Proceedings of Machine Learning Research 1–14

Low-Resource Rhythm Learning of South Asian Beat Structures: Machine Learning Approaches to Nattuvangam

Anonymous Author(s)
Paper under double-blind review

Abstract

Semantic representations of rhythmic structures are important for AI-driven music generation and choreography. South Asian classical dance, such as Bharatanatyam, relies on intricate rhythms that guide choreography and improvisation. These rhythms are expressed through *Nattuvangam*, a vocal and percussive form that uses rhythmic syllables (*Solkattus*) and cymbal cues (*Talam*). Despite its pedagogical importance, *Nattuvangam* is rarely documented in digital form, which limits systematic study and teaching. We present the first curated dataset of *Nattuvangam* recordings that capture diverse *Solkattu* patterns and cyclic *Talam* structures. Each clip is analyzed using handcrafted and learned features, including onset envelopes, inter-onset intervals, tempograms, and Mel-spectrogram embeddings. These representations allow machine learning models to identify, cluster, and retrieve rhythmic motifs across performances. The dataset serves as a pedagogical tool and supports computational exploration of *Solkattu* patterns in relation to *Talam*, revealing the structural principles underlying *Nattuvangam*. This work establishes a foundation for studying *Nattuvangam* as both a standalone and performative art form, bridging cultural teaching with AI-based rhythm analysis in low-resource contexts.

Keywords: Nattuvangam, Structured Beats, Low Resource, South Asian Music, Machine Learning

1. Introduction

The application of deep neural networks to Western music analysis and synthesis has advanced rapidly, driven by large, well-annotated datasets. In contrast, many global musical traditions remain low-resource, lacking extensive digital archives and limiting computational study. A prime example is *Nattuvangam* Priyadharsini (2017), the intricate rhythmic art central to South Indian classical dance, *Bharatanatyam*. Beyond serving as a rhythmic framework, *Nattuvangam* functions as the sonic backbone of a performance: the *Nattuvanar* recites rhythmic syllables (*solkattus*) while striking cymbals (*talam*), guiding the dancer, setting the tempo, and weaving complex rhythmic compositions (*jathis*) Basani and Himabindu (2025).

Unlike many Western traditions, where rhythm often varies freely, *Nattuvangam* is highly structured with precise beat cycles (*tala*), defined tempos, subdivisions, and hierarchical accents. Each *solkattu* fits these cycles, forming complex temporal patterns. Its mathematical rigor supports dance and improvisation, provides a systematic framework for analysis and teaching, and presents unique challenges for computational modeling, yet it remains under-researched. Here are two publicly available videos that illustrate the structure and performance style of *Nattuvangam* nat (a,b).

A major impediment to studying *Nattuvangam* and similar low-resource cultural practices is the "annotation bottleneck". Manually creating structured datasets requires deep

© .

domain expertise, is time-intensive, and can be inconsistent. To address this, we propose a hybrid approach that combines traditional signal processing with the reasoning capabilities of Large Language Models (LLMs). Trained on extensive internet corpora, LLMs have absorbed substantial cultural knowledge from sources such as academic papers, blogs, and performance descriptions, allowing them to act as "cultural experts" that assist in scalable and reproducible annotation of rhythmic syllables (*solkattus*) within *Nattuvangam* performances. Mention where we got those annotation labels- we researched about it, found of what characterize the seq. We settled on a set of salient attributes and went on to annotate.The labels pertain to syllables, beat cycles etc.

In this work, we present a curated dataset of *Nattuvangam* recordings, systematically organized and annotated to capture the rich rhythmic structure of this art form. Alongside dataset creation, we apply classical machine learning approaches to analyze rhythmic patterns, uncover cyclic temporal structures, and gain insights into the underlying rhythm organization, leveraging LLMs primarily to guide annotation of *solkattus*. This dual focus-dataset curation and pattern analysis-provides a comprehensive foundation for computational studies of *Nattuvangam* and enables further exploration of its complex temporal structures and hierarchical rhythms.

This work offers a generalizable framework for computationally analyzing culturally rich, data-scarce traditions.

2. Related Works

The computational analysis of music, or Music Information Retrieval (MIR), has a rich history, with foundational work focusing on understanding musical structure through beat and downbeat tracking Goto and Muraoka (1994); Dannenberg (2005); Goto and Muraoka (2021); Dannenberg and Goto (2008). Early approaches laid the groundwork for analyzing rhythmic structure from audio signals, evolving into sophisticated models capable of recognizing complex rhythmic patterns Krebs et al. (2013). These advances have enabled high-level tasks such as music genre classification Fu et al. (2011); Cao and Tan (2025), automatic music tagging Lyberatos et al. (2025), and music generation Yuan et al. (2024), often leveraging deep learning architectures including CNNs, LSTMs, and attention mechanisms Zhang et al. (2024); Seo et al. (2023); Ajay and Rajan (2023); Du et al. (2024).

