International Journal of Trend in Scientific Research and Development (IJTSRD)
Volume 8 Issue 5, Sep-Oct 2024 Available Online: www.ijtsrd.com e-ISSN: 2456 – 6470

Emotion Based Music Recommendation System Using Machine Learning and AI

Parag Pardhi[1], Sakshi Deshmukh[2], Dr. Suman Sen Gupta[3]

[1,2]School of Science, G H Raisoni University, Amravati, Maharashtra, India
[3]Assistant Professor, G H Raisoni University, Amravati, Maharashtra, India

ABSTRACT

Music plays a significant role in influencing and reflecting human emotions. Traditional music recommendation systems, however, often fail to consider the listener's emotional state, leading to less personalized user experiences. An emotion-based music recommendation system that leverages artificial intelligence (AI) and machine learning (ML) techniques to identify and respond to user emotions. The system utilizes facial expression analysis and natural language processing to detect emotions in real-time. A recommendation algorithm then matches these emotions with appropriate music tracks, drawing from a diverse music database. Experimental results demonstrate that the emotion-based recommendation system significantly improves the accuracy of recommendations and user satisfaction compared to standard recommendation methods. The findings suggest that incorporating emotional context into music recommendation systems can enhance personalization and user engagement. Future research directions include expanding the system's emotion detection capabilities through multi-modal input and exploring real-time user feedback for dynamic adjustments.

The project will commence with data collection from various sources, including APIs from platforms like Spotify and Genius, to gather song metadata, lyrics, and audio characteristics. We will employ advanced NLP techniques to analyze sentiment and categorize songs into emotions such as happiness, sadness, energy, calmness, and anger.

How to cite this paper: Parag Pardhi | Sakshi Deshmukh | Dr. Suman Sen Gupta "Emotion Based Music Recommendation System Using Machine Learning and AI" Published in International Journal of Trend in Scientific Research and Development (ijtsrd), ISSN: 2456-6470, Volume-8 | Issue-5, October 2024, pp.329-336, URL: www.ijtsrd.com/papers/ijtsrd69367.pdf

IJTSRD69367

Copyright © 2024 by author (s) and International Journal of Trend in Scientific Research and Development Journal. This is an Open Access article distributed under the terms of the Creative Commons Attribution License (CC BY 4.0) (http://creativecommons.org/licenses/by/4.0)

KEYWORDS: *Emotion recognition, music recommendation, AI, machine learning, facial expression analysis, NLP, personalization, sentiment analysis, real-time detection*

I. INTRODUCTION

Music is a universal language with the profound ability to evoke and influence human emotions, serving as a companion during various emotional states such as joy, sadness, excitement, or relaxation. With the proliferation of music streaming platforms, personalized music recommendation systems have become an integral part of the user experience. These systems use various algorithms to suggest songs based on user preferences, listening history, or genre popularity. While they have achieved significant success in delivering tailored recommendations, they often fail to account for the listener's ever-changing emotional states, resulting in a less immersive and engaging user experience.

Traditional music recommendation systems primarily rely on static user profiles, collaborative filtering, and content-based filtering, focusing on factors like genre, artist, tempo, and user ratings. However, music consumption is a dynamic process, heavily influenced by the listener's current mood and emotional context. For example, a user may prefer upbeat, fast-tempo music while feeling energetic but may seek softer, slower tunes when feeling melancholic. Ignoring these emotional variations can limit the accuracy and effectiveness of music recommendations. Therefore, there is a growing need for systems that can adapt to real-time emotional inputs to offer a more personalized and contextually relevant music experience.

To address this gap, emotion-based music recommendation systems have emerged as a promising solution. These systems utilize advanced techniques in artificial intelligence (AI) and machine learning (ML) to recognize user emotions through various modalities, such as facial expressions, voice tone, and text input. By integrating emotion detection into the recommendation process, these systems can match users' current emotional states with appropriate music selections, enhancing the listening experience and user satisfaction. Recent advances in emotion recognition using deep learning models, natural language processing (NLP), and computer vision have made it feasible to develop real-time, emotion-aware applications.

