International Journal of Scientific Research in Computer Science, Engineering and Information Technology

ISSN : 2456-3307

Available Online at : www.ijsrcseit.com

Volume 10, Issue 7, May-June-2024 | Published Online : 20th June 2024

Machine Learning for Facial Emotion Recognition and Music Recommendation

Pundalik Chavan*[1], Hanumanthappa H[2], Naveen kumar. B[3], Ramaprasad H C[4], Hari Krishna H[5]

[*1,3,4]Associate Professor, Reva University, Bangalore, India

[2]Assistant Professor, Nitte Meenakshi Institute of Technology, India

[5]Dept of CSE, Ballari Institute of Technology & Management, Ballari

ABSTRACT

This paper proposes a mood-based music player that incorporates real-time mood detection using facial expressions. The popularity of music players and streaming apps has increased due to their accessibility and versatility in daily activities. The proposed system offers an additional feature to traditional music players, using image analysis to predict the user's mood and play songs accordingly. The goal of this system is to enhance customer satisfaction by providing personalized music that can alleviate stress and elevate mood. Psychological research has shown that music has therapeutic benefits for various disorders, making it a crucial tool for addressing mental health concerns. However, existing music recommendation systems do not consider the user's mood, which greatly impacts their perception of a song. By incorporating mood detection, this proposed system aims to bridge this gap and offer a more personalized music experience.

Keywords: Mood detection, Music player, Facial expressions, Mental health, Personalized music.

I. INTRODUCTION

Human emotions are generally categorized as follows: fear, rage, surprise, sadness, and happiness. This broad category of emotions also includes a great deal of other feelings, such cheerfulness, which is a variant of happiness. These feelings are quite nuanced. There are very few facial muscular contortions, and it can be difficult to distinguish between them because even a slight variation can cause a change in expression. Additionally, because emotions are so context-dependent, different people—or even the same people—may exhibit the same feeling in different ways. Even if the attention is limited to the facial regions surrounding the lips and eyes that exhibit the greatest range of emotions, it is still crucial to understand how these movements are identified and categorized. Machine learning and neural networks have been used to these tasks and have achieved satisfactory outcomes. Since machine learning algorithms have shown to be particularly effective in classification and pattern recognition, they can also be used to identify moods.

Recent research has shown that people respond and react to music, and that music has a significant impact on brain activity. In one investigation on the reasons people listen to music, scientists found that the relationship between arousal and mood was significantly influenced by music. Music has been shown to have two major effects on participants: it can improve their mood and increase their self-awareness.

It has been demonstrated that musical tastes are strongly correlated with personality attributes and sentiments. The brain regions that influence emotions and mood are also responsible for controlling the meter, timbre, rhythm, and pitch of music. One of the main components of lifestyle may be interpersonal interaction. It makes flawless details and a wealth of information about people visible, including emotions, body language, voice, and facial expressions.

Approximately 1 billion individuals worldwide are afflicted with mental health illnesses, with 7.5% of Indians reporting having at least one such condition. The second most common cause of death for young adults between the ages of 15 and 29 is suicide. Globally, it is estimated that almost one in seven teenagers between the ages of 10 and 19 suffer from a diagnosed mental illness. These startling figures are made even more troubling by the knowledge that there is a severe shortage of mental health specialists, and the sad reality is that therapy is not generally available to those who need it. Between 76% and 85% of persons with mental illnesses in low- and middle-income nations do not obtain treatment for their conditions. This global dilemma has multiple causes, including genetics, social stigma, abuse and trauma, grief, and a host of other variables. In these situations, music therapy works well.

The suggested method can identify an individual's emotions; if the user is experiencing negative feelings, a playlist with the most upbeat music relevant to the sentiment would be displayed. Additionally, if the feeling is pleasant, a certain playlist will be played. This includes a variety of musical genres to boost happy feelings. We utilized the Kaggle Facial Expression Recognition dataset to detect emotions. Bollywood and Hindi songs have been used to develop a dataset for the music player. Convolutional neural networks are used to implement face emotion detection, providing an accuracy of about 83%.

The goal of this effort is to develop a music player/suggestive system that can recognize the user's face, determine their current mood, and then recommend a playlist depending on that mood.

