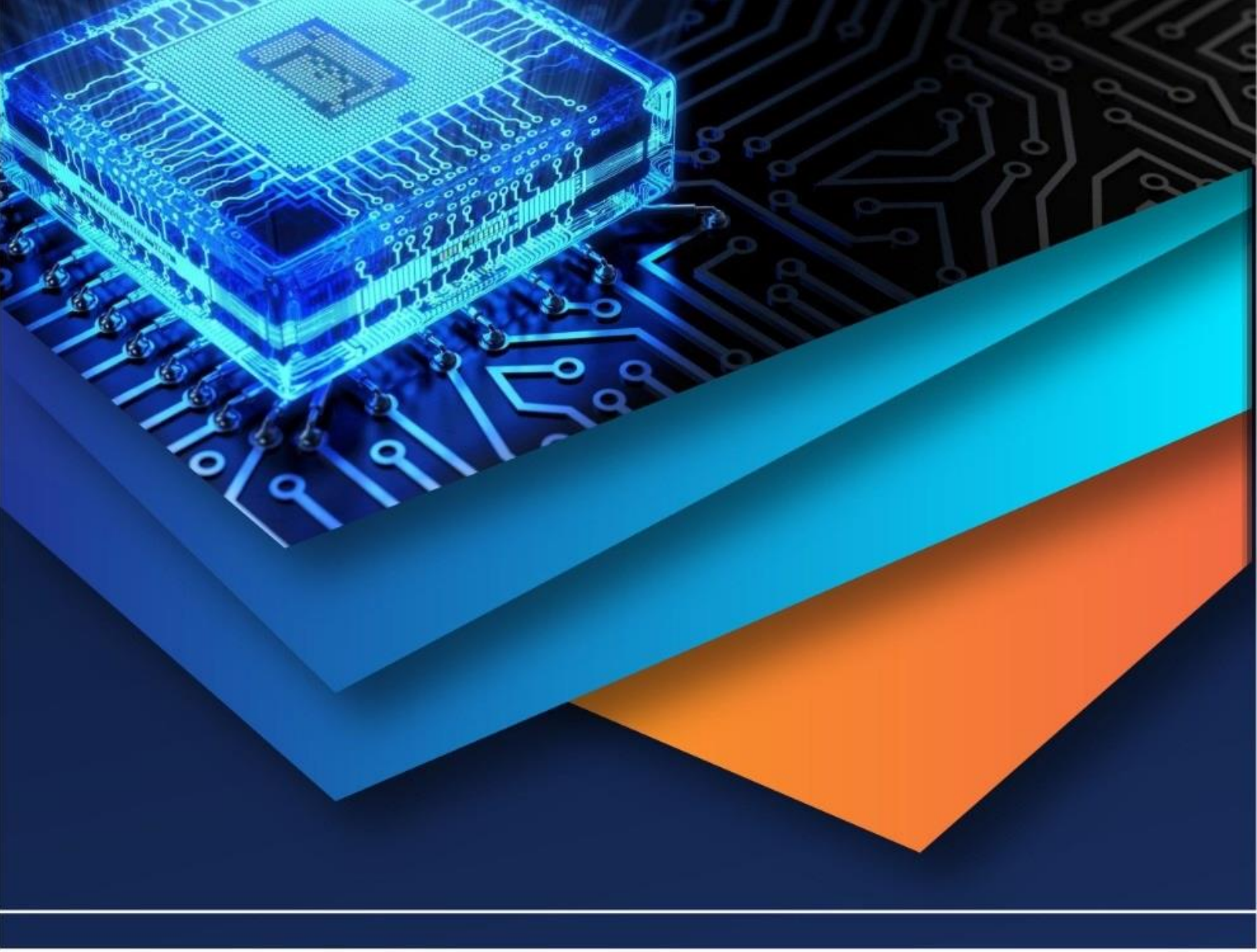
VIII. ACKNOWLEDGMENT

We are very thankful to the Department of Computer Science and Engineering of Adi Shankara Institute of Engineering and Technology for permitting us to work on the topic “Musical Therapy through Facial Emotion Recognizer using ML”. We express our gratitude to Prof. Prabhu M., our project guide who gave us all the guidance and motivation and at any time available to us for doubts solving, grateful for giving us the required help, point by point recommendations and furthermore support to do the venture. We are excited and happy to offer our thanks to the Head of the Information Technology Department Prof. Manesh T I., for his endorsement of this undertaking. We are amazingly appreciative to all staff and the administration of the school for giving every one of us the offices and assets required.

REFERENCES

[1] Shlok Gilda, Husain Zafar, Chintan Soni, Kshitija Waghurdekar,"Smart music player integrating facial emotion recognition and music mood recommendation", Date of Conference: 22-24 March 2017, India.
[2] A. Fathallah, L. Abdi, and A. Douik, “Facial Expression Recognition via Deep Learning,” in 2017 IEEE/ACS 14th International Conference on Computer Systems and Applications (AICCSA), Hammamet, 2017, pp. 745–750, doi: 10.1109/AICCSA.2017.124
[3] H. Bahuleyan, “Music Genre Classification using Machine Learning Techniques,” arXiv:1804.01149 [cs, eess], Apr. 2018.
[4] A. Elbir, H. Bilal Çam, M. Emre Iyican, B. Öztürk, and N. Aydin, “Music Genre Classification and Recommendation by Using Machine Learning Techniques,” in 2018 Innovations in Intelligent Systems and Applications Conference (ASYU), Oct. 2018, pp. 1–5, doi: 10.1109/ASYU.2018.8554016.
[5] S. L. Happy and A. Routray,"Automatic facial expression recognition using features of salient facial patches," in IEEE Transactions on Affective Computing, vol.6, no. 1, pp. 1-12, 1 Jan.-March 2015.
[6] Rahul ravi, S.V Yadhukrishna, Rajalakshmi, Prithviraj, "A Face Expression Recognition Using CNN & LBP",2020 IEEE.
[7] Anuja Arora, Aastha Kaul, Vatsala Mittal,”Mood based music player”,Date of conference : 2019, India
[8] S.Deebika, K.A.Indira, Jesline, "A Machine Learning Based Music Player by Detecting Emotions", 2019 IEEE
[9] Karthik Subramanian Nathan, manasi arun, Megala S kannan," EMOSIC — An emotion-based music player for Android,2017 IEEE
[10] H. I. James, J. J. A. Arnold, J. M. M. Ruban, M. Tamilarasan, And R. Saranya, “Emotion Based Music Recommendation System,” Vol. 06, No. 03, P. 6, 2019.

©IJRASET: All Rights are Reserved | SJ Impact Factor 7.538 | ISRA Journal Impact Factor 7.894 |



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Call : 08813907089 (24*7 Support on Whatsapp)