Design and Implementation of AI Based Efficient Emotion Detection and Music Recommendation System

Saurabh Kumar
School of Computer Science and Engineering
REVA University
Bengaluru, India
saurabhkmr329@gmail.com

Rohan Kumar Nandi
School of Computer Science and Engineering
REVA University
Bengaluru, India
rohannandy2001@gmail.com

Shashwat Roy
School of Computer Science and Engineering
REVA University
Bengaluru, India
shashwatkroy@gmail.com

Argha Sarkar
School of Computer Science and Engineering
REVA University
Bengaluru, India
saurabhkmr329@gmail.com

2023 14th International Conference on Computing Communication and Networking Technologies (ICCCNT) | 979-8-3503-3509-5/23/$31.00 ©2023 IEEE | DOI: 10.1109/ICCCNT56998.2023.10306938

***Abstract*—In recent times, with the development and operation of big data, deep learning has taken more and more attention. As deep learning neural network, a convolutional neural network plays an extremely important part in face image recognition. In this paper, a combination of expression recognition technology of an convolutional neural network and automatic music recommendation algorithm is developed to identify a model that recognizes facial expressions and recommends music according to corresponding mood. The facial expression recognition model established in this paper uses FER2013 with a recognition rate of 62.1. After relating the matching expression, a content- grounded music recommendation algorithm is used to prize the point vector of the song and a cosine similarity algorithm is used to make the music recommendation. This exploration helps to ameliorate the practicality of the music recommendation system, and the affiliated results will also serve as a reference for the operation of the music recommendation system in areas similar as emotion parameter.**

Keywords—facial expression recognition, music recommendation, deep learning, convolution neural network

I. INTRODUCTION

With the arrival of the information age, deep learning is extensively used in image processing, image recognition and especially facial expression recognition. Face recognition has come to an exploration hotspot in the field of human-computer relation, but it still has limitations on the operation of image processing results. Image exploration frequently focuses on perfecting the precision of recognition, and the data in the image lack the use of secondary processing, that is, in the factual product and a life process, the image information has not been fully and efficiently used. In this paper, a deep literacy system is used to design and train a convolutional neural network expression recognition model. The results of image processing are combined with a music recommendation algorithm, and the music that adjusts the mood is recommended by judging the mood has shown by the person. Music data sets are created by crawling the playlists and homemade re-flections of major music websites. The compass of operation of image processing results has been meetly expanded.

II. METHODOLOGY AND SYSTEM USE

In current system the facial recognition is accomplished by applying machine learning methods similar as random forest, a decision tree. This procedure in-volves characteristic extraction which is actually complex processor. The delicacy of these algorithms also actually low. The major disadvantages are listed below.

- No precise results
- Not possible to find the facial expression.
- Hard to identify the outcomes.

In proposed system we are enforcing facial emotion recognition applying con-volutional neural network. By applying CNN algorithm we categorize types of feelings and different songs are recommended. These deep learning methods give the finest precision results. Fig.1 depicts the block diagram.

The advantages of the proposed system is listed below.

- Time saving
- Find the facial expression while driving to prevent accidents.
- Accurate results
- It will give alert alarm or buzzer to the person.

The facial emotion picture dataset is considered fromkaggel.com, the data set is spited into two types one is training and another type is testing. After dividing data set, followed by data cleaning which means the unwanted data are deleted. Resizing and reshaping the pictures into applicable layout to train our model. Apply the pre-processed training dataset is utilized to train our model by applying CNN algorithm and save the model. The outcome of our model is generating the facial feelings as either Happy, Sad, Angry & Neutral. The result of the model is generating the facial feelings as either Happy, Sad, Angry & Neutral along with feelings we recommended the music related to emotion. Predicated on the facial expression of the user, the expression will be categorized and unveiled. On the basis of the classified emotion, system will suggest music automatically that user can listen.

Authorized licensed use limited to: VIT University- Chennai Campus. Downloaded on March 11,2024 at 13:49:12 UTC from IEEE Xplore. Restrictions apply.

