2022 Third International Conference on Intelligent Computing Instrumentation and Control Technologies (ICICICT) | 978-1-6654-1005-2/22/$31.00 ©2022 IEEE | DOI: 10.1109/ICICICT54557.2022.9917956

Music Recommendation System Based on Facial Mood Detection

1st Madhav Lahoti
Computer Engineering
Sardar Patel Institute of Technology
Ichalkaranji, India
madhav.lahoti@spit.ac.in

2nd Sanchit Gajam
Computer Engineering
Sardar Patel Institute of Technology
Nagpur, India
sanchit.gajam@spit.ac.in

3rd Aditya Kasat
Computer Engineering
Sardar Patel Institute of Technology
Mumbai, India
aditya.kasat@spit.ac.in

4th Nataasha Raul
Computer Engineering
Sardar Patel Institute of Technology
Mumbai, India
nataasharaul@spit.ac.in

***Abstract*—A person's mood is mainly affected by the surrounding atmosphere and the type of company they keep around them. Its repercussions are seen in the person's day-to-day life, socially and career-wise, leading to a positive or negative mindset. Our proposed system recommends appropriate songs to the customers of their liking based on their current dominant mood to emphasize this idea. Here, the customer's face is detected using the OpenCV library, and different expressions are analyzed using CNN (Convolutional Neural Network) classifier and FER (Facial Expression Recognizer). In the end, appropriate songs get recommended to the customer based on their dominant mood detected. To make sure the songs recommended are of the customer's liking, the customer selects their playlists at the starting. The system's overall idea is to ensure that the customer gets to listen to their preferred songs to help them relate to their current feelings and help them reduce them to some extent by relaxing and stabilizing their emotions.**

***Index Terms*—OpenCV, Face Detection, CNN, Mood Analysis, Song Recommendation**

I. Introduction

Listening to music helps you increase bonding with your close family and friends, and it can play an important role in sharpening memory and increasing blood flow to the brain. Also, there was a study of cancer patients conducted in which the result stated that listening or playing to music helps reduce the anxiety level of the customer significantly. Another similar study showed that if a person uses music therapy, they feel more calm, relaxed, and less stressed [1]. In support of the above idea, a system detected the facial expression of the given person and extracted the current mood. To confirm that the customers get suggested their preferred songs when they log in for the first time, they are asked about their preferred languages and then asked to select their preferred playlists for the chosen languages. Thus while suggesting songs, the songs (according to the dominant mood detected) are extracted and displayed from the selected playlists only. The above method helps reduce the sadness/depression of the person and makes them calm and freshen their mindset. The application uses the Python3 language in the PyCharm editor. The front end used the Kivy framework. The image recognition process used the OpenCV library by plotting the vector points of the faces by the HAAR Cascade method [2]. The region of the image containing the face structure was resized to 48x48 and passed as input to the CNN. The boundaries for the system were the input of any number of playlists according to the user's liking, then taking seven emotions of facial expressions under consideration [3]. Then, the system will detect only one face at a time for the mood detection - if two or more faces come into the frame or if an obstacle like blocking the camera with a hand/paper, then the 3-second timer again runs till a perfect facial structure is detected [4]. There are certain assumptions made. Users will select the playlists which they like, then their facial structure will be recognized by the camera, and the valid expressions will be detected [5]. Thus the customer would be suggested songs from their playlists.

The entire application comprises three objectives. The first one is Face Detection. The facial structure is analyzed using the Haar Cascade algorithm (a machine learning-based algorithm) [6] along with the OpenCV library. It mainly focuses on removing the background noises and detecting the face even in dim light or nighttime and gives the proper detected facial structure and its facial properties for further analysis. The second one is analyzing the dominant mood from the detected face. [7] There are in total seven types of expressions – Happy, Sad, Neutral, Angry, Surprise, Fear, and Disgust. So, depending on the facial expression and its properties, the detected moods are categorized into percentage form, and the highest percentage output is taken into consideration to recommend the songs. [8] The last one is to suggest songs to the customer according to the dominant mood analyzed. Our database comprises different languages, playlists, and songs based on the expressions. So, depending upon the user's current mood, the system provides suitable and appropriate songs (from his playlist) to reflect their dominant mood. This, in turn, helps the customer to be engaged fully in our application and get what's best for them.