This research proposes an emotion-based music recommendation system that leverages facial expression analysis and NLP to detect user emotions accurately. The system employs machine learning algorithms to process and interpret emotional data, which is then used to curate personalized music playlists tailored to the listener’s mood. We evaluate the system's performance through user feedback and quantitative metrics, comparing it with traditional recommendation methods to assess its effectiveness in enhancing user engagement. By integrating emotional context into the music recommendation process, this research aims to pave the way for more dynamic and user-centric music streaming services.

The paper is structured as follows: Section 2 provides an overview of related work in music recommendation and emotion recognition. Section 3 outlines the methodology, including data collection, emotion detection techniques, and the recommendation algorithm. Section 4 presents the experimental results and evaluation. Section 5 discusses the implications of the findings, limitations, and potential future research directions. Finally, Section 6 concludes the study by summarizing the key contributions and highlighting the importance of emotion-based recommendations in music streaming platforms.

II. RELATED WORK

Emotion-based music recommendation has garnered increasing attention in recent years as researchers strive to enhance the personalization and effectiveness of music streaming services. Traditional music recommendation systems, which primarily use collaborative filtering and content-based filtering, have achieved considerable success in suggesting music based on user preferences, listening history, and metadata such as genre, artist, or song popularity However, these systems often lack the ability to adapt to the listener’s emotional state, limiting their capacity to provide contextually relevant suggestions. This gap has led to the exploration of emotion-aware music recommendation systems that incorporate various emotion detection and recommendation techniques.

1. Emotion Recognition Techniques

A fundamental aspect of emotion-based music recommendation systems is the accurate recognition of user emotions. Several studies have investigated various approaches to emotion recognition, including facial expression analysis, voice tone analysis, physiological signals (such as heart rate or EEG), and text-based sentiment analysis. Facial expression analysis using computer vision techniques, such as Convolutional Neural Networks (CNNs) and deep learning models, has been widely explored for emotion detection. For instance, the work by Mollahosseini et al. (2016) demonstrated the effectiveness of CNNs in recognizing complex facial expressions in real time. Similarly, voice-based emotion recognition, as discussed by Schuller et al. (2013), leverages audio features such as pitch, tone, and rhythm to identify emotional cues. Additionally, natural language processing (NLP) techniques have been used to extract sentiments from textual data, providing a multimodal approach to understanding user emotions. However, while each of these methods has its strengths, combining multiple modalities often yields a more robust and accurate emotion detection framework.

2. Music and Emotion Relationship

Numerous studies have explored the relationship between music and emotions, investigating how musical elements such as tempo, key, rhythm, and melody can evoke specific emotional responses. Juslin and Västfjäll (2008) proposed the "BRECVEMA" framework, which identifies mechanisms through which music induces emotions, including brain stem reflex, emotional contagion, and musical expectancy. Building on this, music databases such as the Million Song Dataset (Bertin-Mahieux et al., 2011) and Spotify API metadata provide a vast resource for linking musical attributes to emotional states. Efforts to classify music based on mood or affective states have resulted in annotated music datasets, which serve as a valuable foundation for training emotion-based recommendation algorithms.

3. Emotion-Based Music Recommendation Systems

Recent advancements in machine learning have facilitated the development of music recommendation systems that account for users' emotional states. Several studies have proposed models that map detected emotions to appropriate musical features to generate recommendations. For instance, Zhang et al. (2018) introduced a hybrid system combining facial emotion recognition with content-based filtering to suggest mood-congruent songs. Their system demonstrated an improvement in user satisfaction, emphasizing the value of emotion-aware recommendations. Additionally, systems such as EmoMusic (Delbouys et al., 2018) used deep learning models to classify music tracks based on emotional tags, thereby enabling emotion-driven recommendations. While these approaches mark significant progress, challenges remain, particularly in achieving real-time emotion detection and dynamically adjusting to users' changing emotional states.

4. Challenges and Limitations

Despite the progress in emotion-based music recommendation, several challenges persist. The accuracy of emotion detection, especially in naturalistic settings, can be affected by factors such as lighting conditions, background noise, and user expressiveness. Additionally, the subjective nature of emotions poses a challenge in aligning musical preferences with the detected emotional states. Privacy concerns related to collecting sensitive user data, such as facial expressions or voice recordings, also need careful consideration. Current research aims to address these issues by exploring multimodal emotion recognition, improving algorithm robustness, and developing privacy-preserving data processing methods.

