

International Journal of Advanced Research in Education and Technology (IJARETY)

Volume 12, Issue 2, March-April 2025

Impact Factor: 8.152



Facial Expression-Driven Song Suggestion System

Lokesh Devathati¹, Chekuri Bhargava Surendra², Chekka Madhava Rao³,

Marri Bhargava Ranga Swamy⁴, Masood Khan⁵

Assistant Professor, Department of CSE, Eluru College of Engineering & Technology, Eluru, India¹

B. Tech Student, Department of CSE, Eluru College of Engineering & Technology, Eluru, India^{2,3,4,5}

ABSTRACT: Many individuals often struggle to identify the type of music they truly want to listen to based on their current mood. These challenges can lead to significant time wasted in searching for the "perfect" song that aligns with their emotional state. In this era of rapid technological advancements, particularly in the domain of Artificial Intelligence, leveraging AI in song recommendation systems can bridge this gap, creating a seamless fusion of technology and user needs.

This paper aims to develop an emotion-based song suggestion system that suggests songs based on the user's emotional state. The process begins with the user capturing a photo of themselves using their laptop's webcam—an action undertaken only with their explicit consent. Using a Convolutional Neural Network (CNN), the system analyzes the image to detect the user's current emotional state. Once the emotion is identified, the system matches it to a curated playlist or category of songs that correspond to the detected mood. The selected song is then played automatically, fulfilling the user's musical preferences effortlessly. The proposed system enhances user engagement by dynamically adapting to their emotions, making it a promising approach for next-generation music recommendation platforms.

KEYWORDS: Expression, Convolutional Neural Network, Detection and Recognition.

I. INTRODUCTION

Music has the power to evoke emotions and influence human behavior. Music streaming services have become increasingly popular in recent years, providing users with access to a vast library of songs from various genres and artists. However, the sheer volume of music available can be overwhelming for users, and finding the right song to match their current emotional state can be challenging. This has led to the development of automatic song recommendation systems, which aim to provide personalized music suggestions based on the user's preferences and behavior. One approach to developing such systems is to use human facial expressions as a means of detecting the user's emotional state. Facial expressions are a non-verbal form of communication that can reveal a person's emotions, and recent advancements in image processing and digital image processing techniques have made it possible to detect and analyze facial expressions automatically. By using computer vision algorithms to analyze facial expressions, it is possible to determine the user's emotional state and recommend music that matches their current mood. Recommendation engines may discover data patterns in the data set by learning user's choices and produce the outcomes that co-relates to their needs and interests. Most of the recommender systems do not consider human emotions or expressions. However, emotions have noticeable influence on the daily life of people. For a rich set of applications including human-robot interaction, computer aided tutoring, emotion aware interactive games, neuro marketing, socially intelligent software apps, computers should consider the emotions of their human conversation partners. Speech analytics and facial expressions have been used for emotion detection. However, in case of human beings prefer to camouflage their expressions, using only speech signals or facial expression signals may not be enough to detect emotions reliably. Compared with facial expressions, using physiological signals is a more reliable method to track and recognize emotions and internal cognitive processes of people. Our motivation in this work is to use emotion recognition techniques with wearable computing devices to generate additional inputs for music recommender system's algorithm, and to enhance the accuracy of the resulting music recommendations. The proposed system has the potential to enhance the user experience of music streaming services by providing personalized music recommendations based on the user's current emotional state. By utilizing facial expression analysis, the system can overcome the challenge of information overload and provide users with a more engaging and satisfying music streaming experience.

1.1 MOTIVATION

Music has been an integral part of human culture and society, providing an outlet for emotional expression, comfort, and entertainment. With the rapid growth of digital music platforms, users are faced with an overwhelming number of music choices, making it increasingly difficult to discover new music that resonates with their emotions.

1.2 PROBLEM DEFINITION

The existing music recommendation systems rely heavily on explicit user feedback, such as ratings or playlists. These methods have limitations, as users may not always provide accurate or up-to-date feedback. Moreover, these systems often neglect the emotional context of music listening, which is a crucial aspect of the music experience. As a result, users may not receive music recommendations that align with their current emotional state. This can lead to a suboptimal music listening experience, where users may not fully connect with the music.

