iJRASET
International Journal For Research in Applied Science and Engineering Technology



INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 11 **Issue:** IV **Month of publication:** April 2023

DOI: https://doi.org/10.22214/ijraset.2023.50825

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 IV Apr 2023- Available at www.ijraset.com

Emotion-based Music Recommender System using Viola-Jones and CNN

Bijal Dharne[1], Naveen Shukla[2], Ankit Hooda[3], Prof. Ichhanshu Jaiswal[4]
Department of Information Technology, Vidyalankar Institute of Technology, Mumbai, India

Abstract: ***The goal of this project is to develop an emotion-based music recommender system that provides personalized music recommendations based on the user's current emotional state. The system incorporates an emotion detector that analyses user's facial expressions, to determine their current mood. Based on the detected emotion, the system recommends music tracks that match the user's current emotional state.***
Index Terms: convolutional neural network, emotion detection, intelligent agent.

I. INTRODUCTION

Music is an excellent way for a person to relieve stress. Choice of music of a person is dependent on the person's mood. Mood of a person can be inferred from the person's emotion and emotions can be inferred from person's facial expressions. Mehrabian [1], states that two-thirds of people communication is carried through non-verbal communication and facial expression constitutes the largest component in this percentage. While only one third of emotion is conveyed in the verbal communication.

According to this research [2], there are 6 basic emotions which are- happiness, sadness, anger, fear, disgust and surprise. Nowadays, emotion detection is considered intrinsic in many applications such as retail, education, medical diagnosis, videogaming etc [3].

According this study [4], participants in sad mood are likely to sad music, those in happy mood prefer to listen happy music and vice versa. To detect emotions using facial expression, intelligent agent is crucial. Input device like a good resolution camera also plays a vital role in emotion-detection systems. If a camera is used as an input device, good illumination is required to ensure the user's facial expressions are detected properly.

II. LITERATURE SURVEY

1) System 1: Music Recommendation Based on Face Emotion Recognition [5]

Presented a system in which only a device (e.g., laptop) with inbuilt camera (or a suitable exterior webcam) is required for detecting emotions and then suitable music scores are recommended to the user. Purpose of this system is to help the user achieve good mood, if the user is showing a negative emotion. This system uses CNN algorithm (with 95.14% of accuracy) for emotion detection and is able to detect 5 emotions. The emotions are: happy, sad, angry, surprise and neutral.

The system has following benefits. It is cross platform so can work on any operating system. The system is easy to use.

This system has following drawbacks. Some important emotions like fear and disgust are not detected by this system. Does not perform well in poor camera resolution and extremely bad light conditions.

2) System 2: Music Recommender System for Users Based on Emotion Detection through Facial Features [6]

Proposes a system which detects emotions, if the user has negative emotion, then suitable playlist will be recommended that contains appropriate types of music needed to improve his mood. Only a device (e.g., laptop) with inbuilt camera (or a suitable exterior webcam) is required for detecting emotions. The user needs to capture his photo and based on the photo; the system detects emotion. This system uses Viola-Jonze algorithm for face detection and Principal Component Analysis (PCA) for detecting emotion. The following system is able to detect 4 emotions which are: happy, sad, natural and surprised.

The system has following advantages. It is easy to use.

The system has following disadvantages. Some important emotions like anger, fear and disgust are not detected by this system. Sometimes the emotions are not classified correctly like happy face without showing the teeth may be classified as neutral, likewise, surprised face when the teeth appear may classified as happy, and when the teeth don't appear sometimes it's classified as sad because of the shape of the mouth and that may justify the cases that their detection was not accurate.

©IJRASET: All Rights are Reserved

3) System 3: A Machine Learning Based Music Player by Detecting Emotions [7]

Proposes a music player which uses camera to capture emotions of the user. The purpose of this system is to relieve the user from the negative emotion if detected and to enhance the positive mood if user already displays a positive emotion. In this music player, CNN is used to detect user's emotions and SVM is used to classify songs stored in the local database.

These are the merits of this music player. It is best suited for physically challenged people because it requires very less manual intervention.

