International Journal of
INTELLIGENT SYSTEMS AND APPLICATIONS IN ENGINEERING
ISSN:2147-6799 www.ijisae.org Original Research Paper

Creation of A Music Recommendation System using Facial Expression Recognition with MATLAB

V. Maruthi Prasad[1], Gutti Naga Swetha[2], K. Md. Riyaz Ali[3]

Submitted: 09/12/2023 **Revised**: 20/01/2024 **Accepted**: 30/01/2024

Abstract: Choosing music from a wide array of options often poses a challenge for individuals. Various recommendation frameworks have been made available for themes such as music, food, and shopping, depending on the user's perspective. The main objective of our music recommendation system is to provide consumers with suggestions that align with their preferences. Examining the user's emotions and facial expressions might provide insights on their present emotional or mental state. The domains of music and video provide enough opportunities to provide customers with a diverse range of choices based on their preferences and stored information. Humans commonly transmit their intentions and the context of what they say through their facial expressions. More than 60% of users feel that their music library occasionally contains too many songs, making it challenging to decide which one to play. It could assist someone in selecting the music they ought to listen to by developing a recommendation system, enabling them to feel less bothered all around. After identifying the most suitable track that aligns with the user's mood, songs are shown to the user based on their mood, eliminating the need for the user to manually search or look for music. The user's picture is obtained via the use of a camera. The user is thereafter recorded on camera as a song from their playlist that most accurately reflects their mood or attitude is played.

Keywords: *Facial expression, Haar cascade technique, Music identification, CNN*

1. Introduction:

Facial expressions are the most common means by which individuals convey their reactions and feelings. A person's disposition may be influenced by music, as has been known for a long time. By observing and identifying the emotion being displayed by a person, as well as playing music that matches that person's mood, one can gradually calm their thoughts and finally have a pleasing effect. The project's objective is to capture an individual's facial expressions as they display emotion. A music player records human emotion via a web camera interface for computers. In order to determine the feeling that the individual is attempting to convey via their facial expressions, the software use picture segmentation and image processing algorithms to extract information from the face of the audience member. The purpose of the idea is to improve the user's disposition by playing music of their choosing while simultaneously snapping a snapshot of them. Since the beginning of time, facial expression recognition has maintained its position as the most accurate method for evaluating expression. Facial expressions provide people with a clear method by which they may infer or interpret the feelings, moods, or thoughts that another person is attempting to transmit to them. Altering one's frame of mind sometimes has the potential to assist one in overcoming unpleasant and discouraging conditions. With the help of expression analysis, it is possible to avoid a great deal of potential dangers to one's health, and it is also possible to take measures that will enhance one's mood.

Due to the fact that music has the capacity to elicit a wide range of emotional states, the relationship between music and human emotions has been widely acknowledged. The area of music recommendation systems has seen a rise in the use of facial expression detection technologies in recent years, which has led to an increase in their popularity throughout the industry. Using facial emotion recognition to modify music selections depending on the listener's facial expressions is a novel approach to music recommendation that incorporates the use of facial identification. It is possible for this technology to analyse the facial expressions of a user in real time and then provide recommendations for music that is most suitable for the user's present emotional state. This article examines the process of developing and implementing a music recommendation system that is based on facial expressions instead of traditional music recommendations. This work investigates the technical components that support this system, the advantages that it offers, and the obstacles that need to be solved in order to increase the efficacy of the system. This study is being conducted with the intention of gaining a full knowledge of the capabilities of such a

1 Asst. Professor, Dept of CST, Madanapalle Institute of Technology & Science, Madanapalle
Email: maruthi.vv@gmail.com
2 Asst. Professor, Dept of ECE, Madanapalle Institute of Technology & Science, Madanapalle
Email: nswethag@gmail.com
3 Asst. Professor, Dept of ECE, Madanapalle Institute of Technology & Science, Madanapalle
Email: riyazalikmd@gmail.com