Furthermore, the sheer volume of music available online can be overwhelming, making it difficult for users to discover new music. The lack of emotional intelligence in existing music recommendation systems exacerbates this problem.

1.3 OBJECTIVE OF THE PROJECT

The objective of the project is to develop a system that can recognize and interpret facial expressions. The system should be able to detect emotional states such as happiness, sadness, or neutral. It should provide personalized music recommendations based on the user's emotional state. The system aims to improve the user experience by providing music recommendations that align with their emotional state. It should increase user engagement by providing a personalized music recommendation experience. The system should be able to integrate facial expression recognition, emotion detection, and music recommendation. It should enable users to discover new music that aligns with their emotional state. The system should demonstrate emotional intelligence by recognizing and responding to the user's emotional state.

II. LITERATURE SURVEY

Music recommender system for users based on emotion detection through facial features, in 12th International Conference on (DeSe) [2019].

Authors: Alrihaili, A. Alsaedi, K. Albalawi.

Music recommendation system that analyzes a user's facial expressions to determine their emotional state and suggests music accordingly. The system utilizes computer vision and machine learning techniques to detect facial features and classify emotions such as happiness, sadness, anger, or neutrality. Based on the detected emotion, the system selects suitable songs from a predefined database to enhance or regulate the user's mood.

- **Emotion Recognition:** The system employs facial recognition techniques to analyze emotions using real-time image processing.
- **Machine Learning Integration:** A machine learning model is trained to classify emotions based on extracted facial features.
- **Music Recommendation:** After identifying the user's emotional state, the system matches it with an appropriate music category to enhance or adjust their mood.
- **Real-world Application:** The research suggests that such a system could improve user experience in music streaming platforms by offering more intuitive and mood-based recommendations..

An intelligent music player based on emotion recognition, in 2nd International Conference on (CSITSS),[2017]

Authors: R. Ramanathan, R. Kumaran, R. R. Rohan.

Intelligent music player that utilizes emotion recognition to automatically select and play music based on a user's facial expressions. The system aims to enhance the user experience by personalizing music recommendations in real-time according to the user's mood.

- **Emotion Recognition System:**
 - Captures facial expressions using a webcam.
 - Uses image processing techniques and machine learning algorithms to classify emotions.
 - Recognizes emotions such as happy, sad, angry, neutral, etc.
- **Music Recommendation Algorithm:**
 - Maps detected emotions to a music database categorized by mood.
 - Suggests and plays songs that match the user's current emotional state.
- **Technology Used:**
 - OpenCV for facial feature extraction.
 - Machine learning classifiers (e.g., Support Vector Machine, Neural Networks) for emotion detection.

- A predefined music dataset labeled by mood.
- Applications:
 - Enhancing music streaming services with emotion-based recommendations.
 - Useful in mental health therapy to influence mood positively.
 - Can be integrated into smart home assistants and mobile applications.

An Music Recommendation System based on Emotion Cornell University, December 2022.

Authors: Ramiz Mammadli , Huma Bilgin and Ali Can Karaca

Intelligent music player that utilizes emotion recognition to automatically select and play music based on a user's facial expressions. The system aims to enhance the user experience by personalizing music recommendations in real-time according to the user's mood.

- Emotion Recognition System:
 - Captures facial expressions using a webcam.
 - Uses image processing techniques and machine learning algorithms to classify emotions.
 - Recognizes emotions such as happy, sad, angry, neutral, etc.
- Music Recommendation Algorithm:
 - Maps detected emotions to a music database categorized by mood.
 - Suggests and plays songs that match the user's current emotional state.
- Technology Used:
 - OpenCV for facial feature extraction.
 - Machine learning classifiers (e.g., Support Vector Machine, Neural Networks) for emotion detection.
 - A predefined music dataset labeled by mood.
- Applications:
 - Enhancing music streaming services with emotion-based recommendations.
 - Useful in mental health therapy to influence mood positively.
 - Can be integrated into smart home assistants and mobile applications.