II. Literature Review

Face detection and image or video detection are popular topics for biometrics research. The main component used widely for this is the OpenCV library. This machine-learning algorithm has numerous uses in various fields like photo processing and real-time operation. The algorithm also suggests a PCA (Principal Component Analysis) [9] program for face recognition. PCA is the analysis part of the element and uses a mathematical way for the process. The total capacity of data retention needs the size of the space, which is then required to display the data economically. A two-dimensional facial image makes a vector (in pixel form) that is one-dimensional and integrated as a whole. The image uses PCA to detect the facial structure and use data of more than one variable quantities in the form of a comprised dataset of the variables [10]. This technique helped us detect the user's facial structure at that exact time with maximum efficiency to detect mood and recommended appropriate songs. There are seven classifications of facial expressions and ways to recognize positive and negative emotions using facial expressions and the development of road-based systems [11]. In a previous study, they used in-depth reading technology to create models with a facial expressions based on perceived emotions. There are applications available that use six emotions, but not seven. Therefore, the seven emotions identified as anger, disgust, fear, joy, sadness, surprise, and neutrality [12] are also divided the listed points of emotional awareness into positive and negative emotions. We used these seven emotions as the reference to classify the detected mood in descending order and take the dominant expression under consideration. A person's face has unusual and straightforward features, so it is difficult to understand and identify facial features. Currently, research on facial expressions is characterized by seven emotions as mentioned above [13]. This paper aims to detect the crucial facial features (like eyes and lips) and then classify them under seven different emotions. This information is passed through a series of filters and processes, identified by a Vector Support Machine (SVM), filtered using Grid Search [14]. The algorithm helps to understand the rules used for class or regression problems. It uses a method referred to as the kernel trick to convert your facts-based totally on those differences and reveals the most efficient boundary among the viable outputs. The test data scans the data and their labels and provides the accuracy of the test data classification in the split report. We used this reference and applied a similar technique using CNN (Convolutional Neural Network) [15] algorithm to recognize and classify emotions from the detected face of the user by predicting the facial expression label and considered as one of the seven moods. Detailed analysis of face recognition techniques using face recognition techniques [16] takes into account the diversity, benefits, disadvantages, and accuracy of different methods [17]. The CNN algorithm requires input into the system in the form of an image to predict the facial expression of the input image in the form of different labels, referred to as emotions like happiness, neutral, anger, disgust, fear, sadness, etc. Thus, recognizing a person's face is an easy task, but doing the same job with a computer is challenging since a person's face will occur if there is a change in face image like complex background or occlusion [18]. This paper gives a review of include-based face acknowledgment procedures. As a biometric innovation, computerized face acknowledgment has various helpful properties using valid methods. With the advancements in the information medium, regardless of any limitation like blurred faces, invalid faces are recognized [19]. Music Information Recovery aims to calculate methods used to help manage and retrieve music data of large datasets [20]. This study uses attitudes to separate music data is logical and confusing. The main objective of this study is to determine by evidence which dividers and combinations of factors work best in classifying emotional states in music data. We used this inference by giving the user freedom to choose their preferred languages and playlist at the starting, thereby recommending appropriate songs from the selected playlist database.

III. Proposed System

The entire system flow is divided into five sub-sections, comprising flow diagrams. Section A) shows the proposed system flow with the basic introduction of each sub-module. Section B) displays the Face Detection and the Mood Analysis based on the detected facial expression. Section C) represents the flow of Mood Statistics of detected Mood in the form of Pi-Chart and Bar Graph. Section D) represents the flow of Favorites Song and the Most Played Songs List.

