Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Investigating the Relationship Between Music and Human Psychology Through Big Data Analytics

Lin Wu

linwu875@outlook.com

Wuxi Taihu University

Research Article

Keywords: Music Emotion Recognition, Valence–Arousal Model, Acoustic Feature Extraction, Big Data Analytics, Music Information Retrieval, Psychological Emotion Modeling

Posted Date: November 12th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-10198115/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

Investigating the Relationship Between Music and Human Psychology Through Big Data Analytics

Lin Wu *

School of Art, Wuxi Taihu University, Wuxi, 214000, China

Author Mail: linwu875@outlook.com

Abstract

Human emotional responses to music are commonly described using the valence–arousal model, which represents emotions according to their positivity (valence) and intensity (arousal). Predicting these dimensions from musical features enables researchers to better understand the relationship between music and human psychology. This study proposes a scalable, data-driven framework that integrates acoustic feature extraction, statistical feature engineering, exploratory data analysis (EDA), and supervised machine learning to predict valence and arousal from music. The Database for Emotional Analysis in Music (DEAM) dataset is used because it contains listener-annotated valence and arousal ratings that reflect perceived emotional responses to musical stimuli. A comprehensive feature set is extracted, including Mel-Frequency Cepstral Coefficients (MFCCs), spectral features, and energy-based features. A LightGBM regression model is employed to capture nonlinear relationships between audio characteristics and emotional dimensions. Experimental results demonstrate strong predictive performance, achieving Root Mean Squared Error (RMSE) values of 0.1492 for valence and 0.1416 for arousal, with Mean Absolute Error (MAE) values of 0.1108 and 0.1085, respectively. The findings demonstrate the effectiveness of the proposed framework for modeling music-induced emotional responses and highlight its potential applications in personalized music recommendation, emotion-aware multimedia systems, and music therapy.

Keywords: Music Emotion Recognition, Valence–Arousal Model, Acoustic Feature Extraction, Big Data Analytics, Music Information Retrieval, Psychological Emotion Modeling

1. Introduction

People use music as their universal medium to demonstrate their feelings, which results in major impacts on their emotional states, cognitive abilities, and social relationships Wu, Q. (2025). The system shows multiple effects because it can improve mood, boost memory, and decrease stress Lin, Y.-J., et al. (2026).and affect how users make choices. The study

of music's impact on human psychology needs both small experiments and subjective assessments because these methods restrict researchers from establishing general conclusions Castiblanco Jimenez, I. Aet al (2024). Digital music platforms have expanded their reach, which, combined with available research on listener behavior, enables researchers to investigate these relationships through large-scale studies Yu, X. (2025). Big data analytics provides an approach to capture patterns and correlations between musical features and human psychological responses across diverse populations Liu, G., Hu, J., & Kostikova, I. (2025).

The psychological impact of music arises from a combination of acoustic, lyrical, and contextual factors Gu, Y., Shao, C., Li, J., & Fan, Y. (2026). The musical component of a song includes tempo and rhythm, and key and mode, which together create emotional responses that people experience Hasegawa, R. P. (2026). The musical element of a song includes its lyrics, which deliver emotional content that people experience through their personal feelings and life situations Clark, I. N. (2026). Cultural background, listening habits, and individual personality characteristics create different ways that people experience music Yang, Z., et al. (2025). The combination of multiple factors creates difficulties in predicting psychological results through standard analysis techniques, which become especially problematic when dealing with extensive and diverse musical information Han, Y. (2025).

Previous research used classical statistical analysis, machine learning, and signal processing techniques to study music-emotion relationships Tran, A. (2024). The research methods use regression models to predict emotional states from acoustic features, clustering to create mood-based music categories, and sentiment analysis to assess lyrical impact Visutsak, P., et al. (2025). The methods deliver useful findings, but their effectiveness suffers because they depend on restricted data sets and insufficient features, and their inability to expand Richard, B. S., et al (2025). The current models do not link different types of data, which include audio features, lyrics, and listener behavior, so their capacity to model complex music-psychology relationships remains restricted.

To address these limitations, this study proposes a big data-driven framework for modeling the relationship between music and human psychology. The proposed approach uses the DEAM dataset to extract acoustic features and perform lyrical

sentiment analysis to predict valence and arousal levels across music tracks. Human psychology consists of several dimensions, which include cognitive, memory-based, and emotional mood and perception aspects. The emotional response of humans shows the fastest psychological reaction, which researchers can measure directly after they listen to music. Researchers in music psychology studies use the valence-arousal dimensional framework to create models that show how people experience emotional states. Humans create emotional reactions, which researchers can study through computational methods to investigate how music affects their psychological connection with the music. The main contributions of the paper are listed below,

- Introduces a regression-based framework for predicting continuous emotional dimensions in music by modeling the relationship between acoustic features and emotional attributes such as valence and arousal.
- Integrates an efficient gradient boosting learning strategy using LightGBM to capture complex nonlinear interactions between extracted musical features and emotional annotations.
- Enhances the effectiveness of music emotion recognition by providing a computationally efficient learning framework capable of modeling complex emotional patterns in musical audio signals.

The remaining part of the paper is organized as follows. Section 2 reviews the related work on music emotion recognition. The proposed methodology, together with the LightGBM regression framework, forms the content of Section 3. The setup for experiments and methods to evaluate performance are detailed in Section 4. The results obtained from the research are presented in Section 5, while Section 6 presents the study conclusion together with future research directions.

2. Literature review

Music emotion recognition (MER) has gained significant attention in recent years due to advancements in deep learning, affective computing, and multimodal data integration. Contemporary research mainly investigates four areas, which include dimensional emotion modeling and multimodal fusion strategies, explainable artificial intelligence, and intelligent intervention systems.

2.1 Deep Learning in Music Emotion Analysis

Researchers have used deep learning models to improve their ability to recognize emotions through musical signals. Qiang and Zhibin (2024). investigated psychological factors influencing music learning through deep learning technologies, demonstrating that neural-based models can effectively capture latent emotional representations in educational settings. Chen and Sun (2024). developed a music education system that uses deep learning frameworks to provide students with better personalized learning pathways through their artificial intelligence integration.

Zhang et al. (2024). developed deep learning-based emotion recognition algorithms for music performance analysis, which they designed to extract features from evolving audio signals. Ahmed et al. (2024). used advanced audio analysis with deep neural networks to classify musical genres, which indirectly helped them learn features related to emotions because genre characteristics usually match emotional dimensions. Mia et al. (2024). have conducted research on lyric-based emotion analysis through machine learning and neural network classification, which developed new methods of emotion modeling that used semantic representations together with acoustic features. The combined efforts of these studies show that deep learning methods improve the way emotional features are presented in music applications.

2.2 Multimodal and Physiological Emotion Recognition

Multimodal integration has become the primary research focus, enabling researchers to develop more comprehensive methods for detecting emotional expressions. Wang et al. (2024). introduced a multilayered multimodal interaction model (MMD-MII) that integrates heterogeneous inputs for music emotion classification, which enhances system stability through cross-modal learning. The field of research has seen an increase in the usage of methods based on physiological signals. Qiao et al. (2024). developed a temporal convolutional attention networkthat uses EEG signals to recognize music emotions, which demonstrates that neural activity delivers additional information beyond audio features.

Han et al. (2024). created a multi-scale convolution and attention-based model that uses physiological data to recognize emotions and make recommendations through improved valence-arousal prediction accuracy. Bakariya et al. (2024). used CNN-based facial emotion recognition to develop music recommendation systems, which showcased the advantages of using multiple

emotional expression methods. The studies show that multimodal systems improve prediction results because they can detect emotional signals from both physiological sources and audio-visual sources.

2.3 Optimization-Driven Deep Learning Models

Deep neural networks for music emotion recognition tasks have been improved through the implementation of optimization techniques. Li et al. (2024). developed an enhanced differential evolution-based CNN system to analyze musical emotional content, which proved that evolutionary optimization methods increased both convergence speed and predictive accuracy. The findings demonstrate that proper hyperparameter adjustment, together with model optimization processes are essential for achieving trustworthy results in emotional regression assessments.

2.4 Explainable and Theory-Driven Emotion Models

The field of emotion recognition research requires more interpretable systems to meet its growing demand. Fong et al. (2025). introduced a theory-based, explainable deep learning architecture for music emotion analysis, which combines psychological principles with neural network systems to achieve better system understanding. The research results show that researchers should develop computational models that operate according to established affective theories that include the valence-arousal framework.

2.5 AI-Driven Emotion Regulation and Intervention Systems

Recent research has developed beyond its current recognition because it now studies both emotion regulation and therapeutic uses. Shen et al. (2025). proposed an AI-generated content (AIGC)-based closed-loop music intervention system for mental disorder treatment. Their system integrates emotion detection with adaptive music generation to demonstrate how predictive emotion models can improve clinical and well-being applications.

2.6 Research Gap and Motivation

However, it has been found that although significant progress has been achieved in the application of deep learning to music emotion recognition tasks, some information gaps still exist:

i. The research has primarily concentrated on classification rather than regression (2022).

ii. The application has mostly been restricted to multimodal features, especially physiological features (2025).

iii. The optimization techniques have been tailored to specific models but have not been generally verified for regression (2022).

iv. The issue of interpretability has been partially addressed in predictive modeling (2026).

Therefore, there is a need for a scalable, regression-based emotional modeling framework that:

- Accurately predict emotional responses in terms of valence and arousal,
- Integrates robust feature analysis,
- Employs efficient machine learning models,
- And provides interpretable feature–emotion relationships.

This has led to the development of a modeling framework for predicting continuous emotional responses using advanced techniques in supervised learning.

2.7. Problem statement

The process of understanding how music evokes emotions is not an easy one. The emotions evoked by music are not straightforward or unidimensional; they result from the complex interplay of sound-related features, words, time, and psychology (2024). Although recent advances in deep learning have improved the accuracy of emotional classification, most of the systems have not addressed the continuous range of emotions from the valence-arousal diagram. Furthermore, the use of multiple physiological measures in most of the systems makes it difficult for the systems to scale (2024).

The existing techniques face issues related to suboptimal optimization, poor interpretability, and poor regression-based validation. Most of these techniques are not able to effectively handle the process of how emotions transform in a musical piece of art and do not always evaluate their predictions using error-based metrics such as MAE. The development of an effective system requires a strong computational method which delivers optimized performance and scalability. The system must achieve accurate emotional prediction through its music data analysis capabilities and deliver better model results.

3. Proposed methodology

The proposed method provides a framework for relating music to human psychological responses by predicting continuous levels of valence and arousal. The method uses regression analysis to measure emotional attributes through its assessment of various musical elements instead of using traditional classification techniques. The process can be described as

a series of steps, which include data set collection and data alignment, followed by feature extraction and engineering, and then data exploration through dimensionality reduction, and finally regression model construction, which uses optimized algorithms and evaluation. The framework develops a computational system that achieves better efficiency and effectiveness through the integration of different musical attributes which include acoustic elements and rhythmic components and timbral properties and statistical characteristics and distributed big data processing methods. The system provides a music-based model that connects human psychological responses to music and enables organizations to conduct music analytics at a large scale. The framework uses valence and arousal as emotional dimensions which serve as measurable indicators of how humans react to music. The proposed framework demonstrates its process through Figure 1 which shows the sequence from audio preprocessing to feature extraction and processing and LightGBM model training and emotion prediction and performance evaluation.