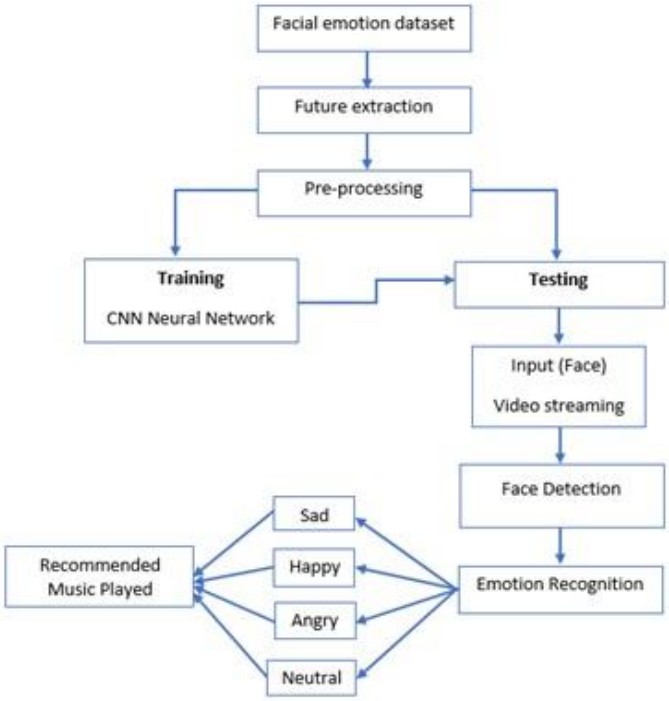


Fig. 1. Block diagram.

A. Data Collection and Preprocessing

Data collection and pre-processing plays a major part in the creating the AI model. For the designing of any AI model the most important part is data. Data is needed to train the model to forecast the results. Choosing the data is the most critical measure as an AI model can only be utilized in a public use if the outcomes have been used earlier. The better the data better is the outcomes. For the better data it should be true else an AI model would be like a GIGO (Garbage in Garbage Out) configuration which is of no usage. For the present AI operation, we collected the data for origins and repositories accessible on the internet.

B. Data Collection and Preprocessing

For working the subsequent problem, we required to collect the image dataset of people showing various feelings. Therefore, upon digging we picked up the FER- 2013 dataset that contained people stating feelings.

The data consists of 48X48 pixel grayscale pictures of faces. There are 7 orders of feelings available in the dataset i.e. Fear, Happy, Sad, Surprise Angry, Disgust and Neutral. The dataset contains of 28,709 samples to train the model on. Some examples of feelings present in the dataset which is shown fig 2.

The classes present in the above dataset have the following distribution:

Happy: 7215 images

Disgust: 436 images

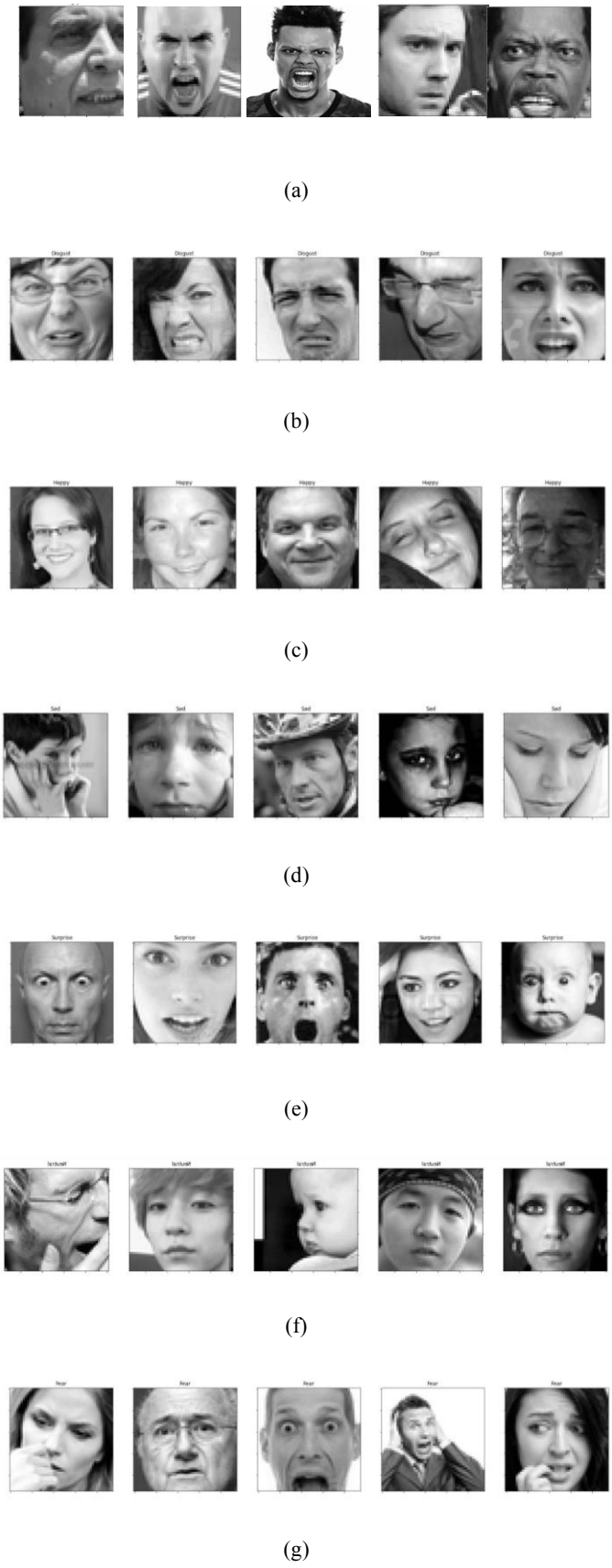
Neutral: 4965 images

Fear: 4097 images

Sad: 4830 images

Angry: 3995 images

Surprise: 3171 images.



(a)

(b)

(c)

(d)

(e)

(f)

(g)

Fig. 2. Different expression (a) Anger (b) Disgust (c) Happy (d) Sad (e) Surprise (f) Neutral (g) Fear

Authorized licensed use limited to: VIT University- Chennai Campus. Downloaded on March 11,2024 at 13:49:12 UTC from IEEE Xplore. Restrictions apply.

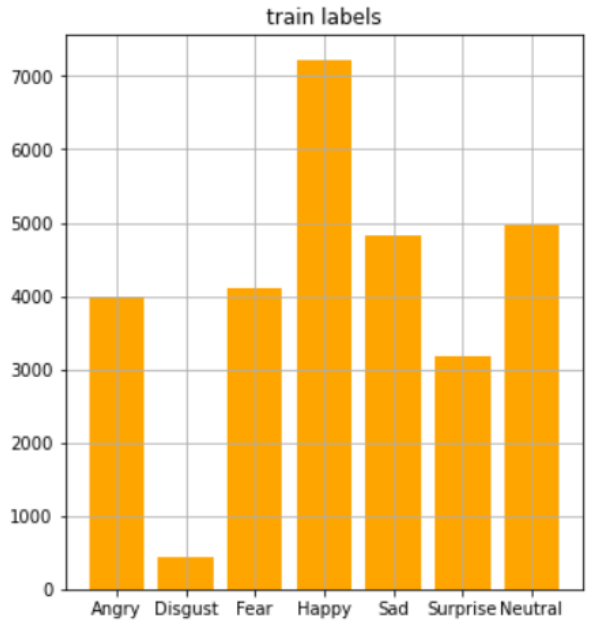


Fig. 3. Comparison of train labels vs different emotions

III. DATA MODELLING

Data Modelling is one of the most vital portion which consists of pre-processing, characteristic selection, and also making the machine learning model. This can be so vital that it can moreover frame the appropriate model, or it can produce the worst model all depends on how the data are reused. Un-alike datasets come with diverse demand of modelling the data and also pre-processing it on the base of null values or the categorical features. All the pictures were resized to 48 X 48-pixel size so that they can be given into the Convolutional Neural Network(CNN).

The architectural diagram is shown in fig. 4.