A. Proposed Flow Diagram

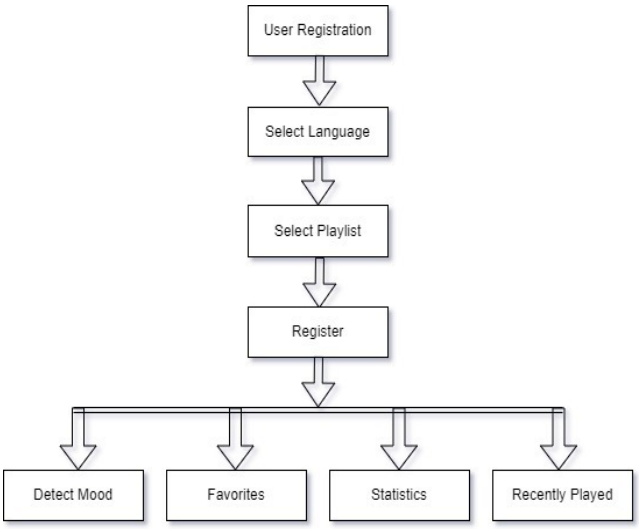


Fig. 1: Proposed Flow Diagram

Fig. 1 represents the entire system flow. The user has to first register into the system. The user has to enter their username (of length less than ten characters) and then select one or more of their preferred languages (from English, Hindi, Marathi, and Punjabi). Then, the user needs to choose one or more

of their preferred playlists from the selected languages. After completing the initial procedure, the user will have access to different application features like - a) Detect their facial mood and get appropriate songs recommended. b) Favorite Song List c) Upto date Mood statistics with Pi-Chart and Bar-Graph. d) Most Recently Played Songs List.

B. Detect Mood

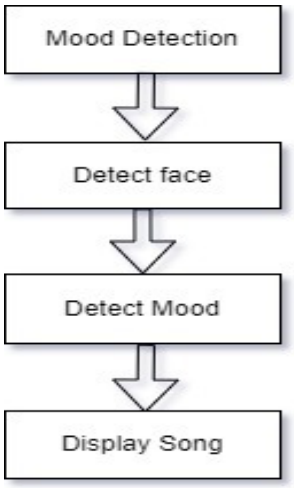


Fig. 2: Detect Mood

Fig. 2 is the block diagram for Detect Mood module. While detecting mood, the user will show their face on the screen. The 3-second timer starts immediately after the user opts to analyze their face. Input frames should not contain multiple faces or 0 faces. If any obstruction is found (like multiple faces, blocking with hands), the 3-second timer will run in a loop until a valid facial structure is detected. Thus, it ensures analyzing the expression with maximum efficiency. Then, percentages of all detected moods in descending order are displayed, and finally, all the preferred songs are displayed, according to the mood's score.

C. Mood Statistics

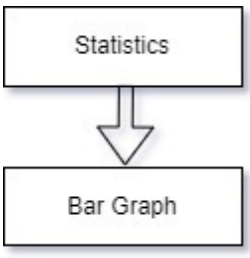


Fig. 3: Mood Statistics

Fig 3. is the block diagram for Statistics. The Mood Statistics uses Pi-Chart (with color code for reference) and a Bar-Graph to plot the frequency of all the seven moods of the user captured until that moment. In this manner, the user can keep track of their mood swings from the beginning until that point. Users can directly go to the Statistics page from the home page and check the mood statistics bar diagram based on the mood history of the user.

D. Favorites and Most Played Song List

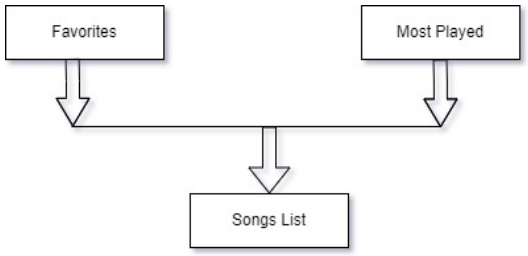


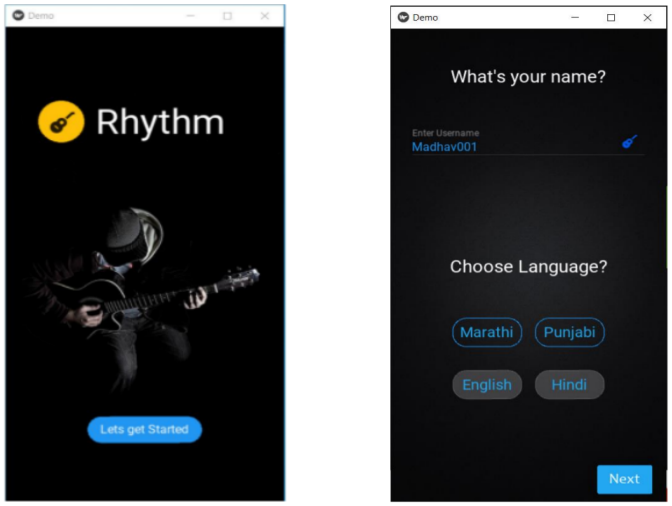
Fig. 4: Favorite Song List

Fig 4. is the block diagram for Favorites and Most Played Songs. The Favorite Song List is all the songs the user had liked until then. The user can dislike any song to remove it from the list. The Most Recently Played Song List is all songs the user had played until that point. It gives the user easy and friendly access to hear their preferred songs. This feature provides the customer easy access to hear their preferred songs. This feature provides the customer easy access to their favorite songs whenever required.

