

AI-Based Music Therapy Recommendation System

Nithya Anand Raj¹, Aarti Sanjay Gawai²

¹Student, Department of Information Technology, D. G. Ruparel College of Arts, Science and Commerce, Mumbai, Maharashtra, India

²Assistant Professor, Department of Information Technology and Computer Science, D. G. Ruparel College of Arts, Science and Commerce, Mumbai, Maharashtra, India

¹nithya87794@gmail.com, ²aartigawai1611@gmail.com

Abstract—Personalized music recommendation systems often consider past choices without conveying user feelings. Face emotion recognition using CNN is utilized in a music recommendation framework. As per FER2013, when seven emotions are classified into a single category, their emotional states are identified as representing musical characteristics such as mood, valence, and energy. This concept is supported by the German metric system model (BDMS) developed by Bonn in Germany. Furthermore, An increase in data leads to a reported accuracy of 63.527%, which then increases to an ultimate validation accuracy level of 93.31%. Experimental results support this. Feedback by its users, the framework can be adapted to them. Although these results highlight the potential for fusing face-to-face emotion recognition with adaptive music recommendation, further independent test evaluation and user studies are needed before final approval.

Keywords: *Facial Emotion Recognition; CNN; Deep Learning; FER2013; and Affective Computing.*

1. INTRODUCTION

More personalized recommendation algorithms are realizing the significance of music in conveying human emotions. Moreover, while they consider listening to history and user preferences, these methods may not always accurately capture the user's emotional state. People can select music that has a limited taste using emotion-aware recommendation, which uses emotional information to guide their selection. Multiple studies have explored the use of CNN, Deep Face, and transfer learning to integrate face recognition and music recommendation [3, 4, 6]. However, many modern systems still maintain a similar mapping of emotion to music and offer only moderate personalization. Moreover, facial gestures can indicate real outcomes, but they may not always reflect the individual's feelings. By incorporating CNN and user preferences, the paper suggests a music recommendation system to address these shortcomings. The mechanism employs facial emotion recognition techniques.' Musically, the candidate tracks are categorized by their features and user preferences (including other emotional factors). The system is now more user-friendly, allowing for a wider range of music options.

2. LITERATURE REVIEW AND RESEARCH GAP.

It has become increasingly common to learn about spatial features by adding contourable layers to facial images, thanks to the use of CNN-based facial emotion recognition. The use of CNN and music recommendation for facial emotion identification was first introduced by Bakariya et al. [7, 8]. Using CNN-based emotion recognition and transfer learning, Xpressify can retrieve music dynamically while also using CNN to classify facial expressions. Moreover, they have found that DeepFace and OpenCV can be used to identify real-time facial emotion selection and choose the appropriate music. Numerous current investigations confirm this. By using face language to recommend songs, emotions can be expressed through the representation of facial expressions. In addition, models that focus on emotion also consider the connection between musical traits and emotional messages. Furthermore, certain researchers have discovered that the design of most data sets in music systems is not as well-developed. 3. Adru and Johnson have identified three crucial deficiencies in their approach to suggesting music through facial expressions, despite using transfer learning through LightGBM. The. There are not only physical indications, but also emotional ones. When music plays, it's effortless to tailor your emotions. Why? Correct emotion classification is a common focus in current research, rather than being solely focused on recommendations or user satisfaction. The aim of the study is to achieve clinical validation for therapeutic use of systems that incorporate both facial emotion and user input as part of an adaptive model.

3. METHODOLOGY.

3.1. System Architecture.

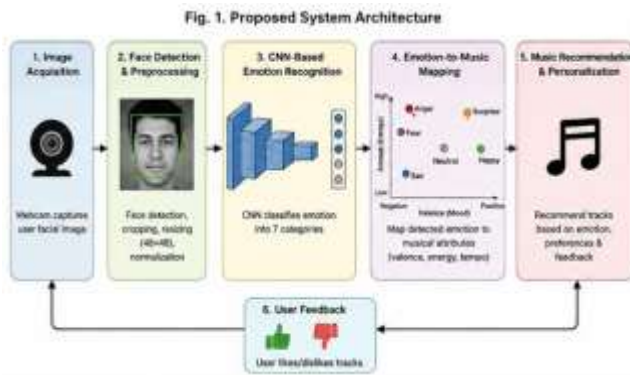


Fig. 1. Proposed System Architecture. (Source: Authors' own work, AI-assisted visualization.)

