Journal of Soft Computing Paradigm (ISSN: 2582-2640)
www.irojournals.com/jscp/

Design of a Music Recommendation Device Using Mini-Xception CNN and Facial Recognition

Chandan Singh [1], V Himayanth[2], Dr. B. Balakiruthiga[3]

Department of Networking and Communication SRM Institute of Science and Technology Chennai, India

Email: [1]cs6637@srmist.edu.in, [2]hv9060@srmist.edu.in, [3]balakirb@srmist.edu.in

Abstract

Due to the emerging developments in Artificial Intelligence and Machine Learning Technologies, various prediction systems are been developed based on human emotions and real time aspects of human psychology as well. Facial recognition system is one such mechanism which is the most vibrant strategy used for predicting human emotions. It is extensively applied in surveillance systems, fault identification and other security related aspects. Based on the human emotions researchers have already proposed several music recommendation systems. This paper aims to propose a Facial recognition-based music recommendation system to treat the psychology patients. This helps to recover the patients from mental stress, anxiety, and depression. The suggested method aims to take into account the limitations of the face recognition system in current frameworks, such as the requirement to lower the processing delay for deep feature extraction and the necessity to design a Mini exception technique based on Deep Convolutional Neural Network (DCNN) architecture. The FER- 2013 image dataset, which consists of 35000 face photos with automated labelling is considered. It is used to determine how well the proposed approach would detect the various emotion classes. In comparison to other states of methods, the Mini exception technique utilised in CNN layers acts as a lightweight system. The proposed solution has a 92% accuracy rate and removes the barrier between the current frameworks. The suggested music is taken from a music database and then further mapped in accordance with the algorithm's output.

DOI: https://doi.org/10.36548/jscp.2023.2.007
Received: 03.05.2023, received in revised form: 03.06.2023, accepted: 18.06.2023, published: 30.06.2023

Keywords: Artificial Intelligence (AI), Machine Learning (ML), Deep Convolutional Neural Network (DCNN), Facial recognition-based Music Recommendation System (MRS).

1. Introduction

Human emotions are special and unpredictable emotions. Real human emotions are conveyed through facial expressions., human emotions can be expressed through behavior in many ways. A challenging area of research in artificial intelligence is the ability to recognize human emotions and recommend music based on those emotions [1].

Deep learning algorithms are often used in music recommendation systems to automatically suggest musical patterns that users like. Systems that suggest music can assist people unwind during lengthy workdays and in a variety of settings. Nowadays, it is more important than ever to recognize human emotions, since scientific developments have revealed the true emotions that people experience every day. Human emotions can be divided into happiness, indifference, anger, contempt, reserve, fear, etc [2].

The first manifestation of human emotions is facial expression, including the movement of facial features such as the eyes, nose, lips, and eyebrows. Research on contemporary techniques, including deep learning algorithms and task learning algorithms for neural networks, has led to increased sensing of the human emotional system. There are many music recommendation datasets available, including 35,000 sentiment-tagged photos in 2013. Facial bone changes related to different emotions are included in the facial data collection [3].

In the facial emotion recognition system, deep convolutional neural network algorithm and recommender algorithm are used to classify facial emotions, such as joy and sadness, accurately. Related studies investigated many limitations of face recognition systems and precisely created a face recognition system based on the Mini-exception model, which solved the problem of feature extraction and similarity [4].

One of a kind are human emotions. Many facial features are associated with the expression of facial emotions. True emotions cause all kinds of physical changes in people. Emotional changes affect physiological changes, such as skin temperature etc. Thanks to current developments in affective computing, Researchers are now able to create predictive models of human emotions [5].

- The proposed method considers the FER 2013 face dataset, which provides 35,000 photos of different emotions labeled by the dataset.
- Advances in deep convolutional neural network (DCNN) design and development of planning and testing methods.
- To safely extract human emotions, lightweight architecture is used.
- The advantage of the proposed method is that it efficiently recognizes emotions by considering the multilevel process and multiple iterations of the CNN Mini-exception algorithm.

A comprehensive review of the available literature is offered as the remaining part of the study in Part II. Section III talks about choosing system resources and diagnostics. Section IV discusses the system's framework and specific system layout procedures. Future improvements are discussed in this article's remaining sections.

2. Background Study

[1] H. -G. Kim et al., (2019) The author developed an approach that makes use of deep leftover bidirectional neural networks that are recurrent. The looks on face change constantly and continuously. To improve the system's efficiency, a collaborative approach is used. A changing spectrometer is used, along with both short- and long-term spectrograms. Recurrent neural networks consider a number of variables and continuously adjust how well it performs in response to environmental changes.

