www.ijbar.org
ISSN 2249-3352 (P) 2278-0505 (E)
Cosmos Impact Factor-5.86

Exploring Emotional Analysis in Music for Insights from the DEAM Dataset and open SMILE Features

Dr. A. Swetha[1*], Sai Madhav[2], D. Bharath Chandra[2], I. Mohammed Adnan Qureshi[2], Mohammed Arshad Noman[2]

[1,2]Department of Computer Science and Engineering (AI&ML), Vaagdevi College of Engineering, Bollikunta, Warangal, Telangana.

*Corresponding Email: swetha_a@vaagdevi.edu.in

ABSTRACT

The emotional impact of music has long intrigued researchers in fields ranging from psychology to artificial intelligence. With the increasing availability of emotionally annotated music datasets and advanced audio processing tools, emotion recognition in music has become a vibrant area of research. The DEAM (Database for Emotional Analysis of Music) dataset, which includes both dynamic and static annotations of valence and arousal values, serves as the primary data source. To capture expressive audio features relevant to emotional states, the openSMILE toolkit is employed, extracting a comprehensive set of low-level descriptors (LLDs) and functionals such as pitch, energy, MFCCs, and spectral properties. The study initially explores the performance of several conventional machine learning algorithms, including Support Vector Machine (SVM), K-Nearest Neighbors (KNN), Logistic Regression Classifier (LRC), and Decision Tree Classifier (DTC). These models are evaluated using metrics such as accuracy, precision, recall, and F1-score to classify emotions based on the extracted audio features. While these models provide baseline performance, their limitations in capturing complex patterns and feature interactions prompt the investigation of more advanced techniques. As a proposed enhancement, the Light Gradient Boosting Machine (LightGBM) classifier is introduced. LightGBM, known for its efficiency and high accuracy in handling large-scale and high-dimensional data, demonstrates superior performance in recognizing emotional content in music. Its gradient boosting framework, coupled with leaf-wise tree growth, allows it to model intricate non-linear relationships between features and target emotional states more effectively than traditional methods. Experimental results highlight that the LightGBM classifier outperforms the baseline models, offering improved classification accuracy and better generalization on unseen data. This project underscores the potential of combining rich audio feature sets from openSMILE with powerful ensemble-based learning approaches for advancing music emotion recognition. The findings can be instrumental in applications such as affective music recommendation systems, music therapy, and multimedia content tagging.

Keywords: Emotional Analysis, DEAM Dataset, Music Emotion Recognition (MER), Acoustic Features, Time-Series Emotion Data

1. INTRODUCTION

Music is a universal form of expression that evokes a wide range of emotions, influencing human behavior, mood, and cognition. In recent years, there has been growing interest in understanding and modeling the emotional impact of music using computational methods. Emotional Analysis in Music is a research area that aims to identify and classify emotions conveyed or induced by musical compositions through machine learning and audio signal processing techniques. This project leverages the DEAM (Database for Emotional Analysis of Music) dataset, which contains both dynamic and static emotion annotations based on valence and arousal dimensions. To extract meaningful acoustic features, the

www.ijbar.org
ISSN 2249-3352 (P) 2278-0505 (E)
Cosmos Impact Factor-5.86

openSMILE toolkit is used, which provides a robust set of low-level descriptors such as pitch, MFCCs, energy, and spectral features.

Audio Pattern	Information	Averaged Song Ratings
	File: 115.mp3 (Quadrant 1) High Valence High Arousal	Valence: 8.4 Arousal: 7.8
	File: 2057.mp3 (Quadrant 2) Low Valence High Arousal	Valence: 3.2 Arousal: 6.8
	File: 488.mp3 (Quadrant 3) Low Valence Low Arousal	Valence: 1.6 Arousal: 3.4
	File: 977.mp3 (Quadrant 4) High Valence Low Arousal	Valence: 7.3 Arousal: 3.5

Fig 1: Sample audio files of DEAM Dataset

A comparative evaluation is performed using multiple classifiers SVM, KNN, Logistic Regression Classifier (LRC), and Decision Tree Classifier (DTC)—as existing systems, with LightGBM proposed as a performance-enhancing classifier due to its gradient boosting capabilities. The emotional context in music is subjective and multi-dimensional, making it difficult to generalize models across different datasets and listener perceptions. Furthermore, most existing systems fail to exploit the full potential of feature-rich datasets like DEAM when coupled with high-performance toolkits such as openSMILE. Therefore, there is a need to develop an efficient, scalable, and accurate classification model that can interpret emotional content in music with improved generalization and minimal loss of information. The integration of the DEAM dataset with openSMILE audio features allows for the extraction of a highly descriptive feature space that encapsulates both temporal and tonal qualities of music. By comparing conventional classifiers with the advanced LightGBM model, the project not only benchmarks existing approaches but also demonstrates the effectiveness of boosting algorithms in learning complex emotional representations. The outcomes of this research can significantly advance the design of emotionally intelligent systems, enhance personalization in music streaming services, and open new avenues in emotion-aware human-computer interaction.