IV. Implementation

This section contains the actual implementation of our proposed system. There were certain assumptions and constraints in carrying out the system's working efficiently. The assumptions made were as follows. 1. The user must have access to a webcam on their application to get their face detected. It ensures that the user has proper conditions to get their face analyzed for mood detection, which is the main crux of the system. 2. The user must be familiar with at least one of the languages (English, Hindi, Punjabi, and Marathi) and their playlists to get recommended appropriate songs after mood detection. 3. The user's facial structure must be in proper shape for their mood to get detected in a valid manner. Thus, there are four languages, each having ten playlists with five songs each for seven moods, which counts to 4*10*5*7 = 1400 songs stored in our database. There is a large dataset for the songs to provide the customers wider range for selecting their preferred playlists. The constraints imposed in our system are as follows. 1. Only one user can log in at a given time into the system. It ensures that only one user is active at any given period. 2. The user must have access to the webcam on the device they are operating to get the current mood detected. Else, if the face of the user is not detected, the mood will not get analyzed and no songs would get suggested to the user. 3. The user must present their face in front of the webcam within the time limit of three seconds. If failed due to some technical reason or any other, the timer again runs for three seconds in a loop until a proper user's face is detected.
The proposed system was made in Tkinter Framework in the PyCharm Framework Editor, using Python as the programming language and MySQL for the database.

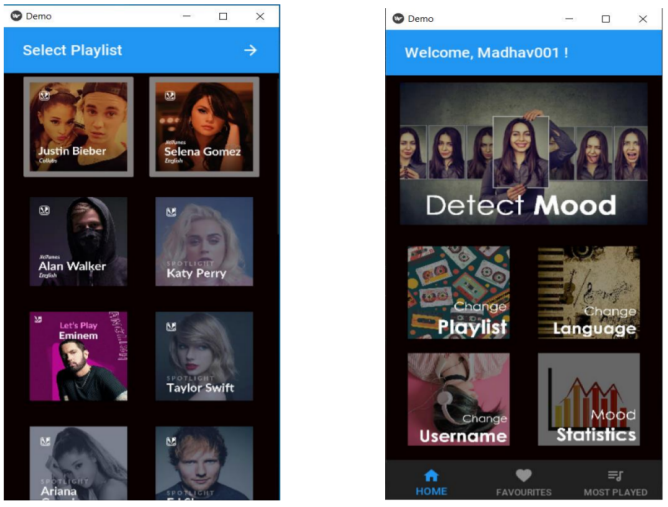
978-1-6654-1005-3/22/$31.00 ©2022 IEEE



(a) Starting Page (b) Filling User Details

Fig. 5: Home Page and User Registration

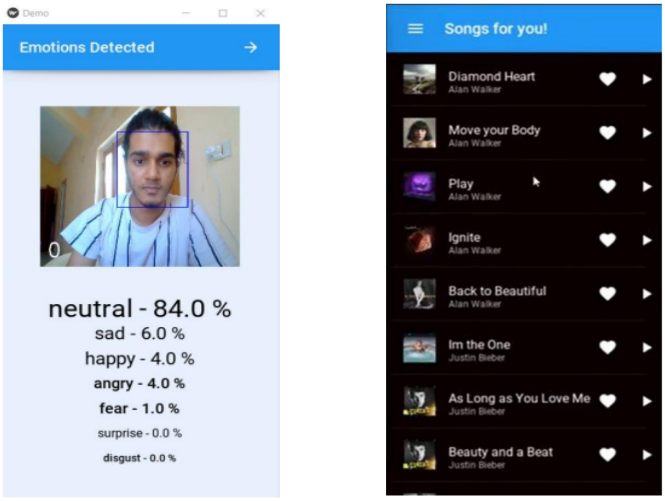
Fig. 5a is the initial home page of the system where the new user logs in for the first time into the system. The user enters a valid username (length of maximum 10, and starting with an alphabet) and selects their preferred languages (from English, Hindi, Marathi, and Punjabi), as shown in Fig. 5b. The user has to fill in all the details, or a warning message will be displayed.