A critical limitation of many computational music analysis methods is their reliance on large, labeled datasets. Low-resource domains, such as Nattuvangam, present a substantial challenge because structured annotationsparticularly of the vocalized rhythmic syllables (*solkattus*)are scarce or nonexistent. This scarcity has motivated strategies such as data augmentation, which expands datasets via pitch shifting or time stretching Sun et al. (2024); Jiang et al. (2025), and transfer learning, where models pre-trained on high-resource tasks are adapted to low-resource ones using parameter-efficient techniques such as adapters Hung et al. (2023); Ali et al. (2024); Mehta et al. (2025). Our work extends this perspective by combining classical machine learning approaches with the knowledge embedded in Large Language Models (LLMs) to assist annotation and facilitate the study of low-resource rhythmic traditions.

Within Indian Classical Music (ICM), prior research has focused predominantly on melodic analysis, identifying repeating patterns (ragas) using spectral features, matrix pro-

files, or symbolic methods Thomas et al. (2016); Nuttall et al. (2021); Shirude and Kolhe (2023), as well as tasks like vocal-instrument separation Murthy et al. (2017). While these studies demonstrate strong analytical interest in ICM, the rhythmic dimension-particularly Nattuvangam-remains largely unexplored. This gap stems primarily from the absence of publicly available, annotated datasets. Our work addresses this by providing a curated dataset with annotated *Solkattus*, enabling both musical analysis and pedagogical exploration of rhythm, tempo, and the underlying musical organization of *Nattuvangam* performances.

3. Background: Nattuvangam and Rhythmic Structure in South Asian Dance

Nattuvangam is the rhythmic and vocal accompaniment for *Bharatanatyam*, a classical South Indian dance.It is performed by the *nattuvanar*, who keeps the beat by striking small bronze cymbals (*talam*) and simultaneously reciting rhythmic syllables called *solkattu*. These syllables, such as *ta di gi na* or *tha ki ta*, indicate the timing and pattern of beats. For example, when the dancer performs a sequence of fast footwork, the *nattuvangam* guides the tempo and emphasizes the correct accents, ensuring the music, rhythm, and dance stay perfectly synchronized.

3.1. Rhythmic Organization: *Tala* and *Jati*

Rhythm in Carnatic music is structured through *tala* Rao (2023), repeating cycles of *angas*—*laghu* (**L**, variable-length), *drutam* (**D**, two beats), and *anudrutam* (**A**, one beat). Seven fundamental *talas* combine with five *jatis* of the *laghu* (*tisra*=3, *chaturasra*=4, *khanda*=5, *misra*=7, *sankirna*=9) to form 35 derived *talas*. For instance, Triputa *tala* with *tisra jati* becomes *Tisra Jati Triputa Tala*. These derived *talas* define the rhythmic framework for compositions. Table 1 lists their nomenclature.

Tala	**Tisra (3)**	**Chatusra (4)**	**Khanda (5)**	**Misra (7)**	**Sankirna (9)**
Dhruva	L_3 D L L	L_4 D L L	L_5 D L L	L_7 D L L	L_9 D L L
Matya	L_3 D L	L_4 D L	L_5 D L	L_7 D L	L_9 D L
Rupaka	D L_3	D L_4	D L_5	D L_7	D L_9
Jhampa	L_3 A D	L_4 A D	L_5 A D	L_7 A D	L_9 A D
Triputa	L_3 D D	L_4 D D	L_5 D D	L_7 D D	L_9 D D
Ata	L_3 L D D	L_4 L D D	L_5 L D D	L_7 L D D	L_9 L D D
Eka	L_3	L_4	L_5	L_7	L_9

Table 1: Seven fundamental *talas* combined with five *jatis* to form the 35-tala system. Each *jati* determines the number of beats in the *laghu*.

3.2. *Solkattu* and Rhythmic Encoding

Solkattu consists of spoken syllables such as *ta*, *di*, *gi*, *na*, and *thom*, which represent percussive strokes aligned with the *tālam*. The structure and phonetics of these sequences provide critical insight into rhythmic perception, articulation, and performance practice Sol.