system, as well as its possible applications in the music business and other sectors. As a result of the broad use of advanced face recognition technology and the growing availability of a variety of music streaming platforms, this technology has the potential to act as an effective solution for providing individualised suggestions for music. The way in which we interact with music has the potential to undergo a total transformation, resulting in an increase in the degree of personalisation and emotional resonance that it has. We are able to improve the listening experience for listeners by adding facial expression recognition technology into music recommendation algorithms. This allows us to personalise the music to the listener's individual interests and feelings. In addition to that, this technology has the potential to be used in other fields, such as the medical field, the educational field, and the entertainment industry. On the other hand, just as with any other rapidly developing technology, there are issues of ethics, privacy, and data security that need to be addressed. We hope that with the assistance of this research, we will be able to recognise these issues and devise solutions that will make it possible to make ethical and secure use of this technology. To make significant contributions to the development of music recommendation systems that are based on facial expressions and to give useful insights for the study that will be undertaken in this area in the future, the primary objective of this research is to accomplish both of these goals.

Fig 1. Facial expressions

A detailed assessment and evaluation of previous research studies relevant to facial expression recognition technologies and music recommendation systems will be carried out by us in order to achieve the objectives of this course of study. In addition to this, we will investigate a variety of machine learning and deep learning techniques, which are often used in the process of developing models for identifying facial expressions of emotion. Furthermore, in order to evaluate the effectiveness of the proposed method, we will conduct an empirical inquiry by collecting data from the individuals who participated in the study. It is possible that the findings of this study may have significant implications for the music industry. These findings may provide music streaming services and music marketers with new ways to personalise their offerings.

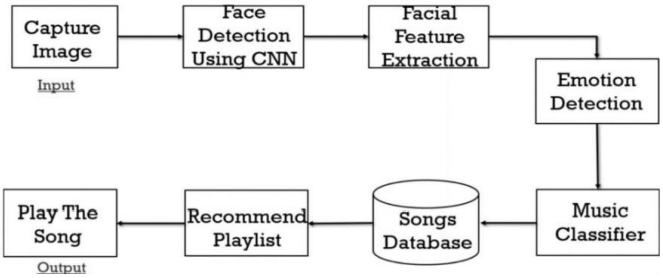

Fig 2. System Overview

This approach has the potential to augment listener engagement, optimise user experience, and eventually, amplify income. Moreover, this technology has pragmatic implications in the field of healthcare, as it may be used to track the emotional condition of patients and provide tailored therapeutic interventions. The primary objective of this study is to enhance comprehension of the possible uses of facial expression detection technology in music and other domains. This will facilitate future investigations and advancements in this promising technological sector.

2. Literature Survey:

The geometric-based feature extraction and the appearance-based feature extraction were presented by Zheng et al. [1] as two major categories for the purpose of face feature extraction. Both procedures included the removal of a number of important facial features, including the lips, eyes, and brows, among others. Nikhil is also included. It was determined by Nikhil et al. [2] that facial expressions may be used to infer the user's personality. Hand gestures, facial expressions, and changes in voice or tone are all common ways for humans to communicate their feelings; nevertheless, the majority of the time, they do convey their feelings via their faces. A player that plays music that evokes feelings allows for more effective use of time. On their playlists, the majority of people like listening to a wide variety of music. There is no improvement in the user's mood when random music is played. Through the use of this technology, users are able to have music played automatically depending on their current state of mind. The web camera takes a picture of the user, and the pictures are stored once they have been snapped. Before anything further, the photos are changed from RGB to binary style. The use of a feature-point detection methodology is the name given to this particular method of encoding the data. One other method that may be used to finish this procedure is the Haar Cascade approach that is utilised by Open CV. When it comes to

machine learning, Haar cascading is a strategy that includes training a classifier by employing a huge number of pictures that are both positive and negative. Both Paul Viola and Michael Jones [5, 6] were the ones who first introduced the idea. Object identification is accomplished by the use of specialised classifiers known as Haar feature-based cascade classifiers. This particular classifier makes use of a machine learning technique that performs a cascade operation in order to recognise items in images based on patterns that are seen in previous photographs. Face recognition and facial expressions are seen and identified accurately in the photo. When the classifier is provided with both positive and negative pictures, the task is said to have been concluded. In the next step, the distinctive characteristics of the picture are extracted. Each characteristic is a unique value that is derived by subtracting the total number of pixels in the white rectangle from the total number of pixels in the black rectangle.This system has the ability to recognise the faces of a wide variety of people in a variety of environments. Because of the use of the integral image, it is possible to calculate the Haar-like feature of any dimension in a manner that is both efficient and constant in time. During the construction of the audio player, Java was used. The database is under its control, and the music is played in line with the user's current state of mind. A study was carried out by Z. Zeng and colleagues that investigated several advancements in the field of human emotion recognition. His primary focus was on the many approaches that may be used to manage records of emotional experiences in either audio or visual format respectively.

