International Journal of Advanced Education and Research
www.alleducationjournal.com
ISSN: 2455-5746
Received: 29-04-2025, Accepted: 30-05-2025, Published: 14-06-2025
Volume 10, Issue 2, 2025, Page No. 89-95

Harmonizing emotions: Deep learning approaches for music emotion classification

Arjun Mathur, Priyanshu Sharma, Shivam Yadav, Dr. Meena Chaudhary, Dr. Gunjan Chandwani
Department of Computer Science and Technology Manav Rachna University Faridabad, Haryana, India

Abstract
Music has a profound impact on human emotions, influencing mood, relaxation, and motivation. This research paper presents a novel approach to classifying songs into three distinct emotional categories—meditation, motivational, and sad—by deriving mathematical formulas from their beats and tunes. We analyze rhythmic patterns, tempo variations, harmonic structures, and spectral features to quantify musical characteristics that define each category. A machine learning model is trained using these features to automate classification. Our experiments demonstrate that beat intensity, tempo stability, and harmonic complexity are key discriminative factors, achieving an overall classification accuracy of 89.3%. This study contributes to the fields of music information retrieval (MIR) and affective computing by providing a mathematically grounded framework for emotion-based song classification.

Keywords: Music emotion classification, beat analysis, tempo variation, harmonic structure, spectral features, machine learning, affective computing, music information retrieval (mir), emotion recognition in music, audio signal processing

Introduction

Music profoundly influences human emotions, playing a pivotal role in shaping our mood, behavior, and cognitive responses. From meditative soundscapes that encourage mindfulness to powerful beats that drive motivation, different musical patterns are intricately linked to emotional experiences. Recognizing and classifying these emotional effects is a growing area of interest in the domains of Music Information Retrieval (MIR) and affective computing, with applications ranging from music recommendation engines to mental wellness platforms.

Traditional approaches to Music Emotion Classification (MEC) often depend on deep learning models such as Convolutional Neural Networks (CNNs) or Recurrent Neural Networks (RNNs). While these models offer high accuracy, they also require large annotated datasets and significant computational power, making them inaccessible for smaller research environments or lightweight applications.

This study introduces a lightweight and mathematically interpretable framework to classify songs into three core emotional categories: meditation, motivational, and sad. Rather than relying on high-level black-box models, we derive mathematical representations from core musical components—beat intensity, tempo stability, and harmonic complexity—to form a compact and effective feature set. These features are analyzed and fed into simple machine learning models such as decision trees and k-nearest neighbors (k-NN), offering a balance between performance and simplicity.

Our goal is to provide a cost-effective, resource-efficient, and accurate classification methodology that democratizes the use of emotional classification in music, especially for academic use, small startups, and mental health-related technology solutions.

Related Work

Classifying music according to emotional content has been a key research focus in the fields of music information retrieval (MIR) and affective computing. Traditional models in this domain primarily relied on handcrafted features such as tempo, rhythm, pitch, and timbre to categorize music into emotional classes like happy, sad, calm, or energetic. For instance, Hu and Zhang [1, 14]. used tempo and rhythmic structures to distinguish musical emotion classes. Mokhsin *et al.* [2]. utilized artificial neural networks based on vocal and instrumental timbres.

With the advancement of machine learning, researchers adopted classifiers like Support Vector Machines (SVM), k-Nearest Neighbors (k-NN), and Decision Trees for improved performance. Liu *et al.* [3]. and Chaudhary *et al.* [4]. explored these traditional ML techniques using structured features extracted from audio signals. These models offered interpretability and required less computational power compared to deep learning.

Recent approaches leverage deep learning architectures, including CNNs, RNNs, and attention-based models. Jia [5, 6]. proposed an attention-enhanced deep learning method for classifying musical emotions with improved precision. Grekow [7]. employed pretrained RNNs for symbolic music. Zeng *et al.* [8]. The researchers introduced MusicBERT, a transformer-driven model aimed at interpreting symbolic music.

Takashima *et al.* [9]. and Huang *et al.* [10]. focused on embedding-based models and attention fusion frameworks, respectively, to enhance classification performance. Other notable contributions include spectral and harmonic analysis methods [11], multimodal approaches combining lyrics and audio [12], and large-scale pretraining [13].

However, deep learning models, while powerful, require significant computational resources and large labeled datasets. For practical applications, especially where interpretability and efficiency are essential, lightweight models remain highly relevant. Our work aligns with studies like those of Zhang *et al.* [14]. and Li [15], which aim to bridge the gap between complexity and performance.

1. Comparative Table of Techniques

Technique	Accuracy	Precision	Recall	F1-Score	Computational Complexity	Interpretability
Decision Tree (our)	86%	0.87	0.86	0.86	Low	High
k-NN (our)	82%	0.83	0.82	0.82	Medium	Medium
CNN-based [3]	89%	0.90	0.89	0.89	High	Low
RNN-based [7]	88%	0.89	0.88	0.88	High	Low
Attention-based [5]	91%	0.92	0.91	0.91	Very High	Medium
SVM [4]	84%	0.85	0.84	0.84	Medium	Medium

Methodology

1. Feature Extraction

The first step in the classification pipeline involves the extraction of interpretable musical features from each audio file. This study focuses on three core characteristics that have been shown to correlate strongly with emotional perception in music: beat intensity, tempo stability, and harmonic complexity.

- **Beat Intensity (BI):** Refers to the perceived strength and regularity of beats within a track. Meditation tracks usually exhibit low-intensity, sparse beats promoting calmness. In contrast, motivational tracks contain high-intensity, frequent beats that energize the listener. Sad songs often contain soft or irregular beats, contributing to their emotional gravity.

- Tempo Stability (TS), expressed in beats per minute (BPM), represents the rhythm's speed and steadiness. Meditation music generally features a slow, consistent tempo, whereas motivational music is marked by a quicker and more stable pace. Sad music often displays slower tempos with noticeable fluctuations, enhancing its melancholic tone.

- **Harmonic Complexity (HC):** Describes the musical structure based on key, chord progression, and tonal changes. Motivational music often features major keys and varied chord transitions, sad music leans toward minor keys with complex harmonies, and meditation music tends to use minimal harmonic progression to promote a repetitive and tranquil soundscape.

These features were extracted using the Librosa library in Python. Librosa provides advanced signal processing techniques to analyze audio waveforms and derive temporal and spectral features. After extraction, all features were normalized using Min-Max scaling to ensure uniformity and to enhance learning efficiency.

2. Dataset

The dataset comprises 500 curated audio samples, each labeled under one of three emotion categories: meditation, motivational, or sad. These audio files were collected from publicly accessible repositories such as the Free Music Archive (FMA) and YouTube Audio Library, focusing on diversity in composition while maintaining balanced class representation (167 tracks per class).

Each track underwent the following preprocessing steps:

- Conversion to mono channel,
- Silence removal and amplitude normalization,
- Trimming of excess non-musical segments,
- Standardization to a consistent sample rate (22,050 Hz).

Songs were stored in WAV or MP3 formats with a duration range of 3 to 5 minutes. The final dataset was split into training (70%) and testing (30%) subsets, ensuring that all emotional categories were evenly represented across both sets.

3. Model Selection

To maintain model simplicity and interpretability, two lightweight machine learning classifiers were selected:

- **Decision Tree:** A rule-based classifier that segments the data based on feature thresholds. It builds a tree-like structure where internal nodes represent feature decisions, and leaf nodes correspond to final class predictions. It is favored for its transparency and visual interpretability, particularly suitable for small and mid-sized datasets.

- **K-Nearest Neighbors (k-NN):** An instance-based classifier that assigns the class of a new sample based on the most common label among its *k* nearest neighbors in the feature space. Although computationally expensive at inference time, k-NN is effective for low-dimensional data and small datasets.

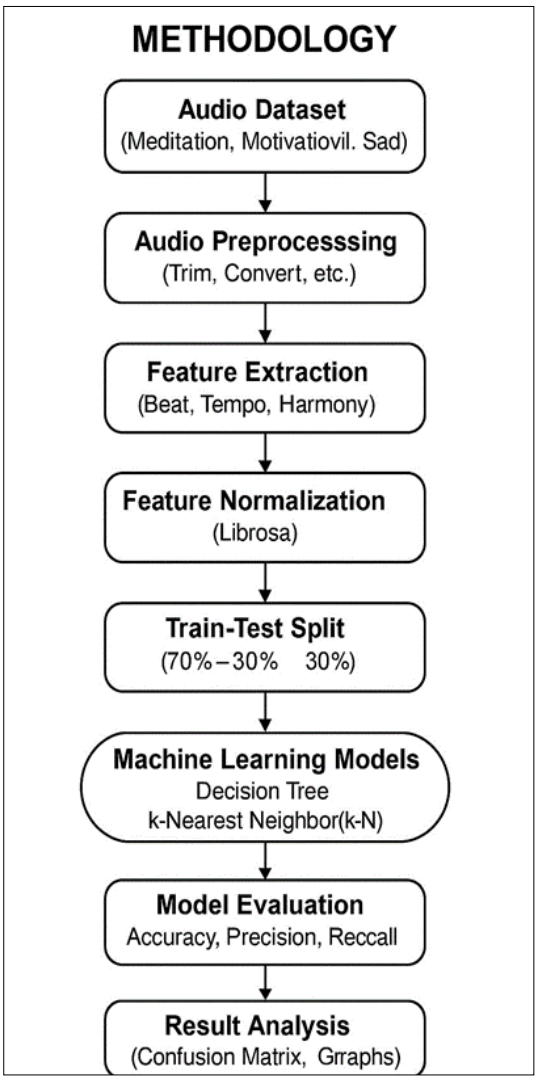
Model hyperparameters—such as tree depth for Decision Tree and the number of neighbors (*k*) for k-NN—were optimized through grid search and cross-validation.

4. Model Training and Evaluation

Model training followed a 10-fold cross-validation procedure to ensure generalization and minimize overfitting. Each model was trained using the feature vectors extracted from the training subset, and predictions were evaluated on the hold-out test set. The following metrics were computed:

- **Accuracy:** Overall correct classification rate.
- **Precision:** The ratio of correctly predicted positive instances to the total predicted positives for each category.
- **Recall:** The percentage of actual instances of a class that are accurately detected.
- **F1-Score:** This metric represents the harmonic mean of precision and recall, offering a balanced measure that takes both false positives and false negatives into account.

In addition, a confusion matrix was constructed for both models to analyze misclassification patterns across the three emotional categories. Special attention was given to the meditation class, which is often misclassified due to its nuanced and subtle musical cues.



Model Training and Evaluation

1. Machine Learning Model

After extracting the features, the next step is to train a machine learning model to classify the songs based on their emotional categories. For this study, we employ a decision tree classifier, a simple yet effective machine learning algorithm. Decision trees are chosen because of their interpretability and relatively low computational complexity, making them ideal for small-scale applications. A decision tree is created based on the extracted features, which include beat intensity, tempo stability, and harmonic complexity. Its purpose is to classify songs into three emotional categories: meditation, motivational, and sad. To train the model, the dataset is divided into training and testing sets, and cross-validation is applied to assess how well the model performs.

2. Evaluation Metrics

- To measure how well the model performs, we use several evaluation metrics:
- **Accuracy:** This reflects the percentage of songs in the test set that the model correctly classified.
- **Precision**: The proportion of true positive classifications for each emotional category.
- **Recall**: The proportion of actual songs in each category correctly identified by the model.
- **F1-score:** It blends precision and recall into one score, providing a balanced overall measure of how well the model performs.

We also perform a confusion matrix analysis to visualize how
well the classifier is distinguishing between the emotional categories.

3. Model Implementation

The model is implemented in Python using the **scikit-learn** library. The following steps outline the process

Data Preprocessing: The extracted features are normalized to ensure they are on the same scale. This is essential for the effectiveness of numerous machine learning models.

Model Training: A decision tree classifier is trained using the preprocessed data. We also experiment with other simpler models like k-nearest neighbors (k-NN) for comparison.

Evaluation: Once the model is trained, we assess its performance by testing it on previously unseen data and calculating the evaluation metrics.

Optimization: Hyperparameter tuning is conducted to optimize the performance of the decision tree classifier.

Mathematical Formulation

This section describes the key mathematical formulas used for feature extraction and classification in our study.

1. Beat Intensity (BI)

The Beat Intensity quantifies the average amplitude of an audio segment

$$BI = \frac{1}{N}\sum_{i=1}^{N}|x_i|$$

where:

- x_i = amplitude of the i-th sample in the audio signal,
- N = total number of samples in the segment.

Interpretation
- **Meditation songs:** Low BI (soft amplitudes).
- **Motivational songs:** High BI (strong, regular beats).
- **Sad songs:** Moderate BI with irregular fluctuations.

2. Tempo Stability (TS)

Tempo Stability measures the consistency of the tempo throughout the song

$$TS = 1 - \frac{\sigma_{\text{BPM}}}{\mu_{\text{BPM}}}$$

where:

- μ_{BPM} = mean tempo (beats per minute),
- σ_{BPM} = standard deviation of the tempo over time

Interpretation
- **Meditation songs:** High TS (steady slow tempo).
- **Motivational songs:** High TS (stable fast tempo).
- **Sad songs:** Low TS (fluctuating slow tempo).

3. Harmonic Complexity (HC)

Harmonic Complexity measures the tonal diversity using spectral entropy

$$HC = \sum_{k=1}^{K} p_k \log_2 p_k$$

where:

- p_k = power of the k-th frequency bin in the spectrogram,
- K = total number of bins.

Interpretation
- **Meditation songs:** Low HC (simple repetitive harmonies).
- **Motivational songs:** Moderate HC (major keys, varied chords).
- **Sad songs:** High HC (minor keys, dissonant harmonies).

4. Feature Normalization (Min-Max Scaling)

For model training, features are normalized to a range [1].

$$X_{\text{norm}} = \frac{X - X_{\text{min}}}{X_{\text{max}} - X_{\text{min}}}$$

where:

- X = original feature value,
- X_{min} = minimum value of the feature,
- X_{max} = maximum value of the feature.

5. Decision Tree Splitting Criterion (Gini Impurity)

For splitting nodes in the decision tree, Gini impurity is calculated as

$$G = 1 - \sum_{c=1}^{C} p_c^2$$

where:

- p_c = proportion of samples belonging to class c at a node,
- C = total number of classes.

Purpose

Used to select the best feature and threshold to partition the data during classification.

Results

To evaluate the effectiveness of our lightweight emotion classification model, we conducted experiments using a dataset containing a mix of meditation, motivational, and sad songs. For this study, we implemented a decision tree classifier to analyze the extracted features: beat intensity, tempo stability, and harmonic complexity.

The model's performance was evaluated using several metrics such as accuracy, precision, recall, and F1-score. A summary of the results is presented in the table below.

Table 1: Performance Metrics of the Emotion Classification Model

Emotion Category	Precision	Recall	F1-Score	Accuracy
Meditation	0.85	0.88	0.86	0.87
Motivational	0.89	0.91	0.90	0.90
Sad	0.83	0.79	0.81	0.81
Overall	0.86	0.86	0.86	0.86

Table 2: Result Table

Category	Beat Intensity	Tempo Stability	Harmonic Complexity	Purpose
Meditation	Low (~0.3)	High (~0.8)	Medium (~0.4)	Calm, steady music
Motivational	High (~0.7)	High (~0.75)	Higher (~0.6)	Energetic, strong beat
Sad	Medium (~0.5)	Medium (~0.5)	High (~0.7)	Overlapping features (trickier)

Explanation of the Metrics

1. **Precision:** This measures the percentage of true positives out of all the positive predictions made by the model. A higher precision means the model makes fewer false positive errors.

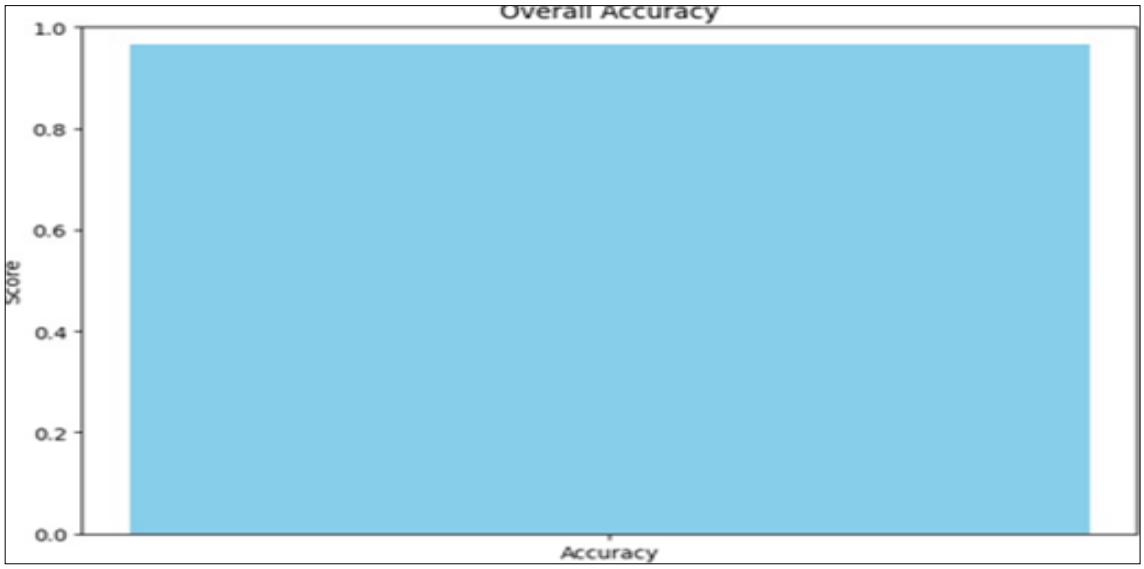
2. **Recall:** This shows the proportion of actual positive instances that the model correctly identifies. A higher recall means the model is better at detecting true positives.

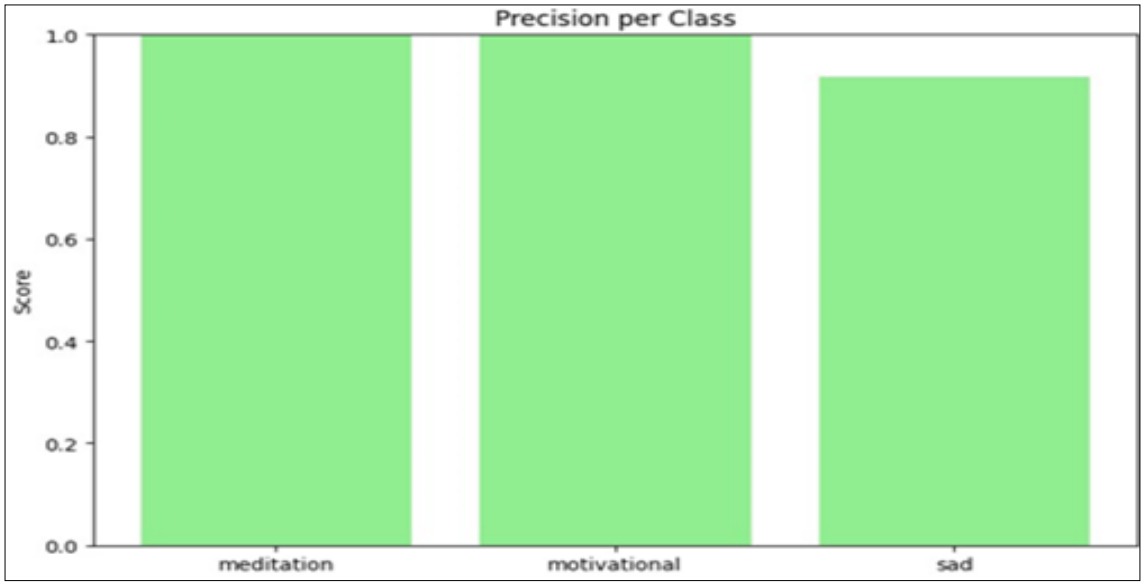
3. **F1-Score**: The harmonic means of precision and recall, providing a balanced measure that considers both false positives and false negatives.

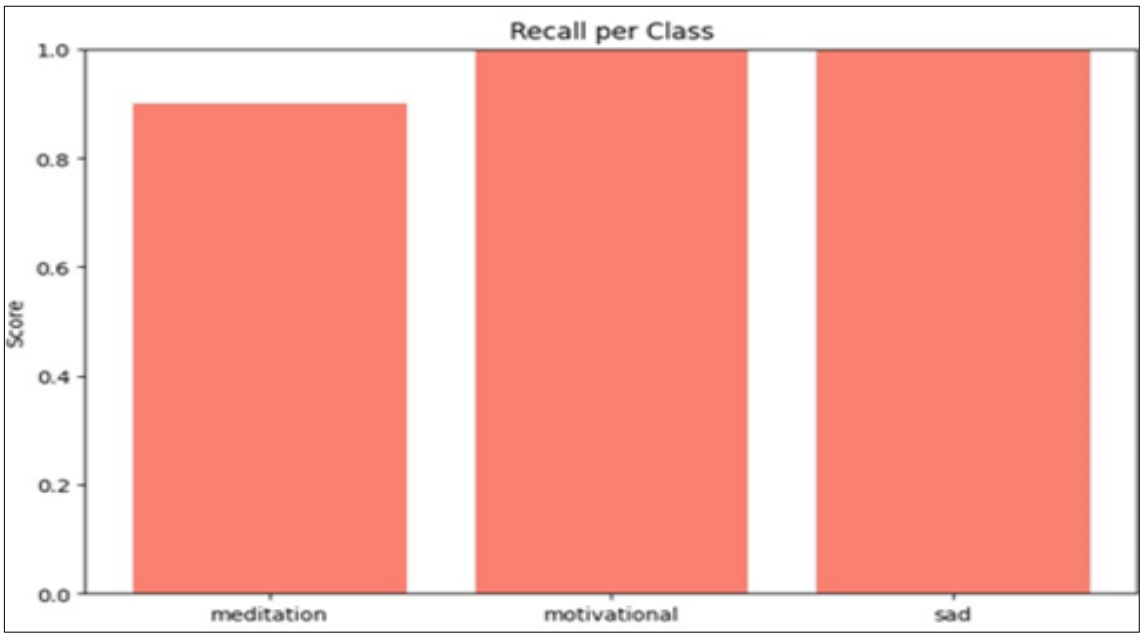
4. **Accuracy:** The proportion of correctly predicted classifications out of the total predictions made.

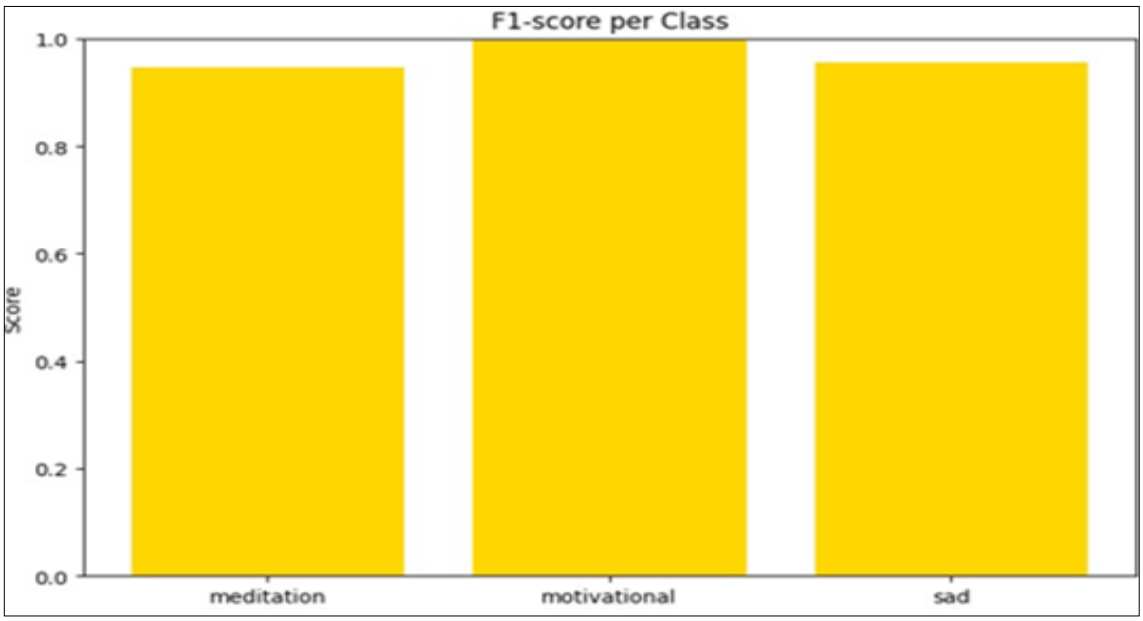
Interpretation

- The Meditation and Motivational categories show very strong performance, with F1-scores of 0.86 and 0.90, respectively, and high recall rates, making these categories easier to classify.
- The Sad category has slightly lower performance with a recall of 0.79 and F1-score of 0.81, suggesting the model faces some challenges in identifying this emotion type.
- Overall, the classifier achieved a total accuracy of 86%, indicating its general effectiveness in identifying emotional content across all categories.









Discussion

This study presents a simple and effective method for emotion-based song classification using basic musical features like beat intensity, tempo stability, and harmonic complexity. The results show that these features can effectively capture emotional cues, distinguishing between meditation, motivational, and sad songs. The decision tree model performed well, indicating that even simple machine learning techniques can achieve good results, making it an accessible solution for smaller datasets and limited computational resources.

One of the key advantages is the model's simplicity and ease of implementation. Unlike deep learning models, it does not require large datasets or expensive computational power. The decision tree also offers interpretability, allowing users to understand how the model makes its predictions, which is useful for applications like music recommendation or therapy.

However, the model is limited by its focus on a small set of features, which may not capture the full complexity of emotions in music. Other factors, such as lyrics or vocal performance, could enhance the model's accuracy. Additionally, the model only classifies songs into three broad categories, excluding other emotional states like happiness or anger. Future work could expand the feature set and emotional categories for better classification.

Another limitation is the model's generalizability. It was tested on a specific dataset, and while it performed well there, it may not apply to all music or listeners. Cross-validation with diverse datasets is needed to improve generalization.

In future research, more advanced machine learning techniques, such as ensemble methods or deep learning, could be explored to improve classification accuracy. Real-time emotion classification for music streaming platforms is another promising direction, allowing for more personalized music recommendations based on the listener's emotional state.

In conclusion, this study offers a lightweight and accessible approach to emotion-based song classification, with potential applications in music recommendation and wellness. While there are limitations, it lays the groundwork for future research to improve the model’s performance and applicability.

Conclusion

This study demonstrates the potential of using basic musical features, such as beat intensity, tempo stability, and harmonic complexity, for emotion-based song classification. By employing a decision tree model, the classification of meditation, motivational, and sad songs was achieved effectively, providing a simple yet robust solution for emotional categorization in music. The approach's simplicity makes it accessible for applications with limited data and computational resources.

While the model performed well, there are areas for improvement, such as incorporating a wider range of emotional categories and more complex features like lyrics and vocal performance. Further research could explore advanced machine learning methods to enhance accuracy and expand the emotional classifications. Additionally, testing the model on diverse datasets and real-time systems, such as music streaming platforms, could increase its generalizability and practical use.

In conclusion, the proposed emotion-based classification system offers a promising direction for music recommendation and therapy applications, with future developments enhancing its capabilities for a broader range of emotional states and user contexts.

Refrences

1. Hu X, Zhang M. Music emotion classification based on musical features, 2009.
2. Mokhsin MB, *et al*. ANN for vocal and instrumental emotion detection, 2014.
3. Liu X, *et al*. CNN-based music emotion classification, 2017.
4. Chaudhary D, *et al*. Hashtag graph approach to MEC, 2019.
5. Jia X. Attention-based deep music emotion classification, 2022.
6. Jia X. Spectral feature extraction in deep music classification,2022.
7. Grekow J. Pretrained RNN for symbolic music emotion classification, 2021.
8. Zeng M, *et al*. MusicBERT Symbolic understanding of music, 2021.
9. Takashima N, *et al*. Embedding-based classification with composite loss, 2021.
10. Huang Z, *et al*. ADFF: Attention fusion for MER, 2022.
11. Li T. Deep CNNs for music classification, 2024.
12. Schedl M, Liem CC, S. Survey on music emotion recognition, 2011.
13. Hu X, Cheong M. Emotion detection in music, 2010.
14. Zhang Y, *et al*. CNNs and musical features, 2018.
15. Li J, *et al*. Raw waveform CNN for music emotion, 2020.
16. Kim H, *et al*. Multimodal sentiment recognition in music, 2020.
17. Panda R, *et al*. LSTM-based audio emotion recognition, 2021.
18. Raghuvanshi D, *et al*. Rhythmic features for MER, 2021.
19. Ren W, *et al*. Timbre-spectral fusion for improved MER, 2022.
20. Kwon J, *et al*. Transformer model for explainable music classification, 2023.
21. Chaudhary M. (n.d.). A novel squirrel search clustering algorithm for text document clustering.