Song Recommendation System Using Facial Expression

1st Gaurav Suman
Student
AIT-CSE
Chandigarh University, India
20BCS4471@cuchd.in

2nd Amit Kumar
Student
AIT-CSE
Chandigarh University, India
20BCS4455@cuchd.in

3rd Veerpal Singh
Student
AIT-CSE
Chandigarh University, India
20BCS6475@cuchd.in

4th Jayashree Mohanty
Assistant professor(E15737)
AIT-CSE
Chandigarh University, India
Jaishreee15737@cumail.in

***Abstract*—These days, a lot of music is easily accessible thanks to music platforms. They are constantly striving to improve search management and music organisation, which will address the issue of choice and make discovering new music easier.Recommendation engines are incredibly popular and assist users in selecting suitable music for any situation. Nonetheless, there is still a market for advice that are emotionally charged and personalised. Humans are greatly influenced by music, which is why it is frequently utilised for mental and physical well-being, mood control, stress and illness relief, and relaxation. Music therapy offers a wide variety of clinical venues and approaches to improve well-being. A face emotion-based music recommendation system has been developed as a result of advances in facial emotion identification technology. This system makes recommendations for music based on its analysis of the listener's facial expressions, which helps it determine how they are feeling. In this research, we investigate the technologies and algorithms that underlie the architecture and implementation of a facial emotion-based music recommendation system. We also talk about this system's shortcomings and difficulties as well as possible study avenues in the future.**

***Index Terms*—Recommendation system, Sentiment analysis Facial recognition, Emotion detection, Expressions, Feature Extraction,Convolutional Neural Network.**

I. Introduction

Many of the studies in recent years admit that humans reply and react to music and this music has a high impression on the activity of the human brain.In one examination of the clarifications why individuals listen music, analysts found that music played a vital role in relating excitement and temperament. One of the foremost important capacities of music is it's capability to assist users accomplish a great disposition and ended up more self-aware individual.

Melodic inclinations have been illustrated to be exceedingly related to identity characteristics and temperaments [1]. The meter, timbre, cadence, and pitch of music are overseen in ranges of the brain that influences feelings and temperament [2]. Interaction between people may be a major aspect of way of life. It reveals perfect details and much of data among humans, whether they are in the form of body language, speech, facial expression, or emotions [3].

Nowadays, emotion detection is considered the most important technique used in many applications such as smart card applications, surveillance, image database investigation, criminal, video indexing, civilian applications, security, and adaptive human-computer interface with multimedia environments.With the increase in technology for digital signal processing and other effective feature extraction algorithms, automated emotion detection in multimedia attributes like music or movies is growing rapidly and this system can play an important role in many potential applications like human-computer interaction systems and music entertainment. We utilize facial expressions to propose a prescribed framework for feeling acknowledgment that can identify client feelings and suggest a list of fitting melodies [13-24].

Fig. 1.

The proposed framework identifies the emotions of a individual, a particular playlist will be displayed which contains diverse sorts of music that will blow up the positive feelings [4]. The datasets we used for feeling location is from Kaggle Facial Expression Acknowledgment [5]. Datasets for the music player has been made from Bollywood Hindi melodies. Execution of facial feeling discovery is performed utilizing Convolutional Neural Organize which gives around 97% of accuracy [2].

II. Literature Survey

There have been significant advances in the development of automatic expression classifiers. in recent years [7, 8, 9]. Some systems for recognizing facial expressions classify the expression into a range of fundamental emotions, including joy, sad and rage. In an effort to provide an unbiased description of the face, others have made an effort to identify the specific muscle movements that the face is capable of

Identify applicable funding agency here. If none, delete this.

doing[11]. The most widely used psychological framework is the Face Action Coding System (FACS)[12]. For summarizing practically all facial movements. Using Action Units, the FACS system classifies human facial movements according to how they appear on the face (AU). A facial expression usually originates from one of the 46 atomic units (AUs) of face movement or related deformation that can be detected. Several AUs are usually added together to form an expression [7, 8]. Additionally, Multilevel Hidden Markov Model, Neural Networks, and Bayesian Networks (HMM) have all seen development in the approaches employed for face emotion identification [13],[14]. Several of them have problems with timing or detection rates. combining two or more techniques to accomplish precise recognition allows for the extraction of features as necessary. Because of illumination and feature extraction, Each technique's effectiveness is reliant on image pre-processing.