This page provides a comprehensive explanation of the many methods of audio and video computing. A variety of emotional states, including surprise, fury, contempt, and fear, are all modelled after this effect, which serves as a template for these emotions. The purpose of this essay was to highlight the challenges involved in the process of building automated and spontaneous affect recognizers in order to facilitate the detection of emotions. In addition to this, it brought to light a few problems that uni-modal emotion detection had either ignored or avoided. An idea was proposed by Parul Tambe and colleagues [7] that would automate the interactions between users and music players, learn about the preferences, emotions, and activities of each individual user, and then deliver song suggestions as a consequence of this information. Additionally, the gadget was able to analyse the users' feelings and predict the musical style by capturing the numerous facial expressions that they displayed. It was proposed by Jayshree Jha and her colleagues that a music player that makes use of visual processing and is based on feelings be developed. The results of this demonstrated how a variety of algorithms and approaches proposed by authors in their study may be used to establish a connection between the music player and the emotional reactions that individuals experience. As a consequence of this, it has made it simpler for users to construct and maintain playlists, and it has also helped to the enhancement of the listening experience for music enthusiasts by recommending the most appropriate song taken into consideration the user's present expression. In 2019, AZM Ehtesham Chowdhury and colleagues [10] presented a novel camera prototype that was developed with the intention of conducting attendance analysis in a more effective manner. in order to construct a model that is more reliable and consistent. It was suggested that another approach may be used to perfectly execute the notion. Through the use of a modus operandi, this strategy will assess the attendance of students. The concepts of facial recognition and detection are the foundation upon which this system is built. When establishing which modus was the most valid, the major factor that was considered was the average accuracy. The year 2019 saw the development of a system by Nandhini R, Duraimurugan N, and others that is capable of accurately and properly identifying and recognising the faces of students in images or videos that were acquired by a camera.

The process of face recognition via the use of Deep Learning has been optimised through the utilisation of a number of different approaches and technologies. The year 2018 saw the development of an autonomous attendance management system by M. Kasiselvanathan and colleagues [15]. This system makes use of face recognition technologies. In order to recognise facial characteristics and determine whether or not a face is present, the system makes use of facial recognition technology. Enhancing the quality of the system has resulted in the development of a face recognition programme that is more effective. Eigen Faces is the algorithm that has been used by the system. The method not only includes the identification of face features, but it also takes into account the spatial elements of facial characteristics, which are defined by dynamic principles.

A strategy for constructing a fully integrated student attendance system that makes use of face identification was proposed by Omar Abdul Rhman Salim and his colleagues [16] in the year 2018. The strategy is predicated on the use of Raspberry Pi, which performs its operations using the Raspbian Operating System. The Raspberry Pi is connected to a camera as well as a screen that is 5 inches in size. A photograph that has been recorded will be transferred to the Raspberry Pi. Facial recognition is processed by the system via the generation of Local Binary Patterns (LBPs), which are created from the inside out. If there is a match between the face in the input photo and the image in the training dataset, then the door will be opened, and the attendance will be recorded and preserved. An method to face

recognition was developed by Xiang-Yu and colleagues in 2017.

This technique offers quick analysis, which helps solve the issue of successfully identifying a face in a variety of different settings. For the purpose of extracting the characteristic from the main database, the Haar-feature classifier was used. In addition, the procedure may also be used for the purpose of having the feature removed. The currently available technologies for investigation have shown that it is possible to obtain private recognition via the use of facial detection. In the year 2017, C.B. Yuvaraj and his colleagues [20] came up with the idea of using image processing methods for the purpose of monitoring personnel attendance. The original purpose of this article was to get an understanding of the facial expressions that were being employed. It was via the use of Haar characteristics that the Viola-Jones approach was improved. It is possible for the algorithm to identify the facial characteristics of a particular student based on the image. Database files are addressed via the provision of solutions that include a planned coaching plan. The Haar cascade method offers a number of benefits when it comes to identifying distinct face traits. Maintaining the numbers of the cameras in a conscientious manner is required in order to successfully complete the commute.

