iJRASET
International Journal For Research in Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH
IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 14 Issue: V Month of publication: May 2026

DOI: https://doi.org/10.22214/ijraset.2026.81815

Facial Expression Recognition Based Music Recommendation System Using Transfer Learning with Xception

Mohit Singh Mahara[1], Manan Verma[2], Niskam Chaudhary[3], Mohd Zaid Ali Khan[4], Rudresh Kaushik[5]
Raj Kumar Goel Institute of Technology & Management, Ghaziabad, India

Abstract: Emotion-aware recommender system is achieving importance in human -computer interaction. Facial expression recognition (FER) provides an effective way to detect human emotion using computer vision technique. This research gives an intelligent music recommendation system that identify human emotion and recommend music using Spotify Api. This model uses transfer learning approach with pre-trained Xception convolutional neural network. The system is implement using Python and its libraries NumPy and Pandas, Spotify Api and OpenCV. The model is trained on FER2013 Balanced dataset available on Kaggle. The model classifies emotions into seven categories happy, sad, neutral, fear, surprise, angry and disgust. Xception based approach achieve higher accuracy compere to traditional machine learning algorithm: Support Vector Machine (SVM), k-Nearest Neighbors (KNN), Random Forest (RF), and Naïve Bayes (NB). The model achieves 92.4% accuracy which demonstration the effectiveness of the transfer learning for music's recommendation system.
Keywords: Facial Expression Recognition, Music Recommendation System, Transfer Learning, Xception, Deep Learning, Spotify API

I. INTRODUCTION

Human emotion is the key for communication decision making and human-computer interaction. Automatic emotion recognition had become important research in artificial intelligence. Facial expression recognition is widely used approach to detect human emotion as it provides significant patterns for human emotional state. FER system analyses the image or video of human face and classify it into happy, sad, neutral, fear, surprise, angry and disgust.

Recent advancement in computer vision and deep learning has improve the performance of Facial Expression recognition system. Traditional machine learning approach like Support Vector Machine (SVM), k-Nearest Neighbors (KNN), Random Forest (RF), and Naïve Bayes (NB) rely on manual feature engineering for emotion classification. However, these approaches struggle to find complex patterns in real world data. Deep learning technique particularly Convolutional Neural Networks (CNNs), automatically learn features and have demonstrate higher performance including FER.

One of the most widely used dataset is FER 2013 dataset introduced in the ICML 2013 facial expression recognition challenge. The dataset contains 35000+ greyscale images of human faces labelled with seven emotion categories. Due to variation in illumination, pose, and occlusion Fer2013 detest become challenging problem for machine learning algorithm and become standard benchmark for evaluation.

In recent years transfer learning has become and effective technique for improving the performance of deep learning model when training data is limited. Pre trained convolutional neural network architectures such as VGGNet, ResNet, and Xception have successfully applied to emotional task. Among these **Xcaption** which is depthwise separable convolutions had shown higher performance in image classification due to effective feature extraction and reduce computational complexity.

At the same time, music recommendation system become an essential component in modern digital entertainment platform. Personalized recommendation system suggests music based on user preferences, mood and contextual information. By integration Facial expression recognition with music recommendation system based on user emotion it can find the suitable music's for it. For example, when user is happy cheerful music can be suggested whereas when user is sad relaxing music can be suggested. Such emotion-based system has application in health, entertainment system and multimedia platform.

In this research we propose Facial Expression Recognition based Music Recommendation System using transfer learning with the Xception model. The system recommends music based on the user facial expression The system is implemented using Python and it's libraries NumPy, Pandas and OpenCV and Spotify Api for music recommendation.

To evaluate the performance of this approach the model is compared with tadeonal machine learning algorithms: Support Vector Machine (SVM), k-Nearest Neighbors (KNN), Random Forest (RF), and Naïve Bayes (NB). Experiment result tells that the proposed model achieves higher accuracy then other traditional models.

The main contribution of this research is summarized as follows.

1) Development of facial expression recognition model using Xception model and transfer learning.
2) Implementation of real- time emotion detection using OpenCV.
3) Integration Spotify Api for emotion-aware music recommendation.
4) Comparative analysis of proposed model with four other traditional models.

II. RELATED WORK

Facial Expression Recognition (FER) has drawn attention in field of computer vision, affective computation and human-computer interaction. Researchers have found many machines learning and deep learning technique to automatically recognise human emotion from facial images. Early study focuses on traditional machine learning algorithm using handcraft feature extraction whereas recent research uses deep learning architecture which is capable of learning complex representation.

One of the earliest and most contribution in FER research was **Facial Expression Recognition Challenge (FER2013)** introduced by Goodfellow et al. This dataset contains more then 35000 greyscale facial images categories into seven emotion classes and has become widely used dataset for emotion recognition research. The challenge highlighted the difficulty due to the variation in illumination, occlusion, and head pose. Since then, FER2013 has been mostly used for training and evaluation both in machine learning and deep learning model.