[2] D. Wang et al., (2021) The author demonstrated a low-dimension dense network-based context-aware music recommendation system. When creating automated systems, it is important to take into account the essential elements of the living style, which includes music. The convolution neural network (CNN) technique is used to identify facial expressions of emotion. Context-aware framework also takes into account how the systems interact.

[3] W. Gong et al., (2021) Proposed a deep musical recommendation algorithm built around dancing analysis of motion in 2021, and its effectiveness is quantified. In order to evaluate music recommendations quantitatively, this work uses an LSTM-AE based technique which analyses the relationships among motion and music. Comparative testing of the two methods reveals how the movement analysis-based approaches perform noticeably better. This research also suggests a quantitative assessment of the best musical subgenre. The proposed motion assessment-based technique obtains an estimated accuracy of 91.3% using the final fusion of joints & limbs data.

[4] I. Agrawal et al., (2021) proposed an approach that examines several topology for convolutional neural networks. This article discusses the recognition of human emotions using a database of facial expressions. The use of a work-efficient combination of hybrid with a convolutional neural network framework is explored to tackle 35,000 human portrait images, significantly cutting down on calculation time. Actual-time constants have been considered in the implementation. The algorithm for identifying emotions under discussion provides a mechanism for extracting information from the world's data collection. For the validation process to be completed accurately, cross validation is crucial.

[5] S. Begaj et al., (2020) Using multifaceted facial information and a neural network design, the innovators were able to discern some of the human emotions associated with facial highlights that were released in the advertising strategy. It can be challenging to predict human sincere emotion, therefore analysing facial expressions may be a challenging task.

[6] Z. Rzayeva et al., (2019) The proposed RAVDESS and ohn-Kanade dataset were both taken into account. Convolution neural organize design is utilized to recognize facial expressions of feeling. There were many limits when thinking about face expression differentiating evidence. To discriminate between different facial traits, special measures are used. A layer of input, filtering convolution layers, a yielding layer, etc. are just a few of the many layers that make up the construction portion of a convolution neural system.

When comparing the various systems that are available, it is evident how the convolution neural organise plan is used in many different situations. Computation using the HOG histogram for angles plus the Haar cascading technique are frequently employed to

determine face expressions. The Haar cascade representation is a commonly used method for identifying facial components.

3. System Design

A. Problem Analysis

- There are some limitations to using images for facial emotion recognition. A specific cancellation environment that can provide characteristic features is used to capture a photograph of a human face. When it comes to identifying facial features, pixels in photographs are considered a key part of pattern recognition.
- In the case of light changes in different environments, these characteristics are immediately affected by specific environmental changes. Facial emotion recognition requires more complementary features to make accurate decisions. Although it has facial bones, features such as eyes, nose, and mouth change over time.
- Recognition of facial emotions is mapped. The eyes, nose, mouth, eyebrows and other facial organs are directly affected by changes in human emotions.
- The systematic approach automatically assesses a multilevel database of facial emotions to analyze these changes and identify specific emotions.
- Image-based facial expressions require additional attention when selecting physiological properties that can be extended to future research projects.

B. FER2013 Dataset

The FER (Facial Emotion Recognition) dataset records approximately 30,000 facial changes, compiled by various volunteers. The shots are recorded as RGB images, along with various emotions recorded and tagged appropriately. The public can access this dataset for analysis. Most 48x48 expressions are labeled. Note the following emotional states: 0 for anger, 1 for fear, 2 for fear, 3 for happiness, 4 for sadness, 5 for surprise, and 6 for neutrality.

Use standard datasets as the basis for analysis.

C. System Architecture

The figure Fig. 1 below is the architecture diagram of the proposed Mini-Xception system for emotion recognition. The starting handling steps of the framework incorporate perusing the input image, resizing the image, and changing over the image to grayscale. To find facial objects, additional processing is applied to the normalized image.