3. Proposed System:

A mood detection model can identify faces in any image and then forecast the emotions those faces will display. Videos and still pictures both work for this. Our recommended system uses the predicted emotion as input after determining the expression from the face to produce recommendations after processing a dataset from a Kaggle. Feature extraction is the process of detecting and extracting the pertinent characteristics of a face that are required for emotion identification. Facial emotion identification generally relies on the analysis of key facial characteristics such as the positioning and morphology of the eyes, eyebrows, mouth, and nose. Additional attributes, such as the skin's texture and colour, as well as the facial structure, may also be used. Haar cascades are widely used for extracting features in face emotion detection. Haar cascades are a collection of distinctive characteristics used for object detection in a picture. Regarding facial emotion identification, these traits are specifically engineered to identify and analyse the eyes, nose, mouth, and other pertinent facial characteristics.

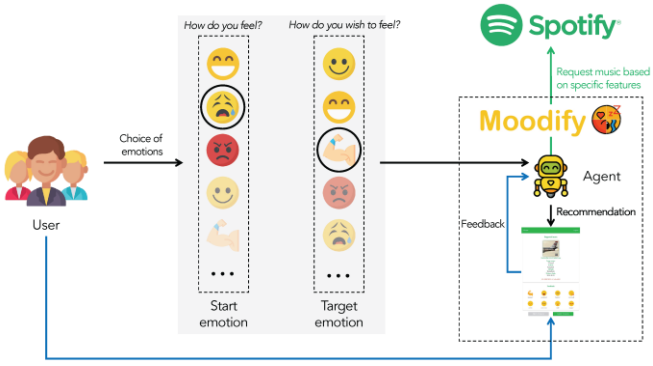

Fig 3. Emotion-Based Music Recommendation

Haar cascades function by systematically examining a picture at various scales and dimensions, in order to detect the existence of certain characteristics. The characteristics consist of rectangular portions in the picture that exhibit contrasting dark and bright areas. Through the integration of several characteristics, it becomes feasible to identify more intricate entities, such as human faces. Following the identification of the features, they may be used to train a machine learning model, such as a Convolutional Neural Network (CNN), for the aim of emotion detection.

4. Data Collection:

The data was gathered using distinct sets of phrases. We categorised the 7 distinct emotions based on the user's facial expressions. Kaggle dataset - The algorithm was trained using the dataset from Kaggle. The collection comprises diverse facial action coded sequences from many people. The labelling serves to provide information on the emotions exhibited by the individual. There is a series of images that depict the target's neutral expression gradually transitioning to the subject's intense emotional expression. Both the initial photo and the final image are used in both our training network and our analysis network. The network does not undergo training using the additional pictures in their original form. The utilisation of the convolutional neural network (CNN) technique proves to be highly effective in the classification of emotions by leveraging facial data. Convolutional Neural Networks (CNNs) are a specialised category of deep learning algorithms that demonstrate exceptional performance in the field of image recognition tasks. To construct the emotion classification module, it is advantageous to employ a Convolutional Neural Network (CNN) algorithm that has undergone training on a sizable dataset containing labelled facial expressions. The ability to recognise patterns in face characteristics that are related with a variety of emotions has been developed by CNN. Through the use of these patterns, it is able to precisely determine the emotional state of the listener at the present moment. It is possible to capture both local and global aspects of the facial expression when using a CNN algorithm for emotion classification, which is one of the many advantages of

using such an approach. For instance, the algorithm may learn to detect particular patterns in the location of the eyes or mouth, in addition to more general patterns in the overall face structure. This is possible because the algorithm is intelligent.