2. LITERATURE SURVEY

Herremans et al. [1] highlighted the potential of deep learning techniques in revolutionizing music and audio technologies, emphasizing how neural networks can automatically learn intricate patterns in audio data. Their work underlines the importance of feature learning and supports the shift from handcrafted features to data-driven models. This supports the motivation for using advanced classifiers like LightGBM in emotion recognition tasks. The emergence of deep models enables richer and more abstract representations, which is essential for handling complex emotion-related patterns in music. Yang et al. [2] explored the influence of individual differences in Music Emotion Recognition (MER) and pointed out that user-specific responses to music are often overlooked. They emphasized that the same musical piece can elicit varying emotional reactions depending on personal experiences. This

insight is crucial for improving personalized emotion recognition systems and highlights the challenges of building generalizable models across different users and cultures. Aljanaki et al. [3] contributed significantly by creating a benchmark dataset for emotional analysis of music, which includes the DEAM dataset. They evaluated various models and showed the importance of consistent annotations and reliable evaluation strategies in music emotion research. Their work provides the foundational dataset used in this project, ensuring that the selected features and models are evaluated on standardized and validated data. Schmidt et al. [4] introduced the use of Conditional Random Fields (CRFs) to model emotion dynamics in music, showing that emotion is not static but evolves over time. Their approach demonstrated how temporal dependencies play a key role in understanding music-induced emotions. This encourages the use of time-aware models and supports the importance of capturing dynamic features from datasets like DEAM. Chua et al. [5] examined emotion prediction using multimodal data from music videos, exploring the relative importance of visual and audio cues. Their findings revealed that audio remains a dominant contributor to emotional perception, validating the focus of this project on audio-only features extracted through openSMILE. They also provided insights into how different sensory modalities can impact affective computing systems. Russell et al. [6] proposed the circumplex model of affect, where emotions are mapped in a two-dimensional space defined by valence and arousal. This model forms the theoretical backbone of the DEAM dataset and is extensively used in labeling emotional states in music. Their work helps frame the classification problem and guides how emotion classes are structured in the current project.

Seashore et al. [7] were among the pioneers in measuring emotions in music, conducting early experimental studies on emotional expressions. They demonstrated that elements such as rhythm, tempo, and intensity influence the emotional character of music. Though dated, this work laid the groundwork for using measurable acoustic features to predict affective states. Meyer et al. [8] argued that musical emotion arises from violations of listener expectations, combining cognitive theories with emotional responses. His theory implies that subtle changes in musical structure can have strong emotional impacts. This justifies the use of fine-grained low-level descriptors like those extracted using openSMILE for emotion analysis. Juslin et al. [9] provided an in-depth explanation of how emotions are expressed and perceived in music, introducing the concept of multiple emotion induction mechanisms. His research also emphasized the importance of considering both subjective listener responses and objective audio features, which this project balances through machine learning classifiers on annotated data. Cespedes-Guevara et al. [10] challenged the notion of music conveying basic emotions, suggesting instead that it communicates affective states. They proposed a constructionist view that aligns with the valence-arousal model. Their perspective supports the DEAM framework and the continuous representation of emotions over categorical labels. Saarikallio et al. [11] compared music emotional responses across cultures, indicating that cultural context significantly influences how music is perceived emotionally. Their findings stress the need for adaptable models in MER and validate the importance of robust and context-independent feature extraction methods like those used in this study. Panda et al. [12] reviewed various audio features for music emotion recognition and categorized them based on their relevance and performance. They recommended features related to timbre, rhythm, and harmony, many of which are available through openSMILE. Their survey supports the use of hybrid and ensemble-based classification models, including LightGBM, to enhance performance. Er et al. [13] investigated chroma spectrograms and visual features for emotion recognition, showing that music's tonal characteristics are key indicators of emotional content. Their experimental results demonstrated the value of combining spectral and deep features for better accuracy. While their method used visual models, their emphasis on chroma aligns with this project's acoustic focus. Gómez-Cañón et al. [14]

called for new standards in MER, focusing on personalization and context sensitivity. They proposed robust evaluation protocols and stressed the role of context in emotion modeling. Their call for better benchmarking supports the use of the DEAM dataset, and their push for robustness aligns with the use of ensemble classifiers like LightGBM. Herremans et al. [15] presented Imma-emo, a multimodal interface for visualizing emotion annotations in music. Their tool enabled synchronized visualization of emotional dynamics across both audio and scores. While this project doesn't utilize visualization tools, their work shows the value of detailed annotations and temporal tracking, both present in the DEAM dataset. Turnbull et al. [16] worked on query-by-semantic-description in music retrieval, proposing systems that allow users to search for music based on emotional or descriptive terms. Their methodology aligns with emotion recognition tasks and shows how semantic tagging can enrich music retrieval systems—one of the potential applications of the project. Aljanaki et al. [17] also explored emotion recognition using a crowdsourcing game to gather subjective emotion labels for music. Their work highlights the variability in emotion annotation and the need for robust learning models. Their use of game-based crowdsourcing supports the reliability of the DEAM dataset and provides context for the diversity in user-generated labels.

3. PROPOSED SYSTEM

The proposed system for "Exploring Emotional Analysis in Music for Insights from the DEAM Dataset and OpenSMILE Features" leverages advanced machine learning techniques to classify emotions in music. It begins with the collection of the DEAM dataset, which provides both audio tracks and emotion annotations for various musical pieces. Audio data is then preprocessed using OpenSMILE to extract key acoustic features such as pitch, MFCCs, and energy levels. These features are cleaned, encoded, and selected for their relevance using dimensionality reduction methods like PCA. The system then employs traditional machine learning classifiers, including SVM, KNN, Logistic Regression, and Decision Trees, as baseline models. The core innovation of the system lies in the implementation of LightGBM, a powerful gradient-boosting machine learning algorithm, which is expected to outperform the traditional models by providing faster, more accurate emotional classification. Finally, model performance is evaluated and compared using various metrics, and feature importance is analyzed to interpret the emotional cues most relevant to music emotion recognition.