Traditional machine learning method dominate FER dataset. Technique such as Support Vector Machine (SVM), k-Nearest Neighbors (KNN), and Random Forest classifiers were applied using handcrafted features like Local Binary Patterns (LBP), Histogram of Oriented Gradients (HOG), and Gabor filters. Dagher et al. [13] demonstrated that SVM combine with facial feature information could achieve moderate accuracy. Similarly, several studies used KNN classifier due to simplicity and effectiveness in small dataset. However, these approaches rely on manually designed feature extraction technique which fails to capture the complex facial patterns and variations.

Random Forest classifier have also used for the FER problem. Ensemble learning approaches like Random Forest combine multiple decision tree to improve the performance. Khan et al. [7] evaluated classical machine learning algorithm; like Random Forest and Naïve Bayes for FER and reported the typically accuracy between 50% and 65% on the FER2013 dataset. Although these algorithms are computational efficient their performance is limited compare to deep learning approach due to inability to learn the hierarchical image features

In the advancement of deep learning, Convolutional Neural Networks (CNNs) become the most used approach for facial expression recognition CNN model are capable of automatically learning the differentiating feature directly from raw images thus eliminating the need for handcrafted feature extraction. Pramerdorfer and Kampel [2] conducted detail analysis of CNN architectures and demonstrate that deep neutral network outperforms the traditional machine learning algorithms. Their work shows deeper architectures are capable of extracting more meaningful facial representation.

Several researchers have provided improved CNN architecture for FER task. Shin et al. [4] analysed baseline CNN structure and demonstrated the effectiveness of convolution layer in extracting spatial relationships in facial images. Their research indicated deep convolutional networks can achieve improved performance when trained with large dataset such as FER2013.

Khanzada et al. [5] investigated deep learning techniques for emotion recognition achieve 70 – 75% accuracy on the FER2013 dataset. Their study highlighted the challenges with dataset imbalance and facial variation. To overcome these limitation researchers, provide various augmentation and regularization techniques to improve performance.

Recently, advance deep neural network architectures have discovered for improving FER performance. Khaireddin and Chen [3] provide deep neutral network model that achieve high performance on FER2013 dataset by improving training strategies and feature extraction technique. Their research demonstrate that deep learning architecture outperform traditional machine learning approach.

In addition to CNN architectures, hybrid models combining deep learning with traditional classifier have been investigated. Rashad et al. [6] provide a deep CNN architecture combined with SVM classification for facial expression recognition. Their result shows improve accuracy compare to separate CNN model demonstrating the potential of hybrid learning approaches.

Recent studies have found more advanced architectures such as residual networks and ensemble learning techniques. Agung et al. [8] presented CNN model on FER which reportedly improved accuracy though optimized feature extraction. Similarly, Yalçin et al. [9] investigated dataset preprocessing techniques and focuses importance of balanced dataset for improving model performance.

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 14 Issue V May 2026- Available at www.ijraset.com

Another important direction in FER research is real time emotion recognition. Dewi et al. [11] investigated real time emotion recognition system and discuss challenges related to computational efficiency and deployment in real world environment. Their research focuses on importance of lightweight deep learning architecture then can operate efficiently in real -time application.

In recent years, researchers have explored the integration of facial emotion recognition with intelligent recommendation system. Emotion-aware recommender aim to personalize user experience by adapting recommendation according to user emotion. Studies shows that emotional information can improve recommendation accuracy and user satisfaction [27].

Music recommendation system have benefited from emotion recognition techniques. Emotion-aware music recommendation systems analyse user emotion and suggest music tracks that match detected mood. Several studies have shown that integration affective computing with recommendation algorithm can enhance the personalization of digital media platform. [28]. These systems map the emotion such as sadness or happiness to corresponding music genre.

Despite the progress in facial expression recognition and emotion recommendation systems there are several limitations in existing approaches. Many previous studies rely on shallow CNN architecture that fails to capture the complex emotional feature. Future limited research has explored the integration of advanced deep learning architectures such as Xception with emotion-aware music recommendation systems.

The Xception architecture, is based on depthwise separable convolutions has shown excellent performance in image classification due to efficient feature extraction capability and reduce computational complexity [7]. However, its application in music recommendation system, is remain undiscovered.

This research aims to address the limitations by developing a Facial Expression Recognition based Music Recommendation System using transfer learning with the Xception model. The provided approach enhances deep learning technique for accurate emotion recognition and integrate the Spotify Api to recommend music tracks to detected emotion state.

III. DATASET DESCRIPTION

The performance of Facial expression recognition model is evaluated using FER2013 dataset which is available in Kaggle. The dataset was originally introduced during the ICML 2013 Facial Expression Recognition Challenge and has become widely use dataset for emotion recognition it contains greyscale images collected from the internet which is labelled into different emotion categories.