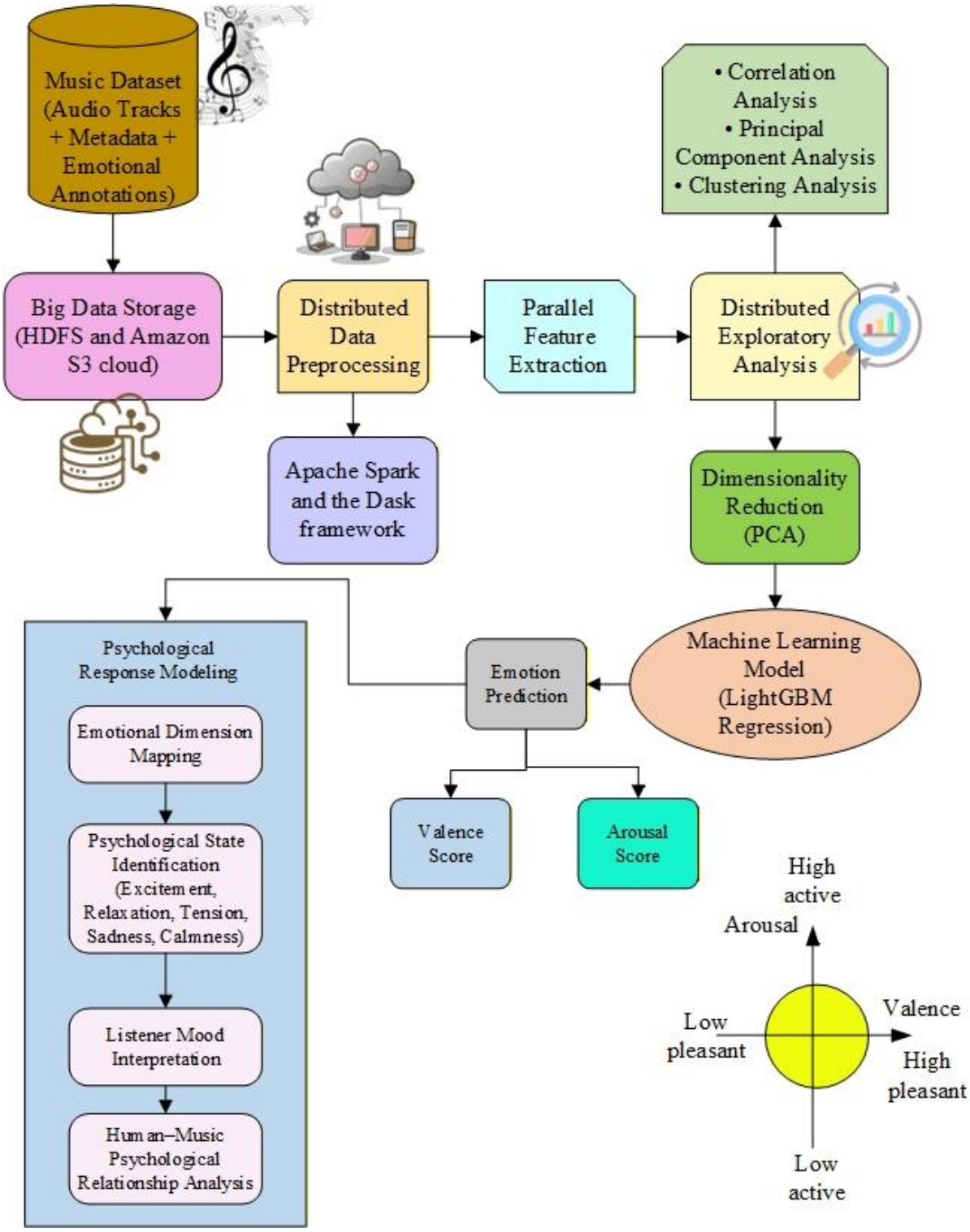


Figure 1: Overall Workflow of the Proposed Music Emotion Prediction Framework

3.1 Dataset Acquisition

The first part of the proposed approach involves obtaining the DEAM dataset [32], which serves as the foundation for studying the relationship between music and human psychology. The DEAM dataset contains listener evaluations of music, which show how listeners experienced

different emotional states through their musical experience. The DEAM dataset contains audio files of various music tracks, pre-computed feature vectors of the music, and continuous annotations of valence and arousal on a per-second basis. The dataset also includes metadata, including the genre, song ID, duration, and tags, which provide more information about the song and its characteristics. The annotations in the dataset are crucial since they will eventually lead to the quantification of the psychological effect.

The present research uses the processed dataset to connect audio files with their corresponding feature vectors and labels through a specific song identification system. The research will expand its capacity to assess how music affects psychological responses by incorporating additional large-scale datasets, which include music streaming data, user interaction information, and social listening pattern data. The research needs to combine multiple datasets because emotional responses to music require multiple dimensions to create accurate models of understanding human emotions.

The process requires accurate data collection and seamless data source integration to establish a scalable system that enables subsequent feature extraction, modeling, and psychological analysis. The data accuracy and completeness requirements exist to create a scalable analysis process together with a predictive model that calculates valence and arousal, which represent the core of music-psychology research.

3.2 Data Inspection, Cleaning, and Storage

The first step after dataset acquisition requires inspection, cleaning, and organization of data to verify its quality, consistency, and analytical scalability. The DEAM dataset contains multiple components which include audio clips feature vectors and per-second valence and arousal annotations despite its comprehensive nature. The assessment process demands thorough verification of all components to establish their complete status and proper functioning. The absence of any musical features or emotion annotations creates gaps which lead to biased results and decreased predictive model accuracy. The process of data cleaning requires the identification of missing entries which are either corrected interpolated or deleted according to established protocols.

The current phase of research requires scientists to establish connections between audio elements and their associated emotional designation. The original annotations may be sampled at

irregular intervals or at different rates compared to extracted audio features. The researchers use valence and arousal scores to create unified data, which they resample at 2 Hz, allowing each audio frame to be linked to its corresponding emotional label. The alignment process establishes the necessary connection that enables researchers to study how music affects emotional reactions throughout different points in time. The cleaned data for studies that examine scalability and big data research purposes gets stored in distributed storage systems which include HDFS and Amazon S3 cloud-based solutions. The system enables multiple users to simultaneously access extensive track collections, which speeds up feature extraction and model training processes while handling streaming logs and user behavior data expansion.

The data processing step conducts complete data inspection, data cleaning, and data storage across multiple locations, which creates high-quality data that maintains uniformity and expands in size for future processing and modeling tasks. The research process establishes a foundation that enables researchers to identify significant features while developing models that produce precise and understandable results, which connect music to human psychological reactions. The data pre-processing stages are shown in Figure 2.

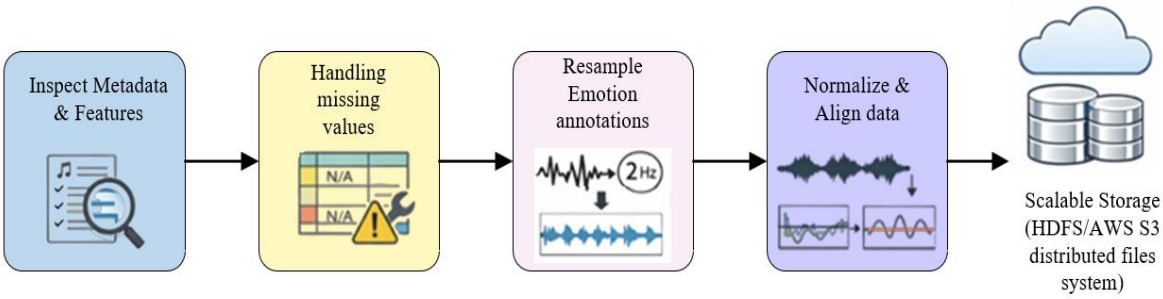


Figure 2: Data cleaning and preprocessing

3.3 Feature Extraction and Engineering

Musical elements, including timbre, rhythm, and spectral composition, create major effects on human psychological responses to music. Acoustic features such as timbre, spectral characteristics and rhythmic properties, which people use to perceive music, create various psychological effects that listeners experience when they hear musical stimuli. The extraction of particular audio descriptors allows researchers to establish connections between musical features and the resulting emotional reactions. The research team extracts multiple acoustic

features that serve as a basis for demonstrating the music's perceptual attributes.

The process starts with dataset cleaning and alignment before proceeding to feature extraction and engineering, which transforms raw audio and metadata into structured multi-dimensional data that researchers use to model human emotional responses. High-quality features are essential for achieving accurate predictions of valence and arousal because these two emotional state measurements exist as continuous variables. Acoustic Feature Extraction starts by calculating MFCCs from every audio frame $x[n]$. The calculation of MFCCs is given in Eq. (1), which serves to provide musicians with a method for recording both the timbral and spectral elements of music:

$$\text{MFCC}_k = \sum_{m=1}^{M} \log\left(\sum_{n=0}^{N-1} x[n] \cdot w_m[n] \cdot e^{-j2\pi kn/N}\right) \cdot \cos\left(\frac{\pi k}{M}(m-0.5)\right) \quad (1)$$

Where $w_m[n]$ represents the Mel-filter bank, N is the FFT length, and M is the number of Mel filters. The audio spectrum brightness and spectral distribution are represented through the extraction of spectral features, which include spectral centroid and spectral bandwidth, together with the MFCCs. The centroid and bandwidth are calculated using Eq. (2) and Eq. (3) respectively.

$$\text{Centroid} = \frac{\sum_{f=0}^{F-1} f \cdot |X(f)|}{\sum_{f=0}^{F-1} |X(f)|} \quad (2)$$

$$\text{Bandwidth} = \sqrt{\frac{\sum_{f=0}^{F-1} |X(f)|(f - \text{Centroid})^2}{\sum_{f=0}^{F-1} |X(f)|}} \quad (3)$$

Where $X(f)$ is the Fourier transform magnitude at the frequency bin f, and F is the total number of frequency bins. Chroma features are extracted to analyze harmonic content, which shows how energy is distributed across twelve musical pitch classes. The entire track's temporal dynamics are summarized through Statistical Feature Aggregation, which provides a statistical summary of time-based measurements. For each feature f_t within the time frame t, the mean μ_f and variance σ_f^2 are computed using Eq. (4) and Eq. (5):

$$\mu_f = \frac{1}{T}\sum_{t=1}^{T} f_t \quad (4)$$

$$\sigma_f^2 = \frac{1}{T}\sum_{t=1}^{T} \left(f_t - \mu_f\right)^2 \quad (5)$$

Where T is the total number of frames in the track. The aggregated statistics show both temporary changes and permanent musical feature changes throughout time. The inclusion of higher-order moments, which include skewness and kurtosis, allows for better representation of dynamic systems. The researchers use Derived and Higher-Level Features to measure specific musical elements in their work. Rhythm Complexity (RC) measures the standard deviation of Inter-Onset Intervals (IOIs) that occur between beats. The RC is calculated using Eq. (6).

$$RC = \sqrt{\frac{1}{N-1}\sum_{i=1}^{N}\left(IOI_i - \overline{IOI}\right)^2} \quad (6)$$

Where IOI_i is the inter-onset interval of beat i and N is the number of beats. Timbre diversity (TD) is calculated using Eq. (7) as the variance across MFCC coefficients:

$$TD = \frac{1}{K}\sum_{k=1}^{K}\sigma^2_{\mathrm{MFCC}_k} (7)$$

The number of MFCC coefficients is represented by the variable K. The system uses NLP techniques to analyze lyrics and produce sentiment scores S_{lyric}, which measures the emotional content of the text. The process of multi-modal integration unifies all track features into one vector X, which represents each track as given in Eq. (8).

$$\mathrm{X} = \left[\ \text{MFCCs, Spectral, Chroma, Tempo, Rhythm, Timbre, } S_{\text{lyric}}\text{, Metadata}\ \right] \quad (8)$$

This complete feature vector includes all acoustic, rhythmic, harmonic, semantic, and contextual elements that exist in music. The system uses this data to create predictive models that analyze valence and arousal through music-based psychological assessments that can be performed at scale and understood by users. The current stage extracts multi-modal features, which it combines to provide models with full musical emotional impact data, which will support big data research and psychological analysis.

Algorithm 1: Feature Extraction and Engineering

Input: Audio dataset $D = \{T_1, T_2, \ldots, T_n\}$
Output: Feature matrix F

1. Initialize feature matrix $F \leftarrow \emptyset$
2. for each music track $T_i \in D$ do
3. Load audio signal A_i

4. Segment A_i into short-time frames
5. Compute MFCC features using Eq.
6. Compute spectral centroid using Eq. (2)
7. Compute spectral bandwidth using Eq. (3)
8. Extract chroma feature representation
9. Compute statistical descriptors of features
10. Mean using Eq. (4)
11. Variance using Eq. (5)
12. Compute rhythm complexity using Eq. (6)
13. Compute timbre diversity using Eq. (7)
14. Perform lyrical sentiment analysis to obtain sentiment score S_i
15. Construct feature vector X_i using Eq. (8)
16. Append X_i to feature matrix F
17. end for
18. Return feature matrix F

3.4 EDA with Mathematical Formulation

After constructing multi-dimensional feature vectors, the next stage is EDA. The current stage focuses on examining musical feature statistical properties and their connection to psychological responses while reducing dimensionality and discovering patterns through clustering. EDA is crucial for accurate modeling of valence and arousal.

a) Distribution Analysis

To understand the central tendency of the emotional responses, the mean valence (μ_V) is computed using Eq. (9):

$$\mu_V = \frac{1}{N}\sum_{i=1}^{N} V_i \quad (9)$$

It calculates the average valence score across all the N music samples, which give a baseline for positive or negative emotional trends. Similarly, the mean arousal (μ_A) is computed with Eq. (10):

$$\mu_A = \frac{1}{N}\sum_{i=1}^{N} A_i \quad (10)$$

The average arousal level measures the emotional response intensity that people show in the dataset. The measurement of emotional score distribution, which includes valence and arousal, is determined

through the equations found in Eq. (11) and Eq. (12):

$$\sigma_V^2 = \frac{1}{N}\sum_{i=1}^{N} (V_i - \mu_V)^2 \quad (11)$$

$$\sigma_A^2 = \frac{1}{N}\sum_{i=1}^{N} (A_i - \mu_A)^2 \quad (12)$$

The current study uses valence and arousal scores together with their mean values to determine which tracks show extreme emotional reactions. The researchers used higher-order statistical methods together with skewness and kurtosis measurements to evaluate the distribution patterns of their data:

$$S_V = \frac{1}{N}\sum_{i=1}^{N} \frac{(V_i-\mu_V)^3}{\sigma_V^3} \quad (13)$$

$$K_V = \frac{1}{N}\sum_{i=1}^{N} \frac{(V_i-\mu_V)^4}{\sigma_V^4} \quad (14)$$

Eq. (13) measures the asymmetry of valence distribution, while Eq. (14) measures how peaked the distribution is compared to a normal distribution, highlighting unusual emotional responses.