A Survey of Music Recommendation Systems and Future Perspectives

Authors: Yading Song, Simon Dixon, Marcus Pearce

Survey of music recommendation systems (MRSs), reviewing existing methods, challenges, and future research directions. It explores how different recommendation techniques are used to personalize music suggestions, enhancing user experience in digital music platforms.

Types of Music Recommendation Systems (MRS):

The authors classify MRS into several approaches:

Content-Based Filtering:

- Analyzes audio features (e.g., tempo, pitch, melody, timbre).
- Suggests similar songs based on extracted musical characteristics.

Collaborative Filtering (CF):

- Recommends music based on user preferences and behaviors (e.g., listening history).
- Includes user-user similarity and item-item similarity methods.

Hybrid Models:

- Combines content-based and collaborative filtering to improve recommendation accuracy.
- Context-Aware Recommendations:
 - Takes into account factors like location, activity, time of day, or mood.
- Emotion-Based Recommendation:
 - Uses facial expressions, voice tone, or physiological data to determine mood-based song suggestions.

Mood Classification of Hindi Songs based on Lyrics;ALC Anthology December 2015.

Author : Braja Gopal Patra, Dipankar Das, Sivaji Bandyopadhyay

"Mood Classification of Hindi Songs based on Lyrics" by Braja Gopal Patra, Dipankar Das, and Sivaji Bandyopadhyay was presented at the 12th International Conference on Natural Language Processing (ICON) in December 2015. aclanthology.org In this study, the authors proposed a mood taxonomy specifically tailored for Hindi songs and developed a mood-annotated lyrics corpus based on this taxonomy. They also annotated the lyrics with positive and negative polarity. Instead of relying solely on audio features for mood classification, the researchers focused on lyrics by extracting a wide range of semantic and stylistic features. They developed a supervised system to identify the sentiment of Hindi song lyrics using these features. The system achieved an average F-measure of 68.30% for

classifying the polarities and 38.49% for classifying the moods of the Hindi lyrics. aclanthology.org The full paper is available for download at the ACL Anthology: aclanthology.org

“Detection of Emotion Intensity Using Face Recognition” Springer Link, February 2021.

Author: hasan Ali Alharb, Mukta Dhopeswarkar and Shubhashree Savant

"Detection of Emotion Intensity Using Face Recognition" by Lhasan Ali Alharbi, Mukta Dhopeswarkar, and Shubhashree Savant was published in February 2021. The study aims to recognize the intensity of emotions by analyzing real-time video footage. The authors prepared a real live video dataset to assess emotion intensity and its percentage. The research focuses on identifying the intensity levels of various emotions through facial recognition techniques. link.springer.com

Challenges in Representation Learning: A Report on Three Machine Learning Contests" was published in 2013.

Authors: Ian J. Goodfellow, Dumitru Erhan, Yoshua Bengio, and colleagues

"Challenges in Representation Learning: A Report on Three Machine Learning Contests" by Ian J. Goodfellow, Dumitru Erhan, Yoshua Bengio, and colleagues was published in 2013. This work discusses three machine learning contests held during the ICML 2013 Workshop on Challenges in Representation Learning, focusing on:

Black Box Learning Challenge: Participants were tasked with developing models without access to the original training data, relying solely on precomputed features.

Facial Expression Recognition Challenge: This challenge aimed to classify facial expressions from images, emphasizing the development of robust feature representations.

Multimodal Learning Challenge: Participants worked on integrating information from multiple modalities, such as images and text, to improve learning performance.

“How convolutional neural networks see the world - A survey of convolutional neural network visualization methods,”[2018]

Authors: Z. Qin, F. Yu, C. Liu, et al.

"How Convolutional Neural Networks See the World: A Survey of Convolutional Neural Network Visualization Methods" by Z. Qin, F. Yu, C. Liu, and colleagues, published in 2018, provides a comprehensive overview of techniques developed to visualize and interpret the inner workings of Convolutional Neural Networks (CNNs). Understanding these methods is crucial for gaining insights into how CNNs process information and make decisions, thereby enhancing their transparency and trustworthiness.