This music player has following deficiencies. It is unable to detect disgust emotion. Songs need to be input separately into the local database so the user has limited choice of songs and the database may consume large amount of user's computer memory.

4) System 4: Emotion Based Music Recommendation System Using Wearable Physiological Sensors [8]

Proposes a system in which a wearable computing device integrated with physiological sensors is used to detect user's emotions. Emotions are detected using physiological signals – GSR (Galvanic Skin Response) signals and photo plethysmography signals. Random Forest classifier (with 71 % accuracy) is used to recommend music based on emotions detected.

This system has the following virtues. Uses physiological signals to track and recognize emotions which is more reliable method compared to using facial expressions.

This system has the following shortcomings. Requires lots of additional hardware like wearable sensors hence the total cost of system becomes expensive. Harder to use compared to other systems in this paper because it uses additional hardware like sensors.

5) System 5: Smart music player integrating facial emotion recognition and music mood recommendation [9]

Presents a music player which detects emotions of the user in real time using camera and generates a playlist based on the emotional state of the user. If user is sad, playlist of serene and soothing music will be recommended. And if the user displays positive mood, a playlist will be recommended to raise the existing positive emotion. It has the following modules – emotion module, music classification module and recommendation module. The emotion module determines the emotional state of the user using CNN (Convolutional Neural Network) with an accuracy of 90.23 %. The music classification module classifies different music based on emotion using relevant and critical audio information using ANN (Artificial Neural Network) with an accuracy of 97.69%. And the recommendation module combines results of both emotion module and music classification module to recommend playlist to the user. This system has the following merits. It is cross-platform so it can work with any operating system.

This system has the following demerits. Only limited to modern American and British English songs. It is only able to detect 4 emotions – angry, happy, sad and calm and not able to detect other basic emotions which are fear, disgust and surprise. Songs need to be manually input in the system hence the choice of songs is limited.

6) System 6: An Intelligent Music Player Based on Emotion Recognition [10]

Presents an intelligent agent that uses camera to capture user's emotion and suggest an appropriate playlist as per the user's emotional state. Whenever the user wishes to generate a mood-based playlist, he must capture his image at that instant. Based on the picture, emotion is detected and music that best matches his emotion is recommended as a playlist. First user emotion is detected by applying Viola-Jones face-detection algorithm for detecting face and a Fischer face model is used for detecting emotion of the user using the captured image. Then the local music database is fetched and K-means clustering is applied on local playlist to classify different music based on emotions. Finally, the music playlist suitable to user's emotion is recommended to him. The overall system has an accuracy of 72.3%. These are the strength of this agent. Music can be played on any system regardless of its music player. Uses Fischer face recognition which proves to be efficient because it is able to work with additional features such as spectacles and facial hair and is relatively invariant to lighting.

Following are the weaknesses of the agent. It uses local music database so there is limited choice of songs available for the user and local music database may consume large amount of user's computer memory. Not able to detect all 6 basic emotions.

III. METHODOLOGY

The proposed system is a music recommender that automatically detects user's emotion using facial expressions through a camera. User's face is captured in real time and then his emotion is classified into one of the seven classes: angry, disgusted, fearful, happy, neutral, sad or surprised.

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



Figure 1.: Examples on the 7 emotions

Firstly, the user enters the singer's name and language through the interface, then clicks a button which is displayed on the interface. The button opens a separate window which displays camera input and user's emotion. After this, the user presses a specific keyboard key, which directs him to a suitable website like YouTube music that searches for music tracks based on the singer's name, language, and detected emotion. The music tracks are then displayed to the user. The following stages are shown in the level 2 DFD diagram below.