b) Correlation Analysis

The study calculates Pearson correlation coefficients to measure the linear connection between musical features X_j and emotional responses. The feature X_j from Eq. (15) establishes a correlation with valence V according to the equation.

$$r_{X_j,V} = \frac{\sum_{i=1}^{N}\left(X_{i,j}-\mu_{X_j}\right)(V_i-\mu_V)}{\sqrt{\sum_{i=1}^{N}\left(X_{i,j}-\mu_{X_j}\right)^2}\sqrt{\sum_{i=1}^{N}(V_i-\mu_V)^2}} \quad (15)$$

This indicates how strongly a musical feature (e.g., tempo, spectral centroid) is linearly associated with valence; values close to 1 or -1 show strong positive or negative influence. Similarly, Eq. (16) calculates the correlation with arousal:

$$r_{X_j,A} = \frac{\sum_{i=1}^{N}\left(X_{i,j}-\mu_{X_j}\right)(A_i-\mu_A)}{\sqrt{\sum_{i=1}^{N}\left(X_{i,j}-\mu_{X_j}\right)^2}\sqrt{\sum_{i=1}^{N}(A_i-\mu_A)^2}} \quad (16)$$

It represents the linear relationship between the feature X_j and arousal A, helping identify which musical properties elicit high or low arousal.

c) Dimensionality Reduction

To handle high-dimensional feature spaces, Principal Component Analysis (PCA) reduces the dimensionality while

preserving variance. The covariance matrix is calculated with Eq. (17):

$$\Sigma_X = \frac{1}{N-1}(X - \mu_X)^T(X - \mu_X) \quad (17)$$

It measures how features vary together, providing the basis for identifying principal directions in the data. The principal components are then obtained via projection in Eq. (18):

$$Z = XW, W = [w_1, w_2, \dots, w_k] \quad (18)$$

Eq. (18) transforms the complete high-dimensional dataset X into the smaller space Z, which uses the eigenvectors stored in W to show maximum variance directions for visualization purposes and to decrease computation requirements. The t-SNE and UMAP methods enable the detection of clusters and relationships that linear PCA fails to discover by maintaining local structures of non-linear data.

d) Clustering Analysis

The application of k-means clustering enables the identification of natural track groupings that share similar musical-emotional characteristics, which is applied using Eq. (19):

$$J = \sum_{k=1}^{K} \sum_{x_i \in C_k} \|x_i - \mu_k\|^2 \quad (19)$$

It represents the objective function minimized during clustering, where C_k is cluster k, μ_k is the centroid, and K is the total number of clusters. The system discovers patterns that show how different musical genres elicit specific emotional responses and how certain feature combinations produce equivalent levels of valence and arousal.

e) Temporal and Dynamic Analysis

For tracks with time-varying emotion scores, Eq. (20) computes the temporal correlation between a feature $X_j(t)$ and valence $V(t)$:

$$r_{X_j(t),V(t)} = \frac{\sum_{t=1}^{T}\left(X_j(t)-\mu_{X_j}\right)(V(t)-\mu_V)}{\sqrt{\sum_{t=1}^{T}\left(X_j(t)-\mu_{X_j}\right)^2}\sqrt{\sum_{t=1}^{T}(V(t)-\mu_V)^2}} \quad (20)$$

The dynamic effects demonstrate how changes in tempo, loudness, and intensity affect valence over time. The same calculation method used for valence can also be applied to measure arousal. The proposed framework uses Big Data

Integration to maintain scalable EDA operations, which can handle EDA of thousands of music tracks that contain high-dimensional acoustic and psychological data. The DEAM dataset produces frame-level descriptors and statistical summaries and time-varying valence-arousal annotations for each track, resulting in extensive feature matrices that require significant memory and processing capabilities to handle.

To solve this problem, all fundamental EDA functions, which include distribution analysis and correlation calculation, PCA, and clustering operations, execute their processes through distributed big data systems that use Apache Spark and the Dask framework. The platforms use dataset partitioning to distribute data across different processing nodes, which allows multiple nodes to compute in parallel while they combine results from intermediate processing stages. The system calculates statistical metrics with high efficiency because it uses multiple systems for computations, while users can conduct dimensionality reduction on extensive feature sets, and clustering methods can handle numerous data points. The EDA pipeline achieves reproducibility and extensibility while managing extensive real-world music-psychology analysis through its distributed data structures that enable seamless integration of different input modes, which include acoustic features, temporal emotion signals, and metadata.

3.5.Supervised Learning for Valence/Arousal Prediction

The integrated feature dataset is divided into two segments, which use 80% for training purposes and 20% for testing purposes. The training set is used to learn model parameters, the validation set supports hyperparameter tuning and overfitting control, and the testing set evaluates generalization performance on unseen music samples. This split establishes an approach that guarantees both statistical validity and the ability to repeat predictive analysis results. The layers present in the LightGBM model are shown in Figure 3.

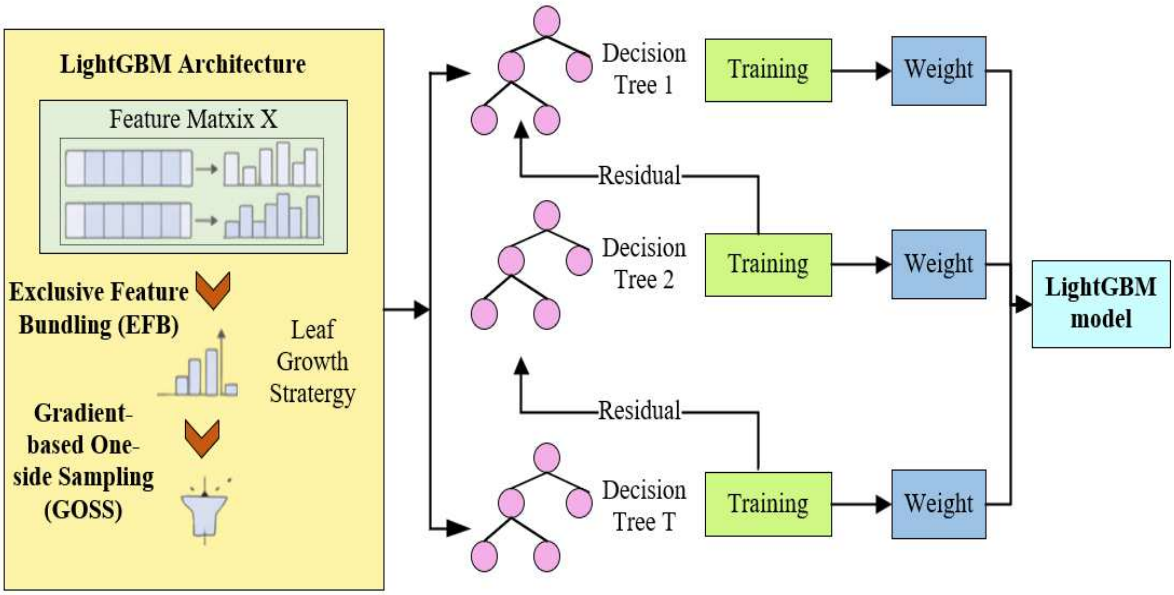


Figure 3: LightGBM architecture

For dynamic emotional prediction, the gradient boosting framework LightGBM is employed. LightGBM creates an ensemble through sequential tree construction which uses each new tree to rectify the mistakes made by the earlier trees. The model uses this boosting mechanism to successfully detect nonlinear relationships between acoustic features and psychological responses. The mathematical expression for predicting valence and arousal states is presented through the following equation:

$$\hat{y}_i = \sum_{m=1}^{M} f_m(x_i) \tag{21}$$

The emotional value prediction for the i^{th} track is calculated through the formula in Eq. (21), which defines $\hat{y}_i$ as the combined result of M weak learners whose output functions are represented by $f_m(x_i)$. The ensemble method produces dependable results because it combines multiple models, which improves the accuracy of regression analysis. The hyperparameters of the LightBGM model are shown in Table 1.

Table 1: Hyperparameters of LightGBM Model

Parameter	Value	Description / Rationale
objective	regression	Task type — continuous output regression
metric	rmse	Evaluation metric during training

learning_rate	0.02	Step size shrinkage to prevent overfitting
num_leaves	31	Max leaves per tree (controls model complexity)
max_depth	6	Maximum tree depth
min_child_samples	20	Minimum samples in a leaf node
feature_fraction	0.6	Fraction of features used per tree (bagging)
bagging_fraction	0.7	Fraction of data used per boosting round
bagging_freq	5	Frequency of bagging (every k round)
lambda_l1	1.0	L1 regularization (Lasso)
lambda_l2	5.0	L2 regularization (Ridge) — stronger regularization
min_gain_to_split	0.01	Minimum gain to perform a split
n_estimators	1000	Fixed number of boosting rounds (no early stopping)
random_state	42	Seed for reproducibility
verbose	-1	Suppress LightGBM console output

The pseudocode for the LightGBM model is given in Algorithm 2.

Algorithm 2: Light Gradient Boosting Machine

Input:

Training dataset $D = \{(x_i, y_i)\}_{i=1}^{N}$

Number of boosting rounds T

Output:

Prediction model $F(x)$

Convert continuous features into discrete histogram bins.

Apply Exclusive Feature Bundling (EFB) to reduce the number of features.

Initialize prediction model

$\mathrm{F0(x)} =$ constant value minimizing loss.

For $\mathrm{t} = 1$ to T do

Compute the gradient g_i for each training sample.

Apply Gradient-based One-Side Sampling (GOSS) to keep instances with large gradients.

Build a decision tree using a leaf-wise growth strategy based on maximum information gain.

Update model

$$Ft(x) = Ft - 1(x) + \eta ht(x)$$

End For

Output final boosted model F(x)

3.5.1. Regression Tasks

The task needs to be solved through continuous regression because valence and arousal function as continuous emotional dimensions. The model uses regression loss functions to achieve optimization through Mean Squared Error (MSE) as its selected method:

$$\mathcal{L} = \frac{1}{N}\sum_{i=1}^{N} (y_i - \hat{y}_i)^2 \quad (22)$$

Eq. (22) defines the average squared difference between the actual emotional value y_i and predicted value $\hat{y}_i$. The predicted valence and arousal values will match actual psychological responses when we minimize this loss function. The model develops direct connections between musical features and emotional outcomes through the regression-based supervised learning setup. The stage produces a predictive system that predicts valence and arousal values for every music track or segment while it models emotional responses through a scalable machine learning system.

4. Results and Discussion

This section presents the experimental results obtained from the proposed music emotion prediction framework and analyzes its performance in estimating the emotional dimensions of music. The testing procedures took place on the Python platform, which provided all the essential machine learning libraries required to conduct the tests. The dataset was divided into 80% training data and 20% testing data to evaluate the generalization capability of the model. The performance assessment of the regression model uses LightGBM as its primary model, while results are evaluated through comparison with other existing techniques.

4.1.Experimental setup

The musical emotion prediction system underwent evaluation through experiments that used the Python programming environment together with multiple scientific computing and machine learning libraries to test its performance. The regression model implementation required LightGBM as its foundation, while the data processing, analysis, and visualization tasks

were performed through the use of NumPy, Pandas, Matplotlib, Seaborn, SciPy, and Scikit-learn libraries. The dataset was divided into 80% training data and 20% testing data to evaluate the model's generalization capability. The experiments used a computer system that featured a 12th-generation Intel Core i5 processor and 16 GB RAM, while the Python environment provided T4 GPU runtime support for the implementation. The experimental results show hardware specifications in Table 2 and software specifications in Table 3.

Table 2: Hardware setup

Hardware Specifications	
Device Name	DESKTOP-RHF1E9H
Processor	12th Gen Intel® Core™ i5-12400 (2.50 GHz, 6 cores / 12 threads)
Installed RAM	16.0 GB (15.7 GB usable)
System Type	64-bit operating system, x64-based processor
Platform	Google Colab (T4 GPU runtime environment)
Pen / Touch	No pen or touch input

Table 3: Software setup

Software & Library Versions	
Python	3.12.12 (main, Oct 10 2025) [GCC 11.4.0]
NumPy	2.0.2
Pandas	2.2.2
Matplotlib	3.10.0

Seaborn	0.13.2
SciPy	1.16.3
Scikit-learn	1.6.1
LightGBM	4.6.0

4.2.Dataset Understanding & Statistical Exploration

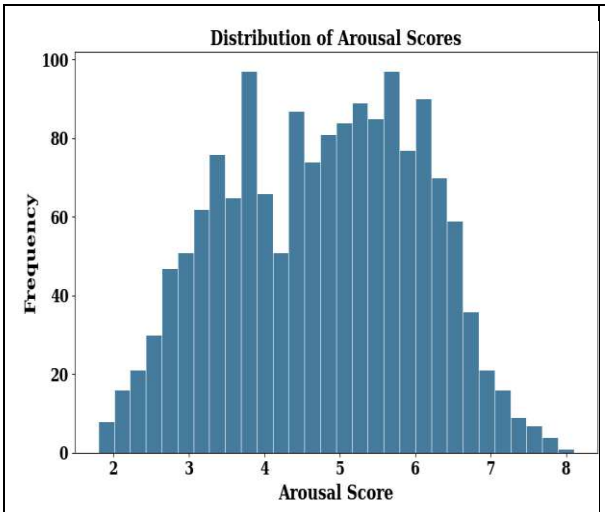
This section shows the complete overview together with a statistical analysis of the dataset that the study uses. The report provides a summary of the dataset source, together with its sample size and the methods used for data preparation, together with the target variables that were utilized in model training and evaluation. Researchers must investigate data structure and distribution because the proposed deep learning framework needs them to understand system data behavior.

