Song playlist generator system based on Facial Expression and Song Mood

Kevin Patel
Computer Science Engineering
Pandit Deendayal Energy University
Gandhinagar, Gujarat
kevin18patel@gmail.com

Rajeev Kumar Gupta
Computer Science Engineering
Pandit Deendayal Energy University
Gandhinagar, India
rajeevmanit12276@gmail.com

Abstract*—Nowadays song becomes important part of daily lifestyle. After emergence of the different music players into the market person are tend to listen music during day-to-day activities as well as listen for relaxing from daily stress. So that this music players are trying to provide best possible recommendation for the user. So that in this paper we are going to proposed new approach for the playlist generation task. In this approach we have proposed method to generate playlist from the emotion of the user and user personal choices of the songs for providing more personalized experience. After introduction of the Convolutional Neural Network object detection, Image classification, Emotion detection tasks reaches great height. In proposed method we have used CNN for Emotion detection task and ANN for the song classification task. With the CNN we have achieve 84% accuracy with FER-13 dataset which contain around 14k facial images. For song classification task we have used different song-features which is extracted from Spotify music player. We have achieved 82% accuracy in song classification task. Currently this system is only with Spotify music player.

Keywords: Artificial Intelligence, Machine Learning, Convolutional Neural Network (CNN) , Artificial Neural Network (ANN) , Emotion Detection , Song mood classification, Song Recommendation

I. INTRODUCTION

Song and Music are becoming part of daily life-style. Music and song are one of the essential parts of life in the modern era. Nowadays human used to listen music in day-to-day life while doing different things. So that there are many Music system was invented and they are used for collecting user data by observing the user activities. After evolution of AI-ML, Music recommendation systems are booming in developer ideas Because if any music player want to stick users as much as they can using personalize experience of listening, thus there are lots of ideas of different methodologies are proposed.

So far many song playlist generator a are there proposed earlier based on the context-of-use [1], based on audio music similarity [2] and many more. At the same time there are many song recommendation systems are proposed based on Similarity of features on audio signal [3], content-based on user grouping [4] and many more.

But there are some challenges are still in the music recommendation systems like cold start problem [5] for new user of the system which don't have any past history, Evaluation of song playlist generator model and song recommendation system [5] and mood and current state of the person wouldn't taken into consideration by this kind of the recommendation system. Snice there is strong relation between the mood of the user and song that he/she want to listen.[5]

After analyzing this issue, we proposed solution with the combination of two idea:

1) First one is User mood prediction using his/her face

2) second is song mood detection using neural network as function of classifier KerasClassifier. So that user can experience personalize recommendation based on his/her mood and his/her listened history.

The rest of the paper is structured as follow: Section 2 highlights the existing work, Section 3 discusses discuss the proposed method used to generate song play list and Section 4 evaluates the performance of the proposed model and Section 5 concludes the paper.

II. RELATED WORK

S. Gilda et al. [6], introduced "Smart Music Player Integrating Facial Emotion Recognition and Music Mood Recommendation". This paper proposed music recommendation system with facial expression and feedback of the user for suggested music. In this Author used CNN for emotion detection. For song recommendation system author uses several audio features like tempo, beat spectrum, tonal mode , pitch ,etc.

Assunção, Willian & Neris, Vania [7], introduced "An algorithm for music recommendation based on the user's musical preferences and desired emotions". This paper presents the selection algorithm of the song based on the user mood and user previous suggested song after effects. In this method user is classified into 20 sample user classes and then get suggestion based on the class in which user is appearing.

A.S. Mali et al. [8], proposed a Mood based Music System which recommend the music based on the mood. This paper proposed with mainly three steps. In the first step face detection performed by Haar cascade classifier, Once the face is detected CNN classifier is used to detect emotion. In the last step Playlist generation task done using calculating BPM - beats per minutes of the song. In this paper author get 66% accuracy of Emotion detection module with FER-13 dataset.

