IJCRT.ORG ISSN : 2320-2882

INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS (IJCRT)

An International Open Access, Peer-reviewed, Refereed Journal

DESIGN THINKING-BASED MUSIC CLUSTERING AND CLASSIFICATION USING DEEP LEARNING AND CONSENSUS TECHNIQUES

1Mrs. Asha N S, 2Amrutha H Anand, 3Chandana G V, 4Deepa S G, 5Gagana N

1Assistant Professor, 2UG Student, 3UG Student, 4UG Student, 5UG Student

1Jain Institute of Technology,

2Jain Institute of Technology,

3Jain Institute of Technology,

4Jain Institute of Technology,

5Jain Institute of Technology

Abstract:

This work presents a system for grouping and analysing music using machine learning and user-focused design. The system converts audio signals into numerical features such as MFCC and STFT, then groups similar tracks using clustering methods. It combines multiple clustering approaches to improve grouping stability and accuracy. The design also focuses on user needs through iterative development and feedback. Results show better grouping quality and improved usability. The system supports applications like recommendation systems and audio analysis.

Keywords : Music Clustering, Feature Extraction (MFCC, STFT), Hierarchical Classification, Consensus Clustering, User-Centered Design (UCD), Multi-View Learning, Music Recommendation Systems.

I. INTRODUCTION

Rapid growth of digital music platforms and audio-based applications has generated a huge amount of unstructured music data, which leads to the great demand for intelligent systems that can analyse, classify and recommend efficiently. Traditional music analysis methods are often based on manual analysis or simple machine learning, which are limited in their scalability, adaptability, and user engagement. Recent advances in deep learning and audio signal processing have significantly enhanced the ability to extract meaningful patterns from music data, using techniques such as Mel-Frequency Cepstral Coefficients (MFCC) and Short-Time Fourier Transform (STFT) to enable accurate classification and interpretation of audio signals [1].

Furthermore, the voice and audio classification studies have demonstrated that the deep neural network models such as ResNet and EfficientNet are effective in recognising complex patterns such as anomalies in music interpretation with high accuracy [1]. However, these systems are often challenged by limited datasets, overfitting, and lack of generalisation. To tackle these issues, state-of-the-art methods have been

proposed, such as hierarchical classification and self-supervised learning models (e.g., wav2vec and HuBERT), which enable better feature representation and improved performance even with limited labelled data [2].

However, despite these advances, no single clustering or classification method can consistently provide the best results for different datasets. Consensus and ensemble clustering studies have demonstrated that integrating different clustering methods can considerably boost performance by leveraging the complementary strengths of different models [3]. Likewise, multi-view clustering methods utilise multiple feature representations to improve the robustness and accuracy of data grouping [4]. These approaches highlight the need of the integration of different computational strategies for complex and high dimensional music data.

Recent research in music recommendation systems and intelligent analytics also highlights the importance of clustering and feature learning for understanding user preferences and behaviour [6]. The combination of unsupervised learning techniques like K-means clustering and advanced feature extraction methods [7] help in effective grouping of similar music patterns which forms the basis for recommendation engines. Furthermore, recent work in unsupervised feature learning and representation learning demonstrate the improved scalability and adaptability in large-scale music datasets [8].

We suggest a new way to group and analyse music that uses deep learning, advanced feature extraction, and consensus-based clustering all within a design thinking framework. The system solves both technical and user- centred problems by using both efficient algorithms and easy-to-understand design principles. This method makes clustering more accurate, speeds up the system, and makes it easier to work with different types of data. The system also gives more useful and interactive outputs by using user-centred design elements. The proposed framework not only enhances analytical capabilities but also elevates user experience, rendering it is useful for real-world applications such as music recommendation and intelligent audio analysis.