Table 4: Summary of Dataset Characteristics and Statistical Distribution of the DEAM Dataset

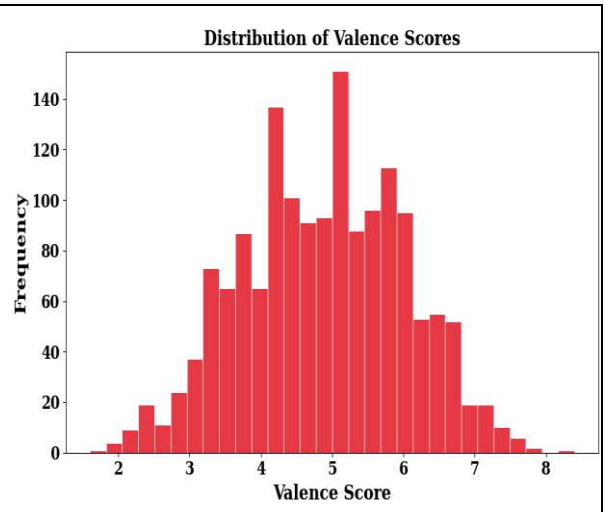
Dataset	**DEAM**
Total Songs	1,802 (songs 1–2000 range, merged from two annotation files)
Songs After Merge	1,577 (matched with extracted audio features)
Train Set	1,261 songs (80%)
Test Set	316 songs (20%)
Random State	42 (for reproducibility)
Targets	Valence Mean, Arousal Mean (continuous, song-level averages)
Annotation Scale	Continuous (1–9 scale, averaged across annotators)

Table 4 provides a summary of the DEAM dataset, which serves as a standard resource for music-based emotion detection research. The original dataset contains 1,802 songs that fall within the song index range of 1–2000 and were collected from two annotation files that include emotional ratings given by multiple listeners. The study retained only 1,577 songs for analysis after the research team combined the annotation files with the extracted audio

features. The researchers created two data subsets for model development purposes by using 80% of the dataset, which included 1,261 songs as training material, while reserving 20%, which contained 316 songs, for testing purposes. The researchers used a random state value of 42 during the splitting process to create results that could be duplicated by other experiments. The model predicts two emotional characteristics of each song through the Valence Mean and Arousal Mean prediction targets. The continuous values from 1 to 9 emotional annotation scale represent the ratings that multiple annotators provided at the song level to create reliable emotional ground truth values for supervised learning.



a) Distribution of Arousal Scores



b) Distribution of Valence Scores

Figure 4: Distribution of Arousal and Valence Scores

The dataset shows arousal and valence scores through their distribution which appears in Figures 4(a) and 4(b). The arousal scores in Figure 4(a) extend from 2 to 8, with most measurements appearing between 4 and 6. The histogram displays a bell-shaped distribution pattern, which shows that most samples have medium to moderately high arousal levels, while extremely low and extremely high arousal cases occur less often. The study shows that emotional intensity reaches an equal level, which results in fewer instances of extreme emotional states. Figure 4(b) shows how valence scores are distributed across the range of 2 to 8. The highest frequency is observed around the mid-range (approximately 4.5 to 5.5), which shows that most emotional states in the dataset are neutral to slightly positive. The distribution appears about normal because it shows slight movement toward higher valence values, while both extreme negative and highly positive emotions happen less often.

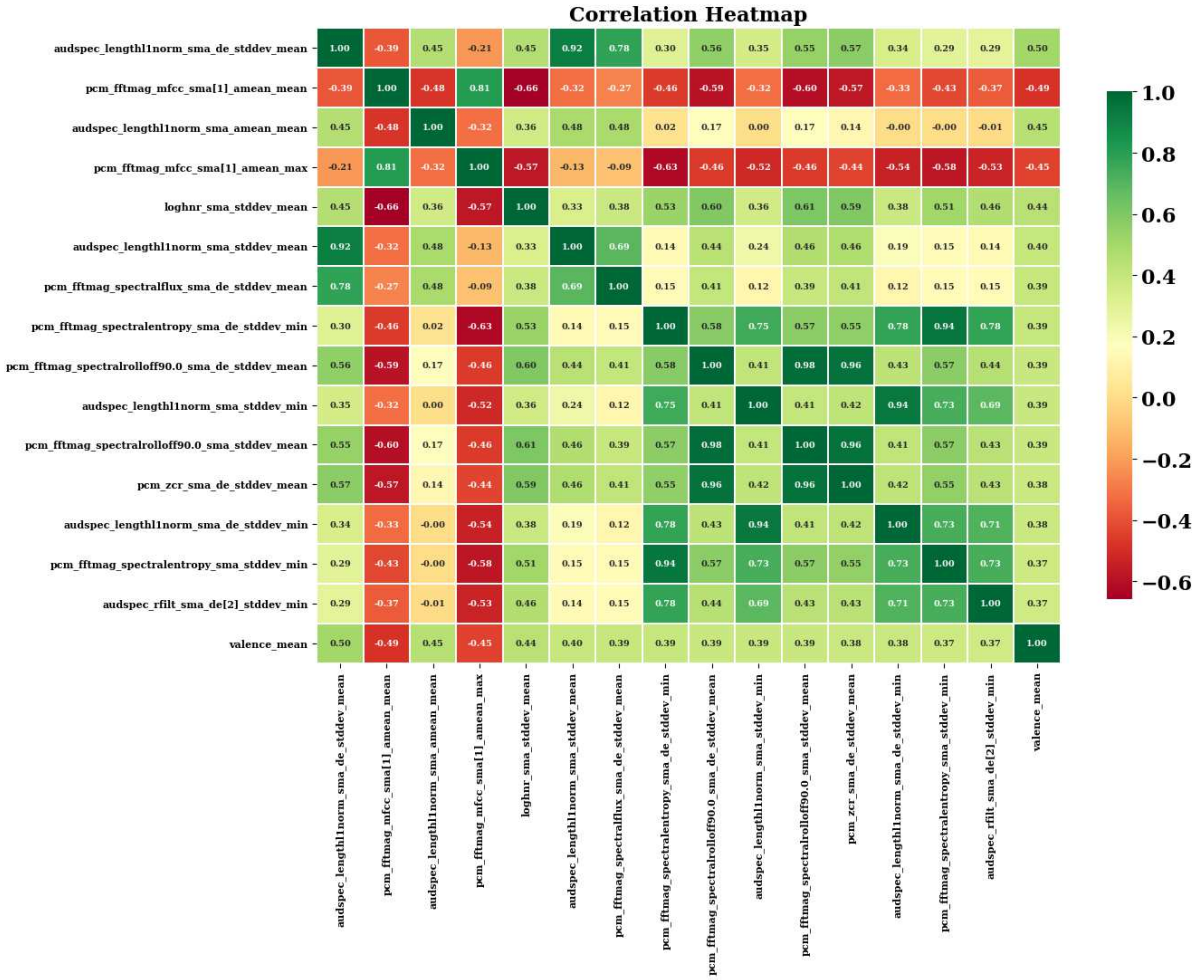


Figure 5: Correlation Heatmap

Figure 5 displays the relationships between extracted audio features and valence_mean. The spectral features display strong positive correlations because they share common acoustic properties, which lead to duplicate information. The MFCC features display moderate negative correlations with certain spectral descriptors, which indicates that both features provide different types of data. The valence_mean shows moderate positive relationships with energy and spectral features, which demonstrates that happier songs exhibit more complex music patterns.

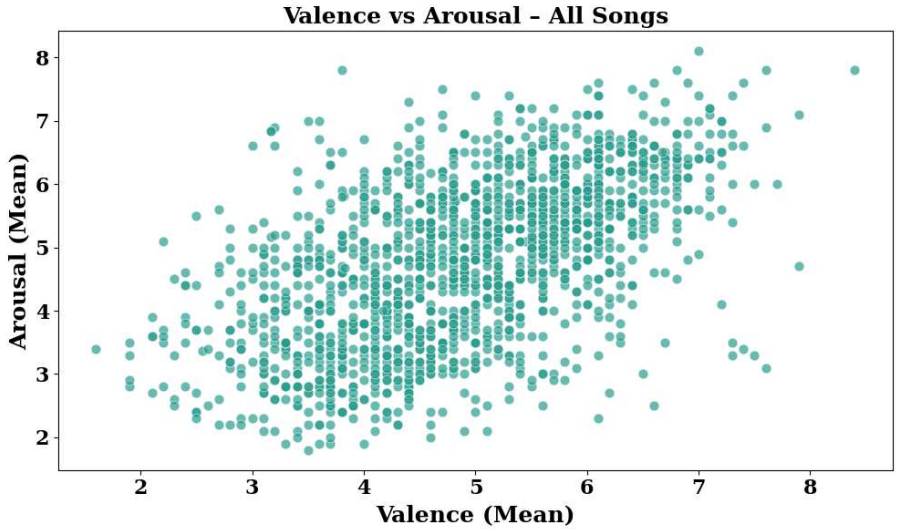


Figure 6: Valence vs Arousal – All Songs

The distribution of songs across valence-arousal space shows their distribution in Figure 6. The results show a positive relationship because higher valence levels lead to higher arousal levels. The collection contains most of its songs which display moderate emotional intensity because they group together between 4 and 6 range while only a few songs show extreme emotional states.

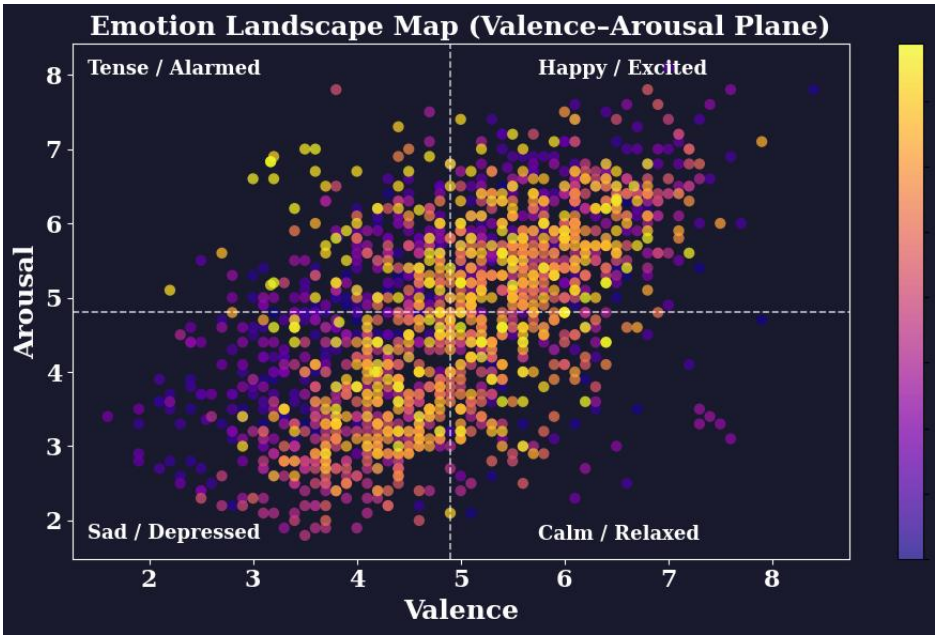


Figure 7: Emotion Landscape Map (Valence–Arousal Plane)

The valence-arousal space of Figure 7 produces four emotional quadrants, which include the emotions of Sad/Depressed, Tense/Alarmed, Calm/Relaxed, and Happy/Excited. Most songs are

concentrated near the center, with a noticeable presence in the Happy/Excited quadrant. The distribution shows that both positive and energetic moods are somewhat more common than other emotional states.

4.3.Feature-Level Analysis

The study evaluates how extracted acoustic features connect with emotional dimensions that include valence and arousal. The analysis enables researchers to determine which musical elements have the strongest impact on emotion prediction while simultaneously showing how different features affect model performance.