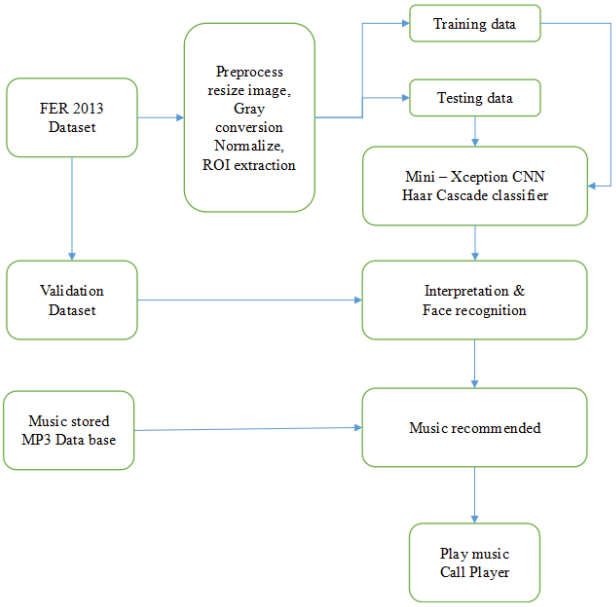

Figure 1. Proposed MRS system design using Mini-Xception System

Fig. 1. The framework design of the proposed MRS is illustrated utilizing the Mini-Xception System.

4. Methodology

A. Haar Cascade Model

The Haar Cascade model can recognize facial features such as lips, nose and eyes. The algorithm takes live photos into account and accurately segments facial components, regardless of image scale or position. Haar's cascade classifier uses multiple trajectories of a large number of image samples to find similar patterns between them. These functions are stored in .XML files.

The bounding box draws attention to the region of pixels that corresponds to the face object pattern. The boundary lines are modeled using the Voila Jones algorithm. Through the use of high-speed processing techniques, the Voila Jones method precisely segments parts of the face. After the facial features are segmented, the dataset is divided into training and test sets. About 20% of the training images are used as test data.

B. Deep CNN

A powerful deep learning method called Deep Convolutional Neural Network (ConvNet or DCNN) is frequently used in computer vision and video processing applications. The information layer, the output layer and the hidden layer constitute a DCNN. Deep learning applications use many pre-trained networks. The prefabricated layers used by the Visual Mathematics Beam Network to prepare the image dataset are included in the CNN model's VGG network. In order to achieve higher accuracy in the image classification process, CNN and VGG Net are powerful designs.

The design includes varying degrees of access to delineate key components and associated spatial highlights. The informative image is sent to a channel with soft convolution to extract unique qualities from the image.

A powerful deep learning convolutional brain network called VGG Net is designed to support 16 additional layers. Characterize using a sequential inclusion extraction procedure. The information image has a range of pixel values from 0 to 255. In the pre-processing step, the average of the information image is compared to the overall average determined from the PictureNet preparation dataset.

The input layer, the convolutional layer and the fully connected layer constitute the CNN architecture.

To reduce the correspondence problem in CNN, a Soft-max layer is used between the fully connected layer and the convolutional layer.

C. Mini-Xception Model

Convolutional neural networks are designed for in-depth analysis of inputs to quickly find characteristic patterns. Many later analyzes with improved layer development still use

robust structures. The Xception model considers a batch normalized augmented convolutional 2D layer added to each layer. After the micro-batch normalization procedure is completed, the maximum pooling layer is available. Depth-separable convolutions provide greater accuracy.

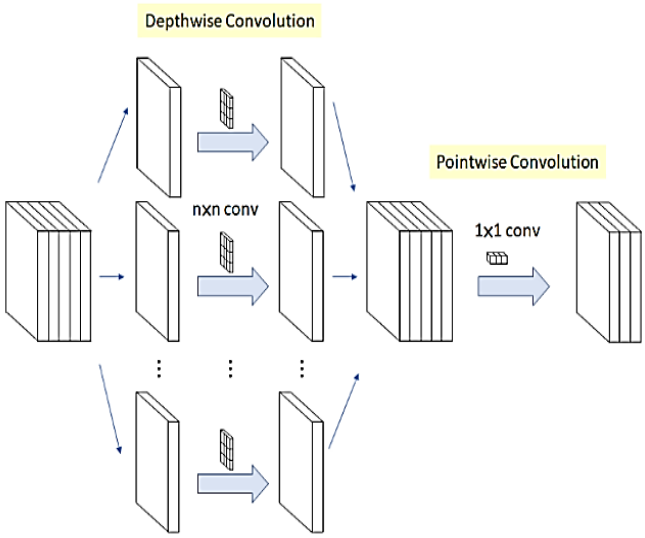

Figure 2. Demonstrates the Mini-Xception System

- The values can be compared more deeply with the training images by inducing spatially varying depth-separable convolutions.
- Compared to conventional neural organize structures, the number of associations is expanded and it is lighter than other sorts of systems. Made strides discovery of non-linearity in input photographs.
- A starter module is a special small configurable design that can be placed wherever a convolutional neural network needs it. The complexity of the input data affects the number of time periods used.