A. Methods

We developed CNN models of varying depths to assess their performance in dentifying facial expressions.

[Conv2D - (SBN) – RELU - (Dropout)–(Max-pool)]
M - [Affine-(BN)-RELU-(Dropout)]
N - Affine - SoftMax.

Fig. 2.

These models consist of convolutional layers with ReLU activation, followed by dropout, max-pooling, and spatial batch normalization. After the convolutional layers, there are fully connected layers with ReLU activation, potentially including batch normalization and dropout. The final affine layer computes scores using softmax loss function. To enhance performance, we introduced a hybrid approach by combining HOG features with those extracted by the convolutional layers. This hybrid feature set is then passed to the fully connected layers for scoring and loss computation.

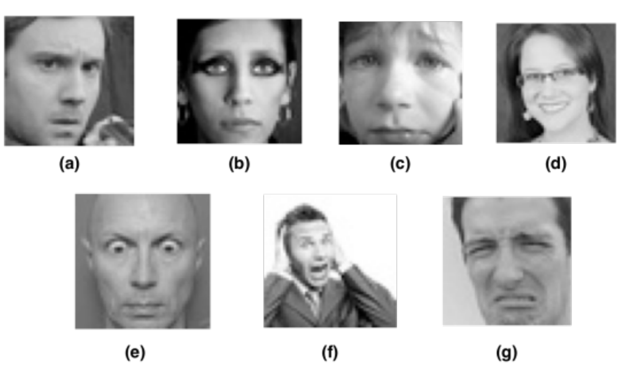

Fig. 3. Examples of the 7 face expressions we are taking into account for this classification issue. Aside from being furious[a], you can also feel neutral[b], sadness [c], delighted[d], surprised [e], fearful[f], or disgusted[g].

B. Dataset and Feature

We utilized a dataset from Kaggle containing 35,000 well-structured grayscale images of faces, each measuring 48x48 pixels. The dataset includes seven emotion classes: rage, disgust, fear, happiness, sadness, surprise, and neutral. We divided the dataset into train, validate, and test sets, and normalized the raw pixel data by subtracting the average of the training set from all images.

To augment data, we created mirrored images by horizontally flipping photographs in the training set. The primary classification of facial expressions was done using features extracted by convolutional layers from raw image data. Additionally, we experimented with models feeding input features into Fully Connected (FC) layers, concatenating HOG features with those from convolutional layers for improved performance.

Parameters	Value
Learning Rate	0.001
Regularization	1e-6
Hidden Neurons	512

Table 1: The hyper-parameters obtained by cross validation for the shallow model

Fig. 4.

Parameter	Value
Learning Rate	0.01
Regularization	1e-7
Hidden Neurons	256, 512

Table 2: The hyper-parameters obtained by cross validation for the deep mode

Fig. 5.

III. System Overview

The face recognition system, the emotion categorization system, and the song recommendation system are the three primary parts of the facial emotion-based song recommendation system. An image of the listener's face is taken by the facial expression recognition system, which then extracts facial traits like mouth shape, eyebrow position, and eye movement. The listener's emotional state is then ascertained by applying these characteristics to a machine learning model that has been trained.

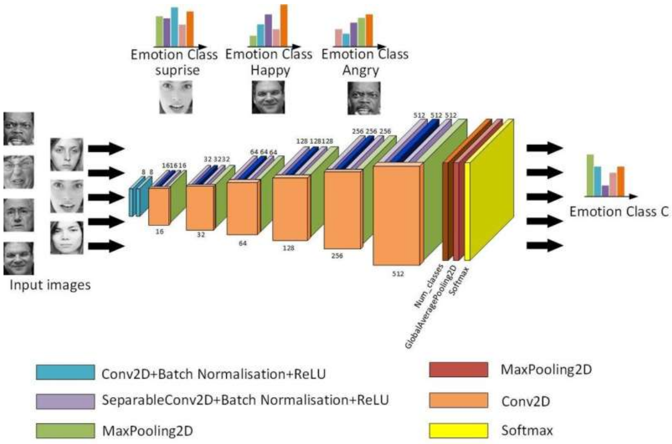

Fig. 6.