V. Patchava, et al. [9], proposed a Sentiment-Based Music Play System which recommend songs based on surrounding sound or voice analysis. In this paper author come up with RASPBERRY-PI based module for mood detection. They first converted listened voice to text and then classifies it using naïve Bayesian classifier for analysing the voice mood. After that they suggest the song based on its BPM value.

Samuvel, D et al. [10], proposed a "Music Recommendation System Based Onn Facial Emotion Recognition". This paper used SVM classifier for emotion classification and they have recommended the song using ANN based on the title, song_id, artist_name, andrelease_by and user_id, song_id and listen time, etc data.

III. PROPOSED METHODS

In this section of the paper we are going to discuss our proposed method for song playlist generator system based on facial experience and past history of user.
We have divide the whole process into 4 part :

A. Develop Model for Facial Emotion Detection
B. Collecting user past history using APIs
C. Song mood classification
D. Creating playlist based on user emotion

A. Develop Model For Facial Emotion Detection
Basically, we can divide emotion detection process into 2 parts:

1. Face-Detection
2. Emotion Detection

1. FACE-DETECTION

In our system the first step is detecting the human faces using camera with the help of openCV library. There are many methods are been proposed for face detection based on different concepts like Knowledge-based, Features-Based, Templete-based , Appearance-based. But Haar Feature-Based Cascade Classifier is one of most effective classifier so we use that for finding the No of faces with it's cordinate loacaton.

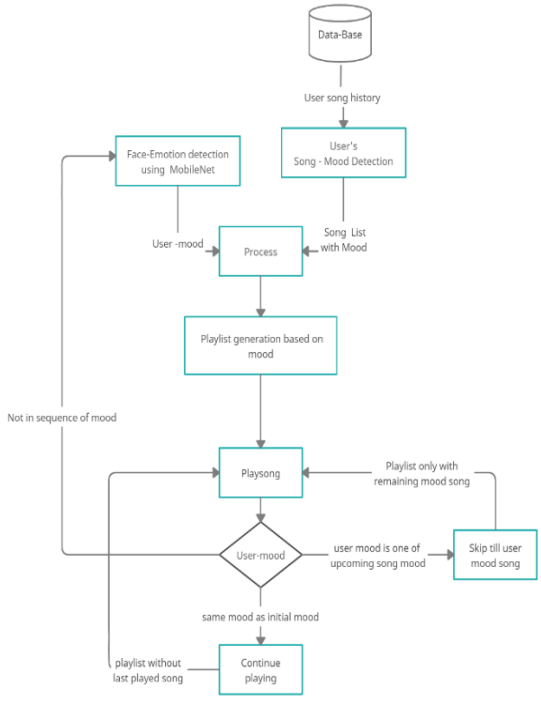

Figure 1: Proposed System Flow-Diagram

2. EMOTION DETECTION

Emotion detection is important task for our system. With the ANN emotion detection was very tough task but after inventions Convolutional Neural Network (CNN) image processing task was comes into picture of researches. Object detection, Image classification, Face detection [13] , Emotion detection field gets new boost for reaching heights. With the help of CNN, we can perform emotion detection task. In this project we are using Convolutional Neural Network (CNN) for emotion detection task. The input of the CNN is 48px X 48px face image which is get by the Haar Cascade Classifier. The CNN will classifies the image into one of the 4 emotions: angry, happy, sad and surprise.

CNN is one of main part of the Neural Network. CNN are useful for different object detection, object recognition, object classification, Face detection, Emotion detection, etc. In CNN Convolutional layer is first layer, which has heavy computation. In this layer the image is convolved by the number of filters or kernel with apply over image using sliding window. The output of the convolutional layer feed into activation layer. This is basically for increase non-linearity. The third layer of CNN is polling layer, which is for downsampling the features.