www.ijbar.org
ISSN 2249-3352 (P) 2278-0505 (E)
Cosmos Impact Factor-5.86

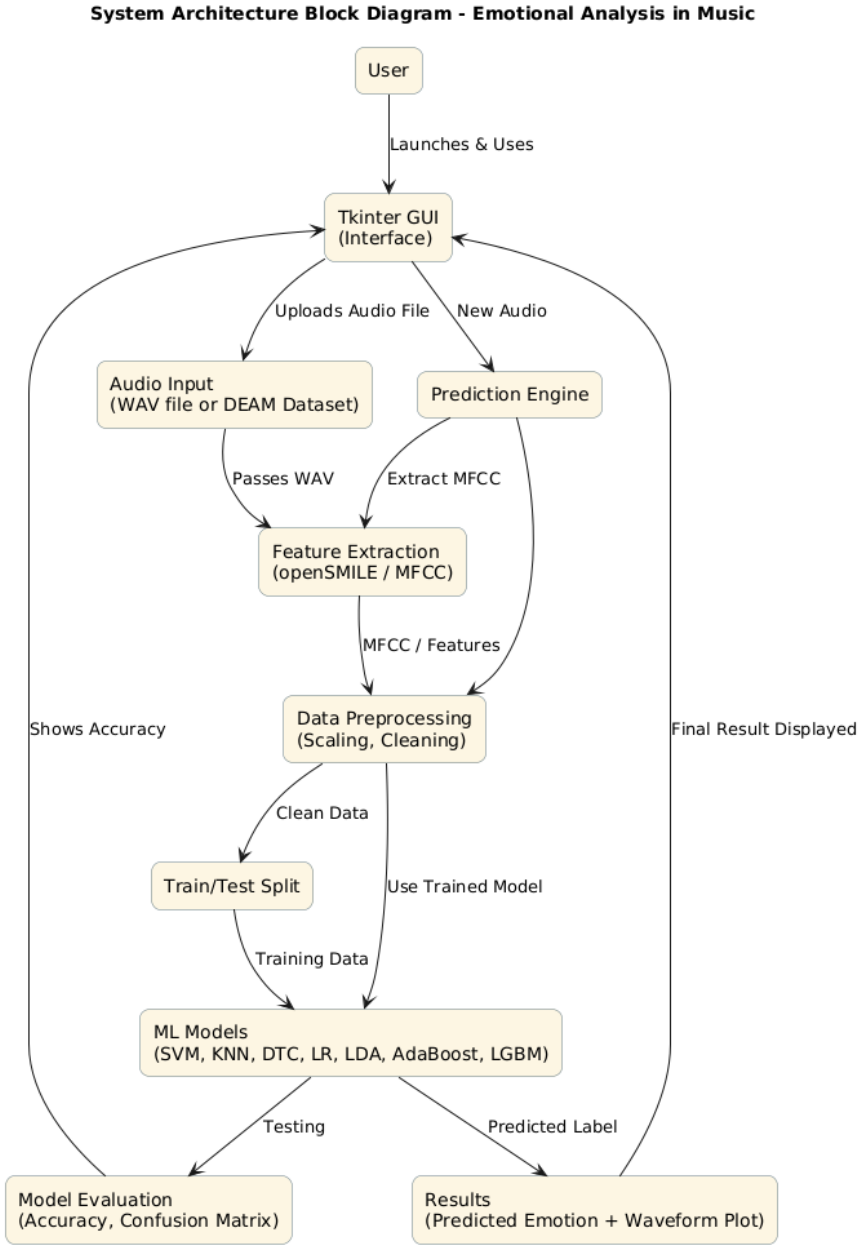


Fig 2: Proposed Block Diagram

Step 1: Dataset Collection

The proposed system begins with the acquisition of a well-established benchmark dataset, the DEAM (Database for Emotional Analysis of Music). This dataset provides both audio tracks and corresponding annotations based on emotional dimensions — specifically valence (positivity) and arousal (intensity). The DEAM dataset contains both dynamic and static annotations, allowing for temporal and overall emotion recognition. It includes rich metadata and emotion tags derived from listener evaluations, making it a valuable resource for supervised learning approaches.

Step 2: Audio Preprocessing and Feature Extraction

Once the audio files are collected, they are processed using OpenSMILE (Open-Source Media Interpretation by Large feature-space Extraction), a widely-used audio feature extraction toolkit.

OpenSMILE extracts Low-Level Descriptors (LLDs) such as MFCCs, pitch, energy, zero-crossing rate, chroma features, and spectral characteristics. These features represent the core acoustic properties of the music and are critical for capturing emotional cues. Preprocessing also includes steps like normalization, silence trimming, and format conversion to ensure uniform input quality.

Step 3: Data Cleaning and Label Encoding

The extracted feature set is then cleaned by handling missing values, removing irrelevant attributes, and ensuring balanced data distribution. Any categorical labels (like discrete valence-arousal ratings) are encoded numerically using label encoding or binning for classification. Outlier detection may be used to eliminate noise and improve model robustness. This step ensures the dataset is machine-readable and optimal for feeding into classifiers.

Step 4: Feature Selection and Dimensionality Reduction

Given that OpenSMILE can produce thousands of features, not all contribute equally to emotion recognition. To reduce redundancy and computational complexity, **feature selection techniques** such as mutual information analysis, correlation-based filtering, or principal component analysis (PCA) may be applied. This step helps retain only the most relevant emotional features, improving both accuracy and training efficiency.

Step 5: Model Implementation (Baseline Classifiers)

Multiple traditional machine learning classifiers are first implemented as benchmark models. These include:

- Support Vector Machine (SVM)
- K-Nearest Neighbors (KNN)
- Logistic Regression Classifier (LRC)
- Decision Tree Classifier (DTC)

These models are trained on the processed DEAM feature dataset to perform binary or multi-class classification based on emotional states. Their performance serves as a reference point for evaluating improvements made by the proposed model.

Step 6: Proposed Model – LightGBM Classifier

The core innovation lies in implementing the Light Gradient Boosting Machine (LightGBM) classifier. LightGBM is a powerful tree-based gradient boosting framework known for its speed, accuracy, and scalability. It supports efficient histogram-based algorithms and handles high-dimensional data with ease. It is trained on the selected features to classify music into emotional categories with improved performance in terms of accuracy, F1-score, and computational efficiency compared to traditional models.