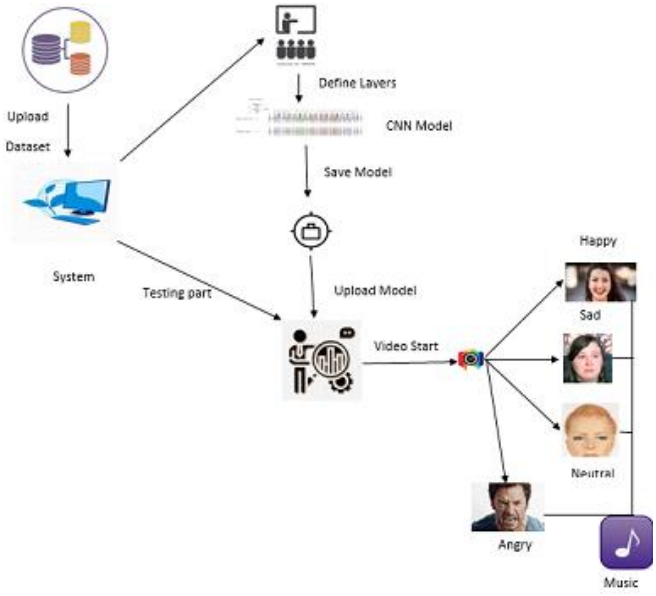


Fig. 4. Architectural diagram of proposed system

A. Relu Layer

The alternate part of this measure will involve the debugged Linear Unit or Relook. We'll step in Relook layers and go through how linearity functions in the environment of Convolutional Neural Networks.

Not compulsory for understanding CNN's, but there is no damage in a quick assignment to enhance your expertise.

B. Pooling Layer

In this proportion, we'll step in pooling and will get to deduce exactly how it generally works. Our nexus then, however, will be a precise type of pooling, maximum pooling. We will step in different paths, however, including mean (or sum) pooling. This portion will conclude with a demo framed utilizing a visual interactive instrument that will surely sort the entire conception out for you.

C. Flattening

This will be a concise breakdown of the smoothing procedure and how we shift from pooled to flattened layers when working with Convolutional Neural Networks.

D. Full Connection

In this portion, everything that we filled in throughout the section will be intermingled together. By mastering this, you will get to visualize an utmost illustration of how Convolutional Neural Networks work and how the" neurons" that are eventually produced learn the group of images. Fig. 5 shows the static view of the representation. The interrelated things between system and user are depicted through fig.6. The sequence diagram and collaboration diagram are represented in fig.7 and fig. 8. Fig. 9 clearly presents the control flow drawn from one system operation to another user operation.

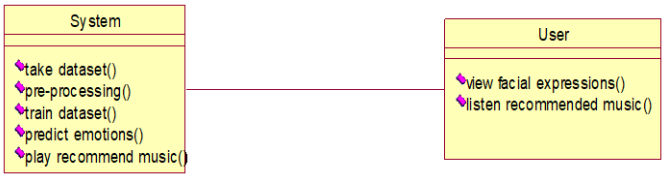


Fig. 5. Class Diagram

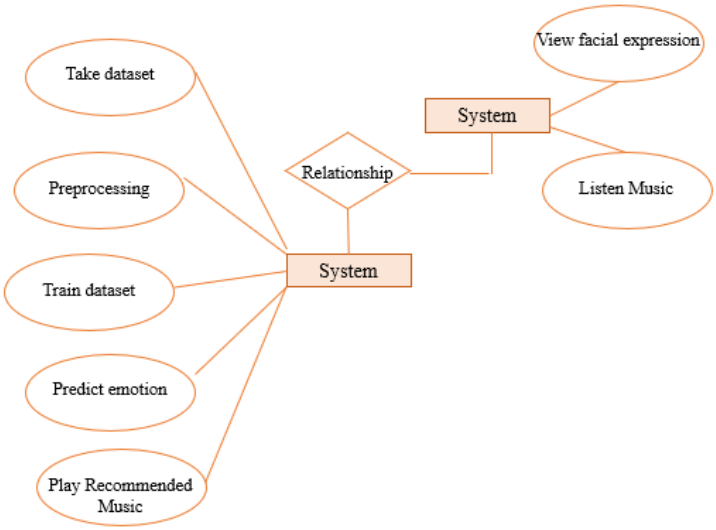


Fig. 6. Entity Relationship Model

Authorized licensed use limited to: VIT University- Chennai Campus. Downloaded on March 11,2024 at 13:49:12 UTC from IEEE Xplore. Restrictions apply.