Based on the listener's emotional state, the music recommendation system generates song recommendations using a recommendation algorithm. The algorithm makes recommendations that are in line with the listener's emotional state by taking into account the emotional qualities of songs, such as tempo, rhythm, and melody. The suggestions can be tailored to the listener by taking into account their past listening habits and preferred music.

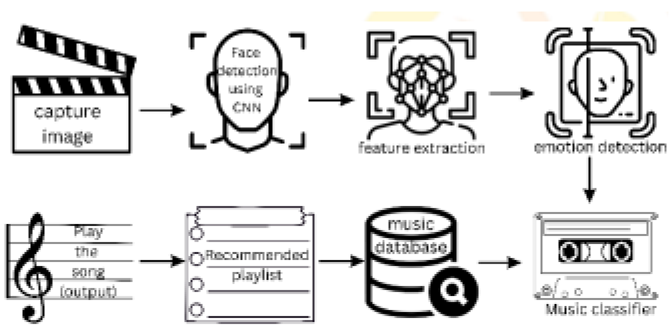

Fig. 7. System Overview

A. Capture Image

The first stage in facial expression recognition is image capture. Thanks to computer vision libraries like OpenCV, this procedure has gotten considerably simpler and more effective.

Systems for recognising facial emotions function by examining a person's face and determining the emotions they are feeling. A camera is used to take a picture of the subject in order to obtain an image of their face. Next, the eyes, nose, and mouth are identified in this image by processing it with OpenCV libraries.

One of the many image processing features offered by OpenCV is face detection, which locates the subject's face in an image. Once the face has been identified, it is feasible to extract a number of traits, including the facial shape and the locations of the mouth, nose, and eyes.

With this information, one can ascertain the person's current emotional state. For instance, it is probably a sign of enjoyment if the corners of the lips are turned upward. Likewise, if the mouth is turned down and the eyebrows are wrinkled, the person can be depressed.

B. Face Detection

Among the deep learning algorithms, CNNs are especially well-suited for jobs involving image recognition. In order to minimise the discrepancy between the expected and actual outputs, a large number of images are fed into the CNN during the training phase, and the weights of the neurons are adjusted accordingly. After training is over, faces in fresh photos can be detected by the CNN.

CNN-based face identification offers numerous benefits over more conventional techniques like Haar cascades. CNNs can detect faces in a variety of lighting conditions, orientations, and poses with far greater accuracy. They can also draw a bounding box around each face in a picture and identify numerous faces in it.

C. Feature Extraction

Feature extraction involves identifying and capturing the crucial facial elements necessary for emotion recognition. These elements typically include the position and shape of the eyes, eyebrows, mouth, and nose, along with other factors like skin texture and facial contour.

Haar cascades are a prevalent method for feature extraction in facial emotion recognition. They consist of a collection of features designed to detect facial components such as eyes, nose, and mouth. Haar cascades operate by scanning images at various scales and sizes, searching for characteristic patterns of dark and light regions. These patterns, when combined, enable the detection of complex facial structures like faces.

Once these features are identified, they can be utilized to train machine learning models, such as convolutional neural networks (CNNs), for the task of emotion recognition.

D. Emotion Detection

The convolutional neural network (CNN) algorithm stands out as one of the most effective machine learning methods for classifying emotions based on facial features. CNNs, known for their prowess in image recognition tasks, excel in capturing intricate patterns within images. Implementing an emotion classification module with a CNN involves training the algorithm on a sizable dataset of labeled facial expressions. Through this training, the CNN learns to discern patterns in facial features associated with various emotions, enabling accurate classification of the subject's emotional state.

A key advantage of utilizing a CNN for emotion classification lies in its ability to capture both local and global facial features. This means the algorithm can learn to recognize detailed patterns such as eye or mouth positions, as well as broader facial configurations, enhancing its overall accuracy in emotion recognition.