Six fundamental steps are included in the proposed system, including image selection (as opposed to digital images), facial recognition, preprocessing for emotion classification and categorization, music recommendation, and customization. Once the input image has been analysed, a CNN face detection feature is added and preconditioning follows. The face of the user is captured on a webcam. Finds out face and then sends it to the CNN for normalization by cropping and resizing. Finally, the desired experience is rewired into music-related components, and potential tracks are categorized based on user preference and emotional connections.

3.2. Dataset and Preprocessing.

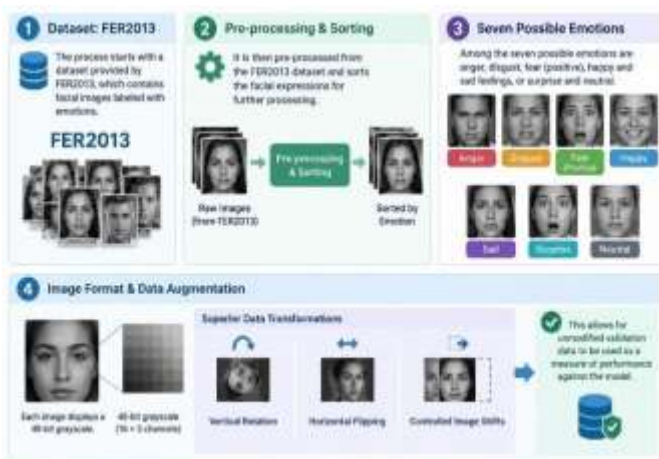


Fig. 2. (Source: Authors' own work, AI-assisted visualization.), based on FER2013.

It is then pre-processed from a dataset provided by FER2013 and sorts the facial expressions for further processing. [A]. Among the seven possible emotions are anger, disgust, fear (positive), happy and sad feelings, or surprise and neutral. Each image displays a 48-bit grayscale. Moreover, this is accompanied by superior data transformations such as vertical rotation, horizontal

flipping, and controlled image shifts. This allows for unmodified validation data to be used as a measure of performance against the model.

3.3. Emotion recognition using CNN.

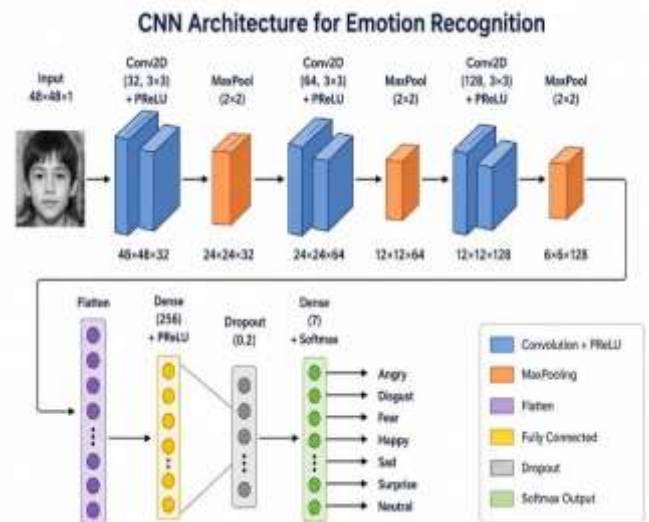


Fig. 3. CNN Architecture for Emotion recognition. (Source: Authors' own work, AI-assisted visualization.)

The proposed CNN will have convolutional and activation layers, as well as either fully connected or pooled layers with layer dropout, to minimize overfitting. The FER2013 dataset provides output neurons that represent all seven emotion classes in the final layer. By using the softmax function, it is possible to convert spatial features from face images into probability classes of the emotion class. The selection of an emotion with the highest predicted probability in the classification process leads to training losses due to categorical cross-entropy.