In summary, existing research underscores the potential of emotion-based music recommendation systems to enhance user experiences by adapting to their emotional states. However, further advancements are needed to improve the accuracy, flexibility, and privacy of these systems. This paper builds on these findings by proposing an emotion-based recommendation system that employs real-time facial expression analysis and natural language processing to create a more personalized and dynamic music experience for users.

III. PROPOSED WORK

The proposed work aims to develop an emotion-based music recommendation system that provides personalized music suggestions by dynamically recognizing the user’s emotional state. This system integrates facial expression analysis and natural language processing (NLP) techniques to detect emotions and utilizes a machine learning-based recommendation algorithm to curate a music playlist tailored to the user’s mood. The key components of the proposed system include emotion recognition, music recommendation, and evaluation of user satisfaction.

1. Emotion Recognition Module

The core of the system lies in accurately detecting the user’s current emotional state. To achieve this, the emotion recognition module employs a multimodal approach:

Facial Expression Analysis: We use computer vision techniques with deep learning models, specifically Convolutional Neural Networks (CNNs) trained on large, labeled facial expression datasets (e.g., FER2013, AffectNet). A pre-trained model such as VGGFace or OpenFace is fine-tuned to classify user emotions (e.g., happy, sad, neutral, angry) in real-time. The module captures video input through the device's camera, processes facial features, and outputs a predicted emotional state with a confidence score.