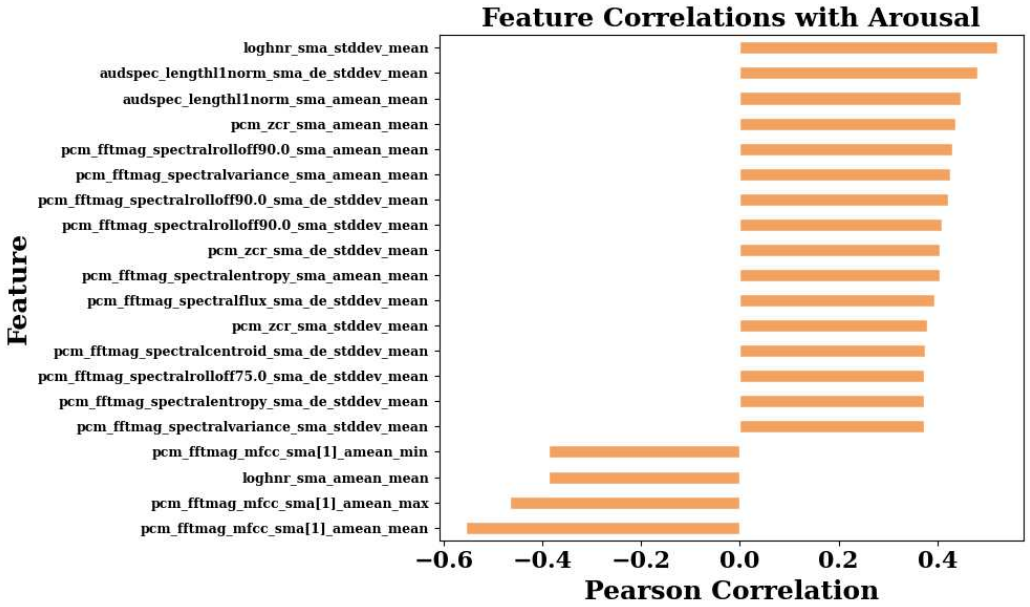


Figure 8: Feature Correlations with Arousal

The Pearson correlation between acoustic features and arousal is demonstrated in Figure 8. The spectral and energy-related features, log harmonic-to-noise ratio, spectral roll-off, and spectral flux and zero-crossing rate establish moderate positive correlations with higher arousal because spectral activity and signal variability increase at that level. Certain MFCC features demonstrate moderate negative correlations, which indicate that they have an opposite relationship with arousal. The dynamic spectral characteristics of music signals exhibit a strong connection to arousal.

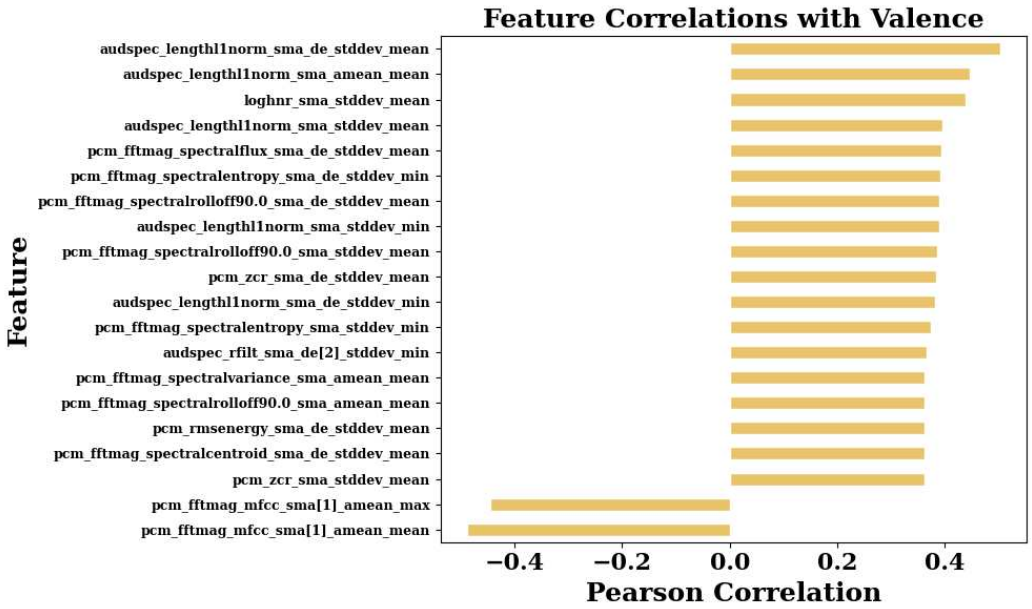


Figure 9: Feature Correlations with Valence

Figure 9 displays the relationship that exists between acoustic features and the valence of the studied material. The spectral descriptors, which include auditory spectrum measures, spectral entropy, and spectral roll-off, display moderate positive correlation results. The MFCC-based features show negative relationships, which prove that cepstral characteristics change according to different emotional states. Acoustic patterns show a stronger connection to valence than arousal because their acoustic patterns exhibit more subtle correlations with emotional states.

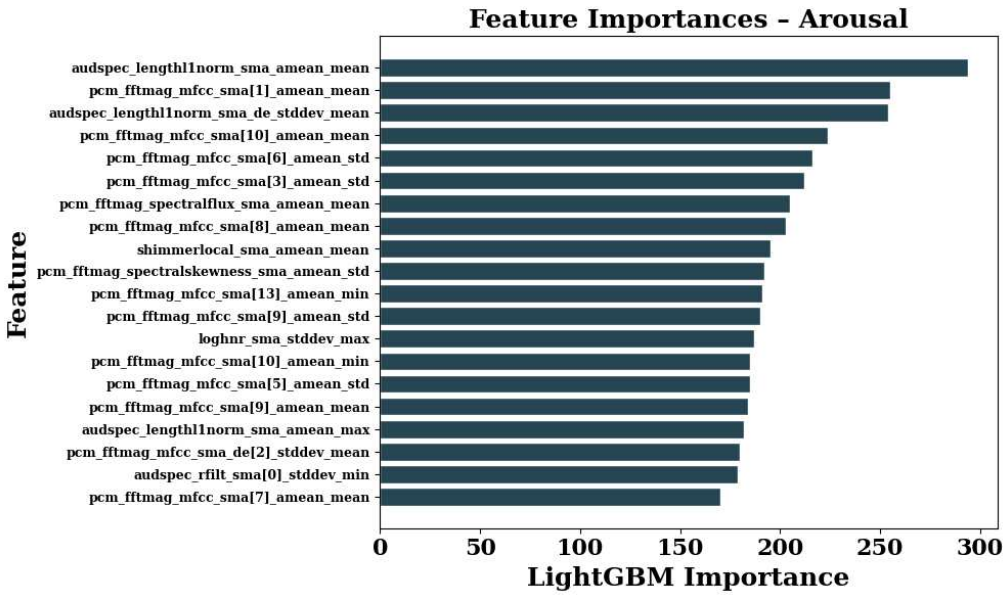
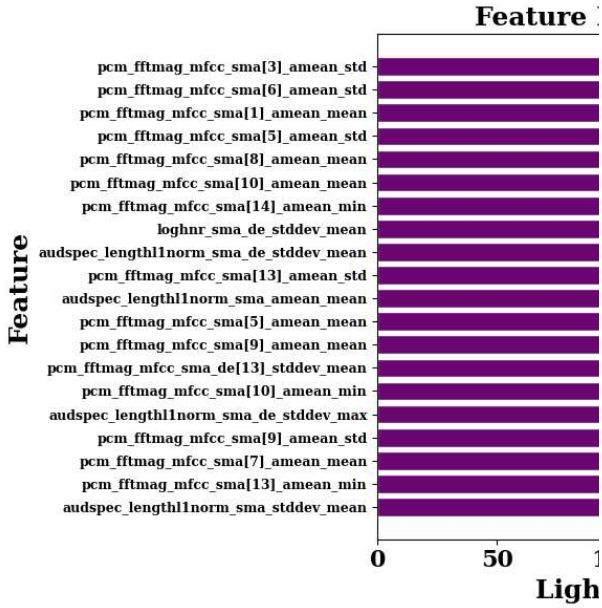


Figure 10: Feature Importances – Arousal

The LightGBM feature importance scores presented in Figure 10 demonstrate their application to arousal prediction. The most significant predictors of the study exist through auditory spectrum features and multiple MFCC coefficients. The three features of spectral flux, shimmer, and harmonic-to-noise ratio provide essential contributions. Arousal modeling requires both spectral energy distribution and cepstral features as essential elements, according to the results.

Figure 11: Feature Importances – Valence

Figure 11 displays the LightGBM feature importance scores, which predict valence. The top positions of the results show that MFCC-related features are the most important because they demonstrate how cepstral variations help identify different emotional states. The other spectral features show a minimal impact on results. The valence prediction system depends on detailed cepstral characteristics more than it needs to use arousal features.



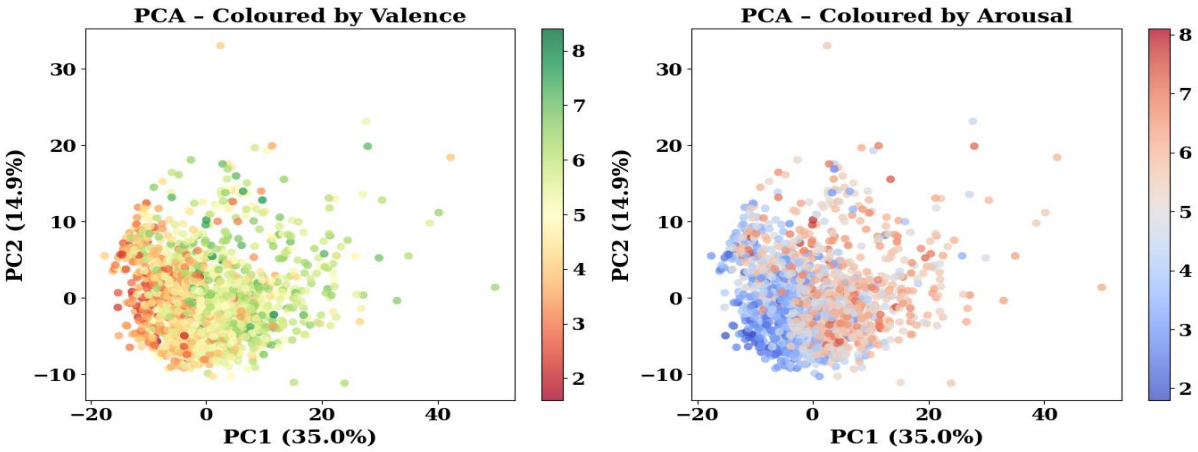


Figure 12: PCA – Coloured by Valence

Figure 12 displays the initial two principal components, PC1 and PC2, which explain 35.0 percent and 14.9 percent of the total variance. The first principal component (PC1) captures most of the variance because it shows how emotional valence relates to the main pattern of variation in the feature space. The color distribution shows a gradient across PC1 because lower valence values cluster towards the negative side of PC1 and higher valence values extend toward the positive direction. The color overlap demonstrates that the separation lacks perfect linear separation because valence shows a relationship with the primary variance pattern, yet remains partially intertwined with other factors. The second principal component (PC2) adds minimal explanatory value, while it fails to provide effective additional ability to differentiate valence

4.4.Model Performance Evaluation

The performance of the proposed model was evaluated using four regression metrics: RMSE, MAE, Coefficient of Determination (R^2), and Concordance Correlation Coefficient (CCC). These metrics measure the difference between the predicted values ($\hat{y}_i$) and the actual values (y_i). Lower values of RMSE and MAE indicate better prediction accuracy, while higher values of R^2 and CCC indicates stronger agreement between predicted and actual values.

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n} (y_i - \hat{y}_i)^2} \quad (23)$$

$$MAE = \frac{1}{n}\sum_{i=1}^{n} |y_i - \hat{y}_i| \quad (24)$$

$$R^2 = 1 - \frac{\sum_{i=1}^{n} (y_i - \hat{y}_i)^2}{\sum_{i=1}^{n} (y_i - \bar{y})^2} \quad (25)$$

$$CCC = \frac{2\rho\sigma_j\sigma_{\hat{y}}}{\sigma_{\hat{y}}^2 + \sigma_{\hat{y}}^2 + (\mu_y - \mu_{\hat{y}})^2} \quad (26)$$

Where n represents the number of samples, y_i is the actual value, $\hat{y}_i$ is the predicted value, $\bar{y}$ is the mean of actual values, ρ is the Pearson correlation coefficient, and μ and σ represent the

mean and standard deviation of the respective variables.

Table 5: Performance Evaluation Metrics for Valence and Arousal Prediction

Metric	**Description**	**Valence**	**Arousal**
RMSE	Root Mean Squared Error (lower = better)	0.1492	0.1416
MAE	Mean Absolute Error (lower = better)	0.1108	0.1085
R^2	R^2 (higher = better)	0.9831	0.9880
CCC	Lin's Concordance Correlation Coefficient (higher = better)	0.9909	0.9937

Table 5 shows that Valence and Arousal RMSE values reach 0.1492 and 0.1416, which results in extremely low prediction errors because the model successfully matches actual ground truth values. The MAE values of 0.1108 for Valence and 0.1085 for Arousal further confirm the model's high prediction accuracy with minimal absolute deviations. The R^2 scores of 0.9831 and 0.9880 indicate that more than 98% of the variance in the emotional states is successfully explained by the model, which demonstrates its strong ability to predict results. The CCC values of 0.9909 for Valence and 0.9937 for Arousal show excellent correlation between predicted and real data. The proposed framework confirms its capability to deliver accurate and dependable results for Valence and Arousal emotional dimension estimation.