Key Visualization Methods Covered:

Activation Maximization: This technique involves identifying input patterns that maximize the activation of specific neurons or layers within the CNN, thereby revealing the features that the network has learned to recognize.

Saliency Maps: Saliency maps highlight the regions of an input image that are most influential in the network's decision-making process, offering a visual representation of the areas deemed important by the CNN.

Deconvolutional Networks and Guided Backpropagation: These methods aim to reconstruct input images from activation patterns, providing insights into the hierarchical feature representations learned by the network.

Class Activation Mapping (CAM) and Gradient-weighted CAM (Grad-CAM): CAM and Grad-CAM generate heatmaps that localize class-specific regions in an image, facilitating the understanding of which parts of the image are associated with particular classifications.

The survey also discusses the applications of these visualization techniques in areas such as model debugging, improving network architectures, and enhancing the interpretability of CNNs in critical applications like medical image analysis.

Face detection using OpenCV with Haar Cascade Classifiers, Becominghuman.ai,[2019]

Author: V. Tabora

"Face Detection Using OpenCV with Haar Cascade Classifiers" by V. Tabora, published on Becoming Human.ai in 2019, provides a practical guide on implementing face detection using OpenCV's Haar Cascade Classifiers. This method is based on the Viola-Jones object detection framework, a robust and efficient algorithm for real-time face detection.

en.wikipedia.org

Key Components of the Viola-Jones Framework:

1. Haar-like Features: These are simple rectangular features that capture the presence of edges or lines in the image. They are reminiscent of Haar wavelets and are used to identify visual patterns.

2.Integral Image: This is a data structure that allows for rapid computation of these Haar-like features at various scales and locations in the image.

3.AdaBoost Learning: A machine learning technique that selects the most critical features and combines them into a strong classifier.

4.Cascading Classifiers: A multi-stage approach where simpler classifiers quickly eliminate non-face regions, and more complex classifiers focus on promising areas, enhancing detection speed.

III. SYSTEM ANALYSIS

3.1 EXISTING SYSTEM :

Existing techniques were using collaboration techniques which will use previous user data to recommend music to user, if there is no input from previous user then this technique will not be useful. This existing technique requires lots of manual work to arrange different music to different categories such as happy, sad or angry etc. Recommendation engines may discover data patterns in the data set by learning user's choices and produce the outcomes that co-relates to their needs and interests. Most of the recommender systems do not consider human emotions or expressions. Previously they have used galvanic skin response (GSR) and photoplethysmography (PPG). Traditional recommendation engines use content – based or collaborative filtering methods and do not consider user emotion state. However, using human emotion state with recommendation engines may increase recommendation engines performance.

3.1.1 DISADVANTAGES OF EXISTING SYSTEM

- Limited Emotion Recognition – Traditional systems do not account for real-time emotions and rely solely on predefined user preferences.
- Manual Data Labeling – Requires extensive human effort to classify songs based on emotional categories.
- Cold Start Problem – Struggles to recommend music for new users with little to no listening history.
- Vocabulary Limitation in Sentiment-Based Approaches– Past methods using text analysis (e.g., Twitter feed sentiment analysis) fail to capture emotions due to:
- Lack of contextual understanding (e.g., emoticons and expressions like ":-)").
- Sarcasm detection challenges, making sentiment-based analysis unreliable.

3.2 PROPOSED SYSTEM

Overcome from the music recommendation problem regarding user current status and then this application we Proposed a framework that involves using CNN and Digital Image Preprocessing to predict the emotion and recommend songs accordingly and will classify/predict the mood by extracting features from face. Based on detected user mood song list will be display/recommend to the user. It compares the emotions in the given list of emotions like happy, sad, angry, surprised, scared, disgust and neutral. Emotional effects of the past recommendations on the user are stored in the system's database and used in future recommendations, as the same musical track's effects can be varied between different users. The proposed system for an automatic song recommendation system based on human facial expressions using image processing and digital image processing techniques consists of three main components: facial feature extraction, emotional state recognition, and song recommendation.