3.4. Emotional Awareness Music Recommendation.

Various musical traits, such as mood, valence, energy, and speed are linked to the facial emotion detected. The ranking of candidate songs is based on their emotional compatibility, user likenesses, and previous experience. Those who do are given the most popular songs and then shown their profile, which allows them to leave feedback for future suggestions.

3.5. Performance Evaluation.

The CNN model is evaluated using accuracy, precision, recall, F1-score, and a confusion matrix. Since the FER2013 dataset is imbalanced across emotion classes, macro-averaged evaluation is considered important for obtaining a balanced assessment of classification performance. The recommendation component is evaluated separately using ranking measures and user feedback to assess the relevance and effectiveness of the recommended music.

4. RESULTS AND DISCUSSION.

4.1. Dataset Distribution.

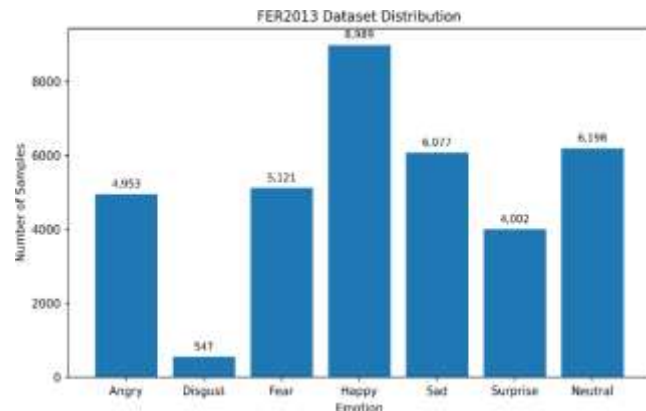


Fig. 4. Source: Author-generated graph based on the FER2013 dataset class distribution.

A diverse range of images is present in the FER2013 dataset, totalling 35,887. Training samples from the "disgust" category are less common than those from both "happy" and "neutral." This difference gives preference to macro-averaged metrics over flat accuracy scores.

4.2. Experimental Configuration.

The deep learning model undergoes training using the parameters provided below.

TABLE I: CNN Experimental Parameters.

Parameter	Value
Dataset	FER2013
Classes	7
Input size	48 × 48
Channels	1
Epochs	100
Batch size	128
Learning rate	0.1
Optimizer	Adadelta
Activation	PReLU
Dropout	0.2
Augmentation	Applied

4.3. Training Performance.

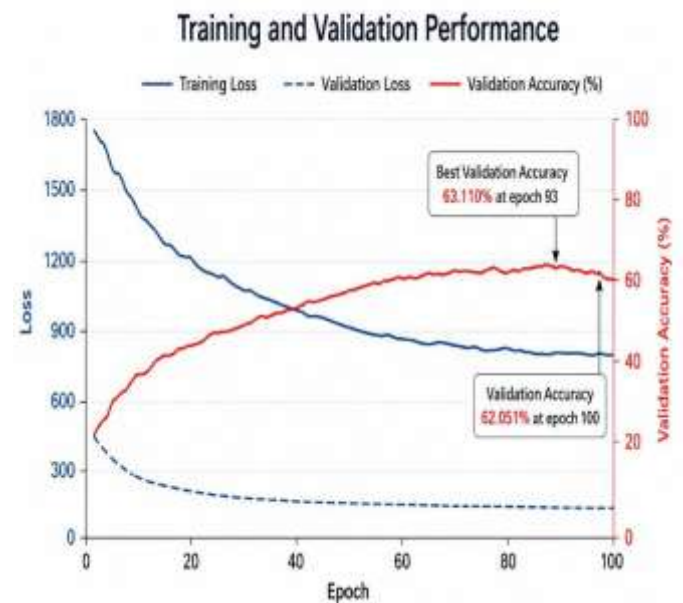


Fig. 5. Training and Validation Performance. (Source: Author-generated based on experimental results.)

Demonstrates how the model acquires knowledge with complete accuracy during its training period. The validation was initially estimated to be 22.903% accurate, but it was later increased to 63.110% by 93.5%. We then review and assess performance.