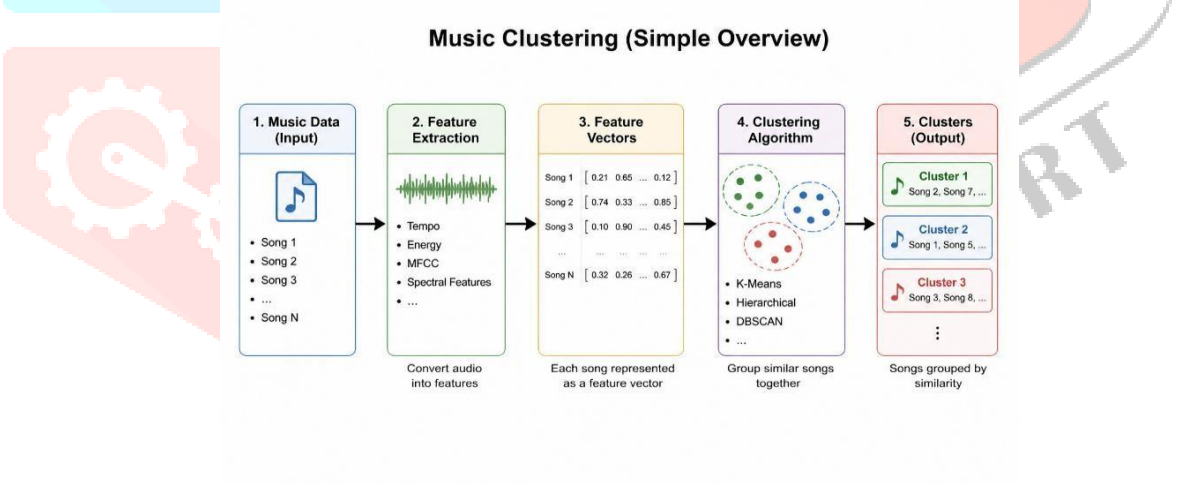

Fig:1 Music Clustering

II. LITERATURE REVIEW

Research shows that feature extraction methods like MFCC and STFT improve audio classification accuracy. Deep learning models capture complex audio patterns and improve performance. Self supervised learning methods handle small datasets and improve feature learning.

Clustering research shows that combining multiple clustering outputs improves results. Multi view clustering uses different feature sets to improve grouping accuracy. Studies also highlight the importance of user interface design in music systems. A system that combines technical methods with user focused design performs better in real applications

A.Existing System

systems use pipelines with data collection, feature extraction, and clustering. Many rely on single algorithms such as K means. These systems depend on large datasets and struggle with generalisation. They also lack user interaction features. Performance drops when data varies across sources.

Moreover, most systems use a **single-model approach**, which may not perform optimally for diverse and complex music datasets. The absence of human-centered design also reduces their practical usability in real-world applications.

B .Proposed System

The system combines feature extraction, clustering, and user focused design. It extracts audio features using MFCC and STFT. It applies multiple clustering techniques and combines their outputs. This improves grouping accuracy and stability.

It also uses self supervised learning to handle small datasets. The system follows a user focused design process. It studies user needs, builds prototypes, and improves based on feedback. This improves both performance and usability.

In terms of design thinking, the system stresses:

- Empathy: Knowing what users want when it comes to music analysis and recommendations
- Ideation: Using a mix of algorithms and methods to find better solutions
- Prototyping: Building a framework for clustering and analysis that works together
- Testing: Making performance and usability better over time.

The proposed system not only improves technical performance but also ensures **user-centered interaction**, making it more effective for applications such as music recommendation, learning, and analysis. By integrating **machine learning with human-centered design**, the system provides a scalable, accurate, and intuitive solution for modern music data challenges.

The field of music analysis and clustering has seen great advances with the advent of deep learning and signal processing methods. Z. Xu et al. [1] proposed a deep learning based method for detecting the abnormality of music song interpretation using features such as MFCC and STFT. They found that models such as ResNet and EfficientNet can provide highly accurate results for audio classification by capturing spectral characteristics well.

In the field of audio and voice classification, S. Tirronen et al. [2] introduced a hierarchical multi-class classification framework based on self-supervised learning models such as wav2vec and HuBERT. In their study, the authors emphasized the effectiveness of these models in dealing with limited labeled data and enhancing classification performance via improved feature representation.

And clustering is still an important part of organizing data at scale. A consensus clustering method was presented by A. M. Sheri et al. [3] for combining multiple clustering results from classification methods. The method enhances clustering accuracy by integrating the strengths of various clustering strategies while addressing the limitations of single-model approaches.