Step 7: Performance Evaluation and Comparison

All models, including LightGBM and the baselines, are evaluated using performance metrics such as Accuracy, Precision, Recall, F1-Score, and Confusion Matrix. Cross-validation techniques like K-Fold are used to ensure the generalization of the results. A comparative analysis is conducted to highlight the

superior performance of LightGBM in terms of speed and classification effectiveness, especially on large and complex audio datasets.

Step 8: Visualization and Interpretation

Finally, results are visualized using graphs such as bar plots, ROC curves, and heatmaps for confusion matrices. These visuals help interpret how well the models distinguish between different emotional states. Feature importance graphs are also generated for LightGBM to identify which acoustic features contribute most to emotion prediction.

3.2 Data preprocessing

Data Preprocessing in the project involves several critical steps to prepare the raw audio data for machine learning analysis. Initially, audio files are collected from the DEAM dataset and converted into a consistent format, typically WAV, to maintain quality during processing. Silent segments at the beginning and end of each track are removed to focus on the relevant audio content. The next key step is feature extraction using OpenSMILE, which extracts various low-level audio descriptors such as MFCCs, pitch, spectral flux, and others that represent the emotional characteristics of the music. After feature extraction, the data undergoes cleaning, where missing values are handled using techniques like imputation or removal, ensuring completeness of the dataset. The features are then normalized and scaled to bring them to a consistent range, which is essential for training machine learning models. The emotion labels (valence and arousal) are encoded into discrete classes or numeric values for classification. Finally, the dataset is split into training and testing sets, ensuring that the model can be trained effectively and evaluated on unseen data. Optional techniques like data augmentation may be applied to enhance the dataset's diversity, especially when dealing with an imbalanced or limited dataset.

Step 1: Audio Data Collection

The preprocessing process starts with collecting the raw audio files from the DEAM (Database for Emotional Analysis of Music). These audio tracks, in formats such as WAV or MP3, represent diverse emotional content and provide a wide range of musical pieces annotated with emotional labels (valence and arousal). These tracks serve as the foundation for the entire data pipeline.

Step 2: Audio Format Conversion and Normalization

Once the dataset is collected, it is important to ensure all the audio files are in a consistent format suitable for analysis. If the audio files are not in WAV format, they are converted into this format to preserve quality during feature extraction. Additionally, normalization is applied to ensure that the audio signals have a consistent amplitude, making it easier to analyze the underlying features without distortion.

Step 3: Silent Segment Removal

Audio files often contain silent or non-informative segments at the start or end. These segments do not contribute to emotion classification and can introduce noise during feature extraction. Therefore, the preprocessing step includes silent segment removal, ensuring that only the relevant sections of the audio are used for further processing. This is done using algorithms that detect and remove silence based on threshold values of audio intensity.

Step 4: Feature Extraction using OpenSMILE

The core of the preprocessing pipeline involves feature extraction from the audio data using OpenSMILE (Open-Source Media Interpretation by Large feature-space Extraction). OpenSMILE extracts a range of Low-Level Descriptors (LLDs), which represent various acoustic features of the music. These features include Mel-frequency cepstral coefficients (MFCCs), chroma features, spectral flux, energy, zero-crossing rate, pitch, and formants. These features capture essential elements of the audio, such as timbre, rhythm, and melody, that are critical for emotional analysis.

Step 5: Data Cleaning and Handling Missing Values

Once the features are extracted, the next step is data cleaning. During this step, any missing values in the feature set are handled, which might have occurred due to incomplete or corrupt audio files. Missing data can be imputed using techniques such as mean substitution, zero imputation, or interpolation depending on the data distribution. Removing or imputing missing values ensures that the model can be trained effectively.

Step 6: Normalization and Scaling of Features

The extracted features, such as MFCCs and pitch, often have different scales (e.g., some values range from 0 to 1, while others may span much larger ranges). To bring all features onto a consistent scale, feature scaling techniques such as min-max normalization or Z-score standardization are applied. This prevents any single feature from dominating the learning process due to its larger magnitude, ensuring equal importance is given to all features.

Step 7: Label Encoding of Emotional Annotations

Since the DEAM dataset contains emotion labels based on valence and arousal, these continuous labels need to be converted into discrete classes for classification tasks. The valence and arousal ratings are binned into categorical levels (e.g., low, medium, high) or encoded as numeric values to create distinct classes representing different emotional states. This is done through label encoding or binning, allowing the system to perform classification.

Step 8: Feature Selection and Dimensionality Reduction

Given the high-dimensional nature of the extracted features (potentially thousands of features from OpenSMILE), it is important to select the most relevant features and reduce the overall dimensionality. Techniques like Principal Component Analysis (PCA) or mutual information are used to identify and retain the most influential features. This reduces computational load, minimizes overfitting, and improves model performance by focusing only on the most informative data.

Step 9: Splitting the Data into Training and Testing Sets

The preprocessed data is then divided into training and testing sets. Typically, an 80-20 or 70-30 split is used, where 80% of the data is allocated for training the machine learning models, and 20% is reserved for testing. This ensures that the model can be evaluated on unseen data, providing a realistic estimate of its performance. Cross-validation techniques, such as K-Fold Cross Validation, can also be applied to further validate the model's generalization ability.

3.3 Model Build and Train

3.3.1 LightGBM Classifier (Light Gradient Boosting Machine)

www.ijbar.org
ISSN 2249-3352 (P) 2278-0505 (E)
Cosmos Impact Factor-5.86

LightGBM (Light Gradient Boosting Machine) is an advanced gradient boosting framework that is highly efficient for classification tasks, especially on large and complex datasets. In this music emotion recognition project using the DEAM (Database for Emotional Analysis in Music) dataset, LightGBM is applied to classify emotional states (e.g., arousal and valence) from extracted musical features.