In this paper we have proposed the CNN for emotion detection with 6 blocks of CNN. Each block consists of Conv2D, BatchNormalization layers. After 2 consecutive blocks we add MaxPolling2D layers with dropout layer for preventing from overfitting problems. We have used Conv2D layers with 64,64,128,128,256,256 filters correspondingly. The convolutional layers are followed by three dense layers with 128, 64 ,4 nodes. For the training of the CNN, we have used FER-13 dataset. The FER-13 dataset consists of 35.9k images of 7 different facial expressions. In this paper we have used only 4 emotion data for training. FER-13 dataset contains 48px X 48px image dataset.

Training_accuracy = 84.91%
Validation_accuracy = 83.44 %
Training_loss = 0.4201
Validation_loss = 0.4532

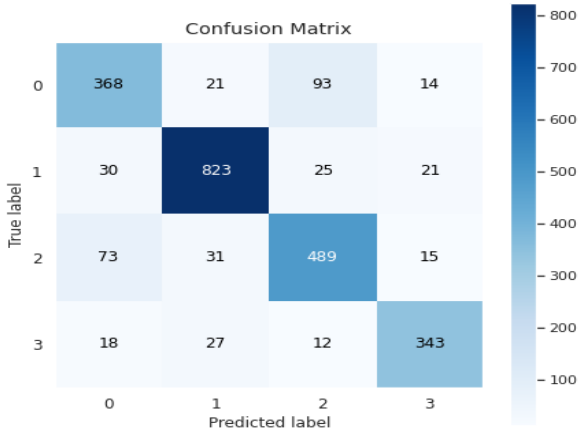

Figure 2: Confusion Matrix for facial expression detection

Here the labels are basically different facial mood

0 : 'Angry' , 1 : 'Happy' , 2 : 'Sad' , 3 : 'Surprise'

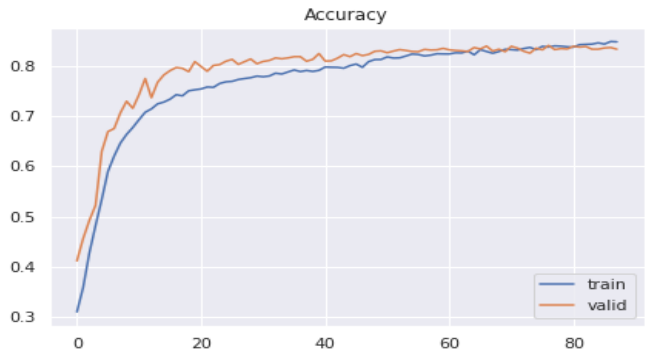

Figure 3: Accuracy Graph for Training and Validation for 100 Epochs

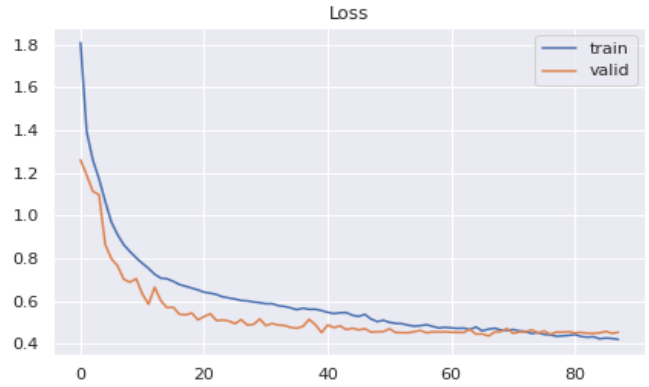

Figure 4: Loss Graph for Training and Validation for 100 Epochs

B. Collecting User Past History Using API

For Collecting user past listened songs, we use Spotify API .We have used Spotipy library to use Spotify API . We have used basically 3 end pointer for fetching different quires results

i. current_user_recently_played(limit=50, after=None, before=None)
used for : getting users past 50 listened song
ii. tracks(id)
used for : getting track with given id
iii. audio_feature(id)
used for : getting audio features with track id

C. Song Mood Classifiacation

In this time, Song and Music are the most effective factor for changing person mood from bad to good. Nowadays people have tendency to listen music during driving, playing, during exercising as well as during relaxing. After emergence of many media players like Spotify, Saavn , etc. listing song is increased rapidly. This companies are investing more and more for giving personalize experience to user so that user tend to stick with that music player.