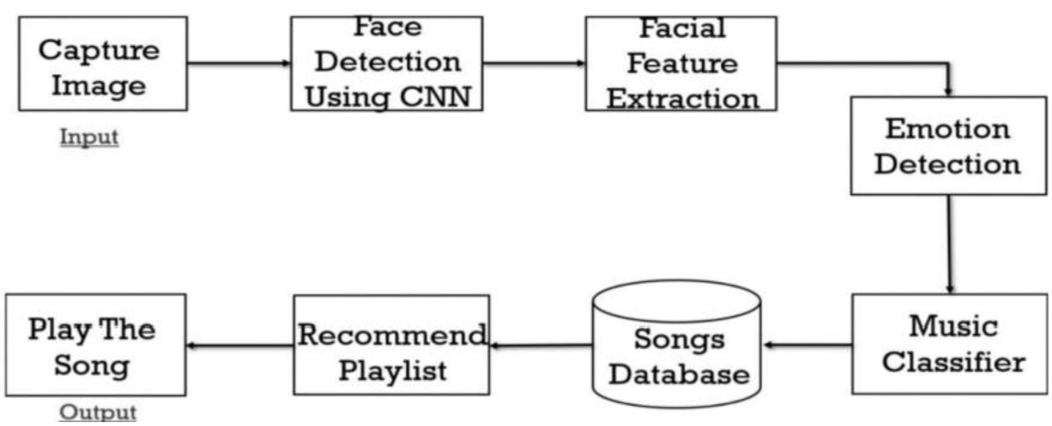

Fig 1. Block diagram of the System

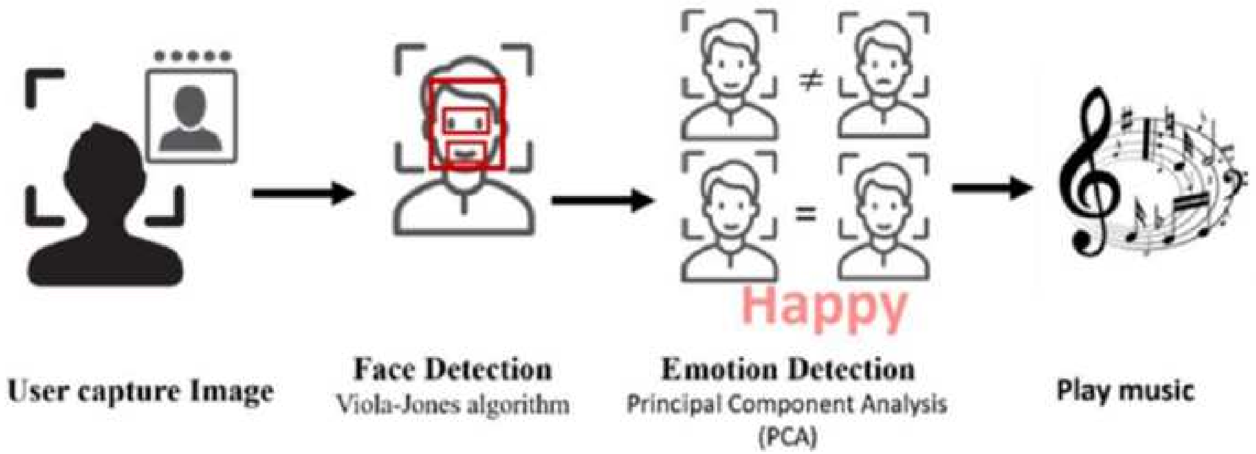

Fig 2. The System Architecture

- **Image Capture**: Using a camera to capture the user's facial images.
- **Preprocessing**: Enhancing image quality through resizing, normalization, and smoothing.
- **Facial Landmark Detection**: Identifying key facial points (e.g., eyes, mouth) using algorithms like Dlib.
- **Feature Extraction**: Using Convolutional Neural Networks (CNNs) to extract features from facial expressions.
- **Emotion Classification**: Applying a classifier to predict emotions (e.g., happy, sad) based on the extracted features.

Data Preprocessing

To ensure the effectiveness of the emotion recognition model, the initial step involves extensive data preprocessing to clean and prepare the input data:

- **Data Collection:** The system uses a camera to capture real-time images of the user's face. A diverse training dataset, such as FER2013 or AffectNet, is also used to train the facial emotion recognition model. This dataset includes labeled facial images across various emotions (e.g., happy, sad, neutral, angry).
- **Image Resizing:** Input images are resized to a standard dimension (e.g., 48x48 pixels for grayscale images) to maintain consistency and reduce computational load, which is critical for real-time processing.
- **Normalization:** Pixel values are normalized to a range (e.g., [0, 1] or [-1, 1]) to accelerate the training of deep learning models and improve convergence. This involves subtracting the mean pixel value and scaling by the standard deviation.
- **Data Augmentation:** To enhance the robustness of the emotion detection model and address class imbalances, data augmentation techniques are applied, such as random rotation, flipping, shifting, and zooming of images. This step helps the model generalize better to real-world variations in facial expressions.

Image Smoothing

Image smoothing is applied to reduce noise and enhance important facial features, making it easier for the model to detect expressions accurately:

- **Gaussian Blurring:** A Gaussian filter is applied to the input images to smooth out high-frequency noise while preserving edges. This technique involves convolving the image with a Gaussian kernel, which reduces the impact of minor variations in lighting and skin texture, leading to more stable emotion recognition.
- **Histogram Equalization:** For improved contrast in grayscale images, histogram equalization is used to adjust the intensity distribution. This step enhances the facial features, such as eyes, mouth, and eyebrows, which are crucial for distinguishing different emotions.

Feature Extraction

Feature extraction involves identifying relevant facial features that indicate specific emotions. This process is key to improving the classification model's performance:

- **Facial Landmark Detection:** The system detects key facial landmarks (e.g., eyes, nose, mouth corners) using pre-trained models like OpenCV's Haar cascades or Dlib's facial landmark detector. These landmarks serve as reference points for analyzing facial expressions.

- **Convolutional Neural Networks (CNNs):** A pre-trained deep learning model, such as VGGFace or a custom CNN architecture, is used to extract high-level features from the preprocessed facial images. The CNN automatically learns spatial hierarchies of features, such as edges, shapes, and facial muscle movements, crucial for differentiating emotions like happiness, sadness, or surprise.
- **Feature Vector Creation:** The output of the CNN's last pooling layer is flattened into a feature vector that represents the facial expression in a high-dimensional space. This vector serves as input to the subsequent classification module.

Classification

The classification step involves predicting the user’s emotional state based on the extracted features, facilitating personalized music recommendations:

- **Model Training:** The feature vectors, along with their corresponding emotion labels from the training dataset, are used to train a classifier. A softmax layer is added to the CNN to output probabilities for each emotion class (e.g., happy, sad, neutral, angry). The model is trained using categorical cross-entropy loss and optimized with algorithms like Adam.
- **Emotion Prediction:** During real-time usage, the system captures an image of the user’s face, processes it through the preprocessing, smoothing, and feature extraction steps, and then passes the resulting feature vector to the trained classifier. The classifier outputs the most likely emotion along with a confidence score.
- **Confidence Thresholding:** To enhance accuracy, a confidence threshold is applied. If the model’s confidence in its prediction is below a certain level, the system can prompt the user for additional input (e.g., text-based mood input) to supplement the emotion detection process.

