International Research Journal on Advanced Engineering and Management
https://goldncloudpublications.com
https://doi.org/10.47392/IRJAEM.2025.0394
e ISSN: 2584-2854
Volume: 03
Issue:07 July 2025
Page No: 2506 - 2511

Recognition of Emotions in Music Using Machine Learning Algorithms

Prof. Sheetal V. Shelke[1], Dr. Mangal Patil[2], Prof. Vinod P. Mulik[3], Prof. Kanchan D. Mahajan[4], Ms. Isha P. Vetal[5], Ms. Nita R. Sonawane[6]

[1]Assistant Professor, Bharati Vidyapeeth's College of Engineering for Women, Pune, India.

[2]Associate Professor, Bharati Vidyapeeth College of Engineering (Deemed to be University), Pune, India.

[3,4]Assistant Professor, Bharati Vidyapeeth's College of Engineering for Women, Pune, India.

[5,6]Student, Bharati Vidyapeeth's College of Engineering for Women, Pune, India.

Emails: *sheetal.shelke@bharatividyapeeth.edu[1], mvpatil@bvucoep.edu[2], vinod.mulik@bharatividyapeeth.edu[3], kanchan.mahajan@bharatividyapeeth.edu[4], ishaa18901@gmail.com[5], nitasonawane1018@gmail.com[6]*

Abstract

Music is vital for entertainment, emotion regulation, and stress relief. With digital platforms like Spotify, classifying large music datasets has become essential. This report introduces a machine-learning framework to detect four emotions Happy, Sad, Calm, and Energetic in Hindi songs. Music has the unique ability to evoke and convey a wide range of human emotions, making it a powerful medium for both artistic expression and practical applications. A curated Hindi music dataset was segmented into 20-second WAV clips, pre-processed with high-pass filtering and volume normalization. Acoustic features extracted included: 13-dimensional MFCCs, 12-dimensional Chroma vectors, Zero-Crossing Rate, Spectral Rolloff. Data was split into training, validation, and testing sets using stratified sampling. Three classifiers were applied: Decision Tree, Random Forest, and XGBoost. XGBoost performed best with 89% accuracy, while Random Forest and Decision Tree achieved 84% and 68%, respectively. Results show ensemble models effectively classify emotions in regional music and support applications like mood-based playlists and smart music systems.

Keywords: *Machine Learning Algoritms, Random Forest, Decision Tree, Gradient Boosting, Emotions, Music files*

1. Introduction

Music Emotion Recognition aims to identify the emotional content embedded within musical compositions. Emotions in music are influenced by various acoustic elements such as melody, harmony, rhythm, tempo, dynamics, and even the semantic content of lyrics[1][2]. Over the years, MER has evolved significantly with the advent of machine learning and signal processing techniques. Researchers have developed a variety of approaches to capture and quantify these emotional cues, making use of both traditional handcrafted features and modern deep learning methods that can learn representations directly from raw data.With the arrival of the digital age, music's historical significance has increased. There has never been a day when so much music has been created and heard. Because of the Internet's widespread use and compact audio formats, which are comparable in quality to CDs, digital music libraries have expanded faster. Emotion recognition enhances our comprehension of human emotions and behavior. These technologies can be utilized by researchers and psychologists to examine emotional reactions in various situations, resulting in a deeper understanding of human psychology and social dynamics. Emotion recognition plays a significant role in delivering customized experiences in diverse fields such as music, video recommendations, and content creation. By comprehending users' emotional preferences, systems can customize content recommendations to align with the emotional context of each individual. Emotion recognition has diverse applications in the

International Research Journal on Advanced Engineering and Management
https://goldncloudpublications.com
https://doi.org/10.47392/IRJAEM.2025.0394
e ISSN: 2584-2854
Volume: 03
Issue:07 July 2025
Page No: 2506 - 2511