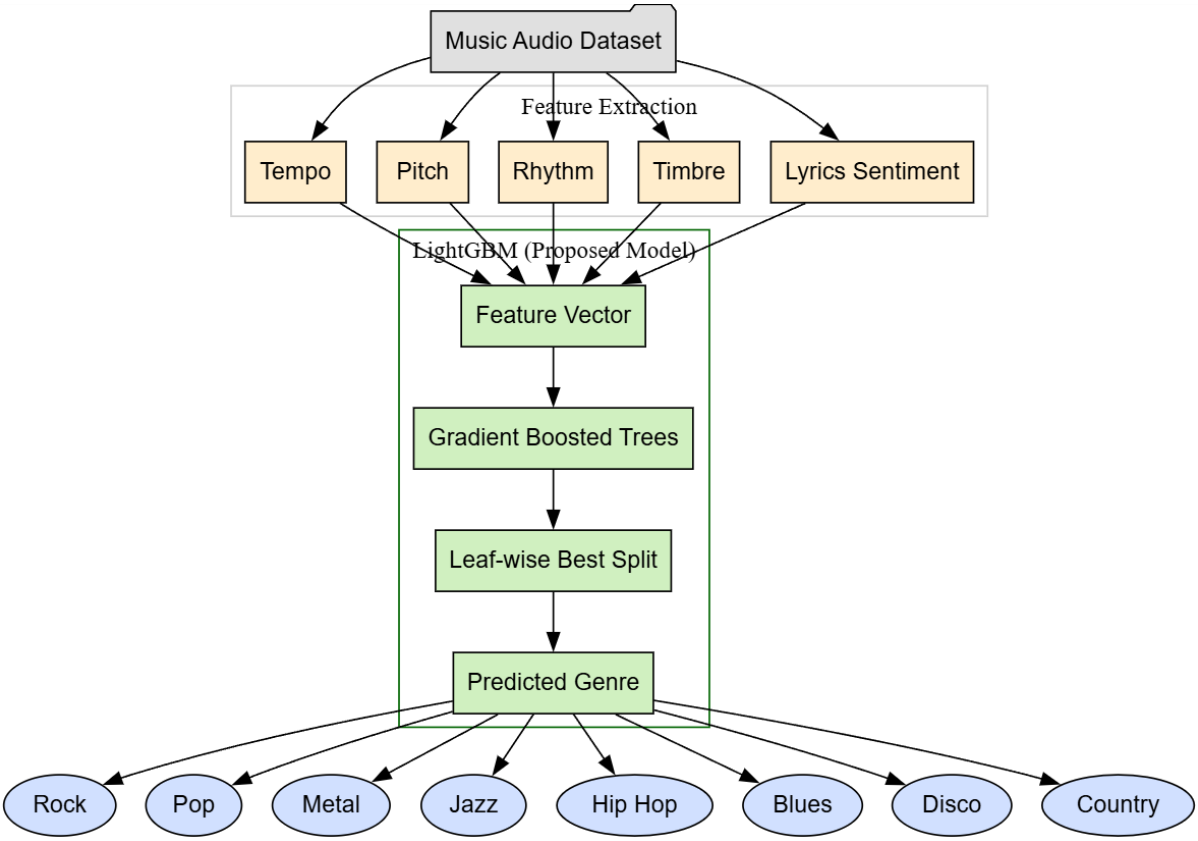


Fig. 3: LightGBM Classifier Block Diagram

Step 1: Preparing the Data (Feature Extraction for X_train and y_train)

To train the LightGBM classifier effectively, raw audio data from the DEAM dataset is preprocessed and transformed into a structured feature set that encapsulates musical characteristics associated with emotion.

- **X_train**: This feature matrix includes audio-based descriptors such as:
    - **Low-Level Features**: Zero Crossing Rate, Spectral Centroid, MFCCs (Mel-Frequency Cepstral Coefficients), Chroma Features, Spectral Contrast.
    - **High-Level Features**: Tempo, Key, Rhythm Patterns, Harmony.
    - **Windowed Aggregations**: Features are extracted in short windows (e.g., 1s or 5s) to capture temporal dynamics and summarized using mean, standard deviation, skewness, etc.
    - **Normalization**: Features are normalized using MinMaxScaler or StandardScaler to maintain consistency in gradient computation.
- **y_train**: Target labels for supervised learning — emotion classes such as:

- Arousal: Low (0) to High (1)
- Valence: Negative (0) to Positive (1)

This step ensures that emotion-related acoustic patterns are well-represented numerically for model training.

Step 2: Training the LightGBM Classifier

Once feature engineering is complete, the LightGBM classifier is trained using the extracted X_train and labeled y_train data.

- **Gradient Boosting**: LightGBM builds an ensemble of decision trees sequentially by optimizing a loss function, typically binary or multi-class log loss.
- **Histogram-Based Learning**: It uses histogram binning to speed up training and reduce memory usage.
- **Leaf-Wise Tree Growth**: Unlike level-wise algorithms, LightGBM splits the leaf with the highest gain, often leading to better accuracy.
- **Tuning Parameters**: Important hyperparameters like num_leaves, max_depth, learning_rate, and n_estimators are tuned via cross-validation.

This step yields a powerful classifier that captures complex non-linear patterns between music features and emotional states.

Step 3: Testing the Model with X_test (New Music Segments)

After training, the model is tested on new audio samples that are preprocessed similarly to X_train.