The first component is facial feature extraction, which involves detecting and tracking facial features such as eyes, nose, and mouth using computer vision techniques. The system will use a face detection algorithm to detect faces in the input image or video stream, and then track facial features using a facial landmark detection algorithm. The facial feature data will be preprocessed and normalized before being passed to the next component.

The second component is emotional state recognition, which involves analyzing the facial features to determine the user's emotional state. This component will use digital image processing techniques to analyze the facial expressions and map them to emotional states such as happy, sad, angry, or neutral. The system will use a combination of feature extraction and machine learning algorithms such as Convolutional Neural Networks (CNNs) to recognize emotions from facial expressions

The third component is song recommendation, which involves selecting songs from a database that match the user's emotional state. The system will use the emotional state recognized in the previous component to select songs that have been tagged with similar emotional labels. The song database will be pre-tagged with emotional labels using music

metadata or crowd-sourced tagging techniques. The system will also incorporate a feedback mechanism that allows users to rate the recommended songs and provide feedback to improve the recommendations in the future.

The proposed system will continuously learn and adapt to the user's preferences over time by utilizing machine learning algorithms that analyze the user's feedback and adjust the recommendation algorithm accordingly. The system will be designed to work in real-time, allowing for seamless integration with music streaming services or other applications. The overall goal of the proposed system is to provide users with personalized music recommendations based on their current emotional state, enhancing the user experience and increasing engagement and satisfaction.

3.2.1 ADVANTAGES OF PROPOSED SYSTEM

- **Personalized Music Recommendations:** Provides users with personalized music recommendations based on their current emotional state, enhancing the user experience.
- **Real-time Emotion Detection:** Uses computer vision and machine learning algorithms to detect emotions in real-time, allowing for seamless integration with music streaming services.
- **Improved User Engagement:** Increases user engagement and satisfaction by providing music recommendations that align with their emotional state.
- **Adaptive Recommendation Algorithm:** Continuously learns and adapts to the user's preferences over time, improving the accuracy of music recommendations.
- **Emotional Intelligence:** Demonstrates emotional intelligence by recognizing and responding to the user's emotional state, providing a more empathetic and supportive user experience.

These advantages highlight the benefits of the proposed Facial Expression Driven Music Recommendation System in providing personalized music recommendations, improving user engagement, and demonstrating emotional intelligence.

IV. SYSTEM DESIGN

4.1 SYSTEM ARCHITECTURE

Below diagram depicts the whole system architecture of “An automatic song recommendation system based on human facial expressions”.

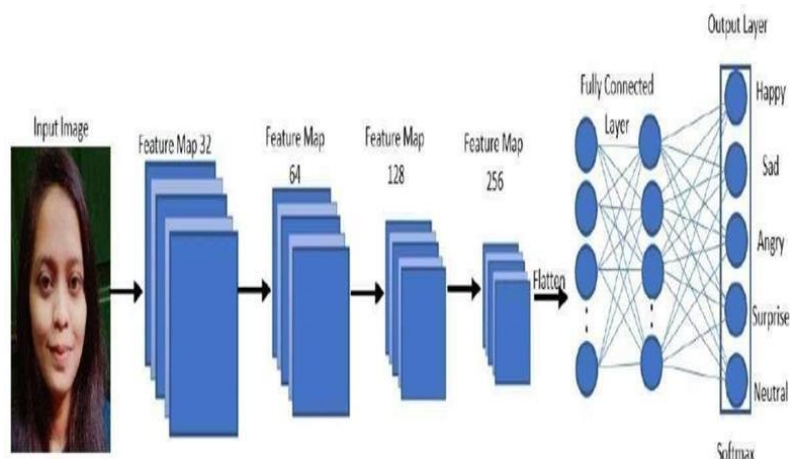


Fig 1: System Architecture

4.2 UML DIAGRAMS

UML stands for Unified Modeling Language. UML is a standardized general-purpose modeling language in the field of object-oriented software engineering. The standard is managed, and was created by, the Object Management Group. The goal is for UML to become a common language for creating models of object oriented computer software. In its current form UML is comprised of two major components: a Meta-model and a notation. In the future, some form of method or process may also be added to or associated with, UML. The Unified Modeling Language is a standard language for specifying, Visualization, Constructing and documenting the artifacts of software system, as well as for business modeling and other non-software systems.

