International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 08 Issue: 11 | Nov - 2024 SJIF Rating: 8.448 ISSN: 2582-3930

A Facial Expression Recognition System forSong Recommendation using CNN

Prof. Ulka Bansode, Kamini Nagare, Suraj Raut , Atharva Pawar, Tejas Sharmale

Department of Computer Engineering, K. J. College of Engineering & Management Research, Pune.

kamikami12267@gmail.com

Abstract: **- The music has special connection with emotion of the person. One's mood can be improved by it in a special way. The classification of the emotion of music is a difficult research area because human perception is subjective. The emotional response of the user is closely related to the music recommendation system because most music is selected based on the listener's mood. Many studies have been conducted to determine how to identify emotions using various methods. These techniques have been useful in evoking the subject's feeling using a variety of devices and other hardware that can be quite expensive and inaccurate. On the other hand, observing the person's facial expression can be quite helpful in accurately identifying their mood or feeling. Hence the main goal of the proposed system is to identify an individual's facial emotions effectively in order to make appropriate music recommendations. The proposed system makes use of Convolutional Neural Networks (CNN) to train facial dataset to recognize various emotional reactions. This trained model is used to detect the mood of the person based on facial expressions and recommend song related to that emotion. The proposed system is also optimizing the results using fuzzy classification. The results demonstrate the effectiveness of the proposed methodology.**

Keywords: **- Music; Emotion; Song Recommendation; Convolution Neural Network (CNN); Mood Identification; Fuzzy classification**

I. INTRODUCTION

Music is very essential in our daily life. As we move towards the digital era, the impact of music grows more significant. One of the most crucial research areas is the classification of music according to facial expression because it is more functionally and content based. Due to the subjectivity of human perception, classifying the emotions that recommends music is a challenging topic of research. The variations in facial patterns and facial expressions reveal information about the person's emotional state and help to control a productive conversation with the person. In human interactions or communication, facial expressions are extremely important. Analyzing facial expressions involves evaluating and observing changes in face featuresas well as a variety of facial gestures. The most important aspect of mood detection is analyzing the situation. A person's issuecannot be solved if they are not recognized. Person's mood or the facial expressions gives better idea to recognize the situation. With the use of this method, a doctor can easily gain a deeper knowledge of the patient. In a similar manner, a psychologist can quickly identify the patient's current state of mind and offer the appropriate treatment. The identification of emotion of a human being is one of the most important and day to day tasks that are performed with reasonable accuracy by all healthy individuals. This has made it possible for us to communicate and work together effectively to complete difficult tasks.

Even while humans and other primates can recognize emotions rather simply, automated systems using computer vision approaches may find it challenging to do so. The expressions on one's face are important for understanding the individual, communicating with humans, and connecting with them. They also play an important part in a patient's medical rehabilitation and serve as a foundation for behavioral research. Mood detection based on the technique of taking face photographs main gives a highly practical method to non-

invasive mood identification. The lack of an effective approach for automatic identification of the emotion is highly complex endeavor that has been one of the most recurring problems in computer vision paradigm. The effective realization of the automatic emotion recognition can be applicable in a variety of different scenariosthat can improve the accessibility and the implementation of a plethora of different use cases. The realization of emotion recognition can enhance recommendation system because emotional responses are connected to a person's behavior and listening preferences.

The Convolutional Neural Network analyzes emotion and computes the accuracy of the system. The Entropy Estimation block examines the accuracy rate, which is used to construct the Entropy List. This Entropy List is then transmitted to the Decision making, which selects a response based on the person's facial expression. The music suggestion module receives this expression and uses it to provide the user with appropriate music depending on the emotional response it has identified. This study focuses on recognizing the user's facial expression at various moments and recommending appropriate music based on mood. The work is divided in different sections. In section 2, the literature survey is given to evaluating earlier efforts. The proposed methodology is broadly described under the section 3. Section 4 discusses the obtained results from the experimental process. And finally, in section 5 conclusions and the potential for future improvements are discussed.