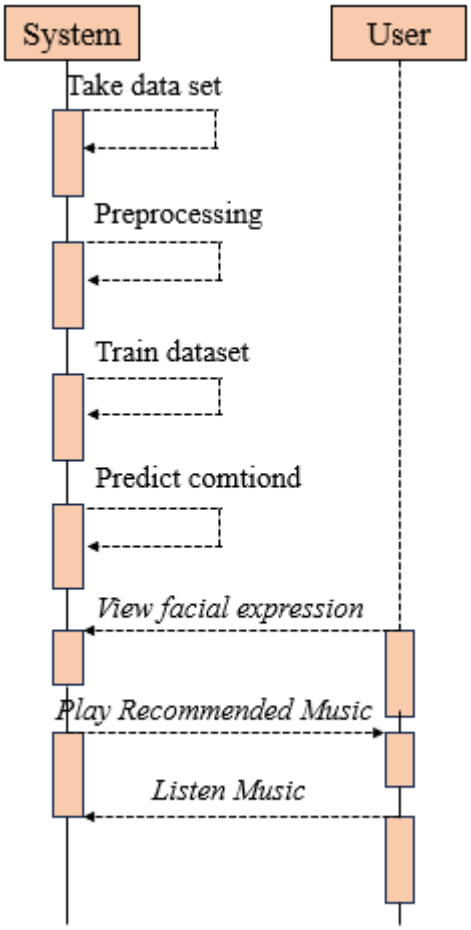


Fig. 7. Sequence Diagram

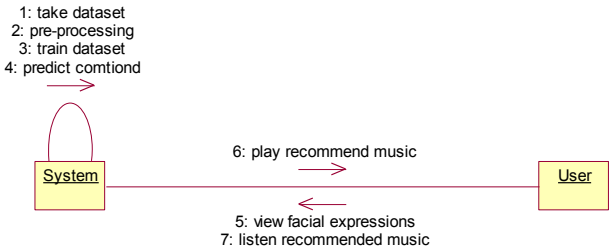


Fig. 8. Collaboration diagram.

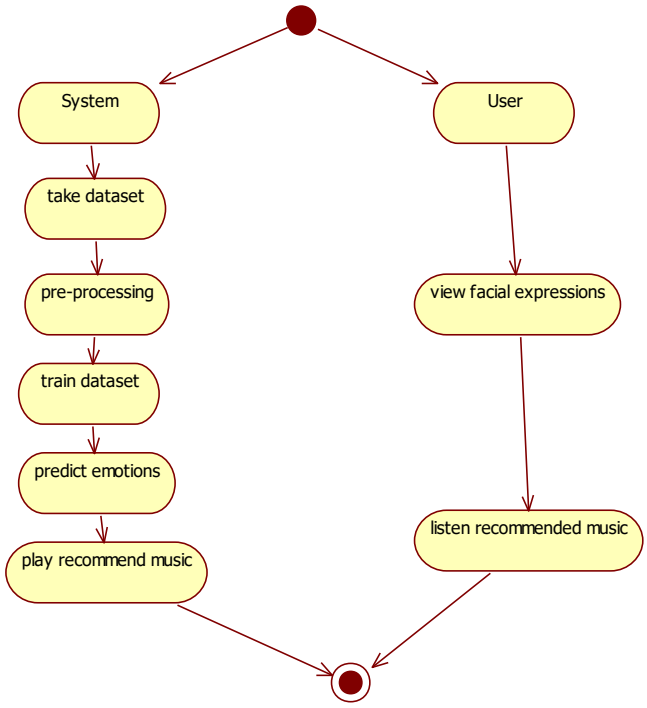


Fig. 9. Activity Diagram

IV. RESULT AND ANALYSIS

The execution window of our project is depicted in Fig 10, where we run the model through which the user facial expression will get capture and fed into the system and after analysing the facial expression its features is being extracted and it shows the corresponding expression on the screen and related song is played from the playlist to enhance the mood of the user. The Predicted emotions is depicted in Fig 11, 12, 13. as Angry, Happy and Sad and related playlist is played to enhance the mood of the user.