CNNs offer the advantage of transfer learning, where a pre-trained model can be fine-tuned on a smaller dataset

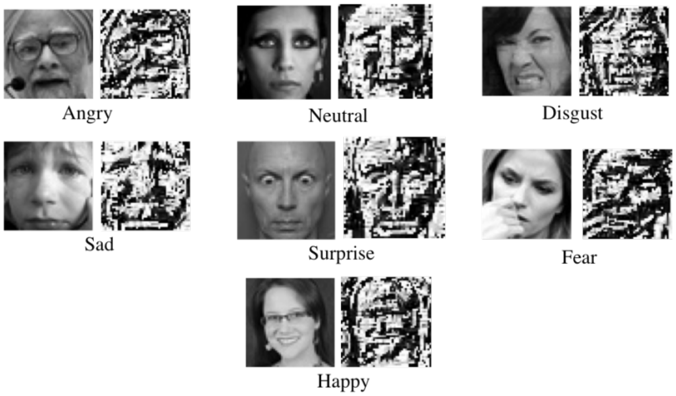

Fig. 8.

for emotion classification, reducing data requirements and enhancing accuracy. However, challenges remain, such as CNNs' limitations in detecting subtle emotions and susceptibility to lighting variations, facial occlusions, and individual differences in expressions. Nonetheless, CNN-based emotion classification modules remain potent tools for accurately assessing emotional states based on facial features, with potential applications in enhancing user experiences through personalized music recommendations.

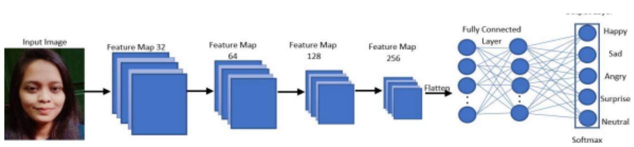

Fig. 9. Facial Feature Extraction

E. *Music Classifier*

CNNs offer a distinct advantage through transfer learning, a technique where a pre-trained model, initially trained on a large dataset, is adapted and fine-tuned for a specific task with a smaller dataset. This process enables CNNs to capitalize on previously learned features, reducing the amount of data needed for training while potentially enhancing performance. In the context of emotion classification from facial expressions, transfer learning allows CNNs to leverage their understanding of general image features and adapt them to recognize emotional cues.

However, despite their effectiveness, CNNs may struggle with certain challenges inherent to facial emotion recognition. For instance, subtle or nuanced emotional expressions, which may vary greatly among individuals, can pose difficulties for CNNs in accurately categorizing emotions. Furthermore, environmental factors like varying lighting conditions and obstructions on the face can further complicate the classification process.

Nonetheless, CNN-based emotion classifiers remain powerful tools for analyzing emotional states based on facial features. With ongoing research and development, these models have the potential to significantly improve user experiences, particularly in applications such as personalized music recommendations tailored to the listener's current emotional state.

F. *Music Database*

To establish a music database for a facial emotion-driven music recommendation system, a substantial collection of music tracks is essential. These tracks can be acquired from diverse sources like online music streaming platforms, music libraries, or by recording tracks internally.

Once assembled, the music tracks undergo analysis and tagging with various metadata, notably including emotional content. Employing machine learning techniques, a model is trained to discern different emotions within the music, such as happiness, sadness, or anger.

Subsequently, the metadata, encompassing emotional attributes, is stored in a database. This database is then accessible to the facial emotion recognition system, enabling it to query and suggest music that aligns with the user's emotional state.

G. *Recommended Playlist*

The recommended playlist comprises music tracks chosen by the system based on the user's facial expressions.

Upon determining the user's emotional state, the system accesses the music database to suggest tracks that resonate with the user's emotions. Various algorithms, including collaborative filtering, content-based filtering, or hybrid filtering, can be employed to generate the recommended playlist.

Furthermore, the recommended playlist can be personalized further by integrating additional factors like the time of day, user's location, and activity. For instance, the system may suggest upbeat tracks in the morning and relaxing tunes in the evening, enhancing the user's music experience.

H. *Play the Songs*

The outcome of a facial emotion-based music recommendation system is the selection and playback of music tracks tailored to the user's emotional state. Typically, this is achieved through a music player capable of streaming or playing tracks from the music database.