5. Results and Discussions

A. Emotion Detection

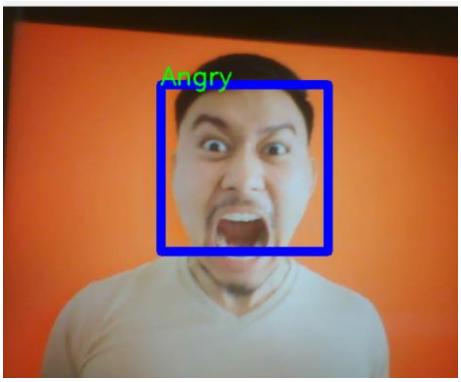

Figure 3. Emotional Intelligence: Angry

Figure 3. Revealing Emotion Detection Results Anger was Detected.

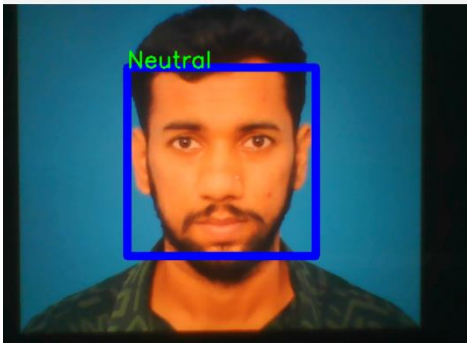

Figure 4. Emotional Intelligence: Unbiased/ Neutral

Fig. 4. Portrays the feeling as neutral. The Haar cascade algorithm is used to detect the face objects at first. With the help of the Haar cascade approach, the bounding box is made, and the Mini-xception model is used to do more thorough analysis. As a result, the facial expression is correctly recognised.

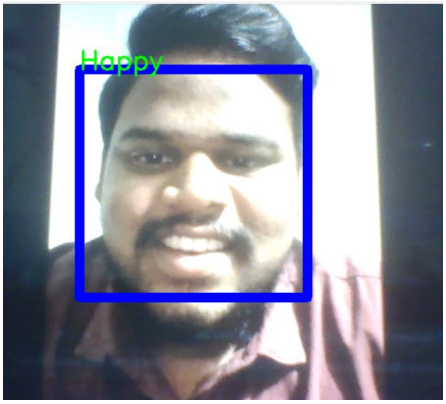

Figure 5. Emotional Intelligence: Happy

Fig. 5. Illustrates happy emotion detection.

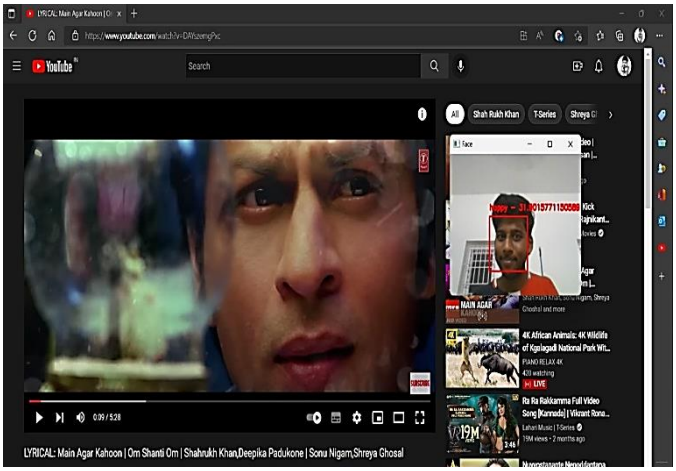

Figure 6. Music Recommended

Fig. 6. displays the outcomes of a suggested Mini-Xception CNN-based music recommendation system. The music meta data is compared and advised using a clever recommendation engine based on the emotions observed.

6. Challenges

The use of large datasets remains a major problem for the proposed models. Larger image data sets require additional processing, which consumes more GPU memory. As a result, system performance suffers. The input photo should be sampled and scaled before processing.

For in-depth extraction and better accuracy, it is recommended to use multiple classes for feature extraction.

7. Conclusion

Human emotions are special. Everyone uses a variety of external cues to express how they feel. Common signs of emotional impact include changes in body language, skin temperature, and facial expression. Traditional approaches to emotion detection and physiological analysis are being improved by new assessments of affective computing and artificial intelligence tools. The main objective of this research is to create such a scenario.