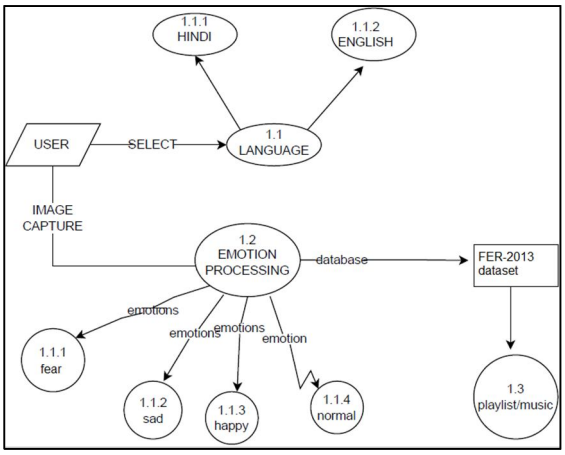


Figure 2: Level 2 DFD

A. Dataset Description

Kaggle dataset titled FER-2013 [11] was used to build a Convolutional Neural Network. FER-2013 dataset is divided into two parts-training dataset and testing dataset. The training dataset consists of 28709 images and the testing dataset consists of 7178 images. Each image in FER-2013 is labelled as one of the seven emotions: angry, disgust, fear, happy, neutral, sad or surprise.

B. Facial Expression Detection

Using input received through the user's camera, Viola-Jones algorithm is to detect facial expressions in the input image. Benefit of using the following algorithm is its robustness and accuracy in detecting objects under varying lighting conditions, poses, and orientations. The algorithm uses a set of Haar features that are designed to capture the variations in texture, brightness, and contrast of the objects. This is an object detection algorithm that detects faces in an image using a set of pre-trained classifiers.

C. Emotion Detection

CNN (Convolutional Neural Network) is used to classify the detected facial expression into one of seven predefined emotion classes. In this case, the Adam optimizer is used to optimize the CNN model. Adam optimizer is a popular optimization algorithm used in deep learning, especially in CNNs, due to its several advantages over other optimization algorithms. These advantages include adaptive learning rate, fast convergence speed, robustness, memory efficiency, and ability to handle sparse gradients.

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

D. Music Tracks are Recommended

If user feels that the emotion displayed in real time is consistent and correct, he needs to enter a specific keyboard key. After this, using a highly popular music platform like YouTube music, the search query is executed based on user's emotional state, singer's name and language. The search query gives the required music tracks.

IV. RESULTS AND DISCUSSION

Hence, we are able to successfully implement a system that uses Viola-Jones algorithm with the help of CNN to detect the user's emotional state depending on that the most appropriate list of music tracks will be recommended to the user. Figure 3 presents a screenshot of the system's separate window that opens after the user presses a specific key. This window detects and displays user's emotional state in real time.

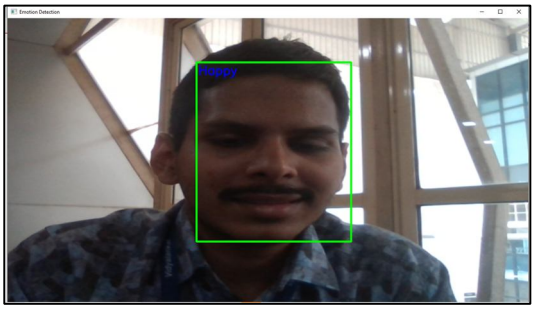


Figure 3: Separate window which shows user's emotional state in real-time

Our system was able to achieve an accuracy of 65.80 % using CNN algorithm with an epoch set to 50. Figure 4 shows the confusion matrix which we obtained while validating our system's algorithm on a test data containing 7178 images of 7 different classes (emotions).