monitoring of mental health, as well as in therapy and interventions. It can aid healthcare professionals in comprehending patients' emotional conditions, identifying mood disorders, and formulating more efficient treatment strategies. It contributes to market research by examining consumer reactions to products and advertisements. Comprehending emotional responses enables businesses to customize their marketing strategies and enhance the efficacy of their campaigns. Emotion recognition in robotics enhances the robots' comprehension and reaction to human emotions. This is especially pertinent in the context of social robots or assistive devices, where the capacity to identify and adjust to human emotions is of utmost importance. To summarize, emotion recognition has a multifaceted impact on technology, psychology, healthcare, and societal interactions. As the field progresses, its influence is expected to expand, resulting in additional breakthroughs and implementations in various fields. Since almost all musical compositions aim to evoke a particular feeling in the listener, music has been categorized and retrieved according to emotion [3][4]. The majority of people believe that love is required for the creation, performance, and enjoyment of music. Music possesses the ability to evoke emotions such as sadness, love, or solace during challenging periods. Studies on music information behavior suggest that individuals also consider emotions when choosing and organizing music [5]. By utilizing affective adjectives such as calm, romantic, sentimental, defiant, fiery, and easy going, individuals can categories their music library based on mood. Teaching computers to identify emotions in music also improves the quality of human-computer interaction. Music that aligns with the user's moods can be played by analyzing prosodic, physiological, or facial cues. A portable device with automatic music emotion recognition (MER), such as an MP3 player or cell phone, can then play the song that most closely matches the user's emotional state. A smart space's background music can be changed to suit the needs of the people utilizing it, be it a restaurant, conference room, home etc [6]. The automatic identification of the perceived emotional content of music has advanced significantly in the field of music information retrieval (MIR). Typically, emotions are divided into a number of classes (such as joyful, furious, depressed, and at ease) and a classifier is then trained using machine learning methods. The timber, rhythm, and harmony components of a musical composition are typically removed to depict the acoustic characteristics of the music. Usually, a subjective test is employed to gather the accurate and reliable data required to train the computer in the model of emotion recognition. Diverse machine learning algorithms, such as support vector machines, have been utilized to examine the relationship between musical characteristics and emotional classifications [7]. The MER approach employs a categorical framework to classify emotions into distinct groups, and subsequently utilizes machine learning techniques to train a classifier. In the dimensional approach to MER, emotions are defined as quantitative values along different dimensions of emotion. In this context, the song is perceived as a singular point within an emotion space. Subsequently, the emotion space offers a fundamental interface for users to organize, investigate, and retrieve musical compositions. The field of emotion recognition is intricate, encompassing numerous challenges and issues that necessitate attention from both researchers and practitioners. There is a scarcity of superior, standardized datasets that can be used to train and evaluate emotion recognition models. This can impede the advancement and assessment of resilient models, as the full range of emotions and contexts is not sufficiently depicted [8]. The analysis of continuously changing emotional states in real-time, particularly in dynamic environments, presents challenges that arise from the requirement for prompt and precise recognition. Latency problems can impact the usability of emotion recognition systems in specific applications.

2. Literature Survey

Recent advancements in emotion recognition have seen a surge in machine learning techniques applied to both EEG and audio signals. Salama et al. (2018) utilized Long Short-Term Memory (LSTM) networks on EEG signals from the DEAP dataset to detect emotional states. Their model achieved high accuracy

International Research Journal on Advanced Engineering and Management
https://goldncloudpublications.com
https://doi.org/10.47392/IRJAEM.2025.0394
e ISSN: 2584-2854
Volume: 03
Issue:07 July 2025
Page No: 2506 - 2511

in classifying valence and arousal levels, though it lacked the capability to fully model temporal features in the data [9]. To address the limitations of traditional CNNs in temporal analysis, Jiang et al. (2021) introduced a Wavelet-Transformed CNN model. This approach allowed for better frequency resolution and achieved 80.65% accuracy, but still fell short in capturing long-term dependencies in EEG sequences [10]. Ozdemir et al. (2021) and Zhang et al. (2020) improved performance by combining CNNs with LSTM networks. These hybrid models successfully extracted spatial and temporal patterns from EEG signals, enhancing classification performance beyond 90% accuracy. However, they often failed to focus on the most emotionally salient regions of the input [11][12]. In a more recent study, Qiao et al. (2024) proposed a CNN-SA-BiLSTM model that incorporated self-attention mechanisms. This approach significantly improved accuracy, with arousal detection reaching up to 96.36%, demonstrating the advantages of integrating attention layers with bidirectional LSTM architectures [13]. From the perspective of music-based emotion recognition, Han et al. (2023) emphasized the potential of lightweight, attention-based models and the importance of multimodal fusion in emotion classification tasks [14]. Spectrogram-based CNNs and BiLSTM networks, as explored by Sarkar et al. (2020) and Liu et al. (2018), further highlighted the effectiveness of using both time-frequency representations and handcrafted audio descriptors in enhancing model accuracy [15][16]. Delbouys et al. (2018) demonstrated the benefits of combining audio and lyrics in a CNN-LSTM architecture, with late fusion techniques leading to an improvement of 3–5% in classification performance [17]. Similarly, Chaki et al. (2020) incorporated attention-augmented LSTM layers to focus on emotionally significant segments of songs, leading to more accurate predictions [18]. In real-time applications, Angusamy et al. (2020) and Srivastav et al. (2022) developed emotion-aware systems using facial expression detection combined with audio analysis, suitable for adaptive music environments [19][20]. Furthermore, Xia and Xu (2022) applied regression models to map audio features into the valence-arousal emotional space, enabling continuous emotion tracking [21]. Recently, studies by Agawane et al. (2024) and Nikhil K. et al. (2024) explored real-time music players that respond to facial expressions. These systems demonstrated promise in linking user emotions to curated playlists, although challenges remain in accurately detecting emotions under variable lighting and facial expression conditions [22].