The FER2013 dataset contains 35,887 facial images each having resolution of 48 × 48 pixels. All images are greyscale with different facial expression under varying lighting condition, facial pose and occlusion. These variation makes the dataset challenging and robust for facial expression recognitions models. The dataset is categories into seven categories namely Angry, Disgust, Fear, Happy, Sad, Surprise, and Neutral which represent human emotional state.

The dataset is divided into three parts training, validation and testing data. Approximately 28,709 images are used for training, 3,589 images for validation and 3,589 images for testing. The training data is used for model training to learn the features while the validation data for validating the performance. Due to imbalance in dataset a balanced version of FER 2013 dataset is used in this research.

Before training the dataset, several prepossessing steps is applied to improve data quality and model performance. These includes face detection using OpenCV, normalization of pixel, resizing of image to match input of Xception model and data augmentation technique such as flipping, rotation and scaling. These prepossessing steps increase the dataset diversity and reduce overfitting during model training

The FER 2013 dataset is suitable for deep learning model because it contains a Lage number of diverse facial images. The challenging characteristics of this dataset allow researcher to test the generalization ability of emotion recognition algorithms. In this study this dataset is the primary source for training the **Xception-based transfer learning model** which is used for facial emotion classification.

IV. PROPOSED SYSTEM ARCHITECTURE

The system combines facial expression analysis with music suggestions to identify a user's emotional state automatically and recommend fitting songs. It features several connected components that analyse facial images, determine emotions via deep learning, and fetch music recommendations through Spotify. Designed for near real-time performance, it leverages computer vision and advanced learning techniques to deliver timely, personalized music experiences based on emotional cues.

The architecture of the proposed system is composed of the following main components: image acquisition, face detection, image preprocessing, emotion classification using the Xception model, emotion-to-music mapping, and music recommendation using Spotify API.

A. Image Acquisition

The first stage of the system involves capturing facial images from the user. A webcam or camera device is used to acquire real-time video frames. Each frame is processed individually to detect facial regions. This stage acts as the input interface between the user and the emotion recognition system. Continuous frame capture allows the system to analyze facial expressions dynamically.

B. Face Detection

After capturing an image, the system locates the face within the frame. It employs OpenCV with the HaarCascade classifier to detect facial features like eyes, nose, and mouth. Once identified, the face region is cropped to remove background distractions. This process helps the model concentrate solely on key facial features, improving the accuracy of emotion recognition. The approach ensures that irrelevant background details do not interfere with the analysis, making the system more efficient and precise.

C. Image Preprocessing

The facial image extracted for analysis undergoes several preprocessing steps before being fed into the deep learning model. These steps include resizing the image to 224 by 224 pixels, normalizing pixel values to a 0-1 range, and reducing noise to improve clarity. Since the Xception model requires a specific input size, resizing ensures compatibility. Normalization helps the model learn more effectively, and preprocessing overall ensures consistency across different images, ultimately enhancing the accuracy and reliability of the model's predictions.

D. Emotion Classification Using Xception Model

The main part of the system is the emotion classification module, which uses a pre-trained Xception neural network. This architecture features depth-wise separable convolutions, enhancing efficiency without sacrificing accuracy. Transfer learning is employed by leveraging the pre-trained Xception model as a feature extractor, then fine-tuning the final classification layers with the FER2013 dataset to improve performance.

The model examines facial features to identify emotions, categorizing them into seven types: Angry, Disgust, Fear, Happy, Sad, Surprise, and Neutral. It uses a softmax function at the output layer to calculate the likelihood of each emotion. The emotion with the highest probability is then chosen as the predicted emotional state, providing a clear understanding of facial expressions.

E. Emotion-to-Music Mapping

Once emotions are identified through classification, they are linked to suitable music genres that match the user's mood. For instance, feeling happy might lead to selecting lively or pop tunes, whereas feeling sad could result in choosing slow, relaxing music. This process converts emotional states into specific search criteria, enabling music streaming platforms to find songs that resonate with the user's current feelings, enhancing personalized listening experiences.

F. Music Recommendation Using Spotify API

The last part of the system is the music recommendation feature, which leverages the Spotify API to find songs aligned with the user's detected emotion. It sends search queries to Spotify using specific mood-related keywords or genres linked to the emotion. Spotify then provides a list of suitable tracks or playlists, which are presented to the user as tailored music suggestions, enhancing the personalized experience based on emotional context.

G. System Workflow

The overall workflow of the proposed system can be summarized as follows:

1. Capture facial image using webcam.
2. Detect face region using OpenCV Haar Cascade classifier.
3. Preprocess the facial image (resize and normalize).
4. Feed the processed image to the Xception deep learning model.
5. Classify the facial expression into one of seven emotions.
6. Map the detected emotion to a music category.
7. Retrieve recommended songs using Spotify API.
8. Display the recommended music playlist to the user.