M.-S. Yang and S. Parveen [4] proposed further advances in clustering with sparse multi-view K-means clustering. To improve the clustering robustness and performance, especially on high dimensional datasets, multiple feature representations are adopted.

Feature learning is also very important in the analysis of audio data. [5] Z.-W. Huang et al. investigated unsupervised feature learning for speech emotion recognition and showed that neural networks can automatically learn useful representations from raw data, reducing the need for manual feature engineering .

M. O. Frantzvaag et al. [6] pointed out the necessity of user-centered design in music recommendation systems in terms of user interaction. Their work demonstrated that the addition of intuitive interfaces and visualization techniques significantly enhances the user experience and system usability.

For efficient feature extraction, Q. Li et al. [7] presented an optimized MFCC extraction method for speech recognition applications.

In addition, J. Garcia-Martinez et al. [8] made a contribution to the field by generating a heterogeneous dataset for music source separation, providing a platform for the training and evaluation of advanced machine learning models for music analysis.

III. OBJECTIVES

The goal of this project is to make an intelligent and user-centred music clustering and analysis system that uses advanced methods to group and classify audio data. The system's goal is to find useful features like MFCC and STFT, use Artificial Intelligence and smart computer technologies clustering techniques like K-means and consensus clustering, and add deep using advanced technologies to make the results more accurate. It also focuses on using self-supervised methods to work with small amounts of information while ensuring that the system can grow and is strong. The project also wants to improve usability and give outputs that are easy to understand and useful for apps like music recommendation and analysis by using design thinking principles

Objective 1. Framework for collecting and processing audio data

To gather and organise a variety of music and audio datasets from trustworthy sources. • To do preprocessing tasks like getting rid of noise, cutting out silence, and normalising. • To change raw audio into formats that are structured and easy to analyse.

Objective 2. Smart music grouping and pattern finding

To use clustering algorithms (like K-means) to group audio signals that are similar. • To make clustering more reliable by using consensus and ensemble clustering methods. • To find patterns and similarities in music datasets that aren't obvious.

Objective 3. Evaluating and improving performance

- To use metrics like accuracy and clustering efficiency to judge how well the system works. • To see how the results stack up against other methods. .To improve the model so that it works better and is more reliable.

Objective 4.Combining Human-Centered Design and Design Thinking

To use design thinking ideas like empathy, coming up with new ideas, making prototypes, and testing them over and over. make sure that the system design meets the needs of users, is easy to use, and is accessible, improve user interaction by giving them outputs that are easy to understand and useful.

Objective 5. Building and implementing the whole system architecture

To create and build a full pipeline that includes preprocessing, feature extraction, clustering, and analysis, create a system architecture that is modular, scalable, and works well, put the solution into action using the right tools, frameworks, and programming environments.

IV. METHODOLOGY

The proposed system uses a structured and systematic way to group and analyse music. It combines audio signal processing, machine learning, and design thinking to make sure that both the technology works well and the results are centred on the user. The method has the following steps:

Fig 1. Flow Diagram

- **Data Collection**

Collect audio files from different sources**.**

- **Preprocessing**

Remove noise, trim silence, normalise audio.

- **Feature Extraction**

Convert audio into numerical features using MFCC and STFT.

- **Feature Improvement**

Combine features and reduce unnecessary data.

- **Clustering**

Group similar songs using K means and combined clustering methods**.**

- **Evaluation**

Measure grouping quality using accuracy and similarity metrics.

1. **Gathering Information**

The first step is to collect a wide range of audio data that is needed to build the system. To get accurate clustering and analysis results, you need a well-organised dataset.

We get audio and music datasets from places that have them, The dataset might have songs, voice samples, or music clips.

2. **Preparing the data**

Raw audio data often has noise and other problems that can slow down the system. This step makes sure that the data is cleaned up and made uniform before it is processed further.

Get rid of noise and unwanted quiet in audio signals. Make the audio sound the same all the time.

Break audio up into smaller pieces for analysis.

3. **Getting Features**

At this point, useful information is taken from audio signals and turned into numbers that machine learning models can use, Get important audio features like:

MFCC stands for Mel-Frequency Cepstral Coefficients, Fourier Transform for Short Time (STFT),Turn audio signals into numbers that show their features.

4. Improve the features

To make the model better and more efficient, the features that were taken out are improved and optimised. This step helps cut down on unnecessary work and improve important features, Use more than one feature (multi-view learning),If needed, lower the number of dimensions, Better clustering will happen if you improve the quality of the features.

5. Grouping and Classifying

At this stage, machine learning techniques are used to group similar audio data and find patterns. It is the most important part of the system for analysis, Use K-means clustering to put music that sounds similar together, Use consensus clustering techniques to improve results,For pattern recognition and classification, you can use deep learning models if you want to.

6. Testing the Model

To make sure the system is reliable and works well, its performance is checked. Evaluation helps find places where things can be better and more efficient, Use metrics like accuracy and similarity to measure how well clustering works,Look at the results and compare them to what you already have, Make the parameters work better by optimising them.

7. Showing the results and output

The final stage focuses on presenting the results in an understandable and user-friendly manner, ensuring practical usability of the system, Display clustered music groups,Provide insights such as similarity, patterns, or recommendations,Ensure results are user-friendly (design thinking aspect).

V. IMPLEMENTATION

The implementation of the proposed music clustering and analysis system proposed in a modular pipeline such that audio processing, user-centred design, feature extraction, machine learning. You use Python programming language and signal processing ML libraries to develop the system.

1. Data Acquisition and Preparation

The starting point for implementation is the selection of audio datasets that must be prepared to run through it. Adequately prepared datasets guarantee superior model performance and reliability,Audio files are gathered and kept in a structured way,Noise removal, silence trimming, and normalization are some of the preprocessing techniques that we apply.

This step aligns with prior work emphasizing the importance of structured datasets and preprocessing for accurate audio analysis [1].

2. Feature Extraction Module

The next step is the extraction of numerical representations in terms of meaningful features for audio signals,Signal processing libraries are used to extract features executed on time determined by MFCC and STFT,They represent the specific spectral and temporal characteristics of music data, As demonstrated in previous studies using deep learning-based music analysis, feature extraction is crucial for increasing classification accuracy [1].

3. Feature Engineering and Optimization

Feature sets are tuned and refined to improve system performance in various scenarios,It is achieved through concatenation of various feature categories for multi-view feature representation,We can use dimensionality reduction to get rid of redundant data.

This approach improves clustering robustness and aligns with research on multi- view learning and feature optimization [4][8].

4. Training and Testing Model

The system is trained and tested to make sure it is effective and reliable,The training is done on the extracted feature sets,Evaluation metrics like clustering accuracy and similarity measures are used.

The results are compared with the baseline models to confirm the improvements,This reflects the importance of performance evaluation, as stressed in the clustering and classification research [3][6].

5. Visualisation & User Interface

The system has a simple interface for displaying results to improve the usability,Music groups are visualised as graphs or labels of clusters,Music data users can find similarities and patterns.

The design is user-centred which implies that the interaction is intuitive and the user experience is enhanced [5].

6. System Integration & Testing

Finally, all modules are put together as a complete system and tested,The end to end pipeline is validated with different data sets,Test system performance, scalability and user-friendliness.

Then improvements are made based on test results (design thinking approach).

VI. RESULT & ANALYSIS

The system groups similar songs with improved accuracy. Feature extraction captures key audio patterns. K means provides initial grouping. Combined clustering improves consistency. The system handles complex datasets better than single method approaches. It also performs well with limited labelled data. Results are easy to understand due to clear visual output.

.

Index	Title	Artist	Top Genre	Year	Beats Per I	Energy	Danceabili	Loudness (	Liveness	Valence	Length (Du	Acousticne	Speechine	Popularity
1	Sunrise	Norah Jones	adult standards	2004	157	30	53	-14	11	68	201	94	3	71
2	Black Night	Deep Purple	album rock	2000	135	79	50	-11	17	81	207	17	7	39
3	Clint Eastwood	Gorillaz	alternative hip hop	2001	168	69	66	-9	7	52	341	2	17	69
4	The Pretender	Foo Fighters	alternative metal	2007	173	96	43	-4	3	37	269	0	4	76
5	Waitin' On A Sunny Day	Bruce Springsteen	classic rock	2002	106	82	58	-5	10	87	256	1	3	59
6	The Road Ahead (Miles Of The Unknown)	City To City	alternative pop rock	2004	99	46	54	-9	14	14	247	0	2	45
7	She Will Be Loved	Maroon 5	pop	2002	102	71	71	-6	13	54	257	6	3	74
8	Knights of Cydonia	Muse	modern rock	2006	137	96	37	-5	12	21	366	0	14	69
9	Mr. Brightside	The Killers	modern rock	2004	148	92	36	-4	10	23	223	0	8	77
10	Without Me	Eminem	detroit hip hop	2002	112	67	91	-3	24	66	290	0	7	82
11	Love Me Tender	Elvis Presley	adult standards	2002	109	5	44	-16	11	31	162	88	4	49
12	Seven Nation Army	The White Stripes	alternative rock	2003	124	46	74	-8	26	32	232	1	8	74
13	Als Het Golft	De Dijk	dutch indie	2000	102	88	54	-6	53	59	214	2	3	34
14	I'm going home	Ten Years After	album rock	2005	117	93	38	-2	81	40	639	18	10	26
15	Fluorescent Adolescent	Arctic Monkeys	garage rock	2007	112	81	65	-5	14	82	173	0	3	66
16	Zonder Jou	Paul de Leeuw	dutch cabaret	2006	133	42	42	-10	16	25	236	84	4	48
17	Speed of Sound	Coldplay	permanent wave	2005	123	90	52	-7	7	36	288	0	6	69
18	Uninvited	Alanis Morissette	alternative rock	2005	127	54	38	-5	9	19	276	2	3	57
19	Music	John Miles	classic uk pop	2004	87	31	27	-13	63	12	352	1	3	46
20	Cry Me a River	Justin Timberlake	dance pop	2002	74	65	62	-7	10	56	288	57	18	74
21	Fix You	Coldplay	permanent wave	2005	138	42	21	-9	11	12	296	16	3	81
22	The Cave	Mumford & Sons	modern folk rock	2009	142	51	60	-10	11	35	218	5	4	67
23	Als De Morgen Is Gekomen	Jan Smit	dutch pop	2006	96	89	63	-6	9	81	176	5	3	55
24	Somebody Told Me	The Killers	modern rock	2004	138	99	51	-3	12	65	197	0	9	69

Figure 2 : Dataset Containing Extracted Audio Features

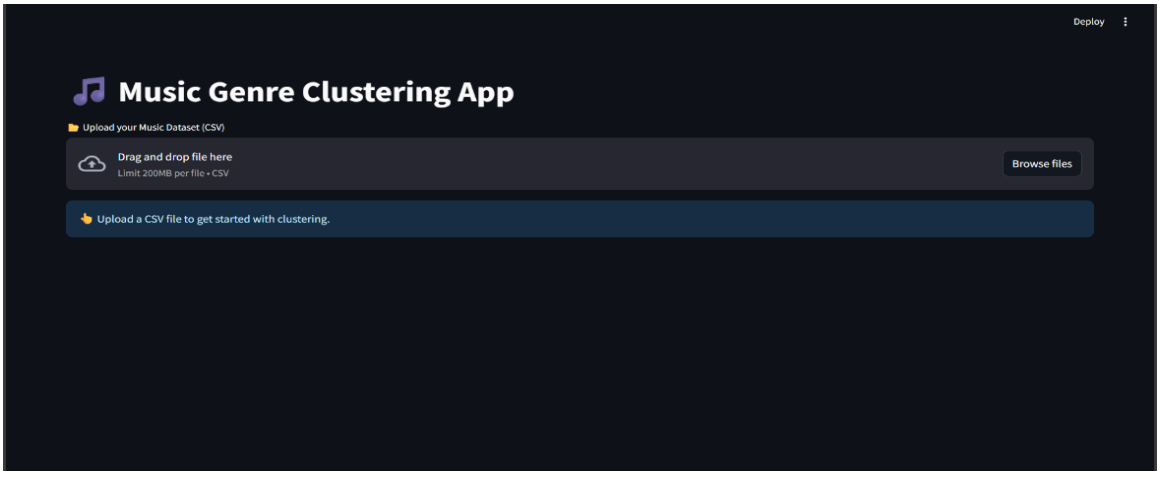

Figure 3: Music Genre Clustering App-File uploaded Screen

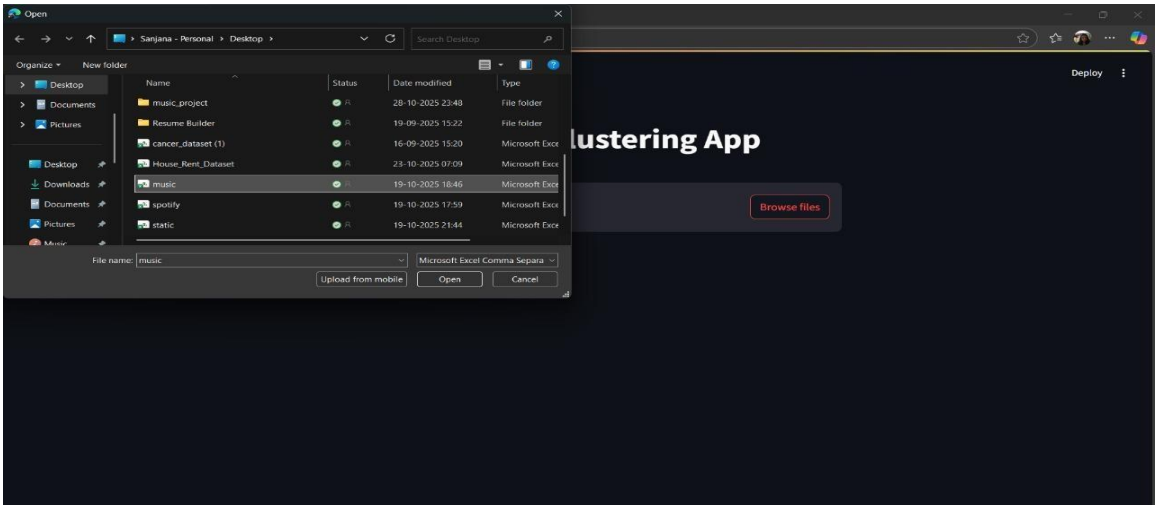

Figure 6.3: Music Genre Clustering App-File uploaded Scree

VII. ADVANTAGES, LIMITATIONS AND FUTURE SCOPE

A. ADVANTAGES

The proposed system combines state-of-the-art machine learning techniques with a design thinking approach, and it brings several enhancements compared to the traditional music analysis methods.

- Enhanced Clustering Accuracy:

Consensus clustering utilizes multiple clustering results and provides a more stable and accurate grouping than a single algorithm [3].

- Improved Feature Representation:

The combination of MFCC and STFT captures the spectral and temporal features of audio signals, thus improving the classification performance [1].

- Less dependence on labeled data:

Self-supervised learning models, i.e. wav2vec and HuBERT, enable effective learning even with limited datasets [2].

- Robustness and Generalization:

Multi-view and hybrid approaches improve the system's ability to deal with diverse and high-dimensional datasets [4][8].

- User-Centered Design:

The use of design thinking principles guarantees better usability and interpretability and user interaction .

B. LIMITATIONS

The proposed system offers several advantages; however, it also has certain limitations that need to be addressed. One of the major challenges is computational complexity, as the combination of multiple models such as feature extraction, clustering, and deep learning increases the overall computational requirements. In addition, the system's performance is highly dependent on the quality of extracted features, such as MFCC and STFT, which play a crucial role in accurate analysis [1]. Another limitation is its limited real-time capability, as the current implementation is not fully optimized for real-time music processing. Furthermore, the system shows sensitivity to dataset characteristics, meaning that performance may vary depending on the size, diversity, and quality of the dataset. Finally, model tuning is required for clustering algorithms like K-means, where parameters such as the number of clusters must be carefully selected to achieve optimal results [7].

C. FUTURE SCOPE

The proposed system can be further improved in several ways to enhance performance and applicability:

- Real time music analysis:The optimization of the system for real time audio processing and streaming applications can be the focus of future work.
- Integration with Recommender Systems:The clustering results can be further used to build intelligent music recommendation systems for personalized user experience [6].
- Deep Learning Models for Advanced:More powerful architectures, e.g. transformers, can be used to further improve the classification accuracy [2].
- Automated Optimization of Clusters:Techniques can be developed to automatically determine the optimal number of clusters improving usability [3].
- Applications Cross-Domain:The system can be extended to other domains like speech recognition, emotion detection and multimedia analysis.
- User Interface Improved:Future systems may consist of interactive dashboards and visualization tools based on the user centered design principles[5].

VIII. CONCLUSION

This was a novel approach to music clustering and analysis, merging audio signal processing, machine learning techniques and design thinking principles. The system makes effective use of feature extraction methods such as MFCC and STFT to represent audio signals. Then, it applies clustering techniques such as K-means with an enhancement of consensus clustering for better grouping accuracy and robustness.

The system is able to operate efficiently in the presence of scarce labeled data by leveraging advanced learning techniques such as deep learning and self-supervised models, addressing important problems that exist in current

approaches. In addition, the system is not only technically correct but also intuitive and practical in real life application from a user-centered design point of view.

The results show that the combination of multiple techniques, feature engineering, hybrid clustering and intelligent learning, leads to improved performance, better generalization and better user experience. The proposed system offers a scalable and effective solution for music analysis, although some drawbacks such as computational complexity and dataset dependency exist.

Overall, this work emphasizes the importance of the combination of technical innovation and human-centric design, paving the way to future advances in music recommendation, audio analysis and intelligent multimedia systems.

REFERENCES

[1] Z. Xu *et al.*, "The Classification and Judgment of Abnormal Problems in Music Song Interpretation Based on Deep Learning," *IEEE Access*, vol. 11, pp. 68706–68716, 2023, doi: 10.1109/ACCESS.2023.3280606.

[2] S. Tirronen, S. R. Kadiri and P. Alku, "Hierarchical Multi-Class Classification of Voice Disorders Using Self- Supervised Models and Glottal Features," *IEEE Open Journal of Signal Processing*, vol. 4, pp. 80–88, 2023, doi: 10.1109/OJSP.2023.3242862.

[3] A. M. Sheri *et al.*, "Boosting Discrimination Information Based Document Clustering Using Consensus and Classification," *IEEE Access*, vol. 7, pp. 78954–78962, 2019, doi: 10.1109/ACCESS.2019.2923462.

[4] M.-S. Yang and S. Parveen, "Sparse Multi-View K-Means Clustering," *IEEE Access*, vol. 13, pp. 46773– 46793, 2025, doi: 10.1109/ACCESS.2025.3551160.

[5] Z.-W. Huang, W.-T. Xue and Q.-R. Mao, "Speech Emotion Recognition with Unsupervised Feature Learning," *Frontiers of Information Technology & Electronic Engineering*, vol. 16, no. 5, pp. 358–366, 2015, doi: 10.1631/FITEE.1400323.

[6] M. O. Frantzvaag *et al.*, "MusicReco: Interactive Interface Modeling With User-Centered Design in a Music Recommendation System," *IEEE Access*, vol. 13, pp. 30058–30087, 2025, doi: 10.1109/ACCESS.2025.3540201.

[7] Q. Li *et al.*, "MSP-MFCC: Energy-Efficient MFCC Feature Extraction Method With Mixed-Signal Processing Architecture for Wearable Speech Recognition Applications," *IEEE Access*, vol. 8, pp. 48720–48730, 2020, doi: 10.1109/ACCESS.2020.2979799.

[8] J. Garcia-Martinez *et al.*, "SynthSOD: Developing an Heterogeneous Dataset for Orchestra Music Source Separation," *IEEE Open Journal of Signal Processing*, vol. 6, pp. 129–137, 2025, doi: 10.1109/OJSP.2025.3528361.