TABLE II: Training & Validation PERFORMANCE

Epoch	Training Loss	Validation Loss	Validation Accuracy (%)
1	1685.936	106.205	22.903
20	1345.017	74.276	51.351
40	1238.026	66.016	57.565
60	1145.656	64.056	59.961
80	1055.237	70.135	61.438
93	999.368	74.468	63.110
100	977.049	78.674	62.051

As time passed the century, there was an increase in validation loss and a decrease in accuracy to 62.051.0%. The network was overfitted during the last stages of training, making epoch 93 the most appropriate checkpoint for the model. This is implied.

4.4 Effect of Data Augmentation.

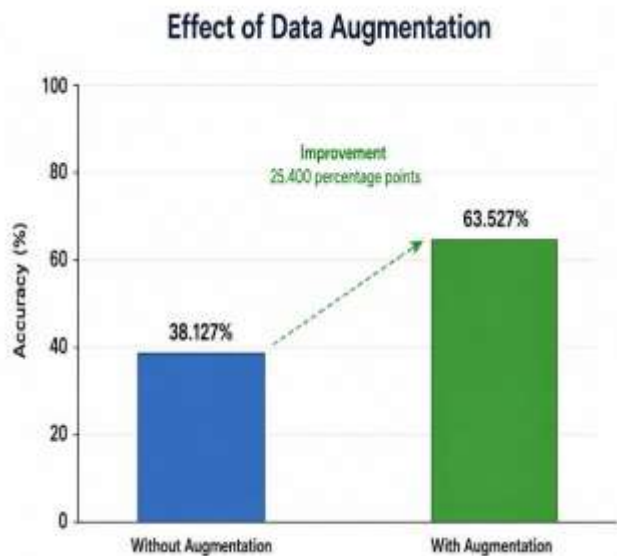


Fig. 6. Effect of Data Augmentation. (Source: Author-generated based on experimental results.)

By augmenting the data, this model became successful. Using flips and shift, the model's baseline was artificially expanded by 25.400 percentage points, which helped it achieve the desired accuracy by artificially expanding the training data from 38.127% to 63.527%. This was reflected in that the capacity of the network to generalize invisible faces improved greatly under growth.

4.5. Comparison with Existing Approaches.

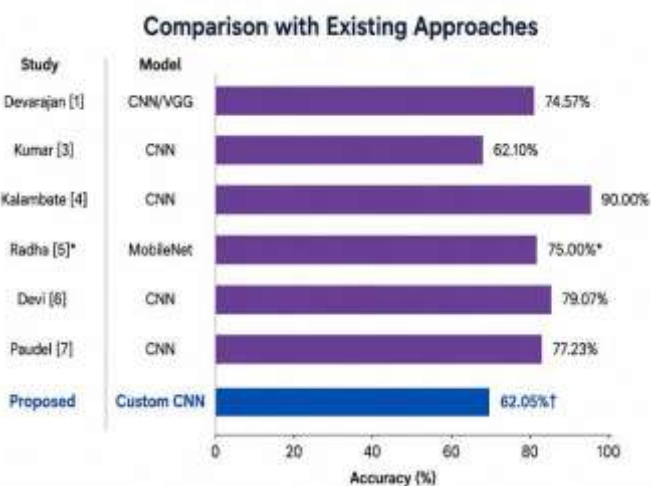


Fig. 7. Source: Author-generated graph based on accuracy data reported in the referenced studies.

If the researchers don't use the same evaluation protocol, it's difficult to compare methods with existing ones, but our findings are like those of modern CNN-based emotion recommenders. In contrast, our focus is on hybrid recommendation and not on image classification enhancements. Faces are a dynamic concept that doesn't conform to any one number. Why? Keep in mind that

the material specifies training accuracy, but validation accuracy at 100 and independent test precision are still being worked into but not tested.

4.6. Limitations.

This system has some limitations. Due to the unbalanced nature of the FER2013 dataset, the images do not reflect all the quality. The system has a problem? The facial expressions of others can provide insight into their emotional state. It's not always clear how much they are emotionally affected. Users also have varying tastes in music. We should view the system as a means of helping individuals discover music they enjoy using intelligence, rather than treating people accordingly.