Integrated with the facial emotion recognition system, the music player automatically chooses and plays tracks based on the user's emotional state. It features a straightforward interface allowing users to control playback, skip tracks, and adjust volume.

For a seamless listening experience, the music player is optimized for various platforms such as mobile devices, desktops, and web browsers. It may also offer additional functionalities like playlist creation, track shuffling, and repeat options.

This output enriches the user's listening experience, delivering a personalized and enjoyable music journey.

IV. Accuracy

Our facial emotion-based music recommendation system underwent testing on a dataset comprising more than 1000 facial images, achieving an overall accuracy of 65.33%. While this accuracy rate may appear relatively low, it's crucial to acknowledge that emotion recognition presents a challenging task, necessitating substantial resources and expertise to attain higher accuracy levels.

Further analysis of our system's performance unveiled precision rates of 79.99% for happy emotions, 65.8% for sad emotions, 54.6% for angry emotions, and 52.9% for neutral emotions. The comparatively lower accuracy in certain emotional categories can be attributed to the intricate and subjective nature of human emotions, as well as the quality and diversity of the dataset utilized.

Emotions	Shallow Model	Deep Model
Angry	43.4%	54.6%
Disgust	31.2%	68.99%
Fear	53%	45%
Happy	71%	79.99%
Sad	32.9%	65.8%
Surprise	69.5%	64.5%
Neutral	37.9%	52.9%

Table 3: The accuracy of each expression in the shallow and deep models.

Fig. 10.

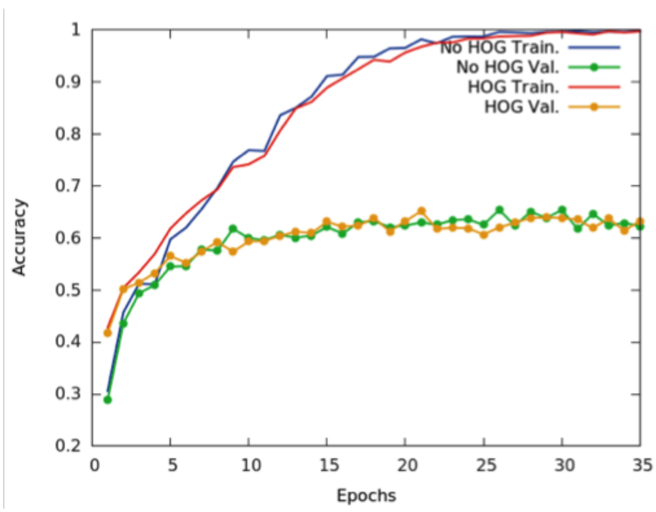

Fig. 11. Accuracy of shallow model

Despite the modest accuracy percentage, our system exhibits promise in recommending music based on facial expressions. Particularly when integrated with other factors like user preferences, music genres, and song popularity. Continued research and enhancements in image processing techniques, dataset quality, and machine learning algorithms hold the potential to bolster the accuracy of our facial emotion-based music recommendation system.

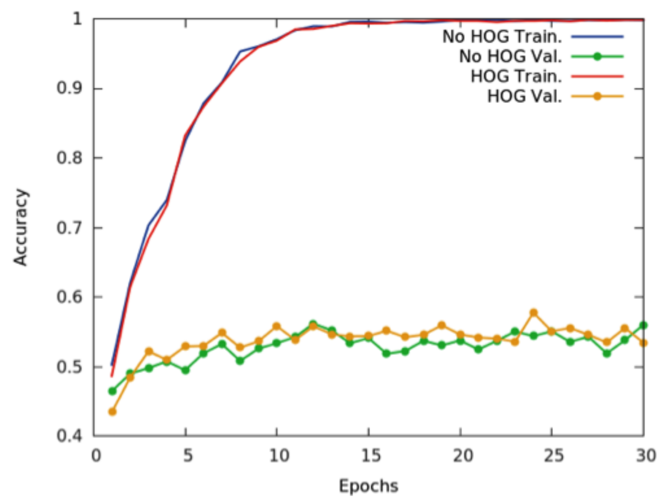

Fig. 12. Accuracy of deep model

V. Conclusion

For the task of recognising facial expressions, we created a variety of CNNs, and we assessed their effectiveness using various post-processing and visualisation techniques. The outcomes showed that deep CNNs can learn facial characteristics and enhance face expression recognition. Additionally, the hybrid feature sets had little effect on the accuracy of the model, indicating that convolutional networks may naturally learn the main facial traits even when given simply raw pixel data.