UML was created as a result of the chaos revolving around software development and documentation. In the 1990s, there were several different ways to represent and document software systems.

The UML represents a collection of best engineering practices that have proven successful in the modeling of large and complex systems. The UML is a very important part of developing objects oriented software and the software development process. The UML uses mostly graphical notations to express the design of software projects.

4.2a GOALS:

The Primary goals in the design of the UML are as follows:

1. Provide users a ready-to-use, expressive visual modeling Language so that they can develop and exchange meaningful models.
2. Provide extendibility and specialization mechanisms to extend the core concepts.
3. Be independent of particular programming languages and development process.
4. Provide a formal basis for understanding the modeling language.
5. Encourage the growth of object oriented tools market.
6. Support higher level development concepts such as collaborations, frameworks, patterns and components.
7. Integrate best practices.

V. RESULTS

The following figures present the sequence of screenshots of the results.

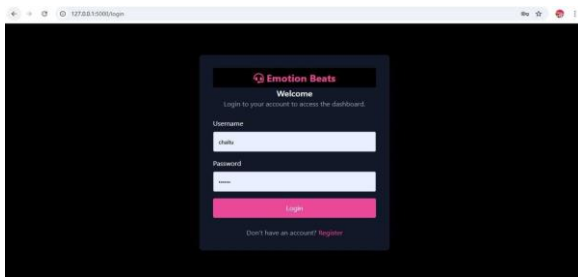


Fig 2a: Login page

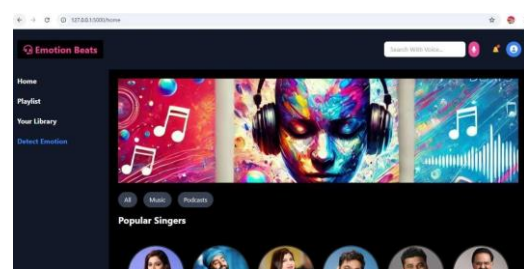


Fig 2b: Home page

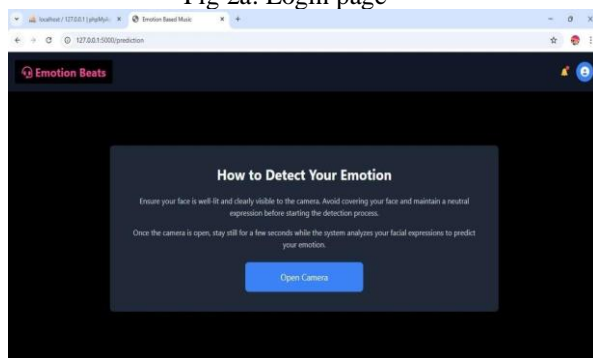


Fig 2c: Emotion Detection page

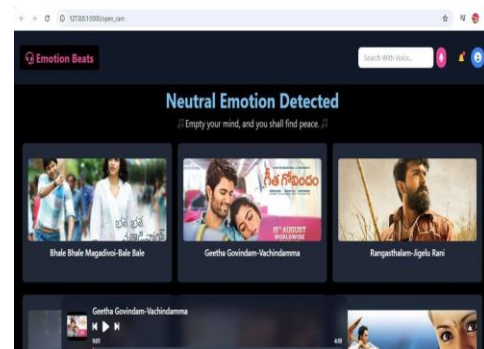


Fig 2d: Prediction song page

VI. CONCLUSIONS AND FUTURE WORK

6.1 CONCLUSIONS

In this work, Deep Learning-Based Facial Expression-Driven Song Suggestion System has transformed the music recommendation process by integrating real-time emotion detection with music selection. Unlike traditional recommendation systems that rely on past user preferences, this system leverages computer vision, deep learning models, and real-time facial emotion analysis to provide instant, mood-based song recommendations.

By using OpenCV and deep learning-based facial expression recognition, the system detects the user's emotion from a live webcam feed and maps it to a music database categorized by emotions such as happy, sad, angry, and neutral. The model, built using Keras and TensorFlow, processes facial expressions and accurately predicts emotions, ensuring that the recommended songs match the user's current mood.