3. Methodology

The methodology for implementing a Music Emotion Classification (MEC) model tailored for songs involves a systematic approach, combining data collection, preprocessing, feature extraction, model development, and evaluation.

3.1 Block diagram

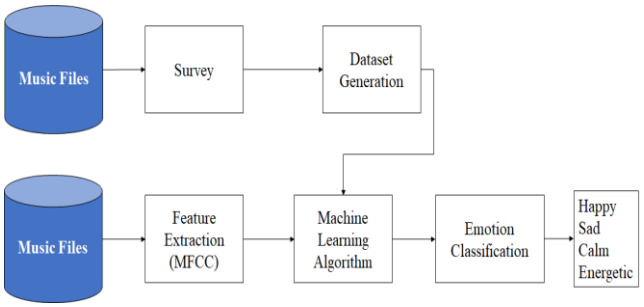

Figure 1 Block Diagram of System

3.2 Block diagram Explanation

3.2.1 Music Files

A curated collection of Hindi songs was compiled from various genres like filmi, devotional, and folk, ensuring diversity in both male and female vocals. Songs were categorized into four emotional classes Happy, Sad, Calm, and Energetic. These were segmented into 20-second non-overlapping clips, forming the base for annotation and classification.

3.2.2 Dataset Generation

Using the 350 emotion-tagged songs, clips were generated via 20-second segmentation using Librosa. Initial labels were assigned based on the song category, then verified against survey responses. Clips lacking at least 70% agreement among listeners or expert reviewers were discarded.

3.2.3 Feature Extraction

Extract important features from the pre-processed audio data. Commonly used features for audio

International Research Journal on Advanced Engineering and Management
https://goldncloudpublications.com
https://doi.org/10.47392/IRJAEM.2025.0394
e ISSN: 2584-2854
Volume: 03
Issue:07 July 2025
Page No: 2506 - 2511

classification include Mel-frequency cepstral coefficients, Spectral Centroid, Spectral Bandwidth, Spectral Rolloff, , etc were extracted

3.2.4 Emotion Classification

The Emotion Classification block applies the trained machine learning model to classify emotions in new, unseen music files. Machine Learning classifiers such as Random Forest, Decision tree, Gradient Boosting were used.

3.2.5 Emotion Output

Finally, the Emotion Output block presents the classified emotions to the user or integrates them into relevant applications. The output emotions, such as Happy, Sad, Calm, or Energetic, can be displayed on a user interface.

4. Results and Discussion

A detailed evaluation of the proposed emotion-recognition framework, beginning with quantitative performance metrics for each classifier on the held-out test set. The report average precision, recall, F1-score, and overall accuracy for Decision Tree, Random Forest, and Gradient Boosting models, highlighting their relative strengths and weaknesses. Hindi-song dataset and inform potential avenues for further refinement.

4.1 Results

Table for comparing the three classifiers Random Forest, Gradient Boosting, and Decision Tree on four evaluation metrics averaged over all emotion classes.