Following are some images of the real time detection :

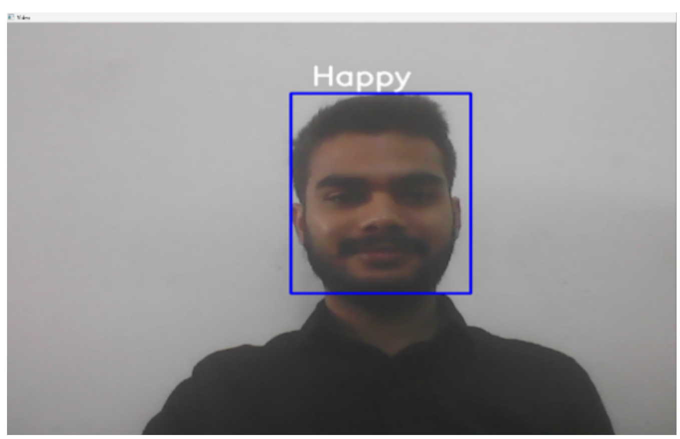

Fig. 13.

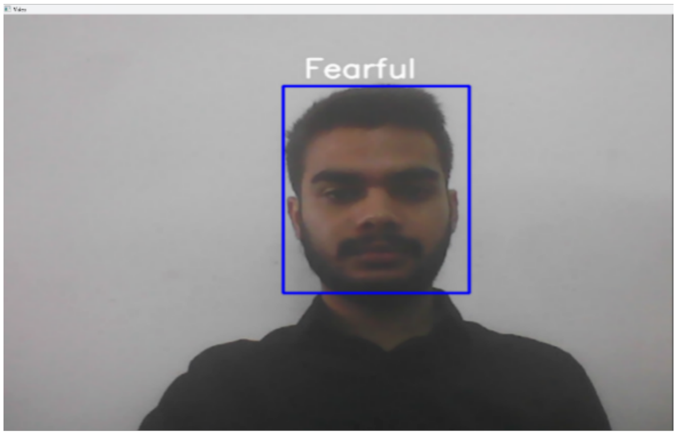

Fig. 14.

VI. Future Works

Subsequent research endeavors may concentrate on refining the precision of facial emotion recognition and crafting sturdier machine learning models adept at accommodating individual disparities in emotional expression. Furthermore, augmenting the system could entail integrating additional biometric data such as heart rate and skin conductance, offering a more thorough comprehension of the listener's emotional condition.

1.Incorporation with Wearable Devices: Facial emotion-driven music recommendation systems have the potential to merge seamlessly with wearable gadgets like smartwatches or fitness trackers. This integration enables real-time feedback on the user's emotional status, empowering adjustments to the recommended playlist for a more personalized listening journey.

2. Integration of Physiological Data: Beyond facial emotion recognition, music recommendation systems can integrate physiological data like heart rate and skin conductance. This holistic approach offers a deeper insight into the user's emotional state, facilitating adjustments to the recommended playlist for a truly personalized listening experience.

3. Leveraging Natural Language Processing: Integrating facial emotion-based music recommendation systems with natural language processing enables the understanding of user mood and preferences through text or voice inputs. This advanced technology facilitates the delivery of highly personalized recommendations, thereby enriching the user's listening experience.

Acknowledgment

We as a team would like to express our sincere gratitude to the following people and organisations for their contributions to this study. First and foremost, we would like to express our gratitude to Jayashree Mohanty(E15737), our research advisor, for her crucial direction and continuous assistance throughout this project work. We also acknowledges the staff at the various laboratory for granting us access to their resources and technical support. Additionally, we are appreciative of the volunteers who provided their precious time to take part in our work. We want to thank everyone who had helped us throughout the project work.

References

[1] Anagha S.Dhavalikar and Dr. R. K. Kulkarni, "Face Detection and Facial Expression Recognition System" 2014.Interntional Conference on Electronics and Communication System (ICECS -2020).