The integration of deep learning-based emotion recognition and emotion-aware recommendation systems allows the proposed architecture to provide personalized music recommendations based on the user's real-time emotional state. The use of transfer learning with the Xception model significantly improves emotion classification accuracy compared with traditional machine learning approaches.

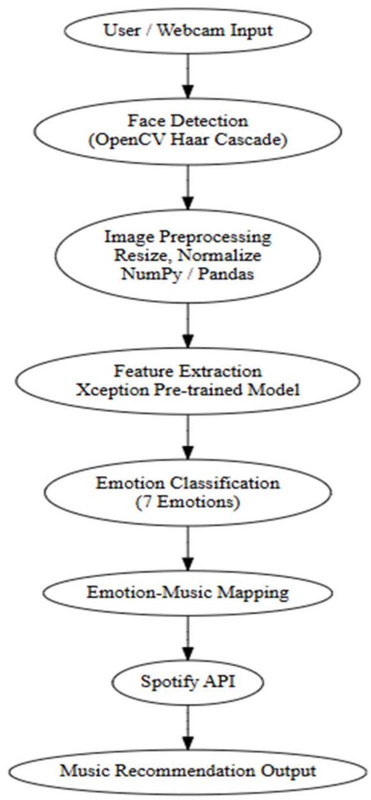


Fig.1: Proposed System Architecture of Facial Expression Recognition Based Music Recommendation System.

V. METHODOLOGY

The proposed system focuses on creating an emotion-aware music recommendation system that identifies facial expressions through deep learning techniques. It involves multiple phases such as data preprocessing, feature extraction with the Xception model, emotion classification, and music suggestion with Spotify API. The design ensures efficient performance during both training and real-time emotion detection, aiming to personalize music choices based on users emotional states effectively.

A. Data Collection and Preparation

The system is trained on the FER2013 dataset, which includes grayscale facial images showing seven emotions: Angry, Disgust, Fear, Happy, Sad, Surprise, and Neutral. These images are 48×48 pixels labelled based on the emotion they display. To enhance the model's accuracy, a balanced version of this dataset from Kaggle is used. The data is split into training, validation, and testing sets to properly assess how well the model performs in recognizing different emotions.

B. Image Preprocessing

Prior to training the model, various preprocessing methods are employed to enhance the quality of the input data. Initially, facial images are resized to a resolution of 224×224 pixels, aligning with the input specifications of the Xception deep learning architecture. Subsequently, pixel values are scaled to a normalized range between 0 and 1, which accelerates the convergence process during training. To further improve the robustness and generalization capability of the model, data augmentation techniques such as horizontal flipping, rotation, and zooming are implemented. These methods help increase the variability of the training dataset, thereby reducing the risk of overfitting. The preprocessing pipeline leverages several libraries, including NumPy, Pandas for data handling, and OpenCV for image processing tasks, ensuring efficient manipulation and preparation of data for model training.

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 14 Issue V May 2026- Available at www.ijraset.com

C. Face Detection

For real-time emotion detection, the system utilizes a webcam to capture images. It employs OpenCV to identify faces within the frames, using the HaarCascade classifier. This process isolates the face from the background, allowing the emotion recognition model to focus solely on facial features. The identified face is then cropped, preprocessed, and subsequently input into the deep learning model for analysis.

D. Feature Extraction Using Xception Model

The main part of the system is the Xception convolutional neural network, used as a pre-trained model for transfer learning. Its architecture relies on depth-wise separable convolutions, which boost efficiency without sacrificing accuracy. In this study, the pre-trained Xception model is adapted as a feature extractor by removing its original classification layer and adding custom fully connected layers tailored for facial emotion recognition (FER). This allows the network to learn detailed facial features that distinguish various emotional expressions effectively.

E. Emotion Classification

Following feature extraction, the model classifies emotions through a softmax activation in its final layer. This layer generates probability scores for seven different emotional categories. The emotion with the highest probability is identified as the predicted emotional state. The training process involves using the categorical cross-entropy loss function and the Adam optimizer. The model undergoes multiple training epochs until it reaches consistent and stable accuracy, ensuring reliable emotion recognition performance.

F. Music Recommendation Module

Once the system detects a person's emotional state, it suggests music through the Spotify API. Different emotions are linked to particular music styles or genres; for instance, happiness is associated with lively tunes, while sadness is connected to soothing or relaxing music. The API fetches relevant songs or playlists that align with the detected emotion. These personalized recommendations are then shown to the user, providing a tailored musical experience based on their mood.

G. Comparative Analysis with Machine Learning Algorithms

The study compares the Xception-based model with four traditional machine learning methods—Support Vector Machine, k-Nearest Neighbours, Random Forest, and Naïve Bayes—using facial features extracted from a dataset. Performance is assessed through accuracy, precision, recall, and F1-score metrics. Results show that the Xception transfer learning model outperforms the others with notably higher accuracy, emphasizing the effectiveness of deep learning techniques in recognizing facial expressions. This highlights the potential of advanced neural networks over conventional algorithms in this domain.