There are many song mood detection methodologies have been proposed in recent years. These methodologies have been based on Gerne analysis, MLP (MultiLayer Perceptron), based on audio lyrics analysis with deep neural network [11]. In this paper we have proposed song mood detection methodology with KerasClassifier. First of all we build function for creating Sequential Model , then we pass that model as function of KerasClassifier. For song mood detection we have used Spotify dataset with different song attributes. After some experimentation and surveys we have finalize 10 attributes which are useful for mood prediction. We have choose Length, Danceability, Acousticness , Energy , Instrumentalness , Liveness, Valence, Loudness , Speechiness , Tempo of the songs. This is just classification problem and we have 10 attribute to pass into neural network and we have chosen basically 4 different mood for out problem Calm , Energetic , Happy and Sad , so that we only need Sequential Neural Network with only 4 layer with 10 , 8, 8, 4 neurons in respective layer.

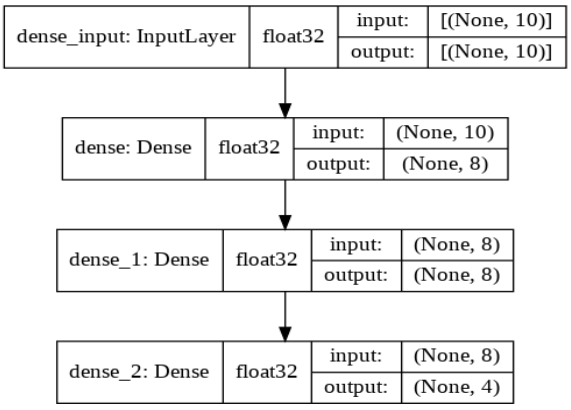

Figure 5: Sequential NN Architecture for Song Mood Detection

We have passed the function that return model to The KerasClassifier along with some initial epochs size, batch size. For finding most effective values of this 2 hyperparameters, we have used GridSearchCV with different values of hyperparameters. After finding the results of GridSeachCV , we have fit that model to the KerasClassifier. We have get the testing accuracy score of song mood classifier around 83%. The confusion matrix of song classifier given below:

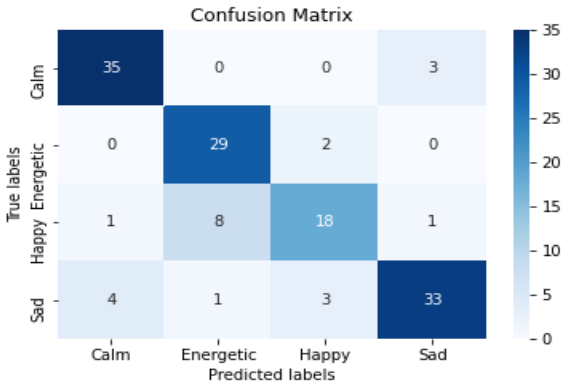

Figure 6: Confusion Matrix for Song Mood Classification system

D. Creating playlist Based On User Emotion

This system firstly detects facial emotion using described method. After that as we have discussed first we

have fetched users past listened 50 songs , then using song mood classifier we labeled each song with it's mood. After that we just separated the whole list into different lists based on song moods. Then we concat different mood specific list based on facial expression as given below so that user can't fell drastically change in songs mood and the end goal of this playlist is to make user more energetic so that he/she can do work with more dedication and energy.

- If user mood ➔Sad
  Song playlist in below order :
  Sad → Calm → Happy → Energetic
- If user mood ➔Angry
  Song playlist in below order :
  Calm → Happy → Energetic
- IF user mood ➔ Surprised
  Song playlist in below order :
  Calm → Energetic
- If user mood ➔ Happy
  Song playlist in below order :
  Happy → Energetic

IV. RESULTS

The result of the system can be given into 3 parameters. 1) Accuracy of facial emotion recognition system 2) Accuracy of song mood classification system 3) mood-based playlist recommendation for users

1) Accuracy of Facial Emotion recognition system:
In this project we just need to classifies the facial expression into 4 moods. We got testing accuracy 83% with loss of 0.4532.