5. Methodologies:

Emotion Extraction Module - There is a camera or webcam that is used to take a picture of the user. Once the webcam stream picture frame has been obtained, it is converted into a grayscale image in order to improve the effectiveness of the face recognition classifier. Following the completion of the conversion procedure, the image is then sent to the classifier algorithm. This algorithm makes use of feature extraction methods in order to separate the face from the whole frame of the web camera feed. In order to ascertain the feelings that the user is expressing, the trained network is provided with certain characteristics. These qualities are derived from the collected facial expression.

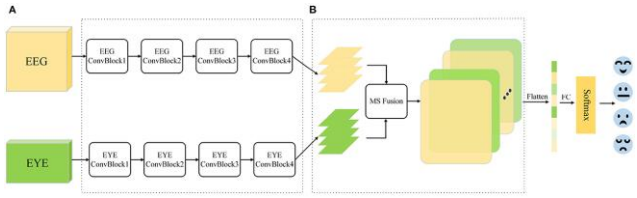

Fig 4. Emotion identification using a combination of EEG and eye movement inputs.

Music Extraction Module – The audio or music based on the emotion expressed by the user has been offered, and the user can select any song based on the emotion of the user that has been extracted. The numerous tunes are created based on the user's unique expressions. Depending on how we feel, certain music are played. The Mood Detection model is capable of identifying and extracting facial features from any given picture, enabling it to accurately forecast the emotional state associated with the detected face. Both still photographs and films may be used for this task. Once the emotion is determined from the facial expression, our recommender system use this anticipated emotion as input. It then generates recommendations by using a Spotify dataset obtained from a kaggle contest. The recommender algorithm will provide a curated selection of the top 40 songs to recommend for a Spotify playlist.

6. Results:

This study proposes a device that automatically plays the songs based on the user expression. To develop this system, we are utilising the MATLAB software. This will enable us to use the MATLAB software's webcam command to extract the user's facial expressions. The system will analyse the expression based on the expressions in the dataset after taking the image. After the face has been removed, specific facial traits are taken and submitted to the trained network in order to determine the emotion the user has exhibited. The outcome is decided based on the expression matching the expression in the dataset. We can ascertain the user's emotion in this way. Then a certain music is played in response to the user's feeling. Here, we had provided many tunes for various emotions.

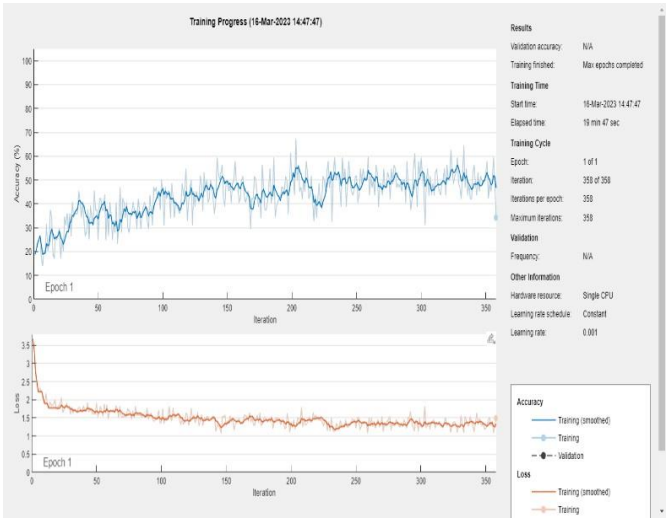

Fig 5. Training time for one epoch

Here firstly we have to train the datasets. This will take some time to load the entire datasets that have different pictures for different emotions.