(a) Selecting Playlist (b) Main Home Page

Fig. 6: Selecting Playlist and Main Home Page

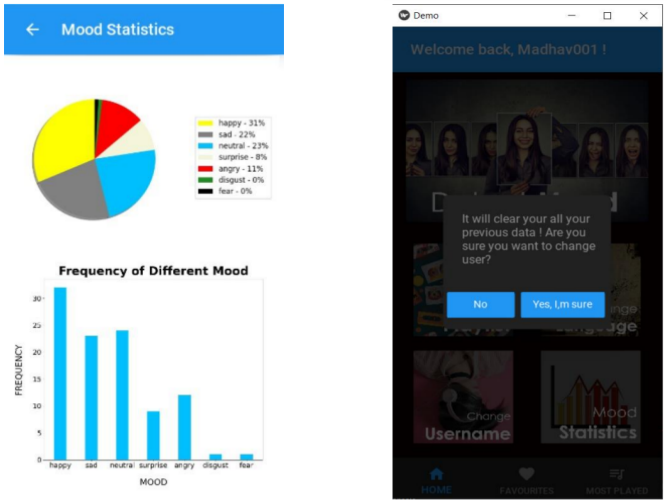
After filling in these details, the user selects their preferred playlists (at least one) based on the languages, as shown in Fig. 6a. The user can scroll down to access all the playlists. If the user does not select any playlist, a dialog box with the warning is displayed. Then, the main home page is displayed (Fig. 6b), which grants access to all main features, which will be explained later in detail.



(a) Facial mood detected (b) Recommending songs

Fig. 7: Facial Mood Detection and Songs Recommendation

The user can select the Detect Mood button to detect their current facial mood within 3 seconds. The webcam locks in each frame and the facial structure with the Haar Cascade algorithm and OpenCV library is analyzed. If obstruction (like blocking with hands or objects) occurs in the webcam, the timer runs for three seconds in a loop until a valid face is detected. Then, each mood's percentages are displayed in descending order (as shown in Fig. 7a). The dominant expression to recommend the songs, as shown in Fig. 7b, is considered. The user can also like and play the songs and access them from the home page under the My Favorites and the Most Played Songs section.



(a) User's Mood Statistics (b) Erasing User Data

Fig. 8: User's Mood Statistics and Erasing User Data

978-1-6654-1005-2/22/$31.00 ©2022 IEEE

The final mood statistics for all the emotions detected by that user are displayed as shown in Fig. 8a. All the moods are analyzed using the Pi-Chart (color code for reference is present). The Bar Graph represents the Frequency of the Mood (the x-axis represents all the seven moods, and the y-axis represents the frequency scale from 0 to 1, using the MatPlotLib Library. If the user wants to change their username, they can do so by going to the home page and clicking on the Change Username button. On doing it, a confirmation dialog box is displayed, as shown in Fig. 8b, stating that the user's entire data will get erased on creating a new user. On proceeding, the user will get directed to the page as shown in Fig. 5b, where the new user can enter his details and continue the same process all over again.

V. Results

We have gathered the results into categories to fulfill the main motive of our proposed system - 1. To detect the face without any obstruction. 2. Properly extract the emotions and classify them under the seven emotions in descending order. 3. Taking the dominant emotion under consideration, suggesting the customer appropriate songs to relate to their mood in a better way. The overall result is calculated in the form of a Bar-Graph and its accuracy is calculated accordingly.