II. LITERATURE SURVEY

A new issue of competency-based music suggestion is investigated by K. Mao et al. [1]. By accounting for voice pitch, intensity, and quality, they were able to model a singer's vocal prowess using a vocalist profile. In order to compute voice quality at runtime, researchers suggested supervised learning to build a speech quality evaluation function. Moreover, they provided a song model that supported vocalist matching. The proposed model is used to build a learning-to-rank technique for music recommendation using human-annotated ranking datasets.

The methods for generating exemplar-based representations for music described in paper [2]. It demonstrated that highaccuracy rates in tag-based music retrieval may be achieved by using the representation to train simple linear Support VectorMachine (SVM) model. The methods described here use labelled training data to build a discriminative classifier for music auto-tagging based on computed feature representations from unlabeled data used as examples in an overly comprehensive lexicon. The dictionary is composed of frame-level feature vectors randomly selected from a broad and diverse variety of unlabeled music clips in order to minimize losing short-time audio information due to the temporal integral and the duplicationof using feature vectors from a limited number of clips. In order to construct a classifier the authors use this dictionary to create feature representations for both the training and test samples. In order to determine the emotional content of music, Yang et al. [3] employed two fuzzy classifiers to evaluate emotional intensity along with the continuous emotional psychologymodel and regression modelling to predict the emotional worth of music.

In paper [4], author proposed a multilayer attention representation-based recommendation technique to distinguish the differences in music preferences among users. Using data such as user attributes and song content, the system mines the preference correlations between users and songs and learns the embedded representations of songs from a multidimensional perspective. It primarily addresses the following issue: to learn song representations through user-based attention networks, and then build song-based attention networks to learn user preference representations depend on the learned song representations, an embedded representation dependent on attention mechanism is proposed to mine different users' differential preferences for multidimensional features of the same song. To improve the accuracy of song recommendations and learn temporal dependencies from listening behavior, a temporal relationship recommendation algorithm that relies on the attention network is proposed. This algorithm will help users distinguish the degree to which various historical behaviorshave influenced their decisions.

Using the Sentimeter-Br2 measure and a specific correction factor based on the user's profile, R. L. Rosa et

al. [5] introduced a personal music recommendation system that relies on a new lexicon-dependent sentiment metric. The authors demonstrate how a low-complexity method that benefits consumer electronic devices can improve the performance of a music recommendation system by combining a lexicon-based sentiment intensity metre with a corrective factor. The correction factor depends on an individual's characteristics, which can be quickly determined through social media. As a result, the obtained sentiment value that is more precise.

Y. Chin et al. [6] developed a novel technique for prediction the probability density function (PDF) of musical emotion across valence-arousal (VA) emotion space. The presented approach not only addresses subjectivity but also accurately captures the feelings expressed in a musical composition. The idea is based on the notion that the PDFs of recorded training pieces may be combined to obtain the emotion distribution of an unknown musical piece using the same set of combination coefficients that can be used to reconstruct the unknown piece in audio space. They specifically used the K-Nearest Neighbor (KNN) model to construct the combination coefficients and subsequently accurately predict the emotion. Two datasets are explored for experiment namely, NTUMIR and MediaEval2013. The paper [7] provides a summary of the fuzzy logic method to emotional recognition of music. A fuzzy system model is created to categories musical compositions according to their arousal levels, with the pace of the songs serving as the system's input and the level of arousal serving as the output. They compared fuzzy logic approach with machine learning and found fuzzy logic gives them optimized result.

Q. Lin et al. [8] provide a novel ANN model for incorporating heterogeneous data in short-term music suggestions. The authors combine graphical data, textual data, and visual data. A high-dimensional representation of the entity, which includes the majority of the heterogeneous information found on online music platforms, is the result of the sum of these embedding findings. Then, RNN deep model is implemented to acquire short-term preferences for the user listening to music. The results show that their system outperforms the current mainstream music recommendation model in terms of recommendation impact. Most importantly, because the majority of real-world data have a long tail, the general recommendation algorithm will primarily suggest well-liked products to consumers.