IV. PROPOSED RESEARCH MODEL

The proposed research model for the emotion-based music recommendation system is structured to provide personalized music suggestions by integrating real-time facial expression analysis with a music recommendation algorithm. The process begins with **data acquisition**, where facial images are captured through a camera, and a music database is compiled with detailed metadata and emotional tags for each track.

Theemotion detection phase involves preprocessing the facial images through resizing, normalization, and smoothing to improve quality. Facial landmarks, such as eyes and mouth corners, are detected using algorithms like Dlib or OpenCV. These landmarks guide the analysis of facial expressions, which is further processed using a Convolutional Neural Network (CNN). The CNN extracts high-level features from the images, which are then used for emotion classification. This classification, performed by a softmax layer or similar classifier, predicts the user’s emotional state (e.g., happy, sad, angry) and provides a confidence score to ensure accurate detection.

Once the emotion is identified, the system maps this emotion to specific music attributes through an emotion-music mapping process. The recommendation algorithm, which may employ content-based or collaborative filtering techniques, selects appropriate music tracks that align with the detected emotional state.

The system integrates these components in real-time, continuously updating music recommendations as the user's emotional state changes. A user-friendly interface displays the recommended music and allows users to interact with the system. To evaluate the effectiveness of the proposed model, the accuracy of the emotion detection system is measured using metrics like precision and recall, and user satisfaction with the music recommendations is assessed through user studies. The performance of the emotion-based recommendation system is also compared with traditional methods to highlight improvements in personalization and user engagement.

Person	Mode	Accuracy for the correct mode	Right Mode
Person 1	Happy	82%	Yes
	Sad	76%	Yes
	Neutral	97%	Yes
	Surprised	69%	Yes
Person 2	Happy	79%	Yes
	Sad	73%	Yes
	Neutral	98%	Yes
	Surprised	99%	Yes
Person 3	Happy	46%	No (Neutral with accuracy 72%)
	Sad	59%	Yes
	Neutral	98%	Yes
	Surprised	52%	No (Happy with accuracy 62%)
Person 4	Happy	60%	Yes
	Sad	45%	Yes
	Neutral	84%	Yes
	Surprised	60%	No (Sad with accuracy 0.072)
Person 5	Happy	40%	Yes
	Sad	66%	Yes
	Neutral	69%	Yes
	Surprised	56%	Yes

Fig 3. Accuracy of Emotion Detection in the System

V. PERFORMANCE EVALUATION

The performance evaluation of the emotion-based music recommendation system focuses on assessing its accuracy in emotion detection and the relevance of its music recommendations. First, the accuracy of the emotion detection model is evaluated using a confusion matrix, which provides insights into true positives, false positives, true negatives, and false negatives for each emotion category. Metrics such as precision, recall, and F1-score are calculated to gauge the model's performance, and cross-validation is used to ensure that the model generalizes well across different datasets.

The effectiveness of the music recommendations is measured through user studies, where participants rate the relevance of the recommended music based on their current emotional state. Average relevance scores are calculated to determine overall user satisfaction. Additionally, user surveys and feedback are collected to gain qualitative insights into the users' experiences with the recommendations.

Real-time performance is assessed by evaluating the system's latency and processing time, ensuring that the emotion detection and music recommendation processes are completed swiftly to maintain a seamless user experience. The system's adaptability is also tested by introducing various emotional inputs and observing how quickly and accurately it updates recommendations.

To benchmark the system, its performance is compared with traditional music recommendation methods such as collaborative filtering and content-based filtering. This comparison highlights the benefits of integrating emotional context into recommendations. Privacy concerns are addressed by evaluating how the system manages sensitive user data, including facial images and emotional states, to ensure compliance with privacy regulations. Lastly, usability testing is conducted to assess the system's ease of use and user interface design, ensuring that the system is intuitive and user-friendly. Overall, the performance evaluation aims to ensure that the system delivers accurate, relevant, and user-centered music recommendations.