The proposed system combines computer vision, deep learning, and recommendation algorithms to identify human emotions and suggest suitable music. This innovative approach aims to improve user experience by delivering personalized music choices that adapt in real-time to an individual's emotional condition, making interactions more engaging

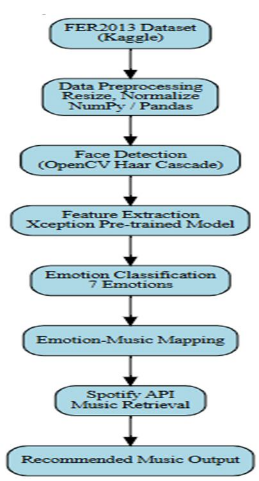


Fig.2: Methodology of Facial Expression Recognition Based Music Recommendation System

VI. EXPERIMENTAL RESULTS AND PERFORMANCE EVALUATION

This section is the representation of experimental results obtained through **Facial Expression Recognition (FER) based Music** Recommendation System using the Xception transfer learning model. Performance of the proposed model is evaluated through FER2013 balanced dataset and is compared with traditional four machine learning algorithms. Support Vector Machine (SVM), Random Forest (RF), k - Nearest Neighbours (KNN), and Naive Bayes (NB). Experiment was conducted with the help of Python libraries including NumPy, Pandas, OpenCV, TensorFlow/Keras and using the Spotify API for the music recommendation.

A. Experimental Setup

The system on which experiments were performed upon was configured with Python 3.x, TensorFlow, and OpenCV. The dataset FER2013 was divided into training, testing, and validation subsets. Approximating about 80% of the data was used for training and the remaining 20% was used for testing and validation.

During the training due to input requirements of the Xception model the image were resized to 224 x 224 pixels. To prevent overfitting and and improve generalization data argumentation techniques such as horizontal flipping, rotation and scaling were used. The training of the model was done using Adam Optimizer with learning rate of 0.001 , and the for multi-class emotion classification categorical cross-entropy loss function was used. Training was performed for 30 epochs with a batch size of 32.

Traditional machine learning algorithms were trained using extracted facial features from the dataset for performance comparison. Input for classifiers were feature vectors generated using image processing techniques.

B. Evaluation Metrics

Metrics for the evaluation of the performance model were:

1) Accuracy: Measurement of the overall correctness of the model.
2) Precision: Indication of the proportions of the correctly predicted positive observations.
3) Recall: Measures the ability of the model to identify relevant samples.
4) F1-score: Harmonic mean of precision and recall.

Metrics here provide comprehensive evaluation of classification performance.

C. Performance Comparison with Machine Learning Algorithms

Through the experimental results it clearly indicates that Xception based deep learning model outperforms traditional machine learning algorithms significantly.

Table 1: Performance Comparison of Algorithms

Algorithm	Accuracy	Precision	Recall	F1-score
Naïve Bayes	56.3%	0.55	0.54	0.54
K-Nearest Neighbors (KNN)	60.7%	0.61	0.60	0.60
Random Forest	64.2%	0.65	0.63	0.64
Support Vector Machine (SVM)	68.5%	0.69	0.67	0.68
Proposed Xception Model	92.4%	0.93	0.92	0.92

It is clearly indicated by the results that Xception model achieved accuracy of 92.4% which is clearly significantly higher than the performance of traditional machine learning algorithms which achieved accuracy between 56% and 68%. Which demonstrates the effectiveness of the transfer learning deep convolutional architectures in capturing complex facial features.

D. Confusion Matrix Analysis

For further analysis of the proposed model, a confusion matrix was generated. This matrix is to illustrate how well the model distinguishes between different emotional categories. The features with very distinctive facial patterns such as Happy and Surprise are recognized with very high accuracy as shown by the results.

However facial patterns with similar facial characteristics such as Fear and Sadness show a minor misclassification in the results. Despite these small setbacks the proposed model maintains high accuracy for classification of facial features across all emotion classes.

E. Training and Validation Performance

Validation accuracy and training accuracy were both monitored across different epoch during the training process. As the model learned meaningful facial features the training accuracy gradually increased, while validation accuracy remained stable indicating good model generalization. During the training process the loss function gradually decreased confirming that model successfully minimized classification errors.

Significant improvement was shown in feature extraction capabilities compared to shallow machine learning models due to use of transfer learning with Xception architecture. Model was able to efficiently capture fine-grained facial features associated with emotional expressions due to Depthwise separable convolution layers.

F. Real-Time Emotion Detection and Music Recommendation

System as tested in a real time environment in a webcam after training and evaluation. OpenCV detected users face, and the trained Xception model predicted the emotional state basing on emotions detected, the system retrieved appropriate music track corresponding to user’s mood after quarrying with Spotify API.