Table 1 Comparison of Classifiers

Classifier	Precision	Recall	F1-Score	Accuracy
Random Forest	0.8912	0.8835	0.8936	0.8924
Gradient Boosting	0.8432	0.8321	0.8436	0.8432
Decision Tree	0.6812	0.6932	0.6837	0.6827

4.2 Discussion

These values indicate that Random Forest correctly identifies over 89 % of the emotion labels while maintaining a very low false-positive rate, reflecting its robustness in combining multiple decision trees to reduce overfitting and exploit diverse feature subsets. Although slightly lower than Random Forest, Gradient Boosting still demonstrates strong classification capability by sequentially correcting errors at each boosting iteration with accuracy of 84%. In contrast, the single Decision Tree yields substantially lower performance accuracy of 68% indicating that a lone tree struggles to generalize across the varied acoustic-feature space of the Hindi-song dataset. In summary, Table no.1 underscores that ensemble approaches (Random Forest and Gradient Boosting) markedly outperform a standalone Decision Tree, with Random Forest providing the highest balanced trade-off between precision and recall for emotion recognition.

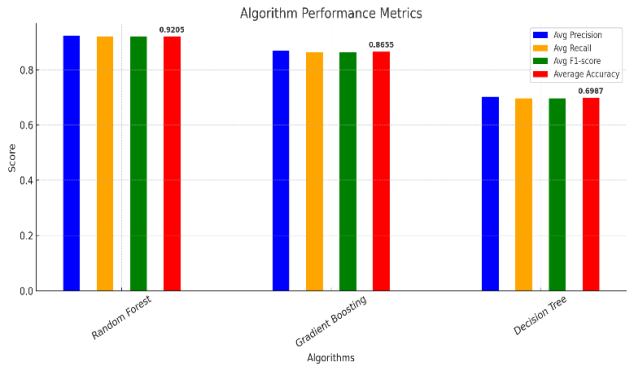

Figure 2 Performance Metrics of the Algorithms

Conclusion

Created an annotated dataset and extracted meaningful features like MFCCs. Developed an MEC model for Hindi songs using Machine Learning. Random Forest achieves the best overall performance, with an average precision of 0.8921, recall of 0.8835, F1-score of 0.8936, and accuracy of 0.8924. These values indicate that Random Forest correctly identifies over 89 % of the emotion labels while maintaining a very low false-positive rate, reflecting its robustness in combining multiple decision trees to reduce overfitting and exploit diverse feature subsets. Gradient Boosting follows closely, with an average precision of 0.8432, recall of 0.8321, F1-score of 0.8436, and accuracy of 84%. Although slightly lower than Random Forest, Gradient Boosting still demonstrates strong classification capability by sequentially correcting errors at each boosting iteration, but it may be more

International Research Journal on Advanced Engineering and Management
https://goldncloudpublications.com
https://doi.org/10.47392/IRJAEM.2025.0394
e ISSN: 2584-2854
Volume: 03
Issue:07 July 2025
Page No: 2506 - 2511

sensitive to noise or outliers, which accounts for its roughly 6 % drop in metrics compared to Random Forest.In contrast, the single Decision Tree yields substantially lower performance—average precision of 0.6812, recall of 0.6932, F1-score of 0.6837, and accuracy of 69% indicating that a lone tree struggles to generalize across the varied acoustic-feature space of the Hindi-song dataset.

Acknowledgements

We would like to thank the researchers as well as publishers for making their resources available and colleagues for their guidance. We are thankful to the authorities of Savitribai Phule University, Pune and concerned members of conference organized by Bharati Vidyapeeth's College of Engineering for Women, Pune for their constant guidelines and support.

References

[1]. Z. Yang and H. Chen, “A survey of music emotion recognition,” Journal of Computer Applications, vol. 30, no. 1, pp. 123–130, 2020.

[2]. W. Zhang and L. Wang, “Emotion recognition for internet music by multiple classifiers,” Procedia Computer Science, vol. 96, pp. 885–893, 2016

[3]. M. R. Jones and T. Kim, “Music emotion recognition from content to context,” Multimedia Tools and Applications, vol. 79, no. 15, pp. 10103–10127, 2020.

[4]. M. Zhao, L. Yang, and Y. Lin, “Audio-based deep music emotion recognition,” IEEE Access, vol. 8, pp. 80670–80679, 2020.

[5]. L. He and F. Sun, “Music recognition and classification algorithm considering genre and emotional mood,” Journal of Intelligent & Fuzzy Systems, vol. 38, no. 1, pp. 849–860, 2020.

[6]. P. Verma and A. Saha, “Machine learning approaches for emotion classification of music: A review,” International Journal of Engineering and Technology, vol. 7, no. 2, pp. 245–250, 2019.

[7]. Dhanapala, S., & Samarasinghe, K. (2024). Music Emotion Classification: A Literature Review. Journal of Research in Music, 2(2), 14–28. https://doi.org/10.4038/jrm.v2i2.27.

[8]. Zhou, L., Yang, Y., & Li, S. (2021). Music-induced emotions influence intertemporal decision making. Cognition and Emotion, 211–229. https://doi.org/10.1080/02699931.2021.1995331.

[9]. S. Salama, M. A. Fahmy, and R. A. El-Khoribi, "Emotion recognition based on EEG using LSTM recurrent neural network," International Journal of Advanced Computer Science and Applications, vol. 8, no. 10, pp. 355–358, 2018.

[10]. R.-N. Jiang, J.-Y. Zhu, and B.-L. Lu, "Wavelet transformed convolutional neural network for EEG emotion recognition," in Proceedings of the International IEEE/EMBS Neural Engineering Conference (NER), San Diego, CA, 2021, pp. 300–305.

[11]. M. Ozdemir, P. J. McDonald, and E. W. Greene, "A robust hybrid CNN-LSTM approach for emotion recognition using EEG," IEEE Transactions on Cognitive and Developmental Systems, vol. 13, no. 3, pp. 945–953, Sep. 2021.

[12]. Y. Zhang, R. Wang, and Y. Ma, "CNN-LSTM architecture for EEG-based emotion recognition with audio-visual stimuli," in Proceedings of the IEEE International Conference on Artificial Intelligence and Neural Networks, Vancouver, BC, 2020, pp. 175–182.

[13]. Y. Qiao, J. Mu, J. Xie, B. Hu, and G. Liu, "Music emotion recognition based on temporal convolutional attention network using EEG," Frontiers in Human Neuroscience, vol. 18, no. 1324897, pp. 1–14, Mar. 2024.

[14]. Han, X., Chen, F., & Ban, J. (2023). Music Emotion Recognition Based on a Neural Network with an Inception-GRU Residual Structure. Electronics, 12(4), 978. https://doi.org/10.3390/electronics12040978

[15]. Sarkar, R.; Choudhury, S.; Dutta, S.; Roy, A.; Saha, S.K. Recognition of emotion in music based on deep convolutional neural

International Research Journal on Advanced Engineering and Management
https://goldncloudpublications.com
https://doi.org/10.47392/IRJAEM.2025.0394
e ISSN: 2584-2854
Volume: 03
Issue:07 July 2025
Page No: 2506 - 2511

network. Multimed. Tools Appl. 2020, 79, 765–783.

[16]. Y. Liu et al., "Lightweight ViT Model for Micro-Expression Recognition Enhanced by Transfer Learning," Frontiers in Neurorobotics, vol. 16, 2022. DOI: 10.3389/fnbot.2022.922761.

[17]. Delbouys, Rémi & Hennequin, Romain & Piccoli, Francesco & Royo-Letelier, Jimena & Moussallam, Manuel. (2018). Music Mood Detection Based On Audio And Lyrics With Deep Neural Net. 10.48550/arXiv.1809.07276.

[18]. J. Chaki et al., "Machine learning and artifical intelligence based Diabetes Mellitus detection and self-management: A systematic review," Journal of King Saud University-Computer and Information Sciences, vol. 34, no. 6, pp. 3204-3225, 2020.

[19]. Angusamy, Punidha & S, Inba & K.S, Pavithra & M, Ameer & M, Athiparasakthi. (2020). Human Emotion Detection using Machine Learning Techniques. SSRN Electronic Journal. 10.2139/ssrn.3591060.

[20]. Srivastav, Mallika & Mathur, Prakhar & Poongodi, T. & Sagar, Shrddha & Yadav, Suman. (2022). Human Emotion Detection Using OpenCV748-751. 10.1109/ICIPTM54933.2022.9754019.

[21]. Yu Xia and Fumei Xu, “Study on Music Emotion Recognition Based on the Machine Learning Model Clustering Algorithm,” Hindawi Mathematical Problems in Engineering Volume 2022, Article ID 9256586.

[22]. S. Agawane, A. Wakhure, A. Bharsakle, and D. Bhosale, "Emotions-Based Music Player," International Journal for Multidisciplinary Research (IJFMR), vol. 6, no. 3, May-June 2024, pp. 1–6.