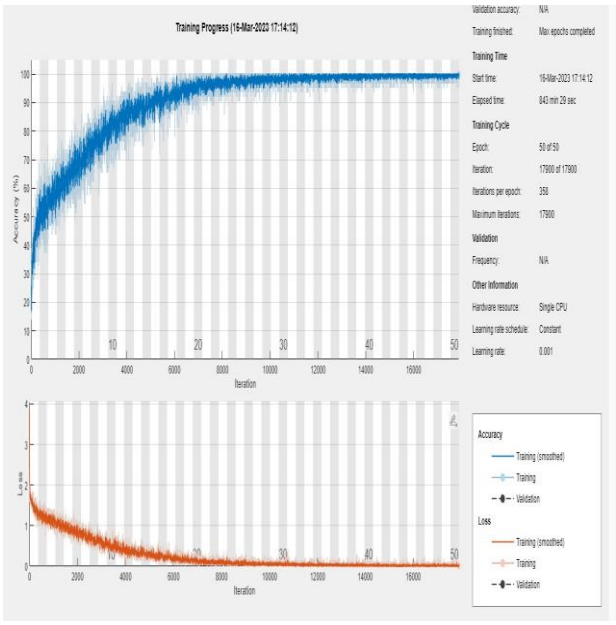

Fig 6. Training time for 50 epoch

When we are designing this system we had returned a program to take the image as the input and the determine the emotion of that picture. After determining the emotion from the picture the song will be played automatically based on the emotion. Here we can give different pictures as input and we can check the output.

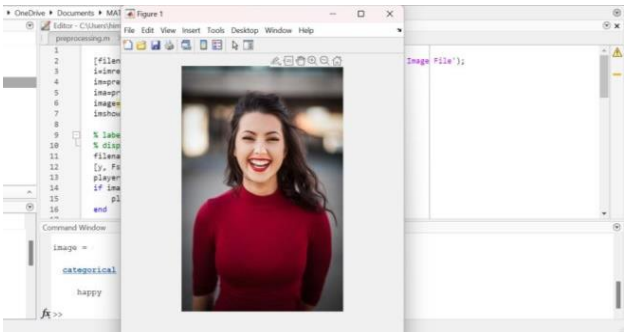

Fig 7. Output for Face emotion based on user input

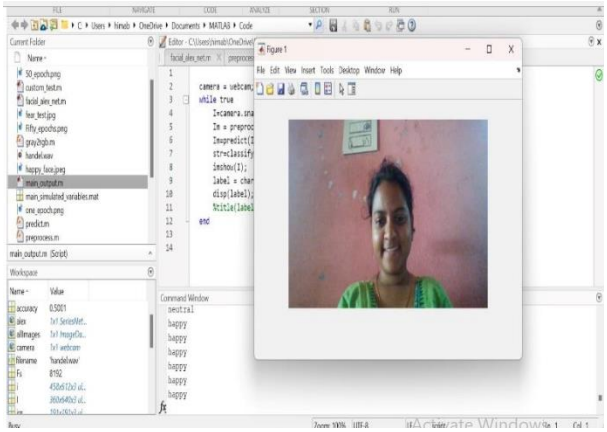

Fig 8. Real Time happy Face Emotion Recognition with music

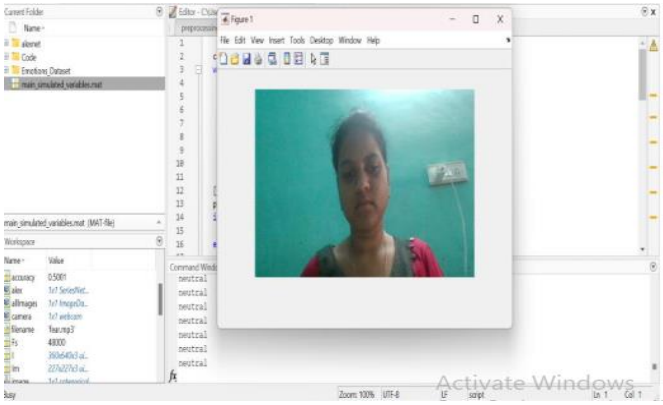

Fig 9. Real Time neutral Face Emotion Recognition with music

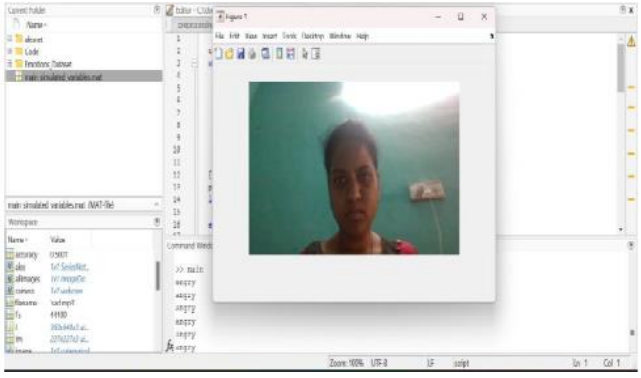

Fig 10. Real Time angry Face Emotion Recognition with music

The above figures demonstrate how the face's distinct features are collected and passed to a trained network to identify the user's displayed emotions. Following that, the emotion is ascertained, and the song is then automatically played in accordance with the emotion.