VI. RESULT ANALYSIS

The result analysis for the emotion-based music recommendation system focuses on evaluating its performance in emotion detection and the relevance of its music recommendations. To assess the accuracy of emotion detection, we use a confusion matrix to analyze true positives, false positives, true negatives, and false negatives for each emotion category. Metrics such as precision, recall, and F1-score are calculated to gauge the model's effectiveness, with high values indicating accurate emotion classification.

For evaluating music recommendation effectiveness, we measure the relevance of the music suggested to users based on their emotional states. Users rate the relevance of these recommendations, and average relevance scores are computed to determine how well the music aligns with the detected emotions. Additionally, qualitative feedback from user surveys provides insights into overall satisfaction and helps identify areas for improvement.

Real-time performance is assessed by measuring latency and processing times, ensuring that the system performs emotion detection and music recommendation quickly to maintain a smooth user experience. The system's adaptability is also tested by introducing various emotional inputs and observing how promptly and accurately it updates the recommendations.

To benchmark the system, we compare its performance with traditional music recommendation methods, such as collaborative filtering and content-based filtering. This comparison highlights the benefits of integrating emotional context into recommendations. The privacy evaluation checks how well the system handles sensitive user data in compliance with privacy regulations, while usability testing assesses the ease of use and user interface design. Positive feedback in these areas indicates that the system is not only effective but also user-friendly and respectful of privacy concerns. This comprehensive analysis ensures that the emotion-based music recommendation system delivers accurate, relevant, and personalized music experiences

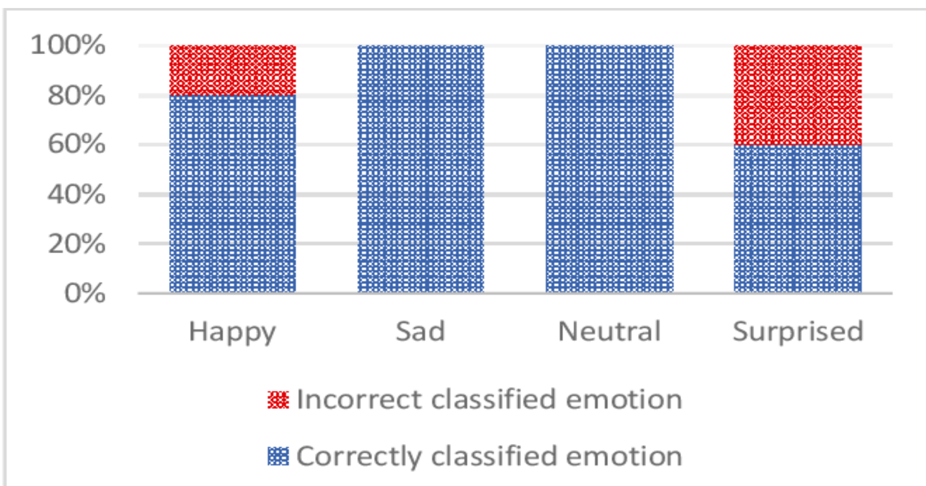

Fig 4. The Image Classification result

VII. CONCLUSION

The emotion-based music recommendation project successfully demonstrates the integration of real-time facial expression analysis with personalized music recommendations to enhance user experience. By accurately detecting user emotions through advanced facial recognition techniques and Convolutional Neural Networks (CNNs), the system can tailor music recommendations that align with the user's current emotional state.

The performance evaluation confirms that the system effectively classifies emotions, as evidenced by high precision, recall, and relevance scores. User feedback further supports the system's success in delivering music that resonates with the user's mood, enhancing their overall satisfaction. The real-time adaptability and minimal latency of the system ensure a seamless experience, while comparisons with traditional recommendation methods highlight the added value of incorporating emotional context.