Table 1: Classification report for emotion detection

	Precision	*Recall*	*f1-score*	*Support*
Angry	*0.76*	*0.74*	*0.75*	*496*
Happy	*0.91*	*0.92*	*0.91*	*899*
Sad	*0.79*	*0.80*	*0.80*	*608*
Surprise	*0.87*	*0.86*	*0.87*	*400*
Accuracy			*0.84*	*2403*
macro avg.	*0.83*	*0.83*	*0.83*	*2403*
weight avg.	*0.84*	*0.84*	*0.84*	*2403*

The number of methods have been proposed in the last few decades for the emotion detection. Table 2 illustrate the comparative analysis of different existing emotion detection algorithm.

Table 2: Comparative Analysis of Different Existing Methods

Author paper	**Dataset**	**Method**	**Accuracy**
Agrawal et Mittal [12]	FER2013[15]	CNN	65%
Kim et al. [13]	MMI [16], CASME II [17]	CNN-LSTM	78.61%, 60.98%
Asad, Maliha [18]	CK+ [19]	SVM	81%
Proposed model	FER2013[14]	CNN	83%

To evaluate the performance of the proposed model, we compared the accuracy of the compared with four existing emotion detection model. Figure 7 illustrate the comparative analysis of accuracy for different existing methods.

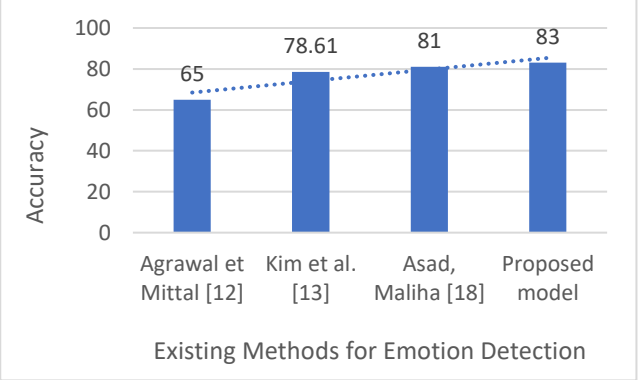

Figure 7: Comparative Analysis of Accuracy For Different Existing Methods

As shows in figure 7, proposed method gives higher accuracy as compare to the existing methods.

2) Accuracy of song mood classification system :

We have achieved 83% of accuracy with 4 classes (Calm song , Happy song , Energetic song , Sad Song) classification. For the training of KerasClassifier we have used Spotify dataset which contain 800+ song with attributes. We have used Neural Network for this Classification process as function with KerasClassifier.

	Precision	*Recall*	*f1-score*	*Support*
Calm	*0.88*	*0.92*	*0.90*	*38*
Energetic	*0.76*	*0.94*	*0.84*	*31*
Happy	*0.78*	*0.64*	*0.71*	*28*
Sad	*0.89*	*0.80*	*0.85*	*41*
Accuracy			*0.83*	*138*
macro avg.	*0.83*	*0.83*	*0.82*	*138*
weight avg.	*0.84*	*0.83*	*0.83*	*138*

Table 3: *Classification report for Song mood classification*

3) Output of the full system

Currently this system show playlist in command prompt as shown below.