7. Conclusion:

One of the most significant areas of research is the detection of emotions based on facial expressions, which has been the subject of a great deal of curiosity in the past. It should come as no surprise that the use of image processing algorithms to identify mood has gotten more difficult with each passing day. Researchers are always looking for solutions to this problem by using a wide variety of characteristics and methodologies to image processing. The use of image processing algorithms in the examination of human subjects and medical research is of utmost significance. The use of image processing algorithms to extract the user's mood and then utilise that emotion to treat the user is resulting in the development of new ways that are constantly being produced as a consequence of this methodology. In the event that a trustworthy algorithm that is capable of reliably identifying a person's emotions is implemented, the industry will make significant progress. In many aspects of life, the ability to recognise one's feelings is becoming an increasingly important skill. An accurate representation of the user's feelings may be obtained by the system. With regard to this particular case, it has been put through testing in a real-time setting. In order to assess the resilient nature of the system that has been designed, it must be tested in a variety of lighting conditions. It was possible for the system to successfully update both its classifier and its training dataset by making use of the user's newly uploaded photographs. A facial landmarks approach was employed in the development of the system, and its performance was evaluated using a number of different situations using the system. It is clearly shown that the classifier obtains an accuracy of more than 80 percent on a constant basis, which is a reasonable degree of accuracy when it comes to the classification of emotions. Furthermore, when the classifier is assessed on a real user, it is possible to demonstrate that it properly predicts the expression of a user in a live environment.

References:

[1] Based on a statistical technique, facial expression and recognition analysis by Londhe RR and Pawar DV was conducted in 2012. Journal of Soft Computing and Engineering International 2

[2] Department of Computer Engineering, Pune Institute of Computer Technology, Pune, India, (IEEE),2017. Smart music player including face expression detection and music mood

recommendation. Husain Zafar, Chintan Soni, Shlok Gilda, and Kshitija Waghurdekar.

[3] Sreenidhi Institute of Science and Technology, Yamnampet, Hyderabad; International Journal of Emerging Technologies and Innovative Research (JETIR), Volume 7, Issue 4, April 2020. Emotion-based music recommendation system. CH. Sadhvika, P. Srinivas Reddy, and Gutta Abigna.

[4] In 2020, P. Singhal, P. Singh, and A.Vidyarthi DCNN in Chest X-Ray interpretation and localisation of Thorax disorders. Engineering in Informatics, Electrical & Electronics, 1(1), 1–7

[5] Learning unique facial dynamics to recognise faces in videos: Hadid A, Pietikäinen M, and Li SZ, 2007, International Workshop on Analysis and Modelling of Faces and Gestures, pp. 1–15 Heidelberg Springer in Berlin

[6] In 2008, Zeng Z, Pantic M, Roisman GI, and Huang TS published a survey of emotion recognition methods in the IEEE Transactions on Pattern Analysis and Machine Intelligence, spontaneous, aural, and visual manifestations 31 39–58.

[7] MoodyPlayer is a mood-based music player, according to Yadav S, Patel AR, Vollal A, Kadam PB, Samant RM, and Yadav S (2016) Int. J. Comput. Appl. 141 0975-8887.

[8] Published in March 2016 is "Comments polarity classification based on emotion analysis and movie recommender system design and implementation," written by Xia Mingxing.

[9] The International Journal of Engineering Research and General Science 3 750-6 published an article by Kabani H, Khan S, Khan O, and Tadvi titled "Emotion-based music player" in 2015.