Unlike conventional recommendation engines that depend on content-based or collaborative filtering, this system provides a more personalized and dynamic music experience by adapting to real-time emotional states. This innovation enhances user engagement, emotional wellbeing, and music discovery, making it a more interactive and immersive experience.

6.2 FUTURE WORK

The Facial Expression-Driven Song Suggestion System can be enhanced in several ways to improve accuracy, personalization, and user experience. One major improvement is multimodal emotion recognition, where facial expressions can be combined with speech emotion analysis to provide more precise recommendations. Additionally, text-based sentiment analysis from chat inputs or social media can further refine song suggestions. The system can also be extended to create emotion-based playlists instead of single-song recommendations, offering users a continuous, mood-adaptive listening experience. Future development could focus on mobile app integration, allowing users to receive real-time music recommendations on Android and iOS devices. Cloud-based deployment using AWS or Google Cloud could enhance scalability, enabling integration with platforms like Spotify for seamless playback. Furthermore, incorporating AI-driven emotion coaching could transform this system into a well-being assistant, suggesting music to reduce stress, boost focus, or enhance relaxation. These advancements will make the system more intelligent, adaptive, and widely accessible, ultimately revolutionizing the way users interact with music based on their emotions.

REFERENCES

- [1] S. Jhahharia, S. Pal, and S. Verma, "Wearable computing and its application," *Int. J. Comp. Sci. and Inf. Tech.*, vol. 5, no. 4, pp. 5700–5704, 2014.
- [2] K. Popat and P. Sharma, "Wearable computer applications: A feature perspective," *Int. J. Eng. and Innov. Tech.*, vol. 3, no. 1, 2013.
- [3] P. Melville and V. Sindhvani, "Recommender systems," in *Encyc. of mach. learn.* Springer, 2011, pp. 829–838.
- [4] N. Sebe, I. Cohen, T. S. Huang et al., "Multimodal emotion recognition," *Handbook of Pattern Recognition and Computer Vision*, vol. 4, pp. 387–419, 2005.
- [5] R. W. Picard, E. Vyzas, and J. Healey, "Toward machine emotional intelligence: Analysis of affective physiological state," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 23, no. 10, pp. 1175–1191, 2001.
- [6] D. Ayata, Y. Yaslan, and M. Kamasak, "Emotion recognition via galvanic skin response: Comparison of machine learning algorithms and feature extraction methods," *IU J. of Elect. & Elect. Eng.*, vol. 17, no. 1, pp. 3129–3136, 2017.
- [7] P. Ekman, R. W. Levenson, and W. V. Friesen, "Autonomic nervous system activity distinguishes among emotions." *Am. Assoc. for Adv. of Sci.*, 1983.
- [8] I.-h. Shin, J. Cha, G. W. Cheon, C. Lee, S. Y. Lee, H.-J. Yoon, and H. C. Kim, "Automatic stress- relieving music recommendation system based on photoplethysmography-derived heart rate variability analysis," in *IEEE Int. Conf. on Eng. in Med. and Bio. Soc. IEEE*, 2014, pp. 6402–6405.
- [9] S. Nirjon, R. F. Dickerson, Q. Li, P. Asare, J. A. Stankovic, D. Hong, B. Zhang, X. Jiang, G. Shen, and F. Zhao, "Musicalheart: A hearty way of listening to music," in *Proc. of ACM Conf. on Emb. Netw. Sens. Sys. ACM*, 2012, pp. 43–56.
- [10] P. Pulivarthy and P. Whig, "Bias and fairness addressing discrimination in ai systems," *Advances in human and social aspects of technology book series*, pp. 103–126, 2024.
- [11] H. Liu, J. Hu, and M. Rauterberg, "Music playlist recommendation based on user heartbeat and music preference," in *Int. Conf. on Comp. Tech. and Dev.*, vol. 1. IEEE, 2009, pp. 545–549.

International Journal of Advanced Research in Education and Technology

ISSN: 2394-2975

Impact Factor: 8.152