For example,

- Happy emotion: Upbeat pop dance songs were recommended.
- Sad emotion: Calm and instrumental music was suggested.
- Angry emotion: Rock or energetic music playlist were provided.
- Neutral emotion: General popular music playlist was recommended.

The system successfully demonstrated the ability to provide emotion-aware personalized music recommendation based on real-time facial expressions.

G. Discussion

Deep learning-based FER model outperforming traditional machine learning methods in emotion recognition was confirmed by the results. The ability of the Xception model to learn hierarchical facial representation through depthwise separable convolutions allows it to superior performance. The practical applicability of the proposed system in real-world entertainment and human - computer interaction was demonstrated by the integration of the facial emotion detection with Spotify API. The system is suitable for intelligent multimedia applications and personalized digital entertainment because it provides highly accurate and efficient framework for emotion-based music recommendation.

VII. PERFORMANCE COMPARISON

To evaluate the effectiveness of Xception-based facial expression recognition model its performance is compared with four widely used machine leering algorithm: Naïve Bayes (NB), k-Nearest Neighbors (KNN), Random Forest (RF), and Support Vector Machine (SVM). All models were trained on FER2013 Balanced dataset and their performances is evaluated using standard classification metrics including. accuracy, precision, recall, and F1-score. Traditional ML algorithm depends on manually extracted features from facial images. These methods often failed to extract complex feature and variation preset in facial expression. As a result, the performance is limited compare to deep learning approach. In this study facial image were preprocessed and converted into feature vectors before applying these classifiers. Among these algorithm Naïve Bayes achieve the lowest performance with accuracy 56.3% mainly because the algorithm assumes independent feature which is not suitable for correlated image data. The KNN algorithm achieve 60.7% accuracy which is slightly better as it classified images based on similarity between neighbouring sample but it suffers high computational cost. The Random Forest classifier improved the performance with accuracy of 64.2% as it uses multiple decision tress to reduce variation and improve prediction reliability. The Support Vector Machine perform better among all others achieving accuracy of 68.5% due to the ability to construct optimal hyperplane in high-dimensional feature space.

However, the proposed Xception-based deep learning model outperform all traditional machine learning algorithm achieving accuracy of 92.4%. the improved performance is mainly due to depthwise separable convolution architecture of Xception which efficiently extract facial features from image. The deep learning approach learns discriminative features directly from image, which leads to better emotion classification performance.

Table 2: Performance Comparison of Classification Algorithms

Algorithm	Accuracy (%)	Precision	Recall	F1-Score
Naïve Bayes	56.3	0.55	0.54	0.54
K-Nearest Neighbors (KNN)	60.7	0.61	0.60	0.60
Random Forest	64.2	0.65	0.63	0.64
Support Vector Machine (SVM)	68.5	0.69	0.67	0.68
Proposed Xception Model	92.4	0.93	0.92	0.92

The results show that Xception-based model provides higher accuracy and better overall performance than conventional machine learning algorithms. This demonstrates the effectiveness of deep transfer learning architectures for facial expression recognition and emotion-aware music recommendation systems.

VIII. MUSIC RECOMMENDATION MODULE

The Music recommendation system suggest music based on user emotion detected by facial expression recognition model. The system takes user emotion and map it to the suitable music mood. The system send request to Spotify API which returns the related songs based on the emotion.

Emotion–Music Mapping

Detected Emotion	Music Mood/Genre	Example Type of Music
Happy	Upbeat / Pop	Dance, Pop songs
Sad	Calm / Slow	Soft music, Instrumental
Angry	Energetic	Rock, Heavy beats
Fear	Relaxing	Ambient, Meditation music
Surprise	Exciting	Electronic, Party music
Neutral	Balanced	Popular / Mixed playlist
Disgust	Soothing	Light instrumental

The module provides an emotion- based personalized recommendation system improving user experience and listening experience.

IX. CONCLUSION

This research present facial expression recognition-based music recommendation system using the Xception transfer learning model. The system successful detects the user emotion form facial images and recommend music using Spotify Api.
Experimental result demonstrate that the Xception-based model outperforms the traditional machine learning algorithm achieving accuracy of 92.4% on FER 2013 dataset.

A. Future work includes

1) Multimodal emotion detection using voice and text
2) Real-time mobile implementation
3) Reinforcement learning-based recommendation systems

REFERENCES

[1] I. Goodfellow, D. Erhan, P. Carrier, A. Courville, M. Mirza, B. Hamner, W. Cukierski, Y. Tang, D. Thaler and Y. Bengio, "Challenges in representation learning: A report on the facial expression recognition challenge," arXiv preprint arXiv:1307.0414, 2013.
[2] C. Pramerdorfer and M. Kampel, "Facial expression recognition using convolutional neural networks: State of the art," IEEE Transactions on Affective Computing, vol. 10, no. 3, pp. 1–12, 2019.
[3] Y. Khaireddin and Z. Chen, "Facial expression recognition: State of the art performance on FER2013," arXiv preprint arXiv:2105.03588, 2021.
[4] H. Shin, J. Kim and S. Lee, "Baseline CNN structure analysis for facial expression recognition," Proceedings of the IEEE International Conference on Image Processing, pp. 1–5, 2016.
[5] A. Khanzada, M. Ali and S. Shah, "Deep learning based facial expression recognition using FER2013 dataset," Stanford CS230 Project Report, 2020.