II. LITERATURE SURVEY

The goal of Lokesh Singh al is to outline the requirements and uses for the facial expression recognition. Comparing Facial expressions in verbal and non-verbal communication.

Though it’s kind of non-verbal communication, expression is extremely important. It conveys the viewpoint, filling, and mental state of the human being. Over the course of two decades, much research has been done to improve human computer interaction (HCI). This work presents the phases of an automatic facial expression identification system, an introduction to facial emotion recognition system, a web app, and a comparative analysis of well- known face expression recognition techniques. The author discussed the advantages and disadvantages of the relevant paper after reviewing it. There is also a description of the facial expression recognition system [1].

Authors Mustafa Sert and Nukhet Aksoy present a novel method for identifying emotional facial expressions using rule-based reasoning and the newly developed facial action unit (AU) detector. Using the Active Appearance Model's (AAM) properties, the author chooses the Support Vector Machine (SVM) algorithm as the binary approach. Using the ADT-AU detector, they discover 17 AUs, either singularly or in combination. By employing the prototypic and major variations of AUs, our rule-based emotion classifier is able to distinguish between six distinct facial expressions of emotion (e.g., surprise, fear, happiness, etc.) [2].

A system that can recognise an emotion on a face automatically is proposed by author Pravin Nagar [3].

Classifying the emotions involves the application of image processing in conjunction with a solution based on Bezier curves. The system receives images of colourful faces as input. Next, a feature point extraction approach based on image processing is applied to extract a subset of selected feature points. Lastly, the curve technique is used to identify the emotion included in the retrieved characteristics, such as the lips and eyes, that were obtained after processing. Through the OpenCV API, the Canny Edge Location calculation was used, and predefined haar cascades records for the mouth and eyes were made. The experiment displays the results of recognition for several facial expressions, including surprise, sadness, smiling, and normal.

A quick and easy expression detection method is presented by Almudena Gil et al. with the intention of functioning on a secondary level in order to provide primary applications, such as exergames, with real-time emotional awareness. The method is based on the extraction of 19 facial landmarks, some of which are defined in the Facial Action Coding System (FACS) and some of which are used in the identification of a newly generated Action Unit (AU). In addition, the innovative concept of Combined Action Units (CAUs) is presented and explained in detail. These are detected as a group of clustered AU's. One of the reasons for the poor results in individual emotion recognition is that it is challenging to conduct a rigorous analysis of the action units when there are few facial markers. The decision tree was trained using the sparse matrix, which is the other. However, extremely good findings are obtained for a quick assessment of an individual's mood. This information can serve as a primary application's valuable knowledge. There is no learning involved in applied emotion based classification; instead, based on logical rules. On a mobile platform, the implementations have been made [4].

Associated with Ashim Sahaa, the author employed the HSV colour model to identify faces in pictures. The high dimensionality of the eigenspace is decreased through the application of principal component analysis, or PCA. The test picture is then projected into the eigenspace, and the expressions can be classified by calculating the Euclidean distance between the test image and the mean of the eigenfaces in the training dataset. The technique uses the face's grayscale photos to categorize five emotions: happiness, surprise, fear, sorrow, and rage. The training dataset is made up of pictures of various people, and tests show that it performs satisfactorily. However, there is some similarity between Sorrow and Fear, which can be improved with more thorough training [5].

Despite having a strong generalization performance, SVMs generally produce less sparse results, according to the authors Gupta, G., and Rathee, N. [6]. Occasionally, this leads to the utilization of nearly all training data as Support Vectors. In comparison to SVM, the results achieved using RVM are sparser, which reduces the amount of Relevance Vectors and, eventually, the processing overhead. The Cohn Kanade Db is used to compare models for facial expression recognition. The face pictures are processed to extract Local Binary Pattern characteristics. They undergo size, brightness, and dimensionality reduction preprocessing before being utilized to train the RVM and SVM models. In specifically, the author compares and evaluates SVM and RVM with regard to regression.