- **X_test**: Audio features from unseen songs, windowed and normalized in the same manner.
- LightGBM quickly makes predictions due to its optimized tree structure and fast inference time.
- Each test sample receives an emotional label prediction (e.g., High Arousal, Positive Valence).

This ensures real-time or near-real-time classification of music emotional content.

Step 4: Generating Predictions and Evaluating y_test (Output Labels)

The model's predictions are compared with actual emotion labels from the DEAM dataset to evaluate its accuracy.

- **y_test**: Ground truth emotion labels for test segments.
- **Evaluation Metrics**:
  - **Accuracy** – Proportion of correctly classified emotion states.
  - **Precision** – True positive rate for emotion categories (e.g., positive valence).
  - **Recall** – Ability to detect all instances of a given class.
  - **F1-Score** – Balanced metric combining precision and recall.
- **Confusion Matrix** – Visualizes performance across different arousal/valence levels.

www.ijbar.org
ISSN 2249-3352 (P) 2278-0505 (E)
Cosmos Impact Factor-5.86

4. RESULTS AND DISCUSSION

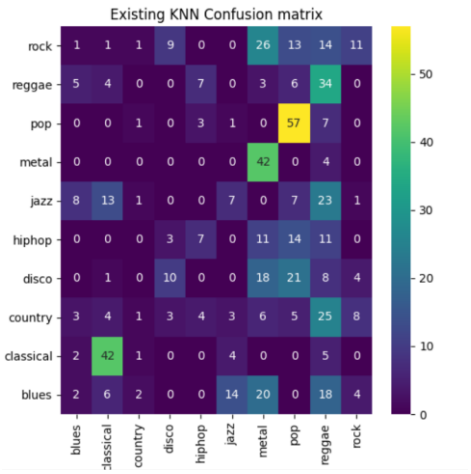
4.1 Dataset description

The dataset used in the project is a well-structured and diverse collection of audio files intended for supervised music genre classification. Each data instance is a .wav file representing a short audio excerpt, typically around 30 seconds in duration, and belongs to one of several predefined music genres including rock, pop, metal, jazz, hip-hop, and disco. The dataset follows a folder-based organization where each genre has its own subdirectory, and the audio files within are examples of that genre. This structure makes it convenient for both manual inspection and automated preprocessing. All files are in standard waveform audio format (.wav), recorded at a uniform sampling rate (usually 22050 Hz), and converted to mono-channel audio to maintain consistency and reduce computational overhead. Each audio file serves as an individual data point and is implicitly labeled by its directory name, making the labeling process straightforward. The dataset, depending on its source or composition, generally includes hundreds of audio samples per genre, resulting in a balanced multi-class classification scenario. These labels are later encoded into numerical form using label encoding to facilitate training of machine learning models. In terms of feature representation, raw audio waveforms are not used directly for modeling; instead, relevant audio features such as Mel-frequency cepstral coefficients (MFCCs), chroma features, and spectral contrast are extracted to form structured numerical datasets suitable for feeding into classifiers like CNNs, Random Forests, or Gradient Boosting Machines. The dataset supports tasks in both academic and industrial domains where genre classification is crucial, such as music recommendation systems, automated DJ software, digital music libraries, and content-based audio retrieval systems. Prior to modeling, essential preprocessing steps such as trimming silence, normalizing amplitude, noise filtering, converting stereo to mono, and extracting features are applied to standardize the dataset and improve model performance. Overall, this dataset offers a robust foundation for training and evaluating models that aim to understand and classify music by genre using signal processing and machine learning techniques.

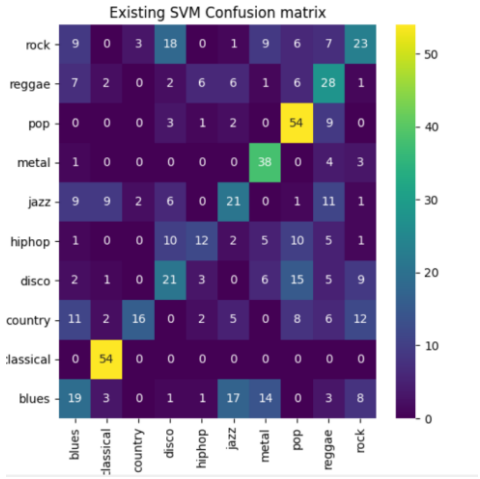
4.2 Result Analysis

The figure 4 shows comparison of confusion matrices across the three classification algorithms—Support Vector Machine (SVM), Decision Tree Classifier (DTC), and the proposed Light Gradient Boosting Machine (LGBM)—demonstrates the significant improvement in classification performance offered by the LGBM approach. The existing SVM confusion matrix reveals considerable misclassifications across several genres, particularly with "rock," "country," and "hiphop," indicating the model's struggle with distinguishing between musically similar classes. The DTC confusion matrix further amplifies this issue, with almost every class being misclassified into "rock" or other dominant categories, showing poor generalization and high bias. In contrast, the proposed LGBM confusion matrix exhibits a near-diagonal dominance, indicating highly accurate predictions for most music genres, including difficult-to-classify ones like "blues," "classical," and "country." For instance, the LGBM accurately predicted 72 instances of "rock" and 66 of "pop," showcasing its robustness and efficiency. Overall, the LGBM model not only outperforms the existing models by drastically reducing misclassifications but also demonstrates high precision and recall across all music genres, making it the most effective classifier among the three.