Figure 8: Sad Face for Testing Output

```
your mood is Sad , we prepare below playlist for you

                                  name                      id       mood
0                               Ambre  4w4fT1QRxtXaTpvSqSqDaM        Sad
1                Thinking out Loud  34gCuhDGsG4bRPIf9bb02f        Sad
2                  Mystery of Love  4HbeGjbt7u3pvwDk1vN7P0        Sad
3                    drivers license  71PN2DXiMsVn7XUKtOW1CS        Sad
4                      Dance Monkey  2XU0oxnq2qxCpomAAuJY8K        Sad
5   i hate u, i love u (feat. olivia o'brien)  7vRriwrloYVaoAe3a9wJHe        Sad
6                The Sound of Silence  3YfS47QufnLDFA71FUsgCM        Sad
7                           Play Date  4DpNNXFMMxQEK17r0ykkWA        Sad
8   i hate u, i love u (feat. olivia o'brien)  7vRriwrloYVaoAe3a9wJHe        Sad
9                    After the Rain  342kvni2b98OzoRX1x8Yqe       Calm
10                Bye Bye Little House  6rY3x2SZINzczbREengsIX       Calm
11                   After the Rain  342kvni2b98OzoRX1x8Yqe       Calm
12                   After the Rain  342kvni2b98OzoRX1x8Yqe       Calm
13                   After the Rain  342kvni2b98OzoRX1x8Yqe       Calm
14                            Liggi  4vwPirpSJCOiIsgxMV4Xcv      Happy
15                            Rylan  6XxPXXqkE41G7MVkpom6F8      Happy
16                     Shape of You  7qiZfU4dY11Wl1zX7mPBI3      Happy
17                         Satisfya  6HGoVbCUr63SgU3TjxEVj6      Happy
18                           Safari  6osKPJp6kQwZcgUqBteJFN      Happy
19  Stereo Hearts (feat. Adam Levine)  0qOnSQQF0yzuPWsXrQ9paz      Happy
20                       SugarCrash!  2ePtv8M1BO9nuuXABqAfEX      Happy
21          Havana (feat. Young Thug)  1rfofaqEpACxVEHIZBJe6W      Happy
22                        Best of You  1ILZ7g9P1ggf1L276HF4uB      Happy
23                            Liggi  4vwPirpSJCOiIsgxMV4Xcv      Happy
24                     Kings & Queens  2tSnkmZL1916IUzMIOC4Qz      Happy
25                             Faded  698ItKASDavgwZ3WjaWjtz  Energetic
26                       The Spectre  2DGa7iaidT5s0qnINlwMjJ  Energetic
27                         Superhero  7EYv3adMrIC5mqurz7fHhD  Energetic
28                              Lily  0lks2Kt9veMOFEAPN0fsqN  Energetic
29                     Drivers License  6KrZtsDSKwb00Dg16xI2W2  Energetic
```

Figure 9: Output for Above Input

V. CONCLUSION

In this project we have try to generate songs playlist based on the facial emotion detection. In this we have used comparatively small datasets with what is in actual real-life scenario. Thus, as we implement it on larger datasets there are many questions arises also. Here playlist generation system is in beginning stage. In this we have implemented songs mood path with respect to user mood is only by taking simple surveys.

For the feature work, still there is lots of space like, we can add more factors like what is person is doing? what is time as well as what is weather of the user place? in what community user belong to?, so that we can generate more better playlist. There are many aspect are still remaining like emotion detection using surrounding audio feature processing, evaluation of the system based on skipping ratio of songs and many other factors , Finding best path of song mood that need to occur according to user mood by evaluating more user data and activity. This all problem can be seen as feature scope of the song playlist generation system.

REFERENCES:

[1] Pauws, Steffen & Eggen, Berry, " PATS: Realization and user evaluation of an automatic playlist generator", Journal of New Music Research, vol. 32, issue 2, 2002, pp. 1-10.

[2] Pampalk, E. and M. Gasser, "An Implementation of a Simple Playlist Generator Based on Audio Similarity Measures and User Feedback" in proceeding of 7th International Conference on Music Information Retrieval, 2006, pp. 1-3.

[3] Adiyansjah, Alexander A S Gunawan, Derwin Suhartono, "Music Recommender System Based on Genre using Convolutional Recurrent Neural Networks", Procedia Computer Science, vol. 157, 2019, pp. 99-109.