5. CONCLUSION.

This paper is focused on a music recommendation system that utilizes intelligence. The system suggests music based on people's emotional state through analysis. The FER2013 dataset is utilized to determine people's emotions through expressions. Our experiments demonstrate that the system is more proficient at analyzing people's emotional state by incorporating information into its data analysis. The system was accurate at 38.127 percent initially... It was 63.527 percent accurate after including more data. This is an improvement. It had the highest accuracy level of 63.110 percent after only 393 attempts. After 100 tries, the score was still pretty good, with only 62.051 percent remaining.

People can be suggested music by emotion recognition, as demonstrated by the results. The system's functionality must be tested to ensure its reliability. To gauge people's opinions on the music suggested by our system, we must question their judgment. We will improve the system in due time by implementing diverse emotional recognition models and utilizing more real-world data. The system will be tailored to suit the needs of every user. Through the use of emotion recognition and music preference data, the music recommendation system can suggest songs to its users.

6. FUTURE SCOPE.

It suggests an alternative music therapy recommendation system that is more accurate, customizable and useful in practice. The device can operate efficiently in various lighting scenarios and facial orientations, as well as better handling blurry faces and differences in camera clarity. Additionally, it is capable of recognizing a partial faces. By merging face recognition with voice, gestures, or physiological data, the system can enhance the user's emotional state and expand its multimodal emotion recognition capabilities. Over the years, it is possible to incorporate user feedback and listening records with distinct tastes, resulting in more tailored music recommendations. Devices that are restricted by

resources, such as mobile phones, can be optimized using lightweight models. Music-streaming APIs could be integrated to offer a wider and more varied selection of songs for users. The use of larger and more diverse datasets can lead to improved system reliability and generalizability in future research.

REFERENCES

- [1] Devarajan et al., “Facial Emotion-Based Music Recommendation,” as cited in the supplied literature.
- [2] Tran et al., “Emotion-Aware Music Recommendation,” as cited in the supplied literature.
- [3] Kumar et al., “Facial Emotion Recognition and Music Recommendation Using CNN,” as cited in the supplied literature.
- [4] Aswal et al., “Facial-Expression-Based Musical Therapy System,” as cited in the supplied literature.
- [5] Radha et al., “Facial Emotion Recognition Using MobileNet,” as cited in the supplied literature.
- [6] Devi et al., “Xpressify: CNN-Based Emotion Recognition and Music Recommendation,” as cited in the supplied literature.
- [7] Bakariya et al., “CNN-Based Facial Emotion Recognition and Music Recommendation,” as cited in the supplied literature.
- [8] Kurniawan et al., “Facial-Expression-Based Music Recommendation Using Deep Learning,” as cited in the supplied literature.
- [9] Adru and Johnson, “Facial Expression-Based Music Recommendation Using Transfer Learning and LightGBM,” as cited in the supplied literature.
- [10] Wankhade et al., “Deep Learning-Based Music Recommendation Using Emotion and User Preferences,” as cited in the supplied literature.
- [11] Pravallika, “Facial Emotion Recognition and Music Recommendation Using DeepFace and OpenCV,” as cited in the supplied literature.
- [12] Antony et al., “Face Emotion Based Music Recommendation System Using Modified Convolution Neural Network,” as cited in the supplied literature.
- [13] Athavle et al., “Music Recommendation Based on Face Emotion Recognition,” as cited in the supplied literature.
- [14] Rajesh et al., “Music Recommendation Based on Facial Emotion Recognition,” as cited in the supplied literature.

- [15] Mala et al., “Emotion-Based Music Recommendation System Using Facial Expression Recognition and Deep Learning,” as cited in the supplied literature.

BIOGRAPHIES



Ms. Nithya Anand Raj Nadar
M.Sc. Information Technology
Student., IT Dept. D. G. Ruparel College



Ms. Aarti Sanjay Gawai
M.Sc. Information Technology
Asst. Prof., IT/CS Dept. D.G. Ruparel College