[10] Londhe RR and Pawar DV (2012) investigated facial expression and recognition using a statistical method. International Journal of Soft Computing and Engineering 2

[11] a cutting-edge music player with a built-in face recognition system, 2015 International Journal of Advanced Research in Computer Science and Software Engineering 5 by Yash Bagadia, Taher Khalil, and Noor UlAin Shaikh

[12] Dubey and Singh (2016) reviewed the use of facial expressions for automatic emotion recognition in IRJET 3 (488–92), an international research journal of engineering and technology.

[13] "Emotional Detection and Music Recommendation System based on User Facial Expression," by Florence, M.Uma, and S. Metilda. IOP Conference Series: Materials Science and Engineering, Vol. 912, No. 6, 2020.

[14] Gilda, Shlok, et al., "Smart music player integrating facial emotion recognition and music mood recommendation." An international conference on wireless communications, signal processing, and networking is the 2017 WiSPNET Conference. IEEE, 2017.

[15] International Journal of Advanced Research in Computer and Communication Engineering, Facial Expression Based Music Recommendation System 2021, DOI: 10.17148/IJARCCE.2021.10682. Raj P, Vinay P, Bhargav S.K., et al.

[16] Mandeep Kaur and Rajeev Vashisht, "Facial Expression Recognition Using a Noval Approach and Its Application," International Journal of Computer and Electrical Engineering, Vol. 3, No. 2, April 2011, 1793-8163.

[17] J. F. Cohn, T. Kanade, J. Saragih, Z. Ambadar, and I. Matthews are among the authors. 2010 CK+, the expanded Cohn-Kanade dataset, contains every action unit and all of the different emotional expressions. 2010 IEEEE Computer Society Conference on Computer Vision & Pattern Recognition Workshops 94–101

[18] Hussain I., Tayyab M., Mirza HT, and Shoaib M. 2017. An extensive review of the research on the use of knowledge and modern technology in the treatment of autistic children In 2017, 1-10 IEEE held the 17th International Conference on Computational Science and Its Applications.

[19] R.K. Dash, T.N. Nguyen, K. Cengiz, A. Sharma Fine-tuned support vector regression model for stock predictions Neural Computing and Applications (2021), pp. 1-15

[20] C.B. Yuvaraj, M. Srikanth, V. Santhosh Kumar, Y.V. Srinivasa Murthy, and Shashidhar G. Koolagudi, “An approach to maintain attendance using image processing techniques”, International Conference on Contemporary Computing (IC3), 10-12 August 2017,Noida, India.

[21] IoT Fog Computing Gateway with High Security and Energy Efficiency Prasad, V.M., Bharathi, B. 14th International Conference on Advances in Computing, Control, and Telecommunication Technologies, ACT 2023, 2023, 2023-June, pp. 2501–2510

[22] K. Seyhan, T.N. Nguyen, S. Akleylek, K. Cengiz, S.H. Islam Bi-GISIS KE: Modified key exchange

protocol with reusable keys for IoT security J. Inf. Secur. Appl., 58 (2021), Article 102788

[23] T.N. Nguyen, S. Zeadally, A. Vuduthala Cyber-physical cloud manufacturing systems with digital-twins IEEE Internet Comput. (2021)

[24] Sujay Patole, Yatin Vispute Automatic attendance system based on face recognition Int. J. Innov. Res. Sci., Eng. Technol., 6 (August 2017)

[25] A Novel Trust Negotiation Protocol for Analysing and Approving IoT Edge Computing Devices Using Machine Learning Algorithm Prasad, V.M., Bharathi, B. International Journal of Computer Networks and Applications, 2022, 9(6), pp. 712–723

[26] B.D. Parameshachari Logistic sine map (LSM) based partial image encryption 2021 National Computing Colleges Conference (NCCC), IEEE (2021, March), pp. 1-6

[27] M.K. Kumar, B.D. Parameshachari, S. Prabu, S. Liberata Ullo Comparative analysis to identify efficient technique for interfacing BCI system IOP Conference Series: Materials Science and Engineering, IOP Publishing (2020, September), Article 012062 (Vol. 925, No. 1, p.

[28] Security in 5G Networks: A Systematic Analysis of High-Speed Data Connections Maruthi Prasad, V., Bharathi, B. International Journal on Recent and Innovation Trends in Computing and Communication, 2023, 11(5), pp. 216–222