Here, a music proposal system based on Mini-Xception CNN is proposed. The proposed technique has an accuracy rate of 92% plus a high error rate of 0.00125. The presented framework was improved by advanced evaluation like highlight extraction, , thorough waterfall analysis using an updated Xception etc.

References

[1] S. Gilda, H. Zafar, C. Soni and K. Waghurdekar, "Smart music player integrating facial emotion recognition and music mood recommendation," 2017 International Conference on Wireless Communications, Signal Processing and Networking (WiSPNET), Chennai, India, 2017, pp. 154-158, doi: 10.1109/WiSPNET.2017.8299738.

[2] A. V. Iyer, V. Pasad, S. R. Sankhe and K. Prajapati, "Emotion based mood enhancing music recommendation," 2017 2nd IEEE International Conference on Recent Trends in Electronics, Information & Communication Technology (RTEICT), Bangalore, India, 2017, pp. 1573-1577, doi: 10.1109/RTEICT.2017.8256863.

[3] B. Lin, M. Liu, W. Hsiung and J. Jhang, "Music emotion recognition based on two-level support vector classification," 2016 International Conference on Machine Learning and Cybernetics (ICMLC), Jeju, Korea (South), 2016, pp. 375-389, doi: 10.1109/ICMLC.2016.7860930.

[4] H. Jun, L. Shuai, S. Jinming, L. Yue, W. Jingwei and J. Peng, "Facial Expression Recognition Based on VGGNet Convolutional Neural Network," 2018 Chinese Automation Congress (CAC), Xi'an, China, 2018, pp. 4146-4151, doi: 10.1109/CAC.2018.8623238.

[5] L. Xu, M. Fei, W. Zhou and A. Yang, "Face Expression Recognition Based on Convolutional Neural Network," 2018 Australian & New Zealand Control Conference (ANZCC), Melbourne, VIC, Australia, 2018, pp. 115-118, doi: 10.1109/ANZCC.2018.8606597.

[6] A. Alrihaili, A. Alsaedi, K. Albalawi and L. Syed, "Music Recommender System for Users Based on Emotion Detection through Facial Features," 2019 12th International Conference on Developments in eSystems Engineering (DeSE), Kazan, Russia, 2019, pp. 1014-1019, doi: 10.1109/DeSE.2019.00188.

[7] B. Verma and A. Choudhary, "A Framework for Driver Emotion Recognition using Deep Learning and Grassmann Manifolds," 2018 21st International Conference on Intelligent Transportation Systems (ITSC), Maui, HI, USA, 2018, pp. 1421-1426, doi: 10.1109/ITSC.2018.8569461.

[8] M. Wang, Z. Wang, S. Zhang, J. Luan and Z. Jiao, "Face Expression Recognition Based on Deep Convolution Network," 2018 11th International Congress on Image and Signal Processing, BioMedical Engineering and Informatics (CISP-BMEI), Beijing, China, 2018, pp. 1-9, doi: 10.1109/CISP-BMEI.2018.8633014.

[9] B. Subarna and D. M. Viswanathan, "Real Time Facial Expression Recognition Based on Deep Convolutional Spatial Neural Networks," 2018 International Conference on Emerging Trends and Innovations In Engineering And Technological Research (ICETIETR), Ernakulam, India, 2018, pp. 1-5, doi: 10.1109/ICETIETR.2018.8529105.

[10] Jun, He, et al. "Facial expression recognition based on VGGNet convolutional neural network." 2018 Chinese Automation Congress (CAC). IEEE, 2018.

[11] J. L. Joseph and S. P. Mathew, "Facial Expression Recognition for the Blind Using Deep Learning," 2021 IEEE 4th International Conference on Computing,

Power and Communication Technologies (GUCON), Kuala Lumpur, Malaysia, 2021, pp. 1-5, doi: 10.1109/GUCON50781.2021.9574035.

[12] K. -C. Liu, C. -C. Hsu, W. -Y. Wang and H. -H. Chiang, "Facial Expression Recognition Using Merged Convolution Neural Network," 2019 IEEE 8th Global Conference on Consumer Electronics (GCCE), Osaka, Japan, 2019, pp. 296-298, doi: 10.1109/GCCE46687.2019.9015479.

[13] G. Yamaguchi and M. Fukumoto, "A Music Recommendation System based on Melody Creation by Interactive GA," 2019 20th IEEE/ACIS International Conference on Software Engineering, Artificial Intelligence, Networking and Parallel/Distributed Computing (SNPD), Toyama, Japan, 2019, pp. 286-290, doi: 10.1109/SNPD.2019.8935654.