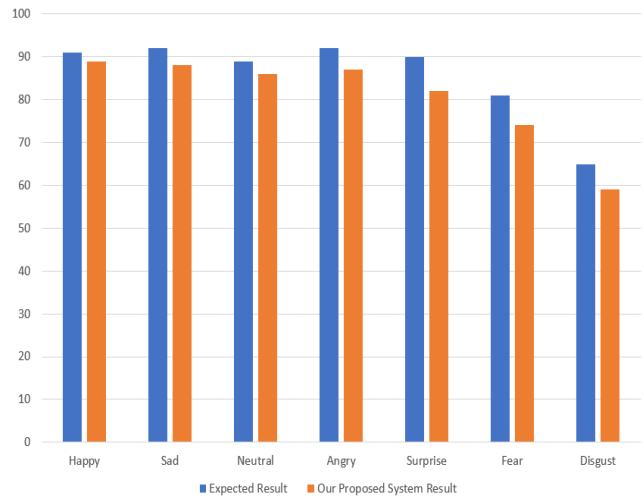


Fig. 9: Expected vs Calculated Mood Graph

Fig. 9 represents the graph of expected (blue bar) v.s calculated (orange bar) mood analysis of 100 individuals taken for testing. There were 50 males and 50 females for testing and were under the age group of 15 - 80 years to test the efficiency of our proposed system for all age groups. The face uses the HAAR Cascade method, which extracts every feature by summing up the pixels beneath the white rectangle by adding the pixels present under the rectangle in black. Here, positive images contain the images that need to be identified by our system. The negative images are the ones that obstruct the required input, which we do not intend our system to detect. The result obtained for the emotion analysis is calculated by CCN (Convolutional Neural Network), which creates feature maps that record a region of the image.

Moods	Expected Result	Calculated Result	Accuracy
Happy	91	89	97.80%
Sad	92	88	95.65%
Neutral	89	86	96.63%
Angry	92	87	94.56%
Surprise	90	82	91.11%
Fear	81	74	91.35%
Disgust	65	59	90.77%

Fig. 10: Mood Accuracy Table

The performance metric used to evaluate our proposed system is accuracy. Fig. 10 represents the accuracy percentage (Calculated/Expected) of all the seven moods, and the average accuracy calculated is 93.98%.

VI. Conclusion

We managed to develop a system that recognizes the person's facial expression and thereby detects the mood/emotion to suggest the appropriate list of songs to uplift the person's spirit. The user can log out of the account by changing their username and can edit their language and playlist preference according to their needs. We have added four languages, each containing 350 songs, for the user to get recommended the songs of their interest. We have managed to overcome the constraints for face detection, by setting the timer again for 3 seconds in a loop until a clear face is detected. We also provided an extra feature of giving the user access to their Favorite Songs (the songs they had liked previously) and the Most Played Songs (the songs they had played previously) to provide ease of access to the user to listen to their preferred songs. Throughout this process of creating the system, we learned about different algorithms like TensorFlow, KivyMD, OpenCV, FER, and MatPlotLib to achieve our main goal and thereby help them to get through their current situation in a better way.

VII. Future Scope

The future scope of our system would be the actual song playing implementation when it is paused or played. For this, there would be a need to get a license from the music company to play their songs which would take some time, thus the future scope. We would include a Customer Care service for the customer to complain or suggest adding some songs/playlists of their liking to the existing data. It can be done through an email with a response time within 24 hours, helping us improve our relationship with the customers and increasing the credibility of our system. Based on the customer's mood, after using the system for some time, we would include some podcasts and motivational talk related to the dominant

978-1-6654-1005-3/22/$31.00 ©2022 IEEE

expressions for the user to listen to and hopefully make them feel relaxed and comforted, according to their current mood. As the proposed system currently runs on a Desktop version, we would also implement it for Android and IOS devices to provide ease of access and usability to our customers.

REFERENCES

[1] A. Baharum, T. W. Seong, N. H. M. Zain, N. M. M. Yusop, M. Omar and N. M. Rusli, "Releasing stress using music mood application: DeMuse," 2017 International Conference on Information and Communication Technology Convergence (ICTC), 2017, pp. 351-355, doi: 10.1109/ICTC.2017.8191001.

[2] D. Tyas Purwa Hapsari, C. Gusti Berliana, P. Winda and M. Arief Soeleman, "Face Detection Using Haar Cascade in Difference Illumination," 2018 International Seminar on Application for Technology of Information and Communication, 2018, pp. 555-559, doi: 10.1109/ISEMANTIC.2018.8549752.

[3] Singhal, Prateek, et al. "A Survey: Approaches to Facial Detection and Recognition with Machine Learning Techniques." Proceedings of Second Doctoral Symposium on Computational Intelligence. Springer, Singapore, 2022.

[4] Arora, Mehul, Sarthak Naithani, and Anu Shaju Areeckal. "A web-based application for face detection in real-time images and videos." Journal of Physics: Conference Series. Vol. 2161. No. 1. IOP Publishing, 2022.