[4] Chen, HC., Chen, A.L.P. "A Music Recommendation System Based on Music and User Grouping", Journal of Intelligent Information Systems,vol. 24,2005, pp.113–132.

[5] Schedl Markus , Zamani Hamed , Chen Ching-Wei , Deldjoo Yashar , Elahi Mehdi,"Current Challenges and Visions in Music Recommender Systems Research",International Journal of Multimedia Information Retrieval,vol. 7(2) ,2018,pp.95-116.

[6] S. Gilda, H. Zafar, C. Soni and K. Waghurdekar, "Smart music player integrating facial emotion recognition and music mood recommendation", International Conference on Wireless Communications, Signal Processing and Networking (WiSPNET), *2017*, pp. 154-158.

[7]Assunção Willian, Neris Vania,"An algorithm for music recommendation based on the user's musical preferences and desired emotions", 17th International Conference on Mobile and Ubiquitous Multimedia,November2018 , pp.205–213.

[8] A.S. Mali, A.A. Kenjale, P.M. Ghatage, A.G. Deshpande, "Mood based Music System," International Journal of Scientific Research in Computer Science and Engineering, Vol.6, Issue.3, 2018, pp.27-30.

[9] V. Patchava, P. Jain, R. R. Lomte, Shakthi Priya G and H. B. Kandala, "Sentiment based music play system",International Conference on Smart Sensors and Systems (IC-SSS), 2015, pp. 1-5.

[10] Samuvel, D. J., Perumal, B., y Elangovan, M."Music recommendation system based on facial emotion recognition",3C Tecnología. Glosas de innovación aplicadas a la pyme. Edición Especial, Marzo 2020,pp. 261-271.

[11] Delbouys, R., Hennequin, R., Piccoli, F., Royo-Letelier, J., & Moussallam, M., "Music Mood Detection Based on Audio and Lyrics with Deep Neural Net",the 19th ISMIR Conference, Paris, France, September 2018 ,pp. 370–375.

[12] Agrawal, A., Mittal, N. , "Using CNN for facial expression recognition: a study of the effects of kernel size and number of filters on accuracy." The Visual Computer , vol. 36 ,issue 2, February 2020, pp 405–412.

[13] D. H. Kim, W. J. Baddar, J. Jang and Y. M. Ro, "Multi-Objective Based Spatio-Temporal Feature Representation Learning Robust to Expression Intensity Variations for Facial Expression Recognition," IEEE Transactions on Affective Computing , vol. 10, no. 2, 1 April-June 2019,pp. 223-236.

[14]Deepak Kumar Jain, Pourya Shamsolmoali, Paramjit Sehdev, "Extended deep neural network for facial emotion recognition", Pattern Recognition Letters, Vol.120 , April 2019,Pp 69-74.

[15] I. J. Goodfellow et al., "Challenges in Representation Learning: A Report on Three Machine Learning Contests", Neural Information Processing, Berlin, Heidelberg, 2013, p. 117-124.

[16] M. Pantic, M. Valstar, R. Rademaker, et L. Maat, “Web-based database for facial expression analysis”, IEEE International Conference on Multimedia and Expo, July 2005, 5 pp,-.
[17] W.-J. Yan et al., “CASME II: An Improved Spontaneous Micro-Expression Database and the Baseline Evaluation” , PLoS ONE, vol. 9, no 1, January 2014, pp. 1-8.
[18] Asad Maliha , Gilani Syed , Jamil Mohsin, “Emotion Detection through Facial Feature Recognition”, International Journal of Multimedia and Ubiquitous Engineering,Vol.12, No.11 ,2017 ,pp.21-30.
[19] P. Lucey, J. F. Cohn, T. Kanade, J. Saragih, Z. Ambadar, et I. Matthews, “The Extended Cohn-Kanade Dataset (CK+): A complete dataset for action unit and emotion-specified expression”, IEEE Computer Society Conference on Computer Vision and Pattern Recognition - Workshops, June 2010, p. 94-101.