Counting solkattu sequences are designed for rapid articulation, clarity, and temporal precision as shown in Table 2. Dental consonants (e.g., [t], [d]) typically mark the onset of units, vowels ([a], [i], [u]) encode intrinsic pitch, duration, and intensity, and terminal retroflexes (e.g., [ṭ]) add auditory distinction without slowing articulation. The "•" denotes a drag of the previous phrase. A key property is permutation flexibility: performers can start with a base unit (e.g., *ta ki ṭa*) and generate multiple valid sequences for longer phrases while maintaining rhythmic clarity. This explains some of the variability observed in mean IOI and spectral flux measures.

Beats	**Primary Syllables**	**Alternate Combinations**
1	ta	–
2	ta ka	–
3	ta ki ṭa	–
4	ta ka di mi	ta ka di na, ta ka jo ṇu
5	ta ka ta ki ṭa	ta din gi ṇa tom
6	ta ri ki ṭa ta ka	ta ki ṭa ta ki ṭa, ta din • gi ṇa tom
7	ta ka di mi ta ki ṭa	ta ki ṭa ta ka di mi, ta ka ta di gi ṇa tom, ta • din • gi • ṇa • tom
8	ta ka di mi ta ka jo ṇu	ta din • gi • ṇa • tom
9	ta ka di mi ta ka ta ki ṭa	ta ka di ku ta din gi ṇa tom, ta • din • gi • ṇa • tom

Table 2: Counting *Solkattu* and Combinatorial Patterns

The *trikāla* (third-speed) representation in Table 3 further illustrates temporal scaling: syllable alignment is systematically maintained across slow, medium, and fast renditions of a phrase, highlighting how performers adapt rhythmic material to different tempos.

Tempo	1	2	3
Slow	ta • • •	ki • • •	ṭa • • •
Medium	ta • ki •	ṭa • ta •	ki • ṭa •
Fast	ta ki ṭa ta	ki ṭa ta ki	ṭa ta ki ṭa

Table 3: Rendering of *Tiśra Jāti Eka Tāla* at Different Tempos

Understanding the phonetic structure and combinatorial flexibility of *solkaṭṭu* is crucial for appreciating how rhythmic patterns are generated and perceived in Bharatanatyam. This knowledge informs how we represent, organize, and interpret motion and audio data in our dataset, ensuring that subsequent analyses remain grounded in practical performance strategies and meaningful musical structure.

4. Dataset: Creation and Characteristics

4.1. Overview

Our dataset comprises a combination of publicly scraped *solkattu* recordings and audio from the Hugging Face repository hf. In total, we collected 31,560 seconds (8.76 hours) of audio, covering diverse performance styles and recording conditions. All clips were preprocessed,

standardized to 10-second segments at a 22 kHz sampling rate, and manually curated to ensure consistency and usability for downstream analysis and modeling.

4.2. Dataset Fields

Each audio segment is annotated with one of seven *tala* labels corresponding to distinct rhythmic cycles commonly used in *Nattuvangam*. Table 4 summarizes the dataset fields, including identifiers, rhythmic structure, and tempo-related attributes.

Field	**Description**
clip_id	Unique identifier for each audio segment.
tala type	The *tala*(s) present in the clip; represents distinct rhythmic cycles. Clips can contain one or multiple *talas*.
solkattu	Vocalized rhythmic syllables corresponding to the *tala*(s).
temp_bpm	Approximate tempo of the clip in beats per minute.
aksharas	Number of beat units per *tala* cycle.
angas	Subdivisions within the *tala*, representing structural segments.
first solkattu	The initial syllable in the *tala* cycle.
boolean_multiple_talas	Boolean indicator of whether a clip contains multiple overlapping *talas*.

Table 4: Description of dataset fields used for rhythmic analysis, detailing identifiers, rhythmic structure (*tala*, *solkattu*, *angas*), and tempo-related attributes.

4.3. LLM-based Annotation

The dataset was annotated for *tala* type, *solkattu* transcription, *aksharas*, *angas*, and tempo (*temp_bpm*), chosen to capture the hierarchical rhythmic structure of *Nattuvangam*. Field selection was guided by prior research, instructional tutorials, and performance analysis. Initial annotation used a large language model (Gemini) with zero-shot prompting on textual and acoustic cues, followed by verification and refinement by one of the authors trained for over two decades in Carnatic music. The prompt used is included in the Appendix A. This demonstrates how LLMs can support structured annotation of low-resource musical datasets while preserving musical validity.

5. Experiments

5.1. Features

To study the structure of *solkattus* and beats, we extracted two feature sets: rhythm-based features for baseline analysis and learned audio features for expressive modeling. Rhythm-based features include onset envelopes, tempograms, inter-onset intervals (IOI), and spectral

flux/energy per beat, capturing beat locations, tempo, micro-timing, and accentuation. Learned features comprise Mel-spectrograms and optional channels such as onset strength, capturing frequency-energy patterns over time. Together, these features represent both coarse and fine-grained temporal and spectral structure underlying *tala* cycles.

5.2. Experiments

To explore low-resource rhythmic learning, we investigate whether representations learned from frequent *tāla* classes (*Eka* and *Triputa*) can transfer to underrepresented tālas (the remaining five rare classes). This simulates a zero-shot transfer scenario, where models trained on abundant data encode temporal and spectral patterns potentially generalizable to rare rhythmic forms.

ROCKET + XGBoost Baseline. We first established a baseline using full-sequence MFCCs processed with ROCKET((RandOm Convolutional KErnel Transform)), a state-of-the-art time-series feature extractor. Features were classified with XGBoost using a GroupShuffleSplit to ensure leak-free evaluation. This baseline as in Table 5 provides a reference for how well models trained on frequent tālas encode discriminative rhythmic patterns.

Class	**Precision**	**Recall**	**F1-score**
Eka	0.52	0.24	0.33
Triputa	0.73	0.90	0.80
Accuracy	0.70 (Weighted Avg: P=0.66, R=0.70, F1=0.66)		

Table 5: Classification Report for ROCKET + XGBoost

The baseline achieves 69.8% accuracy and a weighted F1-score of 0.66. The model successfully identifies the majority *Triputa* class (recall 0.90), while *Eka* shows lower recall (0.24), highlighting the challenge of class imbalance and establishing a foundation for transfer evaluation.

Hybrid Feature Model (MFCC + CLAP(Contrastive Language-Audio Pretraining)). To enhance feature richness and transferability, we combined MFCC statistics (mean and standard deviation) with embeddings from a pretrained CLAP model (**laion/clap-htsat-unfused**). A Random Forest classifier trained on these hybrid features achieved 76.2% accuracy and a weighted F1-score of 0.73, with improved recognition of the minority *Eka* class (F1-score 0.48) as in Table 6.

Figure 1 presents a t-SNE visualization of hybrid test set features, illustrating improved separability between *Eka* and *Triputa*. The clearer clusters suggest that hybrid embeddings capture transferable rhythmic representations, supporting the potential for zero-shot generalization to rare tāla classes.

Key Insights: Models trained on frequent *tālas* capture transferable rhythmic structures, supporting low-resource learning for rarer forms. Integrating statistical and learned audio

Class	Precision	Recall	F1-score
Eka	0.83	0.33	0.48
Triputa	0.75	0.97	0.85
Accuracy	0.76 (Weighted Avg: P=0.78, R=0.76, F1=0.73)		

Table 6: Classification Report for Hybrid Features (MFCC + CLAP)

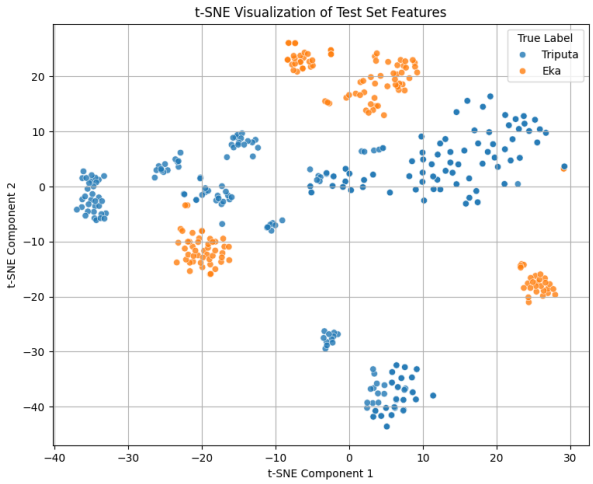

Figure 1: t-SNE embedding of hybrid test set features (MFCC statistics + CLAP embeddings). Colors denote *tāla* class. Hybrid features enhance cluster separability and indicate transferable rhythmic representations.

features improves minority class detection and reveals generalizable rhythmic patterns. t-SNE projections show distinct *tāla* clustering, indicating potential for zero- and few-shot recognition of underrepresented rhythms.

6. Analysis and Discussion

6.1. Tempo Distribution Across Tāla Types

Figure 2 shows the distribution of standardized tempo (BPM) across canonical *Sapta Tālas*. The analysis highlights distinct rhythmic signatures: *Chaturasra Jāti Rūpaka Tāla* and *Khanda Jāti Eka Tāla* exhibit higher medians and broader variance, reflecting flexible pacing, whereas *Tisra* and *Khanda Āta Tālas* have lower, tightly constrained tempos, consistent with longer cycles and denser beat structures. *Chaturasra Jāti Tripuṭa Tāla* demonstrates moderate variability, indicating stylistic diversity across performances. Derived *tālas* such as *Miśra Chāpu Tāla* were excluded to focus on core forms. These patterns explain why models trained on the frequent *Eka* and *Tripuṭa* classes capture a wide range of rhythmic dynamics, enabling transfer learning to rarer tālas. In particular, the tempo elasticity in

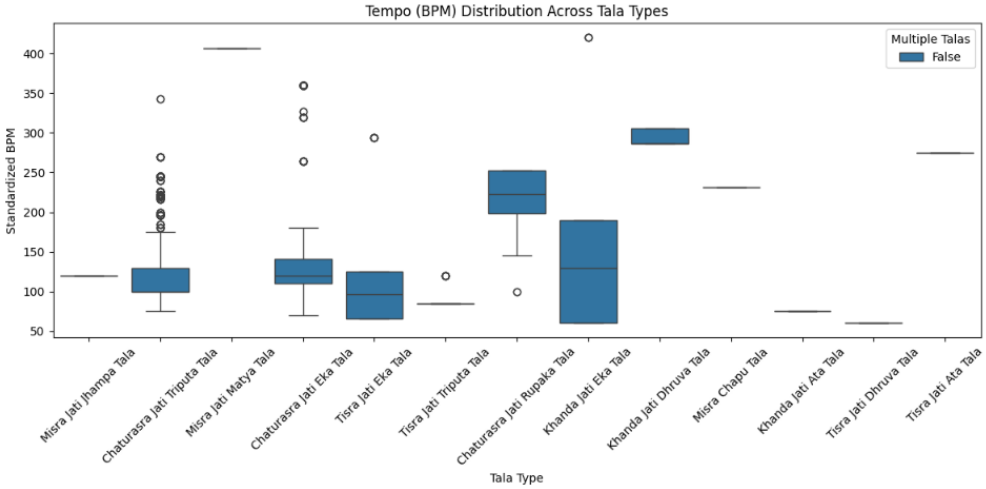

Figure 2: Distribution of standardized tempo (BPM) across canonical *Sapta Tālas*.

these frequent classes provides a rich representation of temporal structure, which complements feature sets like MFCCs and CLAP embeddings, and informs classifier decisions on underrepresented rhythmic cycles.

6.2. Temporal Coverage and Structural Consistency of the First Solkattu

Figures 3a and 3b characterize the temporal structure of the first *solkattu* in our dataset. The coverage plot (Fig. 3a) shows a strongly bimodal distribution, with clips either omitting the start of the cycle or containing the full first *solkattu*. This indicates that most recordings capture rhythmically complete phrases, reflecting pedagogical practice where the first solkattu serves as a foundational template, while partial segments correspond to stylistic variations or improvisation.

The length ratio distribution (Fig. 3b) reinforces this pattern. The peak near 1 shows that the first solkattu typically spans nearly the entire cycle, while a secondary peak around 0.5 reflects partial-cycle renditions due to pedagogical exercises or tempo variations. Rare cases above 1.25 likely result from expressive elongation or annotation inconsistencies.

For modeling, these patterns are informative: (i) the prevalence of full solkattus supports robust learning of internal rhythmic structure, (ii) partial cycles highlight boundary variability critical for generalization, and (iii) the bimodality provides a natural supervisory signal for cycle alignment. Together, these insights link musicological understanding with computational modeling, emphasizing how performance practice shapes data-driven approaches for tasks such as tala classification, solkattu synthesis, and transfer learning to rarer talas.

6.3. Spectral Flux at Onsets Across Tālas

Figure 4 presents the distribution of mean spectral flux at onsets across the seven fundamental *Tālas*, capturing the micro-dynamics and timbral intensity of rhythmic articulation. *Triputa* and *Eka* show higher median flux with greater variability and multiple outliers, in-

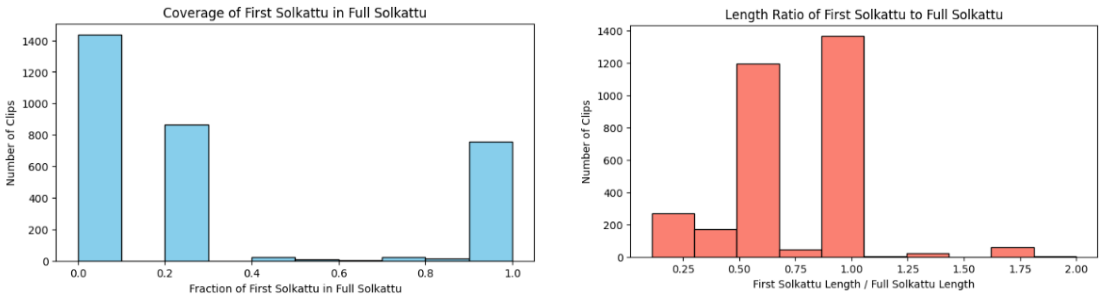

Figure 3: (a) Coverage of first Solkattu across the beat cycle. (b) Distribution of first Solkattu's length ratio relative to the full cycle. The two plots together illustrate both completeness and variability in initial rhythmic phrases.

dicating pronounced dynamic accents and expressive emphasis in their *Solkaṭṭu* execution. In contrast, *Jhampa* and *Aṭa* exhibit lower, tightly clustered flux values, reflecting uniform energy transitions and consistent onset articulation.

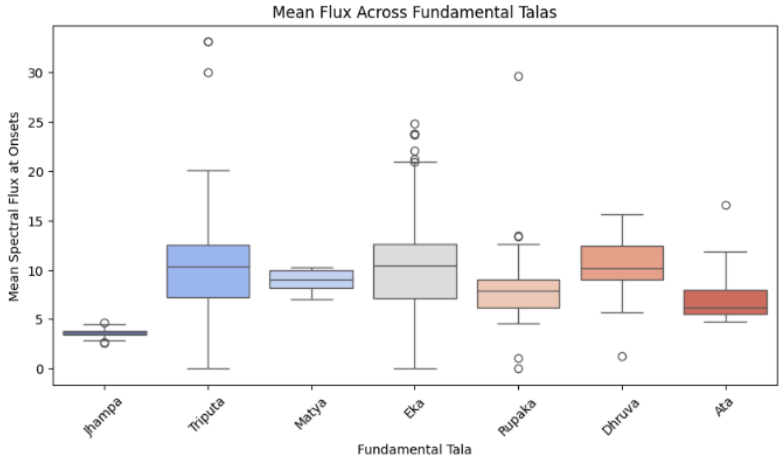

Figure 4: Mean spectral flux at onsets across seven *Tālas*, highlighting differences in rhythmic intensity and micro-dynamic variation.

These patterns complement the temporal analyses (tempo and IOI), demonstrating that *Tāla* characterization is multidimensional: not only does cycle length and timing differ across *Tālas*, but the expressive energy profile of each beat varies systematically. For performers, this informs how different *Tālas* afford dynamic articulation, while for computational models, spectral flux provides a discriminative feature for classifying rhythmic intensity, detecting expressive accents, and improving solkattu-based transfer learning from common to rare *Tālas*.

6.4. Temporal Spacing and Variability Across Tālas

Figure 5 shows the distribution of mean inter-onset intervals (IOIs) across the seven fundamental *Tālas*, capturing both structural and expressive aspects of rhythmic timing. *Dhruva*

and *Triputa* display higher median IOIs with broader interquartile ranges and pronounced outliers, reflecting flexible temporal spacing that accommodates expressive elaboration. In contrast, *Jhampa*, *Aṭa*, and *Mātya* exhibit tightly clustered IOIs, indicative of metrically stable execution with less temporal deviation.

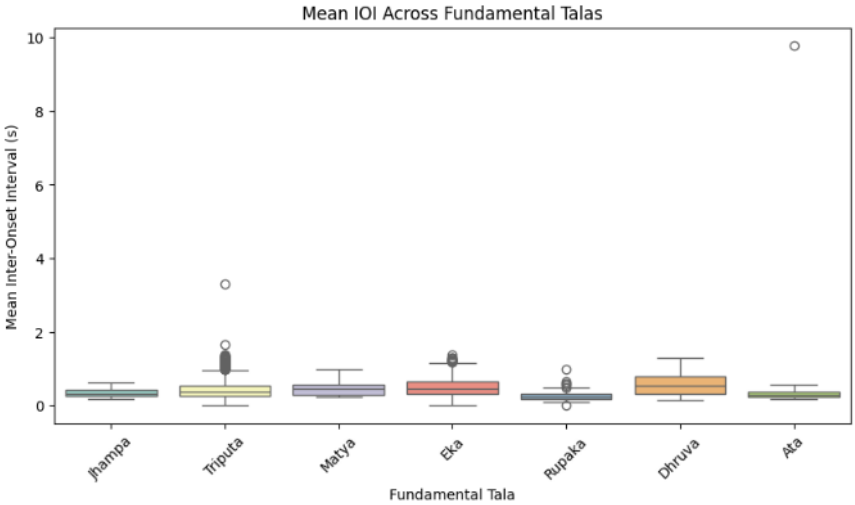

Figure 5: Distribution of mean inter-onset intervals (IOIs) across seven *Tālas*, highlighting differences in rhythmic pacing and temporal variability.

These patterns have dual significance. For performers, they highlight how certain *Tālas* inherently allow expressive timing, while others demand precise adherence to cycle structure. For computational modeling, IOI distributions serve as a quantitative feature for discriminating *Tālas* and capturing temporal dynamics essential for tasks such as solkattu segmentation, rhythmic synthesis, and transfer learning from common to rarer *Tālas*. By linking temporal variability with both musical function and model design, this analysis underscores the importance of IOI as a rhythm-sensitive feature.

6.5. Latent Structure and Clustering of Rhythmic Features

To systematically explore rhythmic variability across *Tālas*, we performed unsupervised analysis on audio clips using mean inter-onset interval (IOI), spectral flux, and tempo.

This approach uncovers latent structure, identifies recurring rhythmic motifs, and examines how expressive variations manifest across different *Tālas*. K-means clustering with PCA projection and t-SNE embeddings capture complementary aspects of the feature space: global cluster organization and fine-grained local similarities (Fig. 6).

The PCA and K-means results reveal clear rhythmic groupings reflecting each *Tāla*'s temporal and dynamic structure. *Triputa* spans multiple clusters, indicating expressive diversity and flexible pacing, while *Jhampa* forms compact clusters with stable timing and energy contours. *Dhruva* and *Eka* occupy slower, irregular regions, and sparse clusters correspond to outliers with extreme tempo or flux values. These patterns align with IOI and spectral flux trends (Figs. 5, 4), highlighting onset timing and energy variation as strong rhythmic discriminators.

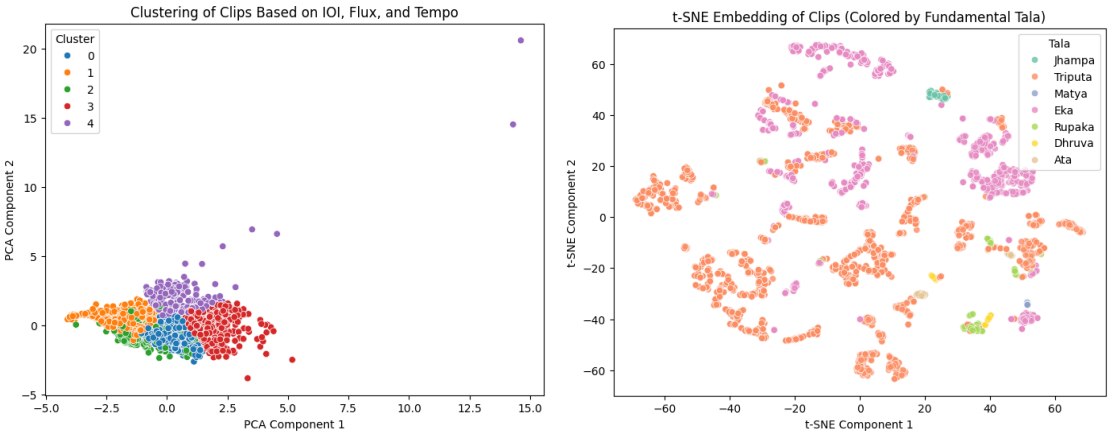

Figure 6: Left: PCA projection of K-means clustered clips based on mean IOI, spectral flux, and tempo (colors indicate cluster membership). Right: t-SNE embedding of the same clips colored by fundamental *Tāla*, highlighting intra- and inter-*Tāla* rhythmic variability.

The t-SNE projection emphasizes intra- and inter-*Tāla* variability. Dense *Triputa* regions reflect rich expressive variation, while compact *Eka* and *Matya* clusters suggest rhythmic uniformity. Partial overlaps among *Jhampa*, *Rupaka*, and *Āta* reveal shared temporal traits, whereas isolated *Dhruva* points highlight distinctive accentuation and pacing.