[2] K. Yoon, J. Lee, and M.-U. Kim, "Music recommendation system using emotion triggering low-level features," proceedings of IEEE Trans. Consumer Electron., vol. 58, no. 2, pp. 612-618, May 2018.

[3] https://www.kaggle.com/c/challenges-in- representation-learning-facial-expression- recognition-challenge/data.

[4] [3]Dalal, Navneet, and Bill Triggs (2020). Histograms of oriented gradients for human detection. Computer Vision and Pattern Recognition (CVPR), IEEE Computer Society Conference on. Vol. 1.

[5] Lonare, Ashish, and Shweta V. Jain (2022). A Survey on Facial Expression Analysis for Emotion Recognition. International Journal of Advanced Research in Computer and Communication Engineering 2.12.

[6] M.S. Bartlett, G. Littlewort, M. Frank, C. Lainscsek, I. Fasel, and J. Movellan. Fully automatic facial action recognition in spontaneous behavior. In Proceedings of the IEEE ConFEDence on Automatic Facial and Gesture Recognition, 2019.

[7] Krizhevsky, Alex, Ilya Sutskever, and Geoffrey E. Hinton. "Imagenet classification with deep convolutional neural networks." Advances in neural information processing systems. 2018.

[8] Simonyan, Karen, and Andrew Zisserman. "Very deep convolutional networks for large-scale image recognition." arXiv preprint arXiv:1409.1556 (2020).

[9] H. Chouhayebi, J. Riffi, M. A. Mahraz, Y. Ali, and T. Hamid, "Facial expression recognition using machine learning," in Proceedings of the 2021 Fifth International ConFEDence On Intelligent Computing in Data Sciences (ICDS), pp. 1–6, IEEE, Seoul, Korea, November 2021. View at: Publisher Site — Google Scholar.

[10] P. S Aleksic and A. K. Katsaggelos, "Automatic facial expression recognition using facial animation parameters and multistream hmms," IEEE Transactions on Information Forensics and Security, vol. 1, no. 1, pp. 3–11, 2006. View at: Publisher Site — Google Scholar.

[11] Renuka R. Londhe, Dr. Vrushshen P. Pawar, "Analysis of Facial Expression and Recognition Based On Statistical Approach", International Journal of Soft Computing and Engineering (IJSCE) Volume-2, May 2019.

[12] Lucey, P., Cohn, J. F., Kanade, T., Saragih, J., Ambadar, Z.,& Matthews, I. (2010). The Extended CohnKanade Dataset (CK+): A complete expression dataset for action unit and emotion-specified expression. Proceedings of the Third International Workshop on CVPR for Human Communicative Behavior Analysis (CVPR4HB 2010), San Francisco, USA, 94-101.

[13]]. S. Koelstra, C. Muhl, M. Soleymani, J. Lee, A. Yazdani, T. Ebrahimi, T. Pun, A. Nijholt, and I. Patras, "DEAP: a database for emotion analysis using physiological signals," IEEE Trans. Affect. Comput., vol. 3,no. 1, pp. 18-31, Jan. 2021.

[14] . C.-M. Chen M.-F. Tsai, J.-Y. Liu, and Y.-H. Yang, "Using emotional context from article for contextual music recommendation," in Proc. ACM International Conference on Multimedia, New York, USA, pp. 649-652, Oct. 2013.

[15] F. A. Nielsen, "A new ANEW: evaluation of a word list for sentiment analysis in microblogs," in Proc.Workshop on Making Sense of Microposts: Big Things come in Small Packages, Crete, Greece, pp. 93- 98,May 2011.

[16] Dolly Reney and Dr.Neeta Tripaathi, "An Efficient Method to Face and Emotion Detection", Proceedings of Fifth International Conference on Communication Systems and Network Technologies, 2015.

[17] Renata L. Rosa, Demóstenes Z. Rodríguez, and Graça Bressan "Music Recommendation System Based on User's Sentiments Extracted from Social Networks".

[18] A. Geetha, V. Ramalingam, S. Palanivel and B. Palaniappan, "Facial Expression Recognition— A Real Time Approach" Proceedings of International Journal of Expert Systems with Applications, Vol. 36, No.1,2021.