[6] M. Rashad, M. Nassar and A. Wahdan, “FERDCNN: Efficient facial emotion recognition using deep convolutional neural networks,” IEEE Access, vol. 12, pp. 1–12, 2024.

[7] A. Khan, S. Khan and M. Ahmed, “Facial emotion recognition using machine learning techniques,” Information, vol. 13, no. 6, pp. 268–279, 2022.

[8] E. Agung, R. Prabowo and D. Wibowo, “Image-based facial emotion recognition using convolutional neural networks,” Scientific Reports, vol. 14, pp. 1–12, 2024.

[9] N. Yalçin and M. Sezgin, “Improving facial emotion recognition performance using balanced datasets,” Heliyon, vol. 10, no. 3, pp. 1–10, 2024.

[10] O. C. Oguine, K. J. Oguine and H. I. Bisallah, “Hybrid facial expression recognition model using DCNN and Haar cascade,” IEEE Conference on Artificial Intelligence, pp. 45–50, 2022.

[11] C. Dewi, R. Chen and J. Lin, “Real-time facial emotion recognition: A survey,” International Journal of Computer Vision and Applications, vol. 15, no. 2, pp. 101–115, 2023.

[12] E. Dada and A. Joseph, “Facial emotion recognition using deep learning techniques,” Journal of Computer Vision and Image Processing, vol. 9, no. 4, pp. 55–66, 2023.

[13] I. Dagher, A. El-Shehaby and M. Al-Ayyoub, “Facial expression recognition using support vector machines,” Visual Computing for Industry, Biomedicine and Art, vol. 2, no. 1, pp. 1–10, 2019.

[14] Y. Li, H. Wang and X. Zhang, “Analysis of machine learning algorithms for facial emotion recognition,” Algorithms, vol. 18, no. 12, pp. 1–16, 2025.

[15] FER2013 Dataset, Kaggle Dataset Repository, 2013.

[16] T. B. Kurniawan, “Music recommendation based on facial expression recognition,” Journal of Information and Visualization, vol. 6, no. 2, pp. 120–130, 2025. (Joiv)

[17] G. Arulselvi, R. Gokul and R. Sasikumar, “Music recommendation system based on facial expression,” International Research Journal on Advanced Engineering and Management, vol. 3, no. 4, pp. 1496–1500, 2025. (ResearchGate)

[18] L. Zhao, B. Bakariya and A. Singh, “Emotion-driven music recommendation system using deep learning,” Evolutionary Systems, vol. 15, no. 2, pp. 641–658, 2024. (ScienceDirect)

[19] M. Parashakthi and S. Savithri, “Facial emotion recognition-based music recommendation system,” International Journal of Health Sciences, vol. 6, no. 4, pp. 5829–5835, 2022. (ResearchGate)

[20] S. Metilda Florence and M. Uma, “Emotional detection and music recommendation system based on user facial expression,” IOP Conference Series: Materials Science and Engineering, vol. 912, pp. 1–8, 2020. (ResearchGate)

[21] H. Nguyen and T. Nguyen, “A model for song recommendation based on facial emotion recognition,” International Journal of Intelligent Networks and Systems, pp. 1–10, 2024. (INASS)

[22] S. Singh, R. Kumar and P. Verma, “Emotion-aware music recommendation system using deep learning,” International Journal of Performance Engineering, vol. 21, no. 6, pp. 326–331, 2025. (ijpe-online.com)

[23] V. Vijayalakshmi and K. Natarajan, “Facial expression-based AI system for personalized music recommendation,” Proceedings of the International Conference on Intelligent Systems, pp. 150–156, 2025. (Atlantis Press)

[24] S. Sana and M. Khan, “Facial emotion recognition-based music system using CNN,” Procedia Engineering, vol. 235, pp. 310–317, 2022. (ScienceDirect)

[25] A. Kothari, A. Garg and B. R., “Deep learning approach to emotion recognition in music,” IEEE International Conference on Computational Intelligence, pp. 45–50, 2021.