The authors of this study, Weihong Deng, Jiani Huet al: The aim of this project is to develop a practical facial emotion recognition model. Highlighting the shortcomings of the available single-label datasets, the authors suggest creating a new database (RAF-DB) containing a variety of social media photos. They contend that this more accurately depicts multilabel classification situations that occur in the actual world. They offer DeepEmo, a deep learning architecture that acquires high-level features for precise recognition, to address this. According to their research, deep learning beats conventional techniques and produces accuracy comparable to that of a human in near-frontal stances.[7]

The authors Vinola, C., and Vimaladevi, K., describe different systems for classifying and recognizing emotions that use techniques to enhance human-machine interaction. Affect detection employs a variety of methods and approaches that enhance the precision and effectiveness of identifying human emotions. They are found in this research using a descriptive and comparative approach. We also go over a number of programs that employ the approaches in various settings to deal with problems immediately. The databases that can be utilized as standard data sets in the emotion detection procedure are also described by the author. As a result, an extensive examination of techniques, databases employed, and applications related to the developing field of affective computing (AC) is conducted [8].

Martinovic, G., and Filko, D. [9] This research suggests a method for identifying human emotions through principal component analysis and neural network analysis of important face regions. This paper's approach to emotion recognition relies on contour analysis—that is, the examination of facial characteristics as they appear in still images. By focusing on particular facial regions, such as the lips and eyes, edge detection is performed to achieve the desired facial traits. These two face regions are therefore used for classification and emotion recognition. Given that the majority of visual cues associated with emotions are found in these two regions, it makes sense to use them as the foundation for emotion recognition. For a classifier that would learn to identify particular emotions, a neural network was used because it allows for greater training flexibility and adaptability to a given classification task. Based on the data, the average success rate for emotion detection is about 70%, meaning that 73 out of 105 test images were correctly categorized.

Zeynab Rzayeva et al. employed the CNN technique in [10], having trained with Cohn-Kanade and RAvDESS datasets.

The suggested method does a great job at identifying macro facial emotions. The results showed that the accuracy for identifying emotions was 88.17±3.18%. In this study, information from the Cohn-Kanade and RAVDESS datasets is utilized to train a CNN model to identify five major facial emotions. In, a dynamic model for 4D video-based facial emotion recognition is explored and contrasted with various methods.

The software Anaconda and Python 3.5 were used to confirm the operation of a certain system [11], and the face identification methods ViolaJones and Haar cascade were applied. Similar to this, the CNN (Convolution Neural Network) model—which was developed with an accuracy of 88%—as well as the KDEF (Karolinska Directed Emotional Faces) and VGG (Visual Geometry Group) 16 datasets were used to validate the performance measurements for face recognition and classification. However, the results showed that the network architecture was a better design than the existing method.

It was attempted to extract emotion from the live feed using one of the works [12] and the system's camera or any previously recorded image. Python 2.7, OpenCV, and NumPy have been used in its implementation. The idea was to develop a system that could predict a person's facial expression by analyzing a photograph. The investigation proved that this procedure is workable and produces accurate outcomes.

There has also been research done on the music recommendation system. One such study [13] proposed a preliminary approach to classify the mood of Hindi music using basic audio feature extraction. MIREX stands for Music Information Exchange. The mood taxonomy (Retrieval Evaluation Exchange) produced an average accuracy of 51.56% when using the 10-fold cross validation.

Moreover, an article [14] states that the results of recent studies on music suggestion are explained from the perspective of musical resources. There is just one assessment index, little feature extraction, and little systematic research on user requests and behaviour, according to the condition of the research. It was discovered that the circumstance played a key role in the system for personalized music selections. In the end, it was shown that giving equal weights to each contextual component considerably reduced the accuracy of the recommendation findings.

Another study [15] claims that their hybrid recommendation system technique's concept will work if their model has had enough training to recognize the labels. The system automatically uses auditory features of music and brain waves to derive user preference data for the automated management of user preferences in the personalized music suggestion service. In their work, a very short feature vector created from low dimensional projection and pre-existing audio data is used to solve problems related to music genre classification. Using a distance metric learning approach, the feature vector's dimensionality was reduced with only a minor performance impact. The suggested user’s choice classifier produced an overall accuracy of in the binary preference classification task for the KETI AFA2000 music corpus 81.08%.

III. DATASET

This study heavily relies on the FACE EXPRESSION RECOGNITION dataset, which is a popular resource in the field of facial expression recognition. Comprising 35,887 grayscale images, each labelled with one of six emotional states such as "angry", "fear," "happy," "sad," "surprise," and "neutral," this dataset has been used as a significant source for training deep learning models in emotion detection from facial expressions.