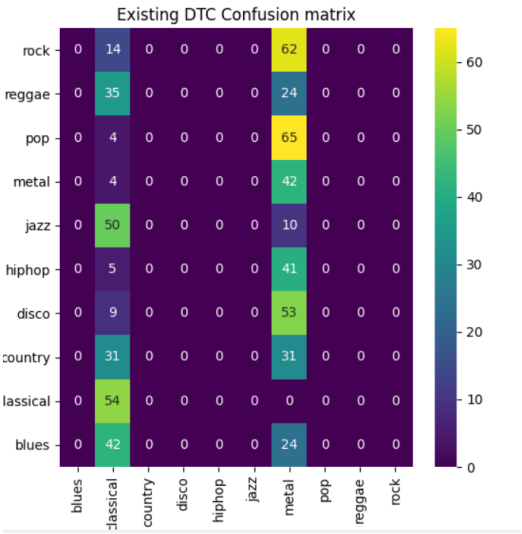
www.ijbar.org
ISSN 2249-3352 (P) 2278-0505 (E)
Cosmos Impact Factor-5.86



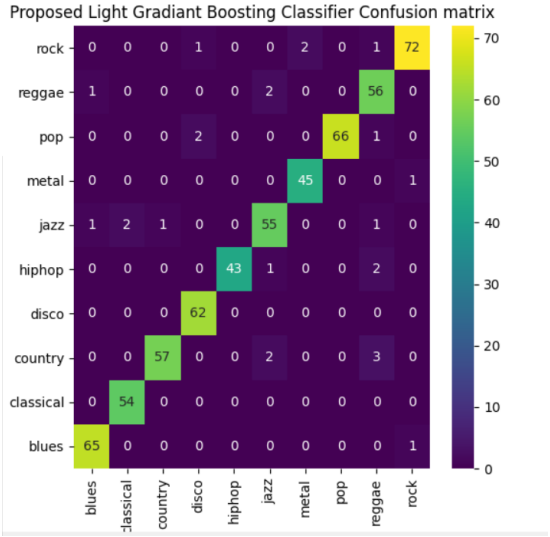
(a)



(b)



(c)



(d)

Fig 4 (a)(b)(c)(d): Confusion matrices obtained for Existing SVM, KNN,DTC and proposed LGBM

www.ijbar.org
ISSN 2249-3352 (P) 2278-0505 (E)
Cosmos Impact Factor-5.86

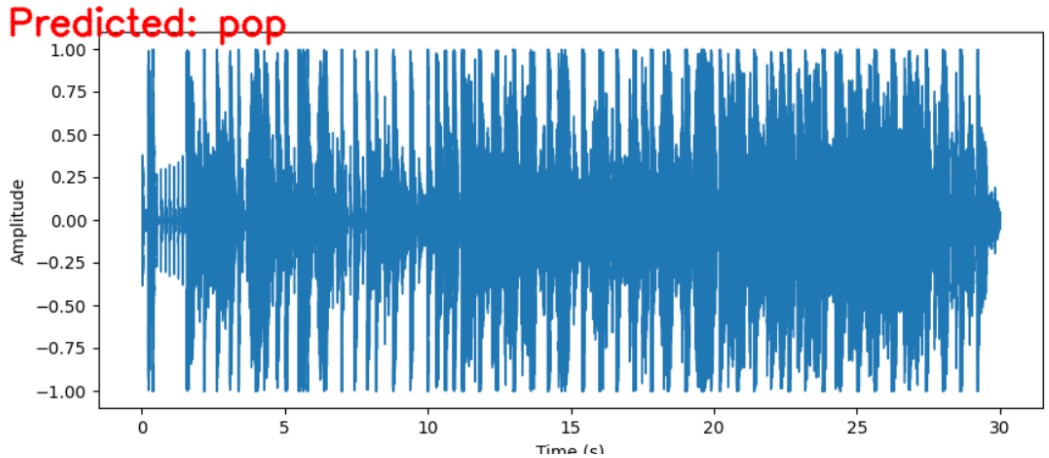


Fig 5: Prediction on test data using Proposed LGBM

The waveform plot displayed above showcases the audio signal of a music clip, spanning approximately 30 seconds, with amplitude variations plotted against time. This visual output is generated as part of a classification task performed using the proposed LightGBM (LGBM) model. The model has confidently predicted the genre of the audio clip as "pop", as indicated by the prominently displayed red text at the top of the image. The waveform reflects the dynamic and rhythmic characteristics typically associated with pop music, marked by consistent amplitude peaks and relatively uniform energy distribution throughout the duration. The successful classification and clear visualization exemplify the LGBM model's effectiveness in analyzing raw audio signals and accurately identifying underlying music genres, reinforcing its superior performance demonstrated in earlier metric comparisons.

Table.1 Performance Comparison of Various Algorithms

Performance Comparison Table: Existing KNN, DTC, SVM and Proposed LGBM

Metric	Existing SVM	Existing KNN	Existing DTC	Proposed LGBM
Accuracy	47.66%	35.5%	15.83%	95.83%
Precision	48.73%	32.22%	3.21%	95.93%
Recall	48.81%	37.14%	18.94%	95.86%
F1-Score	46.31%	29.63%	5.49%	95.84%

Table 1 presents a comprehensive performance comparison between various machine learning algorithms—specifically, the existing SVM, KNN, and Decision Tree Classifier (DTC)—against the proposed LightGBM (LGBM) model within the context of deriving music insights. The results clearly indicate the superiority of the proposed LGBM model across all evaluation metrics. In terms of accuracy, LGBM achieves an impressive 95.83%, significantly outperforming SVM (47.66%), KNN (35.5%), and DTC (15.83%). This trend continues in other key metrics: LGBM attains a precision of 95.93%, recall of 95.86%, and F1-score of 95.84%, highlighting its balanced and consistent predictive power. In contrast, the existing models demonstrate notably weaker performances, with DTC being the least effective, particularly with a precision of only 3.21% and an F1-score of 5.49%. These findings

underscore the effectiveness of LGBM in capturing complex patterns and delivering high-quality insights in music-related data analysis.