S. Mo et al. [9] propose a novel method for music mood classification termed OMPGW, which provides an adaptive time-varying description of music signals with increased spatial and temporal precision. The proposed approach is used to extract audio features and is based on a mixture of three signal processing techniques. Ten various characteristics determined by the proposed OMPGW techniques include spectral centroid, spectral roll-off, spectral flux, spectral bandwidth, spectral contrast, spectral flatness measure, spectrum contrast, subband power, frequency cepstrum coefficient, and coefficient histogram. Five mood-annotated datasets were used in experiments to explain the efficacy of the suggested approach.S. H. Chang et al. explored a CNN model to recommend personalized music system [10]. Based on the audio signal that is present in the song, the CNN approach classifies music. The CNN classifies music based on hidden information extracted from auditory signals in the music. They developed an Android music application to demonstrate their operation.

In order to establish the user's long-term preferences as well as sequential intent in the present session, Y. Guo et al. [11] introduce a session-aware recommendation model that combines RNN and CNN. The results of the studies demonstrate that past data can help improve the precision of suggestions that are tied to sequential events. Nevertheless, the method requires a significant amount of computational power, which makes online practice difficult and has to be improved.

A course instructor recommendation system (FCTR-LFM) based on fuzzy clustering and latent factor model (LFM) was created by D. Yao et al. to improve the personalization and effectiveness of recommendations. Under the direction of pedagogy standards, the main work entails creating a number of strategies for achieving quantitative outcomes of instructor qualities, course features, and teaching performance [12]. These findings will be used to construct the experimental dataset. To present the data, a high-dimensional sparse evaluation matrix is used. In order to address the issues of sparsity reduction in the assessment matrix and teachers' cold starts, a fuzzy

clustering model for teachers is developed. This model enables instructors to automatically cluster based on their attributes. With the enhanced LFM, the evaluation matrix is split into the product of two low-dimensional matrices with implicit components.

The fuzzy approximation theorem and fuzzy inputs as latent spaces are highlighted in paper [13], which also discusses the fundamental concepts of fuzzy logic and its applications. Author also offers an updated analysis of the use of fuzzy logic in musical applications. The final presentation is the Fuzzy Logic Control Toolkit (FLCTK), a collection of tools for producing musical content in the MaxMSP real-time sound synthesis environment. Fourth, he demonstrates how many-to-many musical mappings, algorithmic composition, and applications in sound synthesis can be put to use. Lastly, he discusses some of the compositional elements of Incerta, an acousmatic multichannel work produced in MaxMSP using the FLCTK. The review of the literature suggests that a crucial topic for research is the facial expression-based mood identification. Based on the literature survey below is a summary of the main contributions to our work:

1. Implemented SongRec that is Song Recommendation system using person's emotion using CNN approach.
2. To achieve optimize result Fuzzy classification is explored.
3. To check the reliability of the system it is tested in real-time environment.
4. Our proposed system recognizes facial expressions like angry, disgusted, fearful, happy, neutral, sad, and surprised.

III. PROPOSED SONG RECOMMENDATION SYSTEM USING FACIAL EXPRESSION

Figure 1 show the architecture diagram of proposed song recommended system (SongRec) based on facial expression using CNN. Music is suggested based on emotion by the system. The recommendation system will recognize several emotions including happy, sad, fearful, angry, nutral, surprised etc. The presented approach achieves the prescribed goals through the use of the following steps.

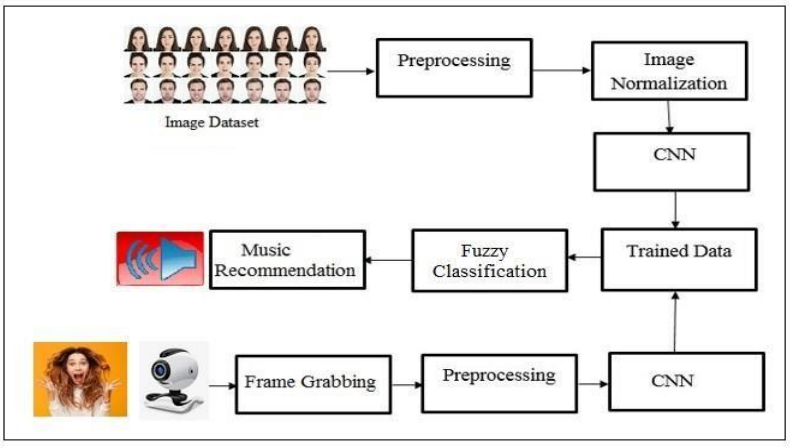

Figure 1. Architecture of song recommendation system using facial expression

The following steps are used by the proposed system for song recommendation.
1)Preprocessing and Image Normalization