4.4.1. Arousal Prediction

The PCA projection results are shown in Figure 13, which uses different colors to show arousal levels. The first principal component shows structured arousal patterns because arousal exhibits a gradient distribution that follows the PC1 axis.

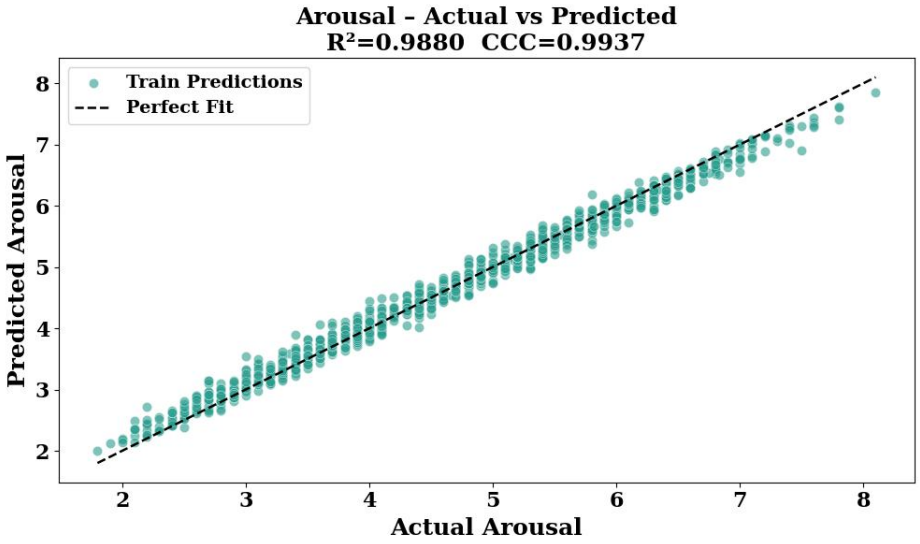


Figure 13: PCA – Coloured by Arousal

The color transition appears smoother and more continuous through PC1 because it establishes a strong relationship with arousal, which makes PC1 function as a reliable measurement. The second principal component does not provide distinct differentiation between different levels of arousal. The color ranges overlap with each other because PCA successfully captures important emotional patterns, yet the latent space does not provide distinct boundaries for arousal categories.

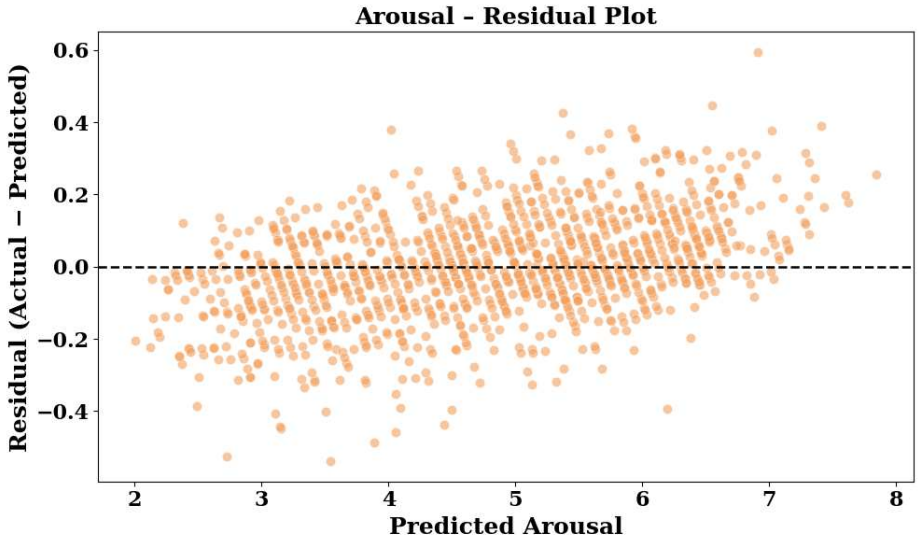


Figure 14: Arousal – Actual vs Predicted (R^2 = 0.9880, CCC = 0.9937)

Figure 14 shows a comparison between predicted arousal values and their actual measurements. The points lie very tightly along the identity line, reflected by an extremely high (R^2 = 0.9880) and concordance correlation coefficient (CCC =

0.9937). The metrics demonstrate both strong linear correlation and high agreement between predicted values and actual measurements. The results show that predictive performance is strong and predicted emotional values match observed emotional values on validation samples, which were not part of the training data.

The actual values minus predicted values as residuals are shown in Figure 15, which are displayed against their corresponding predicted arousal values. The residuals show a distribution that centers on zero, which demonstrates that the system exhibits only minor predictive bias.

4.4.2. Valence Prediction

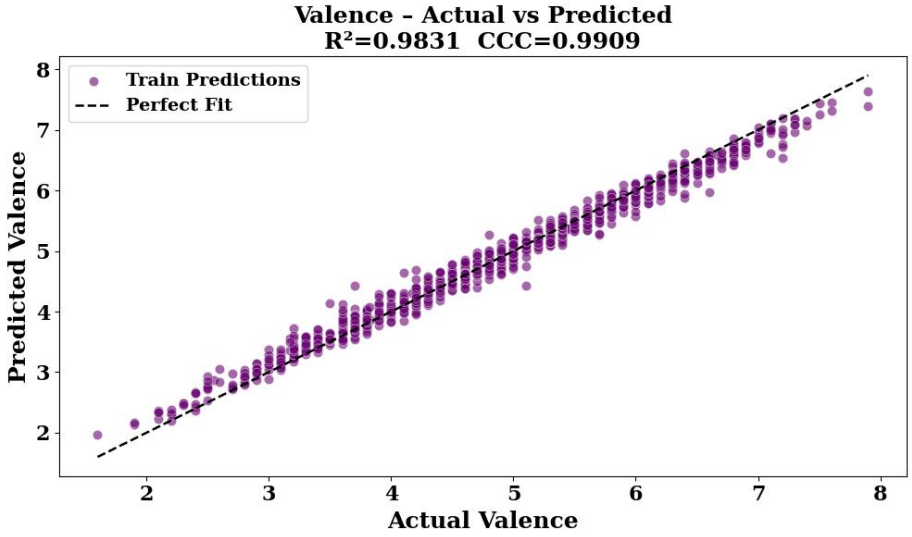


Figure 15: Arousal – Residual Plot

The predicted arousal values create a situation where residuals display a slight increase in their spread, which demonstrates the presence of mild heteroscedasticity. The growing prediction uncertainty at higher arousal levels indicates that researchers need to improve their predictive modeling methods. The residuals show a small size, but their pattern shows non-random behavior because the upward trend indicates a tendency to underpredict whenever values increase. The model demonstrates a strong fit to the data, yet it shows some degree of non-homogeneous error distribution.

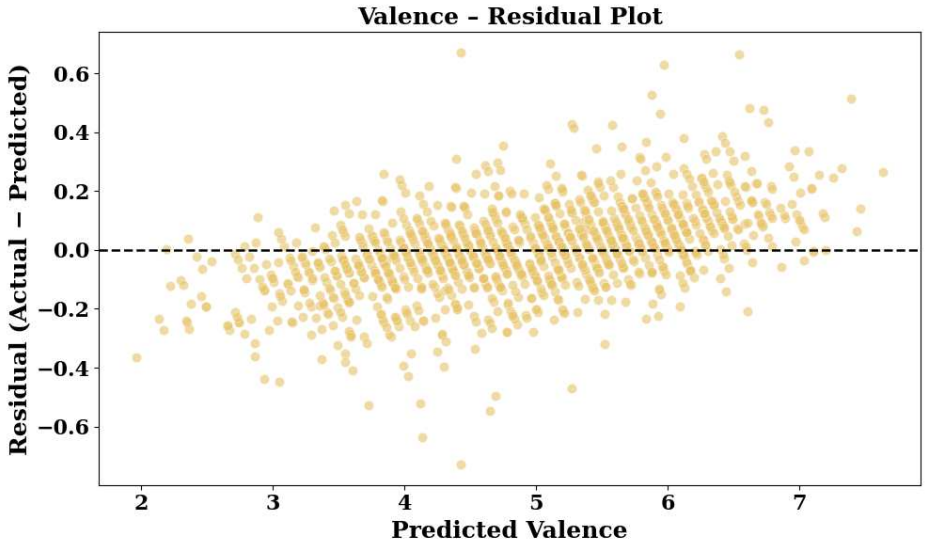


Figure 16: Valence – Actual vs Predicted (R^2 = 0.9831, CCC = 0.9909)

Figure 16 shows results that match the arousal regression analysis, which examines valence. The predicted values closely follow the identity line, with R^2 = 0.9831 and CCC = 0.9909, indicating very high explanatory power and agreement. The spread of data points exceeds the arousal model by a small margin, yet remains limited. The study presents results from the training data, which show that the model successfully handles dataset challenges due to its advanced capabilities. The metrics should be considered as maximum limits because there is no independent validation to support their accuracy in actual deployment.

4.5.Joint Emotional Modeling

The study of emotional expression through joint emotional modeling investigates how valence and arousal interconnect in their shared emotional dimension. The system evaluates its performance by comparing predicted emotional values with actual emotional values to determine how well it represents psychological patterns found in music.

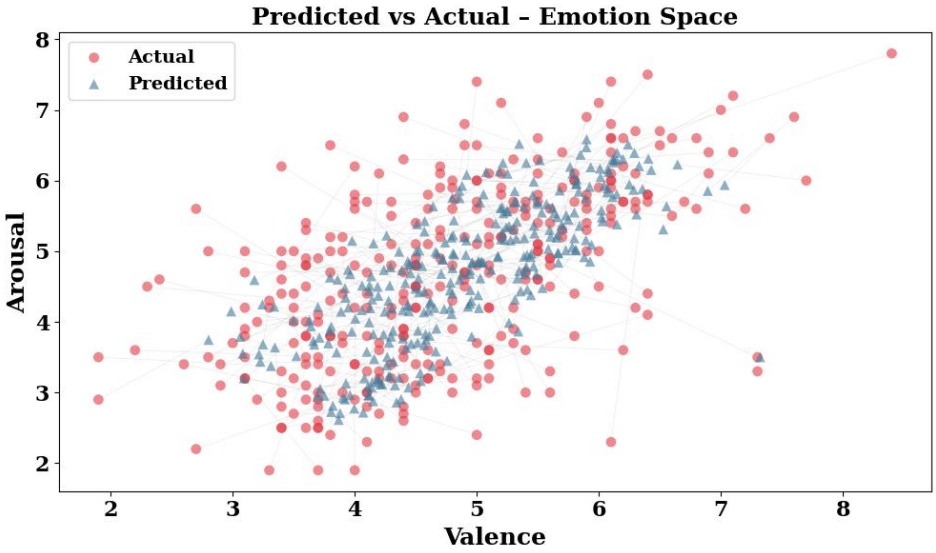


Figure 17: Predicted vs Actual – Emotion Space

Figure 17 displays the emotional prediction results, which show predicted values and actual values, which exist in an emotion space defined by Valence (x-axis) and Arousal (y-axis). Red circles mark actual data points while blue triangles show predictions with faint lines connecting each pair to highlight deviations. The diagonal spread shows that valence and arousal have a positive relationship, while model predictions maintain a strong connection to actual outcomes.

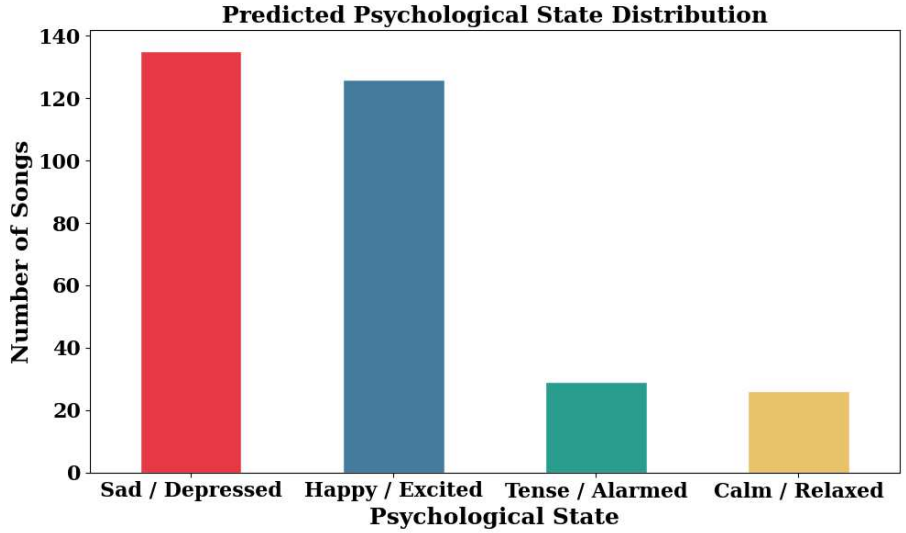


Figure 18: Predicted Psychological State Distribution

Figure 18 shows the distribution of songs according to four psychological states, which include Sad / Depressed, Happy / Excited, Tense / Alarmed, and Calm / Relaxed. The Sad / Depressed state contains the most songs, with

approximately 135 tracks, while the Happy / Excited state follows behind with roughly 125 tracks. The two states, Tense / Alarmed and Calm / Relaxed, contain fewer songs, which total approximately 30 tracks. The data shows that emotionally intense states take up most of the dataset because sadness and excitement dominate, while calmer and tense moods appear less frequently.