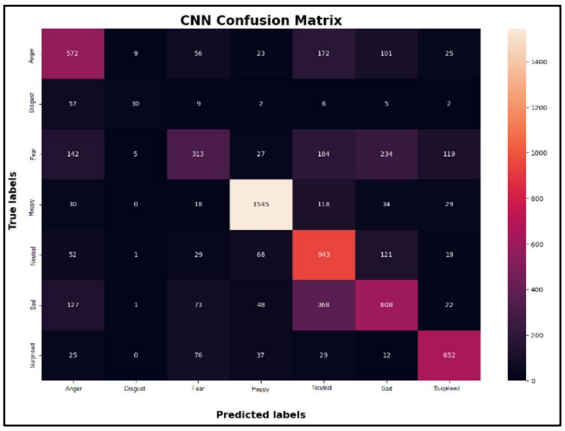


Figure 4: Confusion Matrix

V. FUTURE SCOPE

1) *Integration with Wearables and IoT Devices:* As wearable technology and IoT devices become more prevalent, integrating the emotion-based music recommender system with these devices could enhance the user experience. For example, a smartwatch that tracks the user's heart rate and detects their emotional state could be used to recommend music in real-time.
2) *Incorporating user Feedback:* The system could be further improved by incorporating user feedback and preferences. By allowing users to rate and provide feedback on recommended music tracks, the system could learn and adapt to the user's preferences over time.
3) *Expanding the Music Catalogue:* The system could be enhanced by expanding the music catalogue to include more diverse genres and styles of music. This would provide users with a wider range of options and increase the likelihood of finding music that resonates with their current emotional state.
4) *Multi-lingual Support:* The system could be expanded to support multiple languages, allowing users from different parts of the world to benefit from personalized music recommendations based on their emotional state.