The data was gathered without regard to the size or location restrictions, thus it could have inaccuracies or outliers. Hence, preprocessing of data is necessary before passing it forward for training. Before the training begins, the images of faces with various facial expressions like angry, disgusted, fearful, happy, neutral, sad, and surprised are resized to the dimension of 48 × 48. Data augmentation is the process of creating additional

training samples from the original samples using a variety of techniques [14]. When the model is being trained on batches of images, data augmentation is obtained usingthe ImageDataGenerator class of the Keras framework. By applying different random transformations it create new random transformed batch of images which is used for training model with the rescale ratio of 1./255 for the image assessment for boththe training and testing datasets. This process is used to capture all of the face expressions in images. The ImageDataGeneratorobject is initialized using parameters such as training and testing directory locations, image dimensions set to 48 × 48, batch size of 64, setting up the grayscale color mode, and categorical as the class mode.

2) Training of Convolution Neural Network (CNN)

Preprocessed images are then passed to the CNN for classification. The recognition of facial expression is actually performed in this step of the proposed methodology. This is one of the most important step in song recommendation system.A dataset consisting of the facial expressions collected from different users. This dataset is used to train the CNN model andtraining model is saved with the extension .h5 for testing data. CNN model consist of 3 main layers [15]. Convolution layer is first important layer used to extract different features from the input images using learnable filters. Second layer is poolinglayer which is used to reduce the network's training features and computations. 2×2 kernel is used to perform the pooling operation. Max pooling creates a single output by extracting the block's maximum value. Fully connected layer is classificationlayer. We can change this layer as per our requirement.

The dataset is split into training and testing folders as per standard 80:20 ratio. These folders are again segregated into folders specific to a particular expression, namely, angry, disgusted, fearful, happy, neutral, sad, and surprised. The model is trained on the images using a 64-batch size for 50 epochs. The proposed model consists of stack of three Convolution2D- BatchNormalization-ReLU-MaxPooling2D layers. Dropout is set to 25% while training which is used to avoid overfitting of the model. Adam optimizer is fast adaptive learning optimizer hence speed up the learning process. The Adam optimizer produces results that are better compared to those of conventional optimization methods, takes less time to compute, and needs fewer tuning parameters. Designing parameters of the model are depicted in Table 1.

3) Fuzzy Classification

This stage of the technique involves testing the system for facial expression. The laptop's inbuilt webcam is initialized using the OpenCV library and a test image of the user's face is taken. The Haar Cascade classifier is then used on this image to identify the facial region in the image. Once the facial region has been located, the image is passed through the trained model to find the expressions that match the face the best. The best matched expressions derived from the trained CNN modelare added into a list. The different detected facial expressions are counted in the list. The derived best matched expressions from the captured image are searched for the maximum count, whereas the minimum count is set as 1. The difference between the maximum and minimum count is achieved and segregated into 5 divisions as VERY LOW, LOW, MEDIUM, HIGH, andVERY HIGH and considered as the fuzzy crisp values.

Table 1. CNN model specification

Layer	**Activation**
CONV 2D 32x3x3	RELU
CONV 2D 64x3x3	RELU
MaxPooling2D	
Dropout 0.25	

CONV 2D 128x3x3	RELU
MaxPooling2D	
Dropout 0.25	
Dense 1024	RELU
Dropout 0.25	
Dense 7	Softmax
Optimizer	Adam

Then, using a specified threshold, the expressions are evaluated for their occurrence in the VERY HIGH category. Once an expression is detected, it is then used for the playing of the music. The system retrieves the appropriate music file randomly from the expression specific preselected folders of .mp3 files and starts to play the media.

IV. CONCLUSION

Convolutional Neural Networks and Fuzzy Classification have been used to achieve the research methodology for the goal of mood detection through face expression recognition. The real-time video from a web camera is used to capture the frames that show the person's face. These images are effectively retrieved and used for evaluation after proper preparation. To identify face regions from the preprocessed pictures, the Haar features are used. These images are then passed for facial expression detection. The right music will then be suggested and played based on the categorization and accurate mood detection performed using the fuzzy classification. The experimental evaluation demonstrated the effectiveness of the suggested model for mood detection. On real-time testing data, we achieved 62.88% testing accuracy with MSE and RMSE values of 8.5 and 2.9 respectively.