Facial Expression Recognition (FER), a specialized aspect of computer vision, focuses on the automatic detection and categorization of facial expressions in images or video footage. The process involves several steps, including the collection of labelled datasets, preprocessing of facial images, extraction of distinguishing features, model training using supervised learning methods, performance evaluation of the model, and real-time emotion interpretation. Applications of FER span various fields, such as human- computer interaction, affective computing, and virtual reality, to name a few. By aiding machines in understanding and responding to human emotions, FER contributes significantly to enhancing user experiences and interactions.

Figure 1. Dataset.

The FACE EXPRESSION RECOGNITION dataset plays an instrumental role in the progress of research in facial expression recognition. With its comprehensive collection of grayscale images labelled with seven distinct emotions, it provides a valuable resource for training deep learning models in the field of FER. Researchers using FER datasets like FACE EXPRESSION RECOGNITION can develop more accurate and resilient algorithms. This contributes to advancements in fields such as human-computer interaction, affective computing, and virtual reality. Such progress further improves the ability of machines understand and react appropriately to human emotions, thereby enriching user experiences and facilitating more effective interactions.

IV. PROPOSED METHODOLOGY

This Music Recommendation System is an application that aims to recognize face expression in real time.

It is the model of the new product with three main modules, Face recognition, mood classification and music search.

A. Face detection module:

The project's face detection module analyses and identifies emotions based on facial expressions using computer vision techniques for facial expression recognition. The module recognizes facial landmarks, including the locations of the mouth, nose, and eyes, based on input from a picture or a live video stream. Based on the movement and positioning of these facial landmarks, it then uses machine learning methods, such as Convolutional Neural Networks (CNNs), to categorize the emotions.

B. Mood classification module:

Mood classification module of the project involves prediction of the dominant emotion(s) present in the facial expression, such as happy, sad, angry, fear, surprise or neutral. This module is crucial to the project as it forms the basis for selecting the appropriate mood for music recommendation.

C. Search module:

Search module of the project involves searching for songs on YouTube. This module utilizes the predictions made by Mood detection module, based on the predicted mood(s) of the user. The module uses a standardized query for searching songs in YouTube.

In summary, these three modules work together to provide a seamless and personalized music recommendation Fig 2: Block flow diagram of the suggested task.

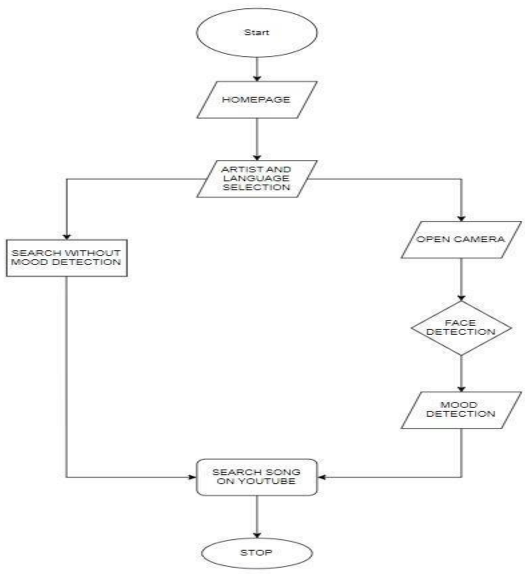

Figure 2. Flow diagram of the suggested task

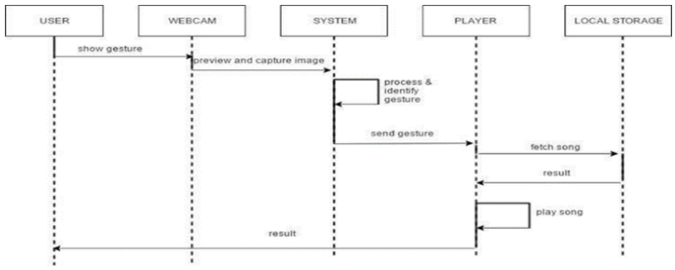

Figure 3. Sequence diagram of the suggested task

In this paper, we have developed a sophisticated face expression recognition system that marks a significant advancement in human-computer interaction by enabling computers to interpret human emotions in real time. At the core of our system is the Python OpenCV library, which facilitates the acquisition of live video streams from webcams, serving as the foundation for the subsequent face detection and emotion recognition processes. Utilizing OpenCV's extensive capabilities, our system adeptly identifies and localizes faces within video frames through a blend of deep learning models and Haar cascades, further enhancing Python OpenCV library is at the core of our systemizing detection accuracy by employing facial landmark detection for consistent face orientation.