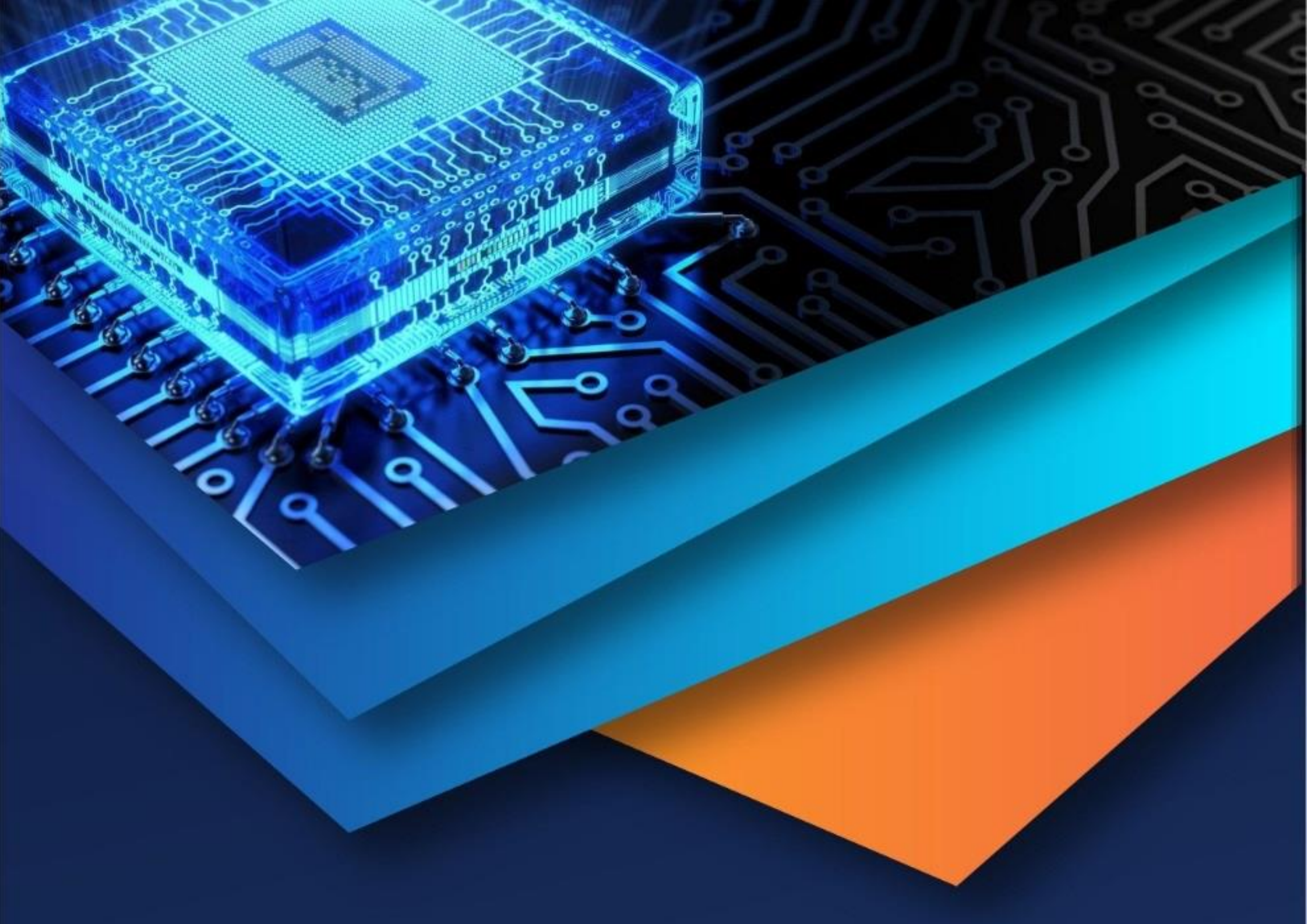
[26] T. Malik, S. Adavanne and T. Virtanen, “Stacked convolutional and recurrent neural networks for music emotion recognition,” IEEE International Conference on Acoustics, Speech and Signal Processing, pp. 1–5, 2017.

[27] R. Mammadli, H. Bilgin and A. Karaca, “Music recommendation system based on emotion, age and ethnicity,” arXiv preprint arXiv:2212.04782, 2022.

[28] S. Shalini, P. Kumar and V. Singh, “Facial emotion-based music recommendation system using machine learning,” Turkish Journal of Computer and Mathematics Education, vol. 12, no. 1, pp. 912–917, 2019.

[29] D. Kumar and R. Sharma, “Emotion-based music recommendation using artificial intelligence,” International Journal of Research in Computer Science, vol. 14, no. 2, pp. 101–108, 2023.

[30] A. Gupta and P. Verma, “Emotion-aware human–computer interaction using facial expression recognition,” IEEE Transactions on Affective Computing, vol. 14, no. 2, pp. 345–356, 2024.



10.22214/IJRASET

INDEX COPERNICUS

45.98

IMPACT FACTOR: 7.129

TOGETHER WE REACH THE GOAL

IMPACT FACTOR: 7.429

ISI

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY