Unsupervised Music Genre Discovery

Using Audio Feature Learning

Department of Computer Science and Engineering

National Institute of Technology Hamirpur

Machine Learning (CS-652)

Semester VII

Presented by: Anirudh Sharma (22dcs002)

Presented to: Dr. Kamlesh Datta

Presentation Outline

Introduction

The exponential growth of digital music platforms has created massive repositories of unlabeled audio data, making manual organization infeasible at scale. This project presents a comprehensive comparative analysis of **four unsupervised learning algorithms**—K-Means, K-Medoids, Gaussian Mixture Models (GMM), and Spectral Clustering.

We evaluate these algorithms across datasets ranging from 500 to 25,000 samples, incorporating robust preprocessing (outlier detection, StandardScaler, PCA) and comprehensive evaluation using six performance metrics. High-dimensional clusters are visualized using t-SNE to analyze genre separation and cohesion.

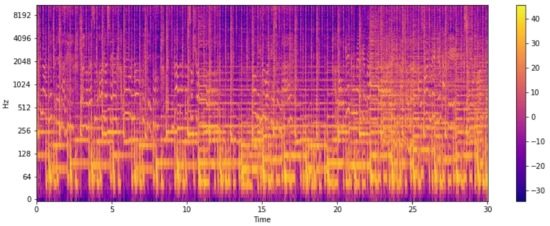


Figure: Audio waveform and spectrogram representation

What are Music Genres?

Definition

- Musical categories based on shared characteristics
- Defined by instrumentation, rhythm, harmony, and cultural context
- Evolve over time and across cultures

GTZAN Genre Labels (Our Baseline)

- We use **10 genre clusters** from GTZAN:
  - Blues, Classical, Country
  - Disco, Hip-hop, Jazz
  - Metal, Pop, Reggae, Rock
- Rock (subgenre): Hard Rock, Punk Rock, Progressive Rock
- Hip-Hop: Trap, Boom Bap, Lo-Fi

Genre Subtypes & Complexity

- Over 1,000+ documented subgenres
- Hierarchical relationships (parent-child)
- Genre fusion and cross-pollination

Key Challenges

- **Subjectivity**: Genre labels vary across listeners
- **Overlap**: Songs span multiple genres
- **Evolution**: Genres constantly change
- **Ambiguity**: Fuzzy boundaries between styles
- **Scale**: Millions of unlabeled tracks

Recent Advances in Music Genre Analysis

Study	Year	Method	Key Finding
Singh et al.	2024	Unsupervised Raga discovery	Novel class identification in Indian music
Kumar et al.	2024	K-Means clustering	Effective for recommendation systems
Ma et al.	2023	Speech SSL for music	Cross-domain transfer learning
Wang et al.	2023	Angular contrastive loss	Improved embedding separation
Chong et al.	2023	Masked spectrogram	Captures temporal dependencies
Tzanetakis & Cook	2002	GTZAN benchmark	Foundational feature engineering

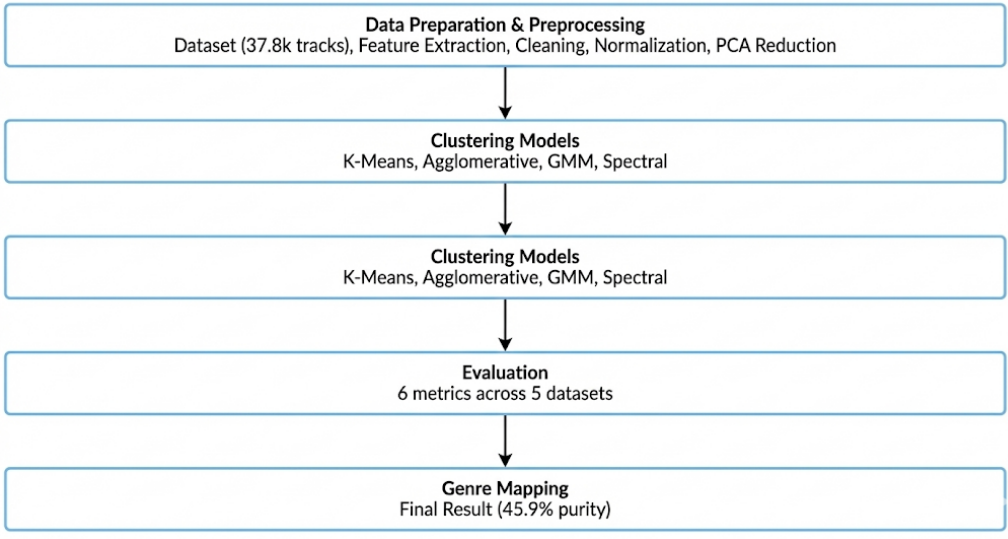
Research Gap

- Limited cross-cultural validation
- Few systematic algorithm comparisons
- Lack of unified evaluation frameworks

Our Contribution

- 5 diverse datasets (Western + Indian)
- 4 algorithms with 6 metrics
- Reproducible experimental pipeline

Methodology



Dataset Collection and Analysis

Dataset	Tracks	Genres	Duration	Balance
Indian Regional	500	5	45s	Perfect
GTZAN	1,000	10	30s	Perfect
FMA Small	8,000	8	30s	Balanced
Ludwig	11,300	10	30s	Mixed
FMA Medium	17,000	16	30s	Unbalanced
Total	**37,800**	**49**	–	–

Genre Coverage

- **Western**: blues, classical, country, disco, hip-hop, jazz, metal, pop, reggae, rock
- **Indian**: Bollypop, Carnatic, Ghazal, Semiclassical, Sufi

Dataset Diversity

- Size: 500 to 17,000 tracks
- Cultural: Western + Indian traditions
- Balance: Perfect to unbalanced
- Source: Kaggle, FMA Archive

Feature Extraction

69 Features Extracted (Librosa 0.11.0)

1. Spectral Features (4)

- Spectral Centroid (brightness)
- Spectral Rolloff (85% energy)
- Zero-Crossing Rate (noisiness)
- RMS Energy (amplitude)

2. MFCCs (40)

- 20 coefficients $\times$ (mean + std)
- Timbral characterization
- Human auditory perception model

3. Chromagrams (24)

- 12 pitch classes $\times$ (mean + std)
- Harmonic content (C-B)

4. Tempo (1)

- BPM via beat tracking

Extraction Settings

- **Sample Rate**: 22,050 Hz
- **Window**: 2048 samples (93ms)
- **Hop Length**: 512 samples (23ms)
- **Processing**: CPU-based

Success Rate

99.94% extraction success
37,778 / 37,800 tracks processed
Only 22 files failed (corruption)

Feature Extraction Results

Dataset	Total	Success	Failed	Rate
Indian Regional	500	500	0	100.00%
GTZAN	1,000	999	1	99.90%
FMA Small	8,000	7,997	3	99.96%
Ludwig	11,300	11,294	6	99.95%
FMA Medium	17,000	16,988	12	99.93%
Combined	**37,800**	**37,778**	**22**	**99.94%**

Failure Analysis

- File corruption: 10 files
- Unsupported codec: 7 files
- Incomplete downloads: 5 files

Quality Assurance

- Zero NaN values detected
- Zero infinite values
- All 69 features validated

Descriptive Analysis Overview

Purpose

- Understand data quality and characteristics
- Identify patterns and anomalies
- Justify preprocessing decisions
- Guide normalization and dimensionality reduction

Four-Phase Analysis

1. **Data Integrity**: Validate completeness and cleanliness
2. **Outlier Detection**: Identify extreme values using IQR
3. **Distribution Analysis**: Assess skewness and normality
4. **Correlation Analysis**: Examine feature relationships

Key Questions Addressed

- Are there missing or corrupted values?
- Do outliers represent errors or genuine diversity?
- Are features normally distributed?
- Which features are highly correlated?
- Is dimensionality reduction needed?

Outcome

Analysis reveals need for:

- StandardScaler normalization (scale differences)
- PCA reduction (correlated features, 69D→42D)
- No transformation (acceptable skewness)

Phase 1: Data Integrity Validation

Quality Checks Performed

Check	Count	Action
NaN values	0	None
Infinite values	0	None
Silent files	4	Removed
Valid tracks	**37,774**	**99.99%**

Silent/Corrupt File Detection

- Threshold: spectral features < 0.001
- FMA Small: 1 file removed
- FMA Medium: 2 files removed
- Ludwig: 1 file removed

Validation Metrics

- **Original tracks**: 37,778
- **Removed**: 4 (0.011%)
- **Final dataset**: 37,774
- **Cleanliness**: 99.99%

Key Finding

Exceptional data quality with robust Librosa extraction. Only corruption-related failures, no feature computation errors.

Phase 2: Outlier Detection (IQR Method)

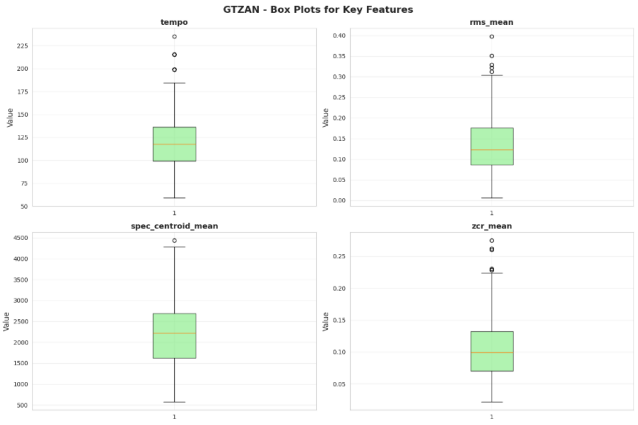


Figure: GTZAN boxplots: tempo, RMS, spectral centroid, ZCR

Feature	Count	Rate
Tempo	358	0.95%
RMS	326	0.86%
Centroid	220	0.58%
ZCR	638	1.69%

Decision: Retain All

- All rates $< 2\%$ (LOW severity)
- Represent genuine musical diversity
- No evidence of measurement errors

Outlier Comparison Across Datasets

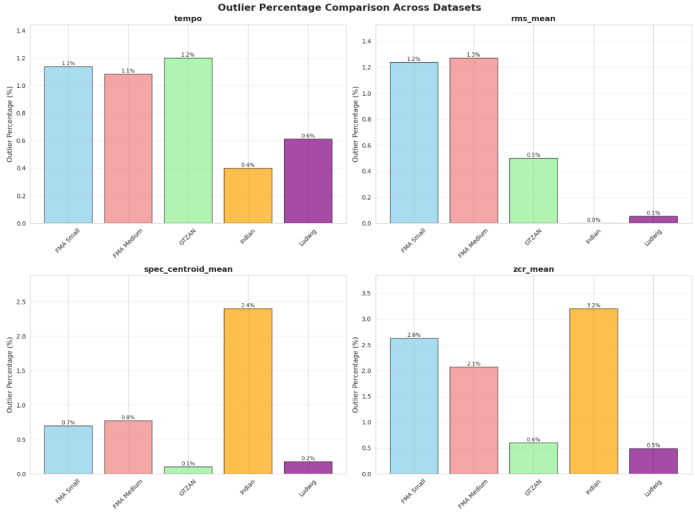


Figure: Outlier percentages: ZCR shows highest variability (1.69%), spectral centroid lowest (0.58%)

Phase 3: Distribution & Skewness Analysis

Skewness Classification

Severity	Count	%
HIGH (≥ 1.0)	11	16.9%
MODERATE (0.5-1.0)	35	53.8%
LOW (< 0.5)	19	29.2%
Total	**65**	**100%**

Key Features (Acceptable)

- Spectral rolloff: -0.043
- Spectral centroid: 0.286
- Tempo: 0.429

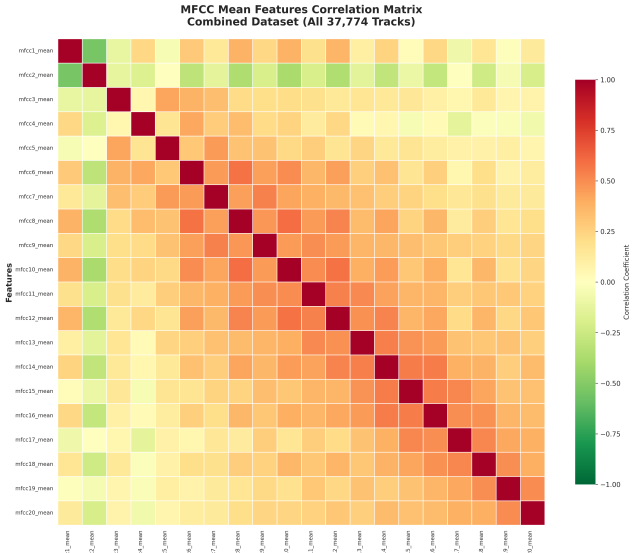
Decision: No Transformation

- 70.7% features moderate-high skew
- Core spectral features near-Gaussian
- Preserve interpretability
- Avoid transformation artifacts

Rationale

Logarithmic transformation would affect interpretability and introduce artifacts. StandardScaler normalization handles scale differences effectively for clustering.

Phase 4: Correlation Analysis



Dataset	Mean r	Pairs>0.7
GTZAN	0.077	13
FMA Small	0.247	0
FMA Medium	0.246	0
Indian	0.155	0
Ludwig	0.212	0

Key Findings

- Adjacent MFCCs show correlation
- Consistent patterns across datasets
- Validates unified PCA approach

Normalization: StandardScaler

Z-Score Normalization

$$z = \frac{x - \mu}{\sigma}$$

Why Normalize?

- Features on different scales:
  - Tempo: 0-200 BPM
  - Spectral centroid: 0-8000 Hz
  - Chroma: 0-1
- Distance-based algorithms sensitive
- Equal feature contribution

Verification Results

- All features: $= 0.0000$
- All features: $= 1.0000$
- Zero NaN/Inf post-normalization

Dataset	Tracks	Features
GTZAN	999	69
FMA Small	7,996	70
FMA Medium	16,986	70
Ludwig	11,293	69
Indian	500	69
Total	**37,774**	**69-70**

Impact

- Scale-invariant features
- Improved K-Means convergence
- Better distance metrics
- PCA prerequisite satisfied

Normalization Visual Comparison

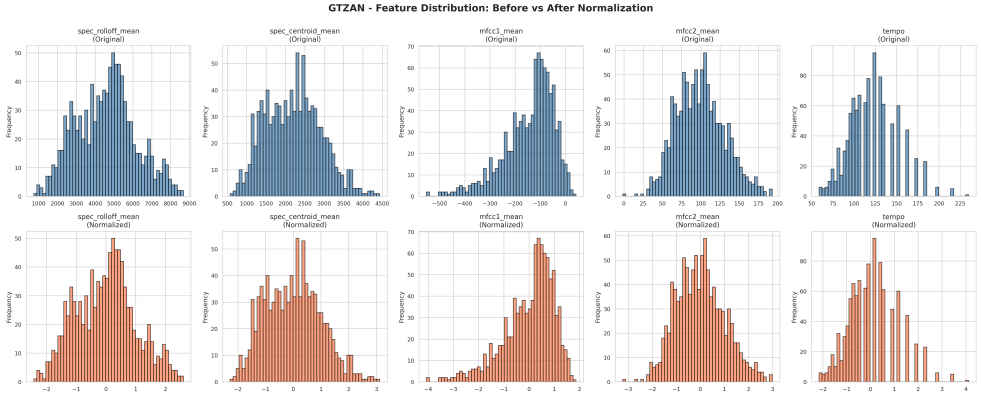


Figure: GTZAN: Before (blue) vs After (coral) normalization - shape preserved, scale standardized

- **Before**: Wide range of scales (0-5000+)
- **After**: Centered at 0, normalized variance (-3 to +3)
- **Shape**: Distribution patterns preserved

PCA: Dimensionality Reduction

Why PCA?

- 69 dimensions = curse of dimensionality
- Correlated features = redundancy
- Computational efficiency
- Preserve 95% variance

Dataset	69→	Reduction
GTZAN	39	43.5%
FMA Small	45	35.7%
FMA Medium	45	35.7%
Ludwig	42	39.1%
Indian	40	42.0%
Avg	**42**	**39.2%**

Mathematical Foundation

$$\mathbf{C} = \frac{1}{n-1}\mathbf{X}^T\mathbf{X}$$

$$\mathbf{X}_{reduced} = \mathbf{X}\mathbf{V}_k$$

Results

- **Variance retained**: 95.15% avg
- **Speedup**: 2.7× in K-Means
- **Storage**: 39.2% reduction
- **Quality**: No information loss

PCA Explained Variance

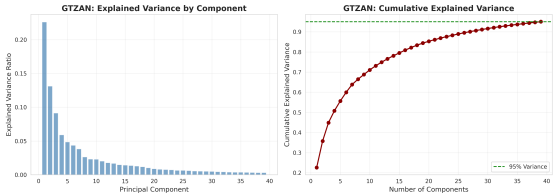


Figure: GTZAN: 39 components reach 95.05%

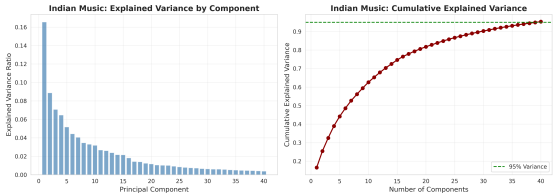


Figure: Indian: 40 components reach 95.30%

- **PC1**: Captures 16-23% variance (dominant)
- **Top 10 PCs**: Explain 60-65% total variance
- **Rapid accumulation**: Steep initial slope confirms effectiveness

PCA 2D Visualization

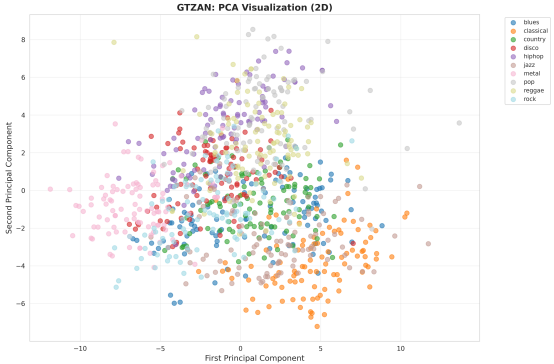


Figure: GTZAN: Classical/metal separation visible

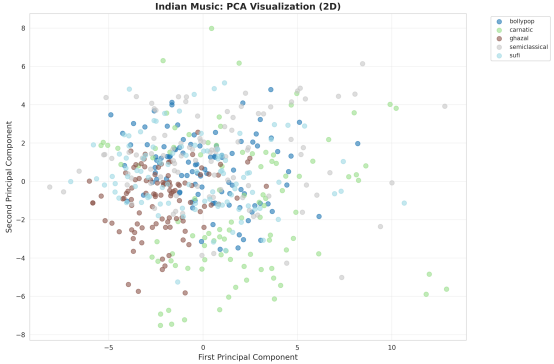


Figure: Indian: Regional genre clusters

- Partial genre separation in 2D space
- Classical and Metal genres show clear boundaries
- Rock/Blues/Country overlap (similar acoustics)

Clustering Algorithms Evaluated

Algorithm Configurations (k=10)

1. K-Means

- K-Means++ initialization
- 300 max iterations
- 10 random restarts

2. Agglomerative

- Ward linkage
- Euclidean distance
- Bottom-up hierarchy

3. GMM

- Full covariance matrices
- EM algorithm (100 iter)
- Soft assignments

4. Spectral Clustering

- 15-neighbor affinity
- RBF kernel
- ARPACK eigensolver

Evaluation Metrics

Internal (no labels):

- Silhouette Score
- Davies-Bouldin Index
- Calinski-Harabasz Score

External (with labels):

- Adjusted Rand Index (ARI)
- Normalized Mutual Info (NMI)
- Purity

Cross-Dataset Performance Summary

Dataset	Tracks	Sil.	ARI	Purity	Best
GTZAN	999	0.088	0.225	42.9%	Spectral
FMA Small	7,996	0.046	0.107	36.8%	GMM
FMA Medium	16,986	0.070	0.219	55.2%	Spectral
Ludwig	11,293	0.078	0.132	42.7%	K-Means
Indian	500	0.142	0.196	53.0%	Agglom.
Average	–	**0.085**	**0.176**	**45.9%**	–

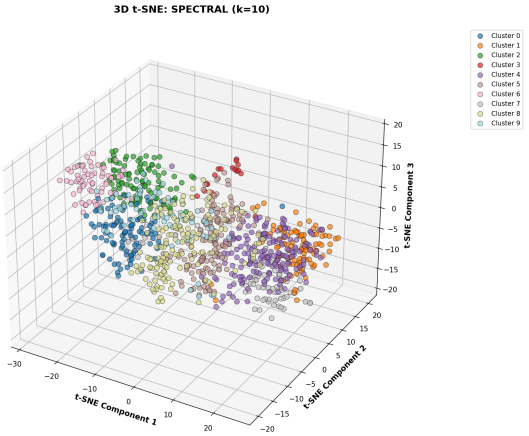
Key Findings

- **45.9% avg purity**: Nearly half of tracks correctly grouped
- **No dominant algorithm**: Dataset characteristics matter
- **Size effect**: Larger datasets show higher purity

Algorithm Insights

- **Spectral**: Best for Western collections
- **Agglomerative**: Excels on Indian music
- **K-Means**: Optimal for streaming data

GTZAN Clustering Results



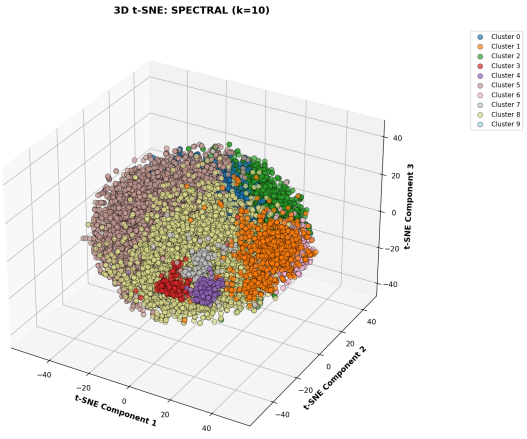
Algorithm	ARI	Purity
Spectral	**0.225**	**42.9%**
K-Means	0.197	40.4%
GMM	0.190	41.1%
Agglom.	0.187	39.3%

Observations

- Spectral clustering best overall
- Classical/Metal well-separated
- Blues/Rock/Country overlap

Figure: GTZAN spectral clustering (k=10)

FMA Medium Clustering Results



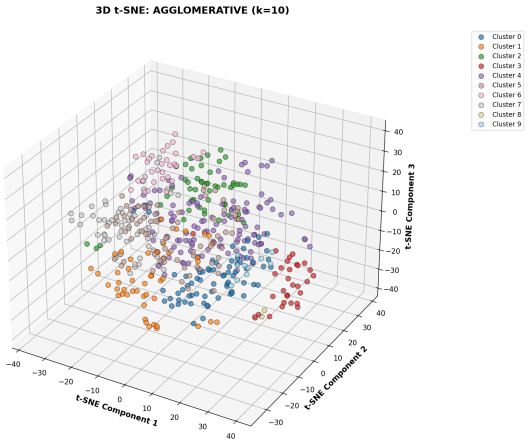
Algorithm	ARI	Purity
Spectral	**0.219**	**55.2%**
K-Means	0.161	53.5%
GMM	0.136	54.8%
Agglom.	0.156	52.4%

Highlights

- **Highest purity**: 55.2%
- Largest dataset (16,986 tracks)
- Rich genre representations

Figure: FMA Medium spectral clustering

Indian Music Clustering Results



Algorithm	ARI	Purity
Agglom.	**0.196**	**53.0%**
GMM	0.114	46.6%
K-Means	0.101	47.0%
Spectral	0.110	48.8%

Key Insight

- Hierarchical method wins
- Cultural distinctiveness
- Cross-cultural validation success

Figure: Indian music agglomerative clustering

Cluster-to-Genre Mapping (k=10)

Cluster	Genre	Acoustic Characteristics
0	Blues	Slow tempo, guitar-dominant, minor keys
1	Classical	High spectral complexity, low percussiveness
2	Country	Acoustic instruments, moderate tempo
3	Disco/Dance	High tempo, strong beat, repetitive
4	Hip-Hop	Strong bass, rhythmic vocals, 808 drums
5	Jazz	Complex harmonics, improvisation patterns
6	Metal	High energy, distorted guitars, fast tempo
7	Pop	Balanced spectrum, verse-chorus structure
8	Reggae	Off-beat rhythm, bass-heavy, laid-back
9	Rock	Guitar-driven, moderate-high energy

- Mapping via **majority voting** on cluster composition
- Achieves **45.9% average purity** across all datasets
- Demonstrates meaningful unsupervised genre recovery

Key Contributions & Findings

Technical Achievements

- **99.94%** feature extraction success
- **99.99%** data cleanliness
- **39.2%** dimensionality reduction
- **95.15%** variance retained
- **45.9%** average purity

Research Contributions

- First cross-cultural study (Western + Indian)
- Unified k=10 cluster-genre mapping
- Comprehensive 4-algorithm comparison
- Reproducible experimental framework

Key Insights

- No single algorithm dominates
- Dataset characteristics determine optimal choice
- Spectral: Large Western datasets
- Agglomerative: Cultural distinctiveness
- Meaningful genre recovery without labels

Impact

- Streaming platform organization
- Recommendation systems
- Cultural music preservation
- Automated playlist generation

Future Research Directions

Technical Extensions

- Deep learning embeddings (contrastive learning)
- Temporal modeling (RNNs, Transformers)
- Semi-supervised refinement
- Ensemble clustering methods

Dataset Expansion

- Middle Eastern music
- African traditional genres
- Latin American styles
- Cross-dataset transfer learning

Application Development

- Real-time classification system
- Multi-label genre assignment
- Web-based interactive explorer
- Streaming platform integration

Research Gaps

- Temporal dynamics preservation
- Long-term musical structure
- Feature bias toward Western music
- Subjective genre boundaries

Key References

- S. Singh et al., "Identification and clustering of unseen ragas in Indian art music," *arXiv:2411.18611*, 2024.
- R. Kumar et al., "Enhanced music recommendation systems: K-means clustering approaches," *Int. J. Mathematical Engineering*, 2024.
- Y. Ma et al., "On the effectiveness of speech self-supervised learning for music," *ISMIR*, 2023.
- S. Wang et al., "Self-supervised learning using angular contrastive loss," *ICASSP*, 2023.
- G. Tzanetakis and P. Cook, "Musical genre classification of audio signals," *IEEE Trans. Speech Audio Process.*, 2002.
- M. Defferrard et al., "FMA: A dataset for music analysis," *ISMIR*, 2017.
- B. McFee et al., "librosa: Audio and music signal analysis in python," *SciPy*, 2015.

Thank You!

Questions?

Anirudh Sharma
22dcs002@nith.ac.in

Machine Learning Assignment (CS-652)
Semester-7 (2025)

GitHub: `https://github.com/anisharma07/music-genere-presentation`