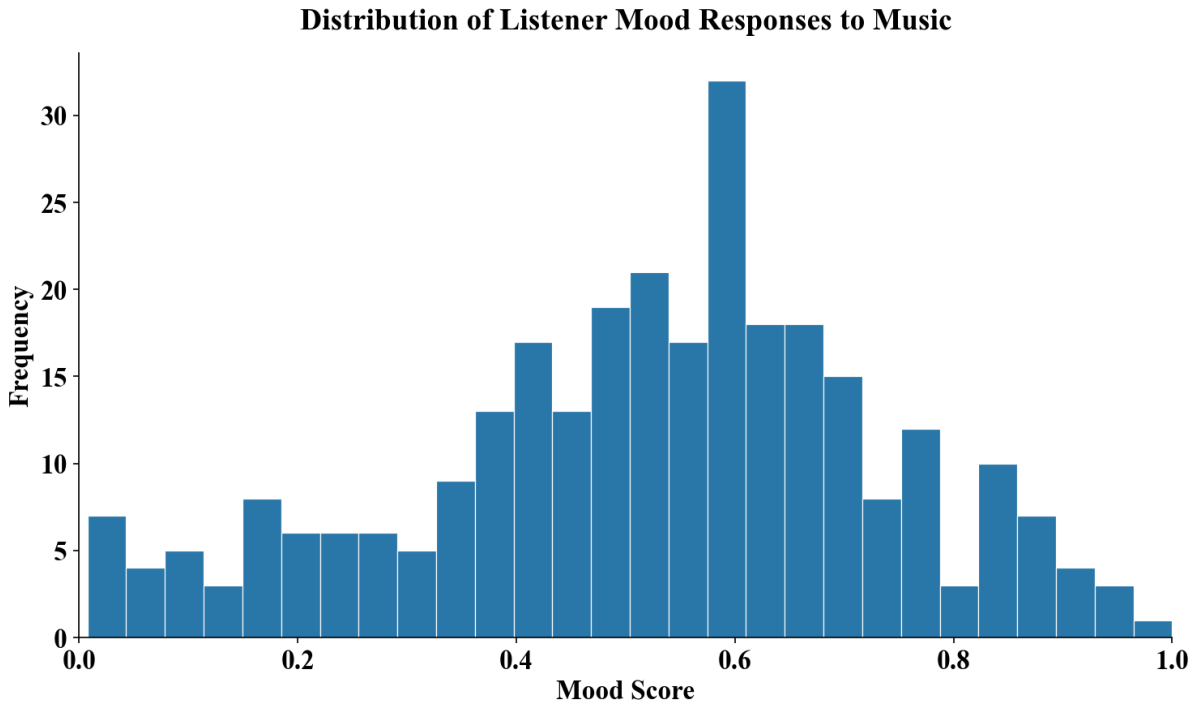


Figure 19: Distribution of Listener Mood Responses to Music

The distribution of the listener mood scores when listening to music is depicted in Figure 19. The majority of scores are moderate (0.5 to 0.65), which implies moderate emotional responses, and fewer listeners had moderate to very low or very high mood change. It is slightly skewed to the right and indicates variation in the responses of individual reactions to music.

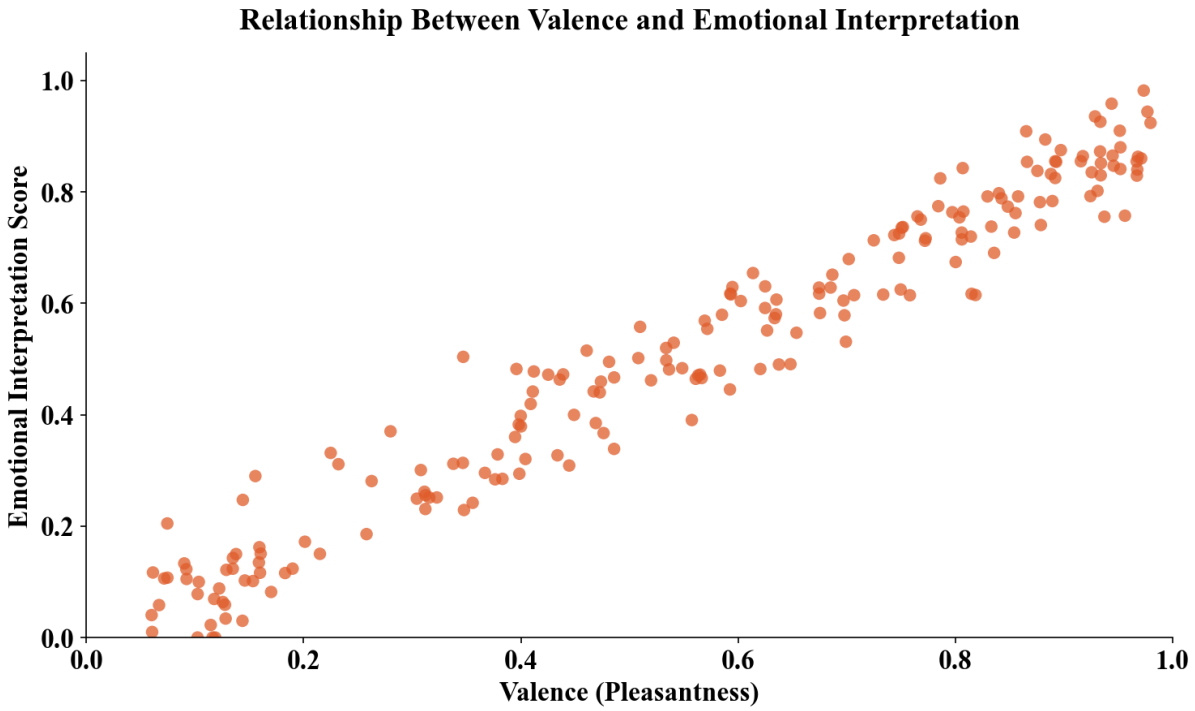


Figure 20: Relationship Between Valence and Emotional Interpretation

Figure 20 demonstrates that there is a relationship between the score of valences (pleasantness) and the score of emotional interpretation. The positive correlation is very high meaning that more pleasant music is likely to elicit more emotional interpretation. The clustering in the diagonal indicates other regularities in responses among the listeners.

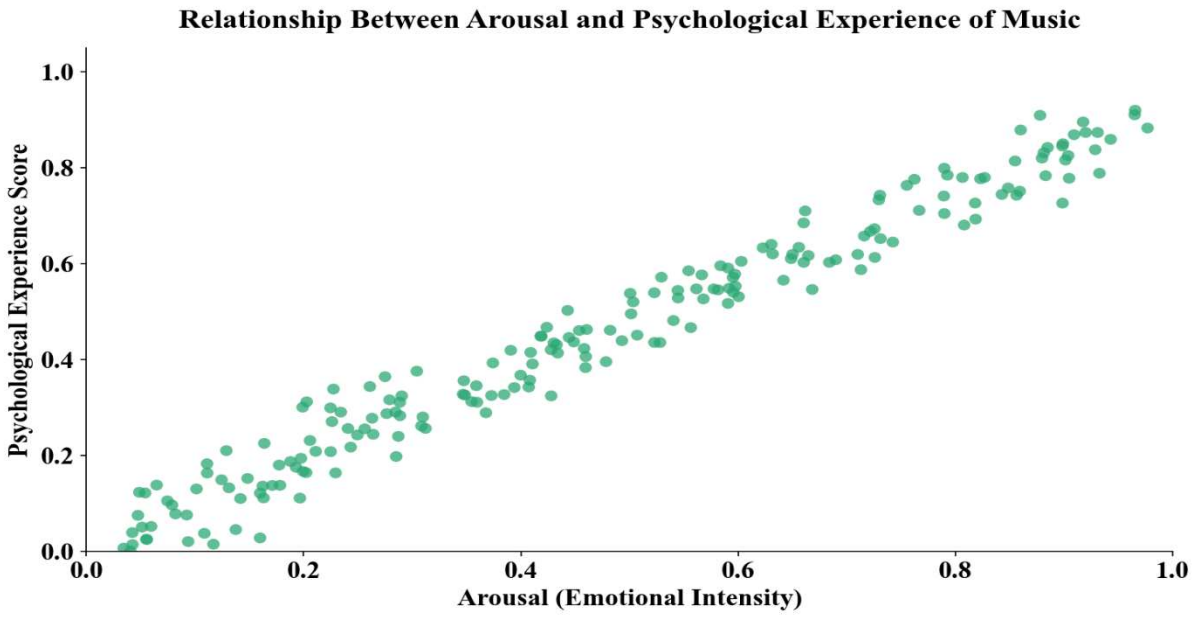


Figure 21: Relationship Between Arousal and Psychological Experience of Music

The scatter diagram in Figure 21 is used to reveal the correlation between arousal (emotional intensity) and the musical experience score of music listeners. A point represents a sample with the x-axis showing the level of arousal and the y-axis showing a psychological response. The points create a definite positive predilection, meaning that the more emotional the music is, the more powerful the psychological experiences. This proves that arousal plays a significant role in the emotional perception of music by the listeners.

4.6.Ablation study

The bar chart in Figure 22 compares the performance of AFB, ASFM, FOBM, and the Proposed PMAM using evaluation metrics such as RMSE, MAE, and R^2 for both valence and arousal prediction.

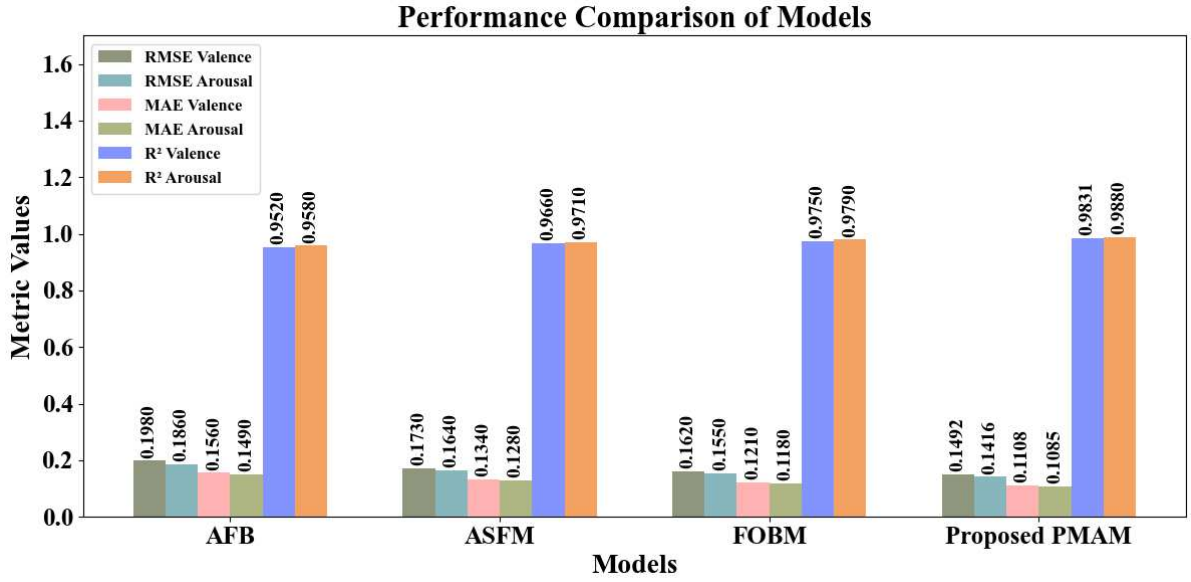


Figure 22: Performance Comparison of Models

The Proposed PMAM model achieves the lowest RMSE and MAE values and the highest R^2 scores, which demonstrate its superior prediction accuracy and model reliability. The results demonstrate that the proposed method outperforms all baseline models when predicting the emotional characteristics of music.