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

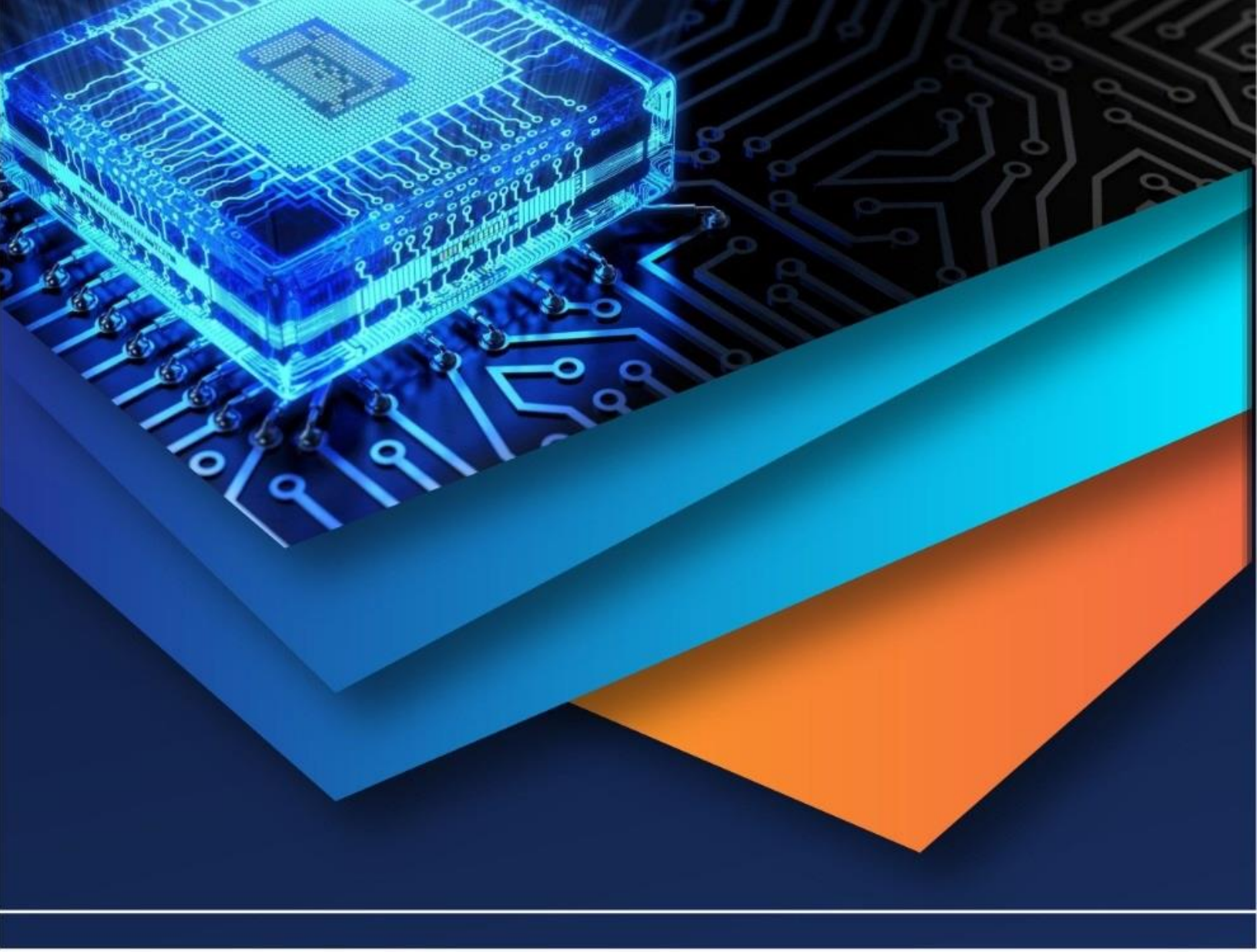
VI. CONCLUSION

Hence, we are able to create a system which is able to detect all 6 basic emotions – fear, disgust, sadness, anger, happiness and surprise. By incorporating an emotion detection system in our application, we are better able to understand the user's emotional state and provide more personalized and effective music recommendations. This could potentially lead to increased user satisfaction and engagement with the system, and ultimately, more successful outcomes in terms of user engagement and retention.

REFERENCES

[1] Mehrabian, "Communication without words," Communication theory, pp. 193-200, 2008.
[2] Cherry, K. (n.d.). The 6 types of basic emotions and their effect on human behavior. Verywell Mind. Retrieved from https://www.verywellmind.com/an-overview-of-the-types-of-emotions-4163976#:~:text=Primary%20emotions%20such%20as%20love,are%20known%20as%20tertiary%20emotions.
[3] 13 surprising uses for emotion AI technology. Gartner. (n.d.), from https://www.gartner.com/smarterwithgartner/13-surprising-uses-for-emotion-ai-technology
[4] Xue C, Li T, Yin S, Zhu X, Tan Y. The influence of induced mood on music preference. Cogn Process. 2018 Nov;19(4):517-525. doi: 10.1007/s10339-018-0872-7. Epub 2018 Jul 12. PMID: 30003367.
[5] M. Athavle, D. Mudale, U. Shrivastav and M. Gupta (2021) Music Recommendation Based on Face Emotion Recognition. Journal of Informatics Electrical and Electronics Engineering, Vol. 02, Iss. 02, S. No. 018, pp. 1-11, 2021.
[6] A. Alrihaili, A. Alsaedi, K. Albalawi and L. Syed, "Music Recommender System for Users Based on Emotion Detection through Facial Features," 2019 12th International Conference on Developments in eSystems Engineering (DeSE), Kazan, Russia, 2019, pp. 1014-1019, doi: 10.1109/DeSE.2019.00188.
[7] S. Deebika, K. A. Indira and Jesline, "A Machine Learning Based Music Player by Detecting Emotions," 2019 Fifth International Conference on Science Technology Engineering and Mathematics (ICONSTEM), Chennai, India, 2019, pp. 196-200, doi: 10.1109/ICONSTEM.2019.8918890.
[8] D. Ayata, Y. Yaslan and M. E. Kamasak, "Emotion Based Music Recommendation System Using Wearable Physiological Sensors," in IEEE Transactions on Consumer Electronics, vol. 64, no. 2, pp. 196-203, May 2018, doi: 10.1109/TCE.2018.2844736.
[9] S. Gilda, H. Zafar, C. Soni and K. Waghurdekar, "Smart music player integrating facial emotion recognition and music mood recommendation," 2017 International Conference on Wireless Communications, Signal Processing and Networking (WiSPNET), Chennai, India, 2017, pp. 154-158, doi: 10.1109/WiSPNET.2017.8299738.
[10] R. Ramanathan, R. Kumaran, R. Ram Rohan, R. Gupta and V. Prabhu, "An Intelligent Music Player Based on Emotion Recognition," 2017 2nd International Conference on Computational Systems and Information Technology for Sustainable Solution (CSITSS), Bengaluru, India, 2017, pp. 1-5, doi: 10.1109/CSITSS.2017.8447743.
[11] Available at: https://www.kaggle.com/datasets/msambare/fer2013.

©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)