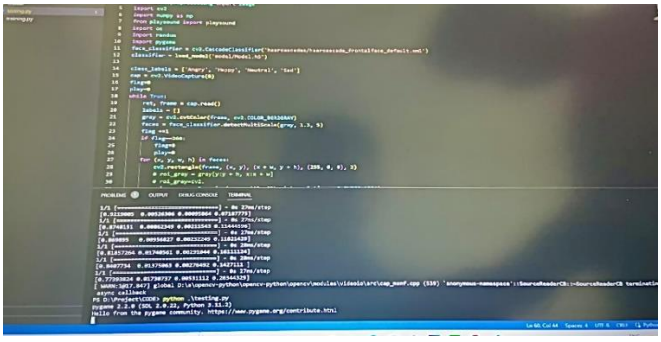


Fig. 10. Execution of the proposed model.

Authorized licensed use limited to: VIT University- Chennai Campus. Downloaded on March 11,2024 at 13:49:12 UTC from IEEE Xplore. Restrictions apply.

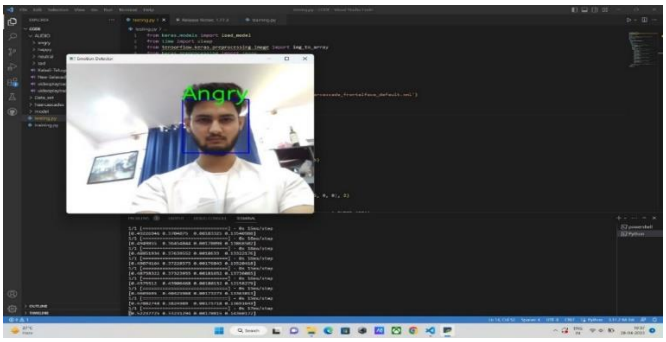


Fig. 11. Detection of Angry emotion.

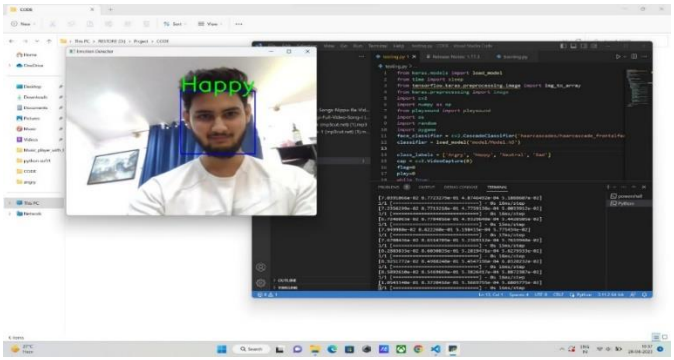


Fig. 12. Detection of Happy emotion.

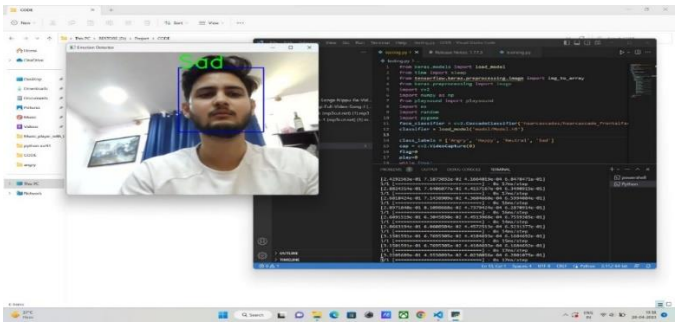


Fig. 13. Detection of Sad emotion.