[5] M. Azimi, "Effects of Facial Mood Expressions on Face Biometric Recognition System's Reliability," 2018 1st International Conference on Advanced Research in Engineering Sciences (ARES), 2018, pp. 1-5, doi: 10.1109/ARESX.2018.8723292.

[6] Choi, Cheol-Ho, et al. "Face Detection Using Haar Cascade Classifiers Based on Vertical Component Calibration." HUMAN-CENTRIC COMPUTING AND INFORMATION SCIENCES 12 (2022).

[7] N. M. Hakak, M. Mohd, M. Kirmani and M. Mohd, "Emotion analysis: A survey," 2017 International Conference on Computer, Communications and Electronics (Comptelix), 2017, pp. 397-402, doi: 10.1109/COMPTELIX.2017.8004002.

[8] Khan, Maliha, et al. "Face Detection and Recognition Using OpenCV." 2019 International Conference on Computing, Communication, and Intelligent Systems (ICCCIS). IEEE, 2019.

[9] S. Sehgal, H. Singh, M. Agarwal, V. Bhasker and Shantanu, "Data analysis using principal component analysis," 2014 International Conference on Medical Imaging, m-Health and Emerging Communication Systems (MedCom), 2014, pp. 45-48, doi: 10.1109/MedCom.2014.7005973.

[10] I. Culjak, D. Abram, T. Pribanic, H. Dzapo and M. Cifrek, "A brief introduction to OpenCV," 2012 Proceedings of the 35th International Convention MIPRO, 2012, pp. 1725-1730.

[11] Lee, Hyeon-Jung, and Kwang-Seok Hong. "A study on emotion recognition method and its application using face image." 2017 International Conference on Information and Communication Technology Convergence (ICTC). IEEE, 2017.

[12] M. Xiaoxi, L. Weisi, H. Dongyan, D. Minghui and H. Li, "Facial emotion recognition," 2017 IEEE 2nd International Conference on Signal and Image Processing (ICSIP), 2017, pp. 77-81, doi: 10.1109/SIPROCESS.2017.8124509.

[13] C. Jain, K. Sawant, M. Rehman and R. Kumar, "Emotion Detection and Characterization using Facial Features," 2018 3rd International Conference and Workshops on Recent Advances and Innovations in Engineering (ICRAIE), 2018, pp. 1-6, doi: 10.1109/ICRAIE.2018.8710406.

[14] Yujun Yang, Jianping Li and Yimei Yang, "The research of the fast SVM classifier method," 2015 12th International Computer Conference on Wavelet Active Media Technology and Information Processing (ICCWAMTIP), 2015, pp. 121-124, doi: 10.1109/ICCWAMTIP.2015.7493959.

[15] R. Chauhan, K. K. Ghanshala and R. C. Joshi, "Convolutional Neural Network (CNN) for Image Detection and Recognition," 2018 First International Conference on Secure Cyber Computing and Communication (ICSCCC), 2018, pp. 278-282, doi: 10.1109/ICSCCC.2018.8703316.

[16] Mahmud, Tanjim, et al. "Face Detection and Recognition System." Intelligent Computing and Innovation on Data Science. Springer, Singapore, 2021. 145-155.

[17] R. Subban and S. Soundararajan, "Human face recognition using facial feature detection techniques," 2015 International Conference on Green Computing and Internet of Things (ICGCIoT), 2015, pp. 940-947, doi: 10.1109/ICGCIoT.2015.7380598.

[18] K. M. Malikovich, I. S. Z. Ugli and D. L. O'ktamovna, "Problems in face recognition systems and their solving ways," 2017 International Conference on Information Science and Communications Technologies (ICISCT), 2017, pp. 1-4, doi: 10.1109/ICISCT.2017.8188594.

[19] Kumar Shukla, Ratnesh, and Arvind Kumar Tiwari. "Comparative Analysis of Machine Learning Based Approaches for Face Detection and Recognition." Journal of Information Technology Management 13.1 (2021): 1-21.

[20] H. Shahmansouri and J. Z. Zhang, "An empirical study on mood classification in music through computational approaches," 2016 3rd International Conference on Systems and Informatics (ICSAI), 2016, pp. 1050-1055, doi: 10.1109/ICSAI.2016.7811106.