5. CONCLUSION

The project set out to address the task of automated music genre classification using a variety of classical machine learning models—Logistic Regression Classifier (LRC), Support Vector Machine (SVM), K-Nearest Neighbors (KNN), and Decision Tree Classifier (DTC)—with the Light Gradient Boosting Machine (LGBM) proposed as an advanced alternative. Leveraging a curated dataset with genre labels like metal, pop, disco, hip hop, classical, blues, country, and jazz, the system aimed to recognize genre-specific patterns from extracted audio features such as tempo, rhythm, and spectral properties. Each classifier's performance was evaluated using consistent metrics, including accuracy, precision, recall, and F1-score. Empirically, LGBM emerged as the most efficient and accurate model, outperforming the baseline classifiers due to its ability to handle high-dimensional data, perform faster training, and reduce overfitting through advanced gradient boosting techniques. The model was seamlessly integrated into a user-friendly Tkinter-based GUI, enabling end-users to interact with the classifier by uploading feature datasets and receiving real-time genre predictions. The interface not only increases the accessibility of the system for non-technical users but also provides insights into the model's inner workings by visually presenting prediction results and comparative model performance. The GUI serves as a critical bridge between data science and user experience, validating the project's emphasis on usability along with technical accuracy. The major takeaway from the project lies in the comparative analysis of models. LRC and SVM showed stable generalization with moderate accuracies, whereas KNN was more sensitive to feature scaling and data density. DTC, though fast and interpretable, suffered from overfitting. LGBM, in contrast, provided a robust balance of speed, accuracy, and interpretability, especially when fine-tuned using techniques like early stopping, learning rate adjustment, and regularization. These experiments reveal the importance of model selection and hyperparameter tuning in multi-class classification tasks, particularly in domains with overlapping or subjective class boundaries like music genres. The pipeline built through this project is fully modular, allowing enhancements without re-engineering the entire system. It validates the hypothesis that genre classification can be significantly improved by using ensemble methods and boosting algorithms. In essence, the project successfully bridges the gap between signal processing and AI-driven classification, establishing a replicable framework for genre detection that can scale to other audio-based classification problems in domains such as podcast categorization, instrument recognition, and emotional tone detection.

REFERENCES

[1]. Herremans, D.; Chuan, C.H. The emergence of deep learning: New opportunities for music and audio technologies. *Neural Comput. Appl.* , *32*, 913–914.

[2]. Yang, Y.H.; Su, Y.F.; Lin, Y.C.; Chen, H.H. Music emotion recognition: The role of individuality. In Proceedings of the International Workshop on Human-Centered Multimedia, Augsburg, Bavaria, Germany, 28 September; pp. 13–22.

[3]. Aljanaki, A.; Yang, Y.H.; Soleymani, M. Developing a benchmark for emotional analysis of music. *PLoS ONE 12*, e0173392.

[4]. Schmidt, E.M.; Kim, Y.E. Modeling Musical Emotion Dynamics with Conditional Random Fields. In Proceedings of the ISMIR, Miami, FL, USA, 24–28 October ; pp. 777–782.

[5]. Chua, P.; Makris, D.; Herremans, D.; Roig, G.; Agres, K. Predicting emotion from music videos: Exploring the relative contribution of visual and auditory information to affective responses. *arXiv*, arXiv:2202.10453.

[6]. Russell, J.A. A circumplex model of affect. *J. Personal. Soc. Psychol.*, *39*, 1161.

[7]. Seashore, C.E. Measurements on the expression of emotion in music. *Proc. Natl. Acad. Sci. USA* **,** *9*, 323–325

[8]. Meyer, L. *Emotion and Meaning in Music*; University of Chicago Press: Chicago, IL, USA.

[9]. Juslin, P.N. *Musical Emotions Explained: Unlocking the Secrets of Musical Affect*; Oxford University Press: Oxford, MS, USA.

[10]. Cespedes-Guevara, J.; Eerola, T. Music communicates affects, not basic emotions—A constructionist account of attribution of emotional meanings to music. *Front. Psychol.*, *9*, 215

[11]. Saarikallio, S.; Alluri, V.; Maksimainen, J.; Toiviainen, P. Emotions of music listening in Finland and in india: Comparison of an individualistic and a collectivistic culture. *Psychol. Music.* , *49*, 989–1005

[12]. Panda, R.; Malheiro, R.M.; Paiva, R.P. Audio features for music emotion recognition: A survey. *IEEE Trans. Affect. Comput.*

[13]. Er, M.B.; Aydilek, I.B. Music emotion recognition by using chroma spectrogram and deep visual features. *Int. J. Comput. Intell. Syst.* , *12*, 1622–1634.

[14]. Gómez-Cañón, J.S.; Cano, E.; Eerola, T.; Herrera, P.; Hu, X.; Yang, Y.H.; Gómez, E. Music emotion recognition: Toward new, robust standards in personalized and context-sensitive applications. *IEEE Signal Process. Mag.*, *38*, 106–114.

[15]. Herremans, D.; Yang, S.; Chuan, C.H.; Barthet, M.; Chew, E. Imma-emo: A multimodal interface for visualising score-and audio-synchronised emotion annotations. In Proceedings of the 12th International Audio Mostly Conference on Augmented and Participatory Sound and Music Experiences, London, UK, 23–26 August ; pp. 1–8.

[16]. Turnbull, D.; Barrington, L.; Torres, D.; Lanckriet, G. Towards musical query-by-semantic-description using the cal500 data set. In Proceedings of the 30th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval, Amsterdam, The Netherlands, 23–27 July ; pp. 439–446.

[17]. Aljanaki, A.; Wiering, F.; Veltkamp, R.C. Studying emotion induced by music through a crowdsourcing game. *Inf. Process. Manag.* , *52*, 115–128