The crucial preprocessing stage involves resizing, grayscale conversion, and normalization of facial images, leveraging OpenCV's preprocessing tools to optimize image quality for emotion analysis. The heart of our system lies in the emotion recognition phase, where a TensorFlow-implemented Facial Expression Recognition (FER) model, trained on a diverse dataset of labelled facial expressions, accurately discerns and predicts emotions by analyzing subtle facial cues. Visualization of recognized emotions is achieved through OpenCV's drawing functions, providing real-time feedback by annotating video streams with emotion labels.

A cornerstone of our project is the real-time processing capability, which ensures instantaneous emotional feedback through optimized algorithms and parallel processing techniques, thereby minimizing latency and maximizing throughput. This research not only demonstrates the integration of cutting-edge technologies in image processing and machine learning but also paves the way for enhancing human-computer interaction by imbuing machines with the ability to understand and react to human emotions, showcasing the potential for myriad applications across various domains.

In addition to the components outlined, the system integrates sophisticated techniques for optimizing real-time processing, ensuring seamless analysis and interpretation of facial expressions. By leveraging efficient algorithms and parallel processing methods, the system achieves low latency and high throughput, facilitating rapid and continuous emotion detection in live video streams or webcam inputs. This emphasis on real-time processing underscores the system's practicality and effectiveness in various applications requiring immediate feedback and interaction with users. The system uses frame-by-frame analysis to enable real-time processing. The facial regions are located and extracted using face detection and facial landmark detection algorithms for each separately managed video frame.

V. Implementation Results

The real-time facial emotion recognition system underwent rigorous testing, revealing outstanding precision and robust performance. Its ability to accurately categorize various facial expressions, spanning from neutral to surprise, anger, happiness, and sadness, positions the system as a promising tool for practical applications. The achieved high precision underscores the system's capacity to discern different emotional states swiftly and reliably in real-time, a crucial aspect for tasks like human-computer interaction and affective computing.

Beyond precision, the system's demonstrated resilience suggests its adaptability to diverse surroundings and demographic variations. This adaptability is essential for deploying the system effectively in dynamic and unpredictable environments, underscoring its potential real-world usefulness.

The successful categorization of a wide array of emotions establishes the system as an asset for applications demanding nuanced emotional understanding. This includes virtual reality environments, educational technology, and mental health monitoring, where comprehensive knowledge of human emotions is indispensable.

The following are the real time facial expressions results:

Figure 4. System detected as Angry face

Figure 5. System detected as Happy face

Figure 6. System detected as Sad face

Figure 7. System detected as surprise face

Figure 8. System detected neutral face

In conclusion, the findings confirm the remarkable efficacy of the real-time facial emotion recognition system, establishing it as a cutting-edge technology with significant implications for affective computing and human-computer interaction. The detailed precision assessments and proposed enhancements offer invaluable guidance for the continuous refinement and advancement of this influential technology. As a pre-trained facial emotion model, it's reported in various blogs that the system achieves an accuracy of 81.03% based on the descriptions of the pre-trained datasets. This accuracy may fluctuate depending on the factors such as lighting conditions and image visibility.

VI. CONCLUSION

In addition, Viby's success lies in its ability to adapt and evolve, continually refining its algorithms to better understand and cater to the complexities of human emotion. By incorporating feedback mechanisms and user interaction data, the system can iteratively improve its recommendations, ensuring a personalized and enriching music experience for each user.

Furthermore, as technology advances and our understanding of emotions deepens, there is vast potential for Viby to expand its repertoire beyond music. Imagine a future where Viby not only suggests songs but also offers therapeutic interventions, personalized playlists for specific emotional needs, and even real-time mood tracking to provide timely support and guidance.

Moreover, collaborations with experts in psychology, music therapy, and artificial intelligence could further enhance Viby's capabilities, fostering interdisciplinary innovation and pushing the boundaries of what is possible in emotion- driven technology.