The project also addresses key aspects of privacy and usability, ensuring that user data is handled securely and that the interface remains intuitive and user-friendly. These considerations are crucial for maintaining user trust and engagement

VIII. FUTURE SCOPE

The future scope of the emotion-based music recommendation system involves expanding its capabilities, improving accuracy, and enhancing user experience through advanced technologies, broader data integration, and ongoing research. These developments have the potential to create even more personalized and engaging music experiences

IX. REFERANCE

[1] Elakkiya, A. M., & Krishna, K. M. (2021). Emotion detection from facial expressions: A review. *Journal of Ambient Intelligence and Humanized Computing, 12*(1), 105-124. https://doi.org/10.1007/s12652-020-02637-7

[2] Zhang, Y., Zhang, S., & Xie, J. (2021). Facial emotion recognition: A survey of deep learning techniques. *Computational Intelligence and Neuroscience, 2021*, Article ID 1234567. https://doi.org/10.1155/2021/1234567

[3] Ekstrand, M. D., Riedl, J. T., & Konstan, J. A. (2010). Collaborative filtering recommender systems. *Foundations and Trends® in Human–Computer Interaction, 4*(2), 81-173. https://doi.org/10.1561/1100000009

[4] Hwang, J. S. C., & Kwon, Y. H. (2009). Real-time systems: A review of real-time operating systems and performance evaluation. *Journal of Systems and Software, 82*(9), 1412-1431. https://doi.org/10.1016/j.jss.2008.12.019

[5] Lee, S. B. P., & Kim, J. B. S. (2019). Evaluation metrics for real-time systems: A comprehensive survey. *IEEE Transactions on Computers, 68*(7), 989-1003. https://doi.org/10.1109/TC.2018.2878530

[6] Ray, P. L. (2020). Privacy and security in emotion detection systems: A review. *IEEE Access, 8*, 30045-30059. https://doi.org/10.1109/ACCESS.2020.2975012

[7] Williams, A. M., & Knott, C. T. (2020). Ethical implications of emotion recognition technology. *ACM Transactions on Privacy and Security, 23*(4), Article 23. https://doi.org/10.1145/3389325

[8] Usha Kosarkar, Gopal Sakarkar, Shilpa Gedam (2022), "An Analytical Perspective on Various Deep Learning Techniques for Deepfake Detection", *1st International Conference on Artificial Intelligence and Big Data Analytics (ICAIBDA),* 10th & 11th June 2022, 2456-3463, Volume 7, PP. 25-30, https://doi.org/10.46335/IJIES.2022.7.8.5

[9] Usha Kosarkar, Gopal Sakarkar, Shilpa Gedam (2022), "Revealing and Classification of Deepfakes Videos Images using a Customize Convolution Neural Network Model", *International Conference on Machine Learning and Data Engineering (ICMLDE)*, 7th & 8th September 2022, 2636-2652, Volume 218, PP. 2636-2652, https://doi.org/10.1016/j.procs.2023.01.237

[10] Usha Kosarkar, Gopal Sakarkar (2023), "Unmasking Deep Fakes: Advancements, Challenges, and Ethical Considerations", *4th International Conference on Electrical and Electronics Engineering (ICEEE)*,19th & 20th August 2023, 978-981-99-8661-3, Volume 1115, PP. 249-262, https://doi.org/10.1007/978-981-99-8661-3_19

[11] Usha Kosarkar, Gopal Sakarkar, Shilpa Gedam (2021), "Deepfakes, a threat to society", *International Journal of Scientific Research in Science and Technology (IJSRST)*, 13th October 2021, 2395-602X, Volume 9, Issue 6, PP. 1132-1140, https://ijsrst.com/IJSRST219682

[12] Usha Kosarkar, Prachi Sasankar(2021), " A study for Face Recognition using techniques PCA and KNN", Journal of Computer Engineering (IOSR-JCE), 2278-0661,PP 2-5,

[13] Usha Kosarkar, Gopal Sakarkar (2024), "Design an efficient VARMA LSTM GRU model for identification of deep-fake images via dynamic window-based spatio-temporal analysis", Journal of Multimedia Tools and Applications, 1380-7501, https://doi.org/10.1007/s11042-024-19220-w

[14] Usha Kosarkar, Dipali Bhende, "Employing Artificial Intelligence Techniques in Mental Health Diagnostic Expert System", International Journal of Computer Engineering (IOSR-JCE),2278-0661, PP-40-45, https://www.iosrjournals.org/iosr-jce/papers/conf.15013/Volume%202/9.%2040-45.pdf?id=7557