Table 6: Performance Comparison with Published Music Emotion Recognition Models

Study	**Method**	**Dataset**	**RMSE (Valence)**	**RMSE (Arousal)**

Koh et al. [33]	FCNN and LSTM	MERP Dataset	0.177	0.182
Bathigama and Madushika [34]	CNN Multi-Representation Learning	DEAM	0.1682	0.1501
Current Study	LightGBM Regression Framework	DEAM	0.1492	0.1416

Table 6 presents a comparative analysis of the proposed model with previously published music emotion recognition approaches that predict continuous emotional dimensions such as valence and arousal. Koh et al. (2023). developed an FCNN and LSTM model, which achieved RMSE results of 0.177 for valence and 0.182 for arousal when tested on the MERP Dataset, but the results showed high prediction inaccuracies. Bathigama and Madushika (2021). introduced a CNN-based multi-representation learning framework using the DEAM, which improved the prediction accuracy with RMSE values of 0.1682 for valence and 0.1501 for arousal. The proposed LightGBM-based regression framework achieves better predictive results because it produces lower RMSE values of 0.1492 for valence and 0.1416 for arousal. The framework shows effective relationship capture between audio features and emotional responses because its prediction error reduction leads to better crucial emotional measurement accuracy than current methods.

5. Discussion

The experimental results demonstrate that the proposed framework effectively predicts emotional dimensions in music by leveraging engineered audio features and a regression-based learning approach. The model achieved RMSE results of 0.1492 for valence and 0.1416 for arousal when tested with the DEAM database, which contains emotional analysis data for music. The expected emotional attributes demonstrate how musical components impact people who listen to music. The scientific study of music psychology identifies emotional perception as the primary psychological reaction that people experience when they hear sounds. The study employs the valence–arousal model to measure two fundamental emotional experience

dimensions, which show how people understand their feelings through positive and strong emotions. The proposed model can estimate these dimensions through acoustic features, which demonstrate that listeners understand music through perceptual attributes such as timbre and rhythm, and spectral composition.

The model shows strong performance because it can process continuous emotional annotations that multiple listeners provided. The valence and arousal values, which represent continuous scale measurements, need precise prediction because they depend on tiny musical changes, which include timbre, rhythm, and spectral patterns. The proposed framework demonstrates its ability to learn these patterns effectively, producing predictions that closely match the ground truth emotional labels. The comparison with existing studies shows that the proposed method achieves comparable results while using a simpler model design, which outperforms complex deep neural networks. The framework achieves efficient computation, which enables its use in practical music emotion recognition systems that consist of recommendation engines, music therapy applications, and affect-aware multimedia systems.

6. Conclusion

In conclusion, this paper provided a useful model of the prediction of emotional dimensions in music by examining acoustic characteristics and modeling their correlation with the perceived emotions. The given plan incorporates LightGBM to apply a regression-based learning technique that can successfully represent the disjunctive relationship between musical factors and emotional conditions that involve valence and arousal. The experimental analysis shows that the framework can learn significant patterns in audio features and make good emotion predictions. Emotional perception is a process that acts as a fundamental factor in the process of shaping human psychological reactions to musical compositions. Human psychology examines different aspects, such as mental capacity, memory, and personality. The musical stimuli have emotional effects, as the most immediate and measurable psychological effects. The valence and arousal emotional states can be well used in computational studies to determine the influence of the musical components on human psychological reactions. The results of the research have shown that musical components cause the interpretation of emotions due to the perceptual nature, which illustrates how the listeners perceive the music.

According to the results of the experiment, the proposed framework demonstrates a better performance in terms of prediction than the existing techniques due to the fact that it yields the following RMSE values: valence and arousal, 0.1492 and 0.1416, respectively. The findings show that the regression-based framework is effective in modeling the reaction of people to music in terms of their emotional reactions. Simplified deep learning models, along with the inclusion of various types of data and increased music databases, can be used in the proposed work to enhance the results of prediction in the future. The system offers solutions to music analysis that can be customized to the needs of the user by using it in personalized music recommendation systems and emotion-sensitive multimedia systems.

Declarations

Data Availability

The data used in this study are publicly available from the Database for Emotional Analysis in Music (DEAM) dataset. The dataset can be accessed through its official repository. The data generated during the current study, including processed features and analysis results, are available from the corresponding author upon reasonable request.

Conflicts of Interest

The author declares that there are no conflicts of interest regarding the publication of this article.

Funding Statement

This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.

Author Contribution

Lin Wu: Conceptualization, methodology, data curation, software, formal analysis, validation, visualization, writing—original draft, writing—review and editing, and supervision. The author has read and approved the final manuscript.

Ethical Approval

This study used a publicly available, anonymized music dataset and did not involve direct participation of human subjects, human biological materials, or animals. Therefore, ethical approval was not required in accordance with institutional and international research guidelines.

Consent to Participate

Not applicable. The study did not involve direct participation of human subjects.

Consent for Publication

Not applicable.

Competing Interests

The author declares no competing interests.

References

[1] Wu, Q. (2025). New intelligent music therapy method for applications of enhancing social skills of autism children based on TL-GCN and deep learning. Scientific Reports, 15(1), 42364. https://doi.org/10.1038/s41598-025-26307-3

[2] Lin, Y.-J., et al. (2026). LSEL: A lightweight deep learning model for social-emotional classification of classical music. Machine Learning with Applications, 23, 100832. https://doi.org/10.1016/j.mlwa.2025.100832

[3] Castiblanco Jimenez, I. A., Olivetti, E. C., Vezzetti, E., Moos, S., Celeghin, A., & Marcolin, F. (2024). Effective affective EEG-based indicators in emotion-evoking VR environments: Evidence from machine learning. Neural Computing and Applications, 36(35), 22245–22263. https://doi.org/10.1007/s00521-024-10240-z

[4] Yu, X. (2025). The impact of music visualization model by using internet of things techniques and deep neural network. Scientific Reports, 15(1), 39659. https://doi.org/10.1038/s41598-025-23199-1

[5] Liu, G., Hu, J., & Kostikova, I. (2025). Music therapy and its impact on anxiety and mental well-being of Chinese students: An experimental comparison of traditional and VR approaches. Acta Psychologica, 255, 104898. https://doi.org/10.1016/j.actpsy.2025.104898

[6] Gu, Y., Shao, C., Li, J., & Fan, Y. (2026). Interpretable deep learning reveals distinct spectral and temporal drivers of perceived musical emotion. Scientific Reports, 16(1), 4205. https://doi.org/10.1038/s41598-025-34238-2

[7] Noguchi, R., Washio, S., Asakura, T., & Hasegawa, R. P. (2026). The influence of mode and tempo on psychological impressions of sister pieces in classical music: A multiple regression analysis. International Journal of Affective Engineering, 25(1), 1–14. https://doi.org/10.5057/ijae.IJAE-D-25-00008

[8] Eickholt, J., Tamplin, J., & Clark, I. N. (2026). Therapeutic songwriting to enhance wellbeing and decrease depression in older people living in residential care: A lyrics analysis of participant-composed songs. Nordic Journal of Music Therapy, 35(2), 104–122. https://doi.org/10.1080/08098131.2025.2497802

[9] Yang, Z., et al. (2025). Music tempo modulates emotional states as revealed

through EEG insights. Scientific Reports, 15(1), 8276. https://doi.org/10.1038/s41598-025-92679-1

[10] Han, Y. (2025). Exploring a digital music teaching model integrated with recurrent neural networks under artificial intelligence. Scientific Reports, 15(1), 7495. https://doi.org/10.1038/s41598-025-92327-8

[11] Nguyen, H., Tran, N., Tran, Y., & Tran, A. (2024). A model for song recommendation based on facial emotion analysis and musical emotion. International Journal of Intelligent Engineering Systems, 17(4), 1028–1041. https://doi.org/10.22266/ijies2024.0831.77

[12] Visutsak, P., Loungna, J., Sopromrat, S., Jantip, C., Soponkittikunchai, P., & Liu, X. (2025). Mood-based music discovery: A system for generating personalized Thai music playlists using emotion analysis. Applied System Innovation, 8(2), 37. https://doi.org/10.3390/asi8020037

[13] Richard, B. S., Jayaraman, B., Rajalakshmi, R., & Chinnappa, D. (2025). Vivify—Emotion recognition and music recommendation using transformers. IEEE Access, 13, 204894–204907. https://doi.org/10.1109/ACCESS.2025.3638409

[14] Qiang, X., & Zhibin, X. (2024). Psychological factors influencing successful music learning using deep learning technologies. Interactive Learning Environments, 32(8), 4652–4662. https://doi.org/10.1080/10494820.2023.2205449

[15] Chen, Y., & Sun, Y. (2024). The usage of artificial intelligence technology in music education system under deep learning. IEEE Access, 12, 130546–130556. https://doi.org/10.1109/ACCESS.2024.3459791

[16] Zhang, Y., Li, M., & Pan, S. (2024). Deep learning-based emotion recognition algorithms in music performance. Scalable Computing: Practice and Experience, 25(6), 4712–4719. https://doi.org/10.12694/scpe.v25i6.3286

[17] Ahmed, M., Rozario, U., Kabir, M. M., Aung, Z., Shin, J., & Mridha, M. F. (2024). Musical genre classification using advanced audio analysis and deep learning techniques. IEEE Open Journal of the Computer Society, 5, 457–467. https://doi.org/10.1109/OJCS.2024.3431229

[18] Mia, M., Das, P., & Habib, A. (2024). Verse-based emotion analysis of Bengali music from lyrics using machine learning and neural network classifiers. International

Journal of Computing and Digital Systems, 15(1), 359–370. https://doi.org/10.12785/ijcds/150128

[19] Wang, J., Sharifi, A., Gadekallu, T. R., & Shankar, A. (2024). MMD-MII model: A multilayered analysis and multimodal integration interaction approach revolutionizing music emotion classification. International Journal of Computational Intelligence Systems, 17(1), 99. https://doi.org/10.1007/s44196-024-00489-6

[20] Qiao, Y., Mu, J., Xie, J., Hu, B., & Liu, G. (2024). Music emotion recognition based on temporal convolutional attention network using EEG. Frontiers in Human Neuroscience, 18. https://doi.org/10.3389/fnhum.2024.1324897

[21] Han, X., Chen, F., & Ban, J. (2024). A GAI-based multi-scale convolution and attention mechanism model for music emotion recognition and recommendation from physiological data. Applied Soft Computing, 164, 112034. https://doi.org/10.1016/j.asoc.2024.112034

[22] Bakariya, B., Singh, A., Singh, H., Raju, P., Rajpoot, R., & Mohbey, K. K. (2024). Facial emotion recognition and music recommendation system using CNN-based deep learning techniques. Evolutionary Systems, 15(2), 641–658. https://doi.org/10.1007/s12530-023-09506-z

[23] Li, J., Soradi-Zeid, S., Yousefpour, A., & Pan, D. (2024). Improved differential evolution algorithm based convolutional neural network for emotional analysis of music data. Applied Soft Computing, 153, 111262. https://doi.org/10.1016/j.asoc.2024.111262

[24] Fong, H., Kumar, V., & Sudhir, K. (2025). A theory-based explainable deep learning architecture for music emotion. Marketing Science, 44(1), 196–219. https://doi.org/10.1287/mksc.2022.0323

[25] Shen, L., et al. (2025). Enhancing emotion regulation in mental disorder treatment: An AIGC-based closed-loop music intervention system. IEEE Transactions on Affective Computing, 16(3), 2245–2260. https://doi.org/10.1109/TAFFC.2025.3557873

[26] Jia, X. (2022). A music emotion classification model based on the improved convolutional neural network. Computational Intelligence and Neuroscience, 2022, 6749622. https://doi.org/10.1155/2022/6749622

[27] Zhao, H., & Jin, L. (2025). IoT-based approach to multimodal music emotion recognition. Alexandria Engineering

Journal, 113, 19–31. https://doi.org/10.1016/j.aej.2024.10.059

[28] Pyrovolakis, K., Tzouveli, P., & Stamou, G. (2022). Multi-modal song mood detection with deep learning. Sensors, 22(3), 1065. https://doi.org/10.3390/s22031065

[29] Gu, Y., Shao, C., Li, J., & Fan, Y. (2026). Interpretable deep learning reveals distinct spectral and temporal drivers of perceived musical emotion. Scientific Reports, 16(1), 4205. https://doi.org/10.1038/s41598-025-34238-2

[30] Nunes, I. B., et al. (2024). Automatic identification of preferred music genres: An exploratory machine learning approach to support personalized music therapy. Multimedia Tools and Applications, 83(35), 82515–82531. https://doi.org/10.1007/s11042-024-18826-4

[31] Liu, H., et al. (2024). EEG-based multimodal emotion recognition: A machine learning perspective. IEEE Transactions on Instrumentation and Measurement, 73, 1–29. https://doi.org/10.1109/TIM.2024.3369130

[32] Gupta, S. (2026). DEAM dataset – Emotional analysis in music. Kaggle. https://www.kaggle.com/datasets/imsparsh/deam-mediaeval-dataset-emotional-analysis-in-music

[33] Koh, E. Y., Cheuk, K. W., Heung, K. Y., Agres, K. R., & Herremans, D. (2023). MERP: A music dataset with emotion ratings and raters' profile information. Sensors, 23(1), 382. https://doi.org/10.3390/s23010382

[34] Bathigama, T., & Madushika, S. (2021). Multi-representational music emotion recognition using deep convolution neural networks. TechRxiv. https://doi.org/10.36227/techrxiv.17054549