V. CONCLUSION

In this paper, we proposed a model of facial expression recognition grounded on a convolutional neural network. After training the model on FER2013 dataset, we got recognition accuracy of 62.1. On the base of the state that facial expression and emotion were both predicted, the content-grounded recommendation algorithm was applied to automatically recommend music for users. Com-pared with the current algorithms that only recommend music according to the users' former listening preferences, the algorithm proposed in this paper increases the user's emotion recognition, so that the recommended music can more meet the users' listening requirements. Thus, this algorithm has relevantly promising use demand. Although we have made some achievements, still some challenges need to be answered. For illustration, the precision of expression emotion recognition needs to be enhanced. In the follow-up work, the recognition rate of labels with low recognition will be enhanced, and the music recommendation algorithm will be further optimized and upgraded.

ACKNOWLEDGMENT *(Heading 5)*

The author would like to acknowledge REVA university for infrastructural support.

REFERENCES

[1] Ramanathan, R., Kumaran, R., Rohan, R.R., Gupta, R. and Prabhu, V., 2017, December. An intelligent music player based on emotion recognition. In 2017 2nd International Conference on Computational Systems and Information Technology for Sustainable Solution (CSITSS) (pp. 1-5). IEEE.

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

[3] Ayata, D., Yaslan, Y. and Kamasak, M.E., 2018. Emotion based music recommendation system using wearable physiological sensors. IEEE transactions on consumer electronics, 64(2), pp.196-203.

[4] Degtyarev, G.L., Garkushenko, V.I., Fayzutdinov, R.N., Gubaydullin, S.I. and Spiridonov, I.O., 2019, October. Robust mixed H2/H_infinity control synthesis for electrooptical device line-of-sight stabilization system. In 2019 12th International Conference on Developments in eSystems Engineering (DeSE) (pp. 73-76). IEEE.

[5] Carrier, P.L. and Courville, A., 2013. Challenges in representation learning: Facial expression recognition challenge. Kaggle Competition, https://www. kaggle. com/c/challenges-inrepresentation-learning-facial-expression-recognitionchallenge/data.

[6] Preema, J.S., Rajashree, S.M. and Savitri, H., 2018. Review on Facial Expression Based Music Player. International Journal ofEngineering Re-search & Technology (IJERT), ISSN-2278-0181, 6(15).

[7] Guidel, A.Y.U.S.H., Sapkota, B. and Sapkota, K., 2020. Music recommendation by facial analysis.

[8] Kundu, M. and Kumar, D.N., 2021, May. Route Optimization of Unmanned Aerial Vehicle by using Reinforcement Learning. In Journal of Physics: Conference Series (Vol. 1921, No. 1, p. 012076). IOP Publishing.*)*

[9] Tabora, V., 2019. Face detection using OpenCV with haar cascade classifiers. línea]. Available: https://becominghuman. ai/face-detection-using-opencv-with-haarcascade-classifiers-941dbb25177.[Último acceso: 12 08 2019].

[10] Qin, Z., Yu, F., Liu, C. and Chen, X., 2018. How convolutional neural network see the world-A survey of convolutional neural network visualization methods. arXiv preprint arXiv:1804.11191.

[11] Chankuptarat, K., Sriwatanaworachai, R. and Chotipant, S., 2019, July. Emotion-Based Music Player. In 2019 5th International Conference on Engineering, Applied Sciences and Technology (ICEAST) (pp. 1-4). IEEE.

[12] Nordén, F. and von Reis Marlevi, F., 2019. A Comparative Analysis of Machine Learning Algorithms in Binary Facial Expression Recognition.

[13] Singhal, P., Singh, P. and Vidyarthi, A., 2020. Interpretation and localization of Thorax diseases using DCNN in Chest X-Ray. Journal of Informatics Electrical and Electronics Engineering (JIEEE), 1(1), pp.1-7.

Authorized licensed use limited to: VIT University- Chennai Campus. Downloaded on March 11,2024 at 13:49:12 UTC from IEEE Xplore. Restrictions apply.