V. REFERENCES

[1] K. Mao, L. Shou, J. Fan, G. Chen, and M. S. Kankanhalli, "Competence-Based Song Recommendation: Matching Songs to One's Singing Skill," in IEEE Transactions on Multimedia, vol. 17, no. 3, pp. 396-408, DOI: 10.1109/TMM.2015.2392562, 2015

[2] P. Jao and Y. Yang, "Music Annotation and Retrieval using Unlabeled Exemplars: Correlation and Sparse Codes," in IEEE Signal Processing Letters, vol. 22, no. 10, pp. 1771-1775, DOI: 10.1109/LSP.2015.2433061, 2015.

[3] YH. Yang, CC. Liu, and HH. Chen, "Music emotion classification: A fuzzy approach", In Proceedings of the 14th ACM international conference on Multimedia, pp. 81-84, 2006.

[4] S. Juan, "Variational Fuzzy Neural Network Algorithm for Music Intelligence Marketing Strategy Optimization," Computational Intelligence and Neuroscience. pp. 1-10, 10.1155/2022/9051058, 2022.

[5] R. L. Rosa, D. Z. Rodríguez and G. Bressan, "Music recommendation system based on user's sentiments extracted from social networks," in IEEE International Conference on Consumer Electronics (ICCE), 2015, pp. 383-384, DOI: 10.1109/ICCE.2015.7066455, 2015.

[6] Y. Chin, J. Wang, J. Wang, and Y. Yang, "Predicting the Probability Density Function of Music Emotion Using Emotion Space Mapping," in IEEE Transactions on Affective Computing, vol. 9, no. 4, pp. 541-549, DOI: 10.1109/TAFFC.2016.2628794, 2018.

[7] Y. Ospitia-Medina, S. Baldassarri, C. Sanz, J. R. Beltrán and J. A. Olivas, "Fuzzy Approach for Emotion

Recognition in Music," *2020 IEEE Congreso Bienal de Argentina (ARGENCON)*, Resistencia, Argentina, 2020, pp. 1-7, 2020.

[8] Q. Lin, Y. Niu, Y. Zhu, H. Lu, K. Z. Mushonga and Z. Niu, "Heterogeneous Knowledge-Based Attentive Neural Networks for Short-Term Music Recommendations," in IEEE Access, vol. 6, pp. 58990-59000, DOI: 10.1109/ACCESS.2018.2874959, 2018.

[9] S. Mo and J. Niu, "A Novel Method Based on OMPGW Method for Feature Extraction in Automatic Music Mood Classification," in IEEE Transactions on Affective Computing, vol. 10, no. 3, pp. 313-324, DOI: 10.1109/TAFFC.2017.2724515, 2019.

[10] S. H. Chang, A. Abdul, J. Chen and H. Y. Liao, "A personalized music recommendation system using convolutional neural networks approach," IEEE International Conference on Applied System Invention (ICASI), 2018, pp. 47-49, DOI: 10.1109/ICASI.2018.8394293, 2018.

[11] Y. Guo, D. Zhang, Y. Ling, and H. Chen, "A Joint Neural Network for Session-Aware Recommendation," in IEEE Access, vol. 8, pp. 74205-74215, DOI: 10.1109/ACCESS.2020.2984287, 2020.

[12] D. Yao and X. Deng, "Teaching Teacher Recommendation Method Based on Fuzzy Clustering and Latent Factor Model," in IEEE Access, vol. 8, pp. 210868-210885, DOI: 10.1109/ACCESS.2020.3039011, 2020.

[13] R. Cádiz, "Creating Music With Fuzzy Logic", Frontiers in Artificial Intelligence, Vol. 3, 10.3389/frai.2020.00059, 2020.

[14] L. Perez and J. Wang, "The effectiveness of data augmentation in image classification using deep learning," In: arXiv preprint arXiv:1712.04621, 2017.

[15] K. Simonyan and A Zisserman, "Very Deep Convolutional Networks for Large-Scale Image Recognition", In: arXiv:1409.1556v6,2015.