In essence, Viby represents more than just a music recommendation system; it embodies a fusion of art, science, and compassion, poised to revolutionize how we interact with technology to enhance our emotional well- being.

VII. REFERENCES

[1] Dubey, M., Singh, P. L. (2016). Automatic emotion recognition using facial expression: A review. International Research Journal of Engineering and Technology.

[2] Sert, M., & Aksoy, N. (2016). Recognizing facial expressions of emotion using action unit specific decision thresholds. In Proceedings of the 2016 ACM on International Conference on Multimodal Interaction (pp. 49-54).

[3] Nagar, P. (2015). Emotion recognition from facial expression based on Bezier curve. International Journal of Advanced Information Technology, 5(3/4/5/6).

[4] Eckert, M., Gil, A., et al. (2016). Fast facial expression recognition for emotion awareness disposal. In 2016 IEEE International Conference on Consumer Electronics (ICCE) (pp. 1-6).

[5] Dea, A., Sahaa, A., et al. (2015, December). A human facial expression recognition model based on eigenface approach. In 2015 International Conference on Advanced Computing Technologies and Applications (ICACTA) (pp. 202-207).

[6] Gupta, G., & Rathee, N. (2015). Performance comparison of support vector regression and relevance vector regression for facial expression recognition. In 2015 International Conference on Soft Computing Techniques and Implementations (ICSCTI) (pp.141146).

[7] Deng, W., Hu, J., et al. (2015). DeepEmo: Real-world facial expression analysis via deep learning. 2015 IEEE Visual Communications and Image Processing Conference (VCIP), (pp. 1-4).

[8] Vinola, C., & Vimaladevi, K. (2015). A survey on human emotion recognition approaches, databases and applications. Electronic Letters on Computer Vision and Image Analysis, 14(2), 1-20.

[9] Filko, D., & Martinović, G. (2013). Emotion recognition system by a neural network based facial expression analysis. Automatika Journal for Control, Measurement, Electronics, Computing and Communications, 54(2), 223-235.

[10] Rzayeva, Z., & Alasgarov, E. (2019, April). Facial emotion recognition using convolutional neural networks. In 2019 IEEE 13th International Conference on Application of Information and Communication Technologies (AICT) (pp. 1-5).

[11] Hussain, S. A., & Al Balushi, A. S. A. (2020, July). A real-time face emotion classification and recognition using deep learning model. Journal of Physics: Conference Series, 1432(012087).

[12] Puri, R., Gupta, A., Sikri, M., Tiwari, M., Pathak, N., & Goel, S. (2020). Emotion detection using image processing in Python.

[13] Patra, B., Das, D., & Bandyopadhyay, S. (2013). Automatic music mood classification of Hindi songs.

[14] Lee, J., Yoon, K., Jang, D., Jang, S., Shin, S., & Kim, J. (2018). MUSIC RECOMMENDATION SYSTEM BASED ON GENRE DISTANCE AND USER PREFERENCE CLASSIFICATION. Kaufman Jaime C., University of North Florida, A Hybrid Approach to Music Recommendation: Exploiting Collaborative Music Tags and Acoustic Features, UNF Digital Commons, 2014.

[15] Priya, D. (2019, April 3). Face Detection, Recognition and Emotion Detection in 8 lines of code! [Blog post]. Towards Data Science. Retrieved January 12, 2024, from https://towardsdatascience.com/facedetectionrecognition-and emotion-detection-in-8-linesof-codeb2ce32d4d5de.

[16] Pundalik Chavan, H. Hanumanthappa, E. G. Satish, Sunil Manoli, S. Supreeth, S. Rohith & H. C. Ramaprasad "Enhanced Hybrid Intrusion Detection System with Attention Mechanism using Deep Learning" Published in SN Computer Science ,ISSN: 2662-995X, 5:534, May2024.(Q2 Scopus Journal).

[17] Pundalik Chavan, Anooja Ali, Ramaprasad H C, Ramachandra H V, Hari Krishna H and E G Satish "Analysis of Wireless Networks: Successful and Failure Existing Technique" Published in International Conference on Advances in Data-driven Computing and Intelligent Systems (ADCIS 2023) ISBN 978-81-955020-2-8. DOI:https://doi.org/1056155/978-81-955020-2-8-75.