Together, these latent structures provide a foundation for generalizable rhythmic modeling. They suggest that models trained on frequent *Tālas* can transfer to rarer ones and that learned representations capture both canonical and expressive rhythmic patterns, supporting classification, synthesis, and pedagogy in *Solkaṭṭu*-based performance analysis.

7. Conclusion and Relevance to the Workshop

Our work aligns with the workshop's focus on human-centric AI for music, addressing low-resource rhythmic traditions. By analyzing *solkattus*, we uncover combinatorial and temporal structures that reveal meaningful patterns in the culturally rich, scarce data of *Nattuvangam*.These insights highlight AI's role as a collaborative tool for understanding, preserving, and augmenting human musical expertise, benefiting both performers and educators. Limitations include dataset size, performer variability, and lack of multimodal features. Future work could expand the corpus, integrate motion or gesture data, and capture expressive timing and dynamics to improve interpretability and modeling of complex rhythmic patterns.Overall, this study demonstrates how AI can uncover structured, interpretable patterns while respecting traditional knowledge, laying the groundwork for richer, multimodal analyses of rhythmic performance.

Acknowledgments

Generative AI tools were used solely to refine the writing and presentation of this manuscript, including grammar correction, clarity improvements, and formatting consistency.

Appendix A. Details of the LLM Annotation Prompt

To facilitate consistent and comprehensive annotation of our *Nattuvangam* dataset, we employed Gemini using the following prompt. The LLM was instructed to analyze each audio clip thoroughly and populate an Excel sheet with relevant rhythmic metadata. This approach enabled scalable, semi-automated annotation while maintaining accuracy, which was subsequently verified by domain expertise and extensive reference to tutorials and prior research.

```
I will give you audio clips one by one, for each, I want you to analyze the
    complete video and maintain an excel sheet where you append each audio file.

For each I want you to append the following information:
clip_id (just the video file name), tala type, solkattu, temp_bpm, aksharas, angas
    , first solkattu, boolean_multiple_talas.

A little background for you on the video files. These are Nattuvangam clips I have
     scraped which I want to annotate with the fields I have provided you.

If there are multiple talas or any other column needs multiple value, add them as
    comma separated value.

I want you to do a deep dive on each file, do not just look at the beginning of
    the audio and annotate, look at the whole sample.
Use the standard solkattu for all videos. Please stay consistent.
```

References

Speaking time, being time: Solkaṭṭu in south indian performing arts by douglass fugan dineen faculty advisor: Dr. david p. nelson. URL https://digitalcollections.wesleyan.edu/_flysystem/fedora/2023-03/22103-Original%20File.pdf.

Huggingface dataset. URL https://huggingface.co/datasets/vibhuti16/bharatnatyam_adavus.

Nattuvangam — nattuvangam basic practice — kala prayoga — sri sai nrithyalaya, a. URL https://www.youtube.com/watch?v=_adBbr-GYLk.

Nattuvangam — adhi thala — jathi recitation — episode 6 — sri sai nrithyalaya, b. URL https://www.youtube.com/watch?v=Ze_8NFO1bZw.

Abhinav Ajay and Rajeev Rajan. Music genre classification using attention-based cnn-feature fusion paradigm. In *2023 Annual International Conference on Emerging Research Areas: International Conference on Intelligent Systems (AICERA/ICIS)*, pages 1–5. IEEE, 2023.

Syeda Farhana Ali, Md Omar Faruk, Md Shiful Islam Piash, Mohammad Marufur Rahman, Md Reasad Zaman Chowdhury, and Sarwar Hossain. Application of transfer learning in low-resource language processing: A case study on bangla numeral recognition. In

International Conference on Machine Intelligence and Emerging Technologies, pages 395–408. Springer, 2024.

Parijatha Reddy Basani and U Himabindu. Dancing through disciplines: Interdisciplinary dimensions of indian classical dance forms. *Home*, 1(3):78–92, 2025.

Maodie Cao and Jie Tan. Music genre classification using artificial neural network and spectral feature analysis. In *2025 4th International Conference on Distributed Computing and Electrical Circuits and Electronics (ICDCECE)*, pages 1–6. IEEE, 2025.

Roger B Dannenberg. Toward automated holistic beat tracking, music analysis and understanding. In *ISMIR*, pages 366–373. London, 2005.

Roger B Dannenberg and Masataka Goto. Music structure analysis from acoustic signals. In *Handbook of signal processing in acoustics*, pages 305–331. Springer, 2008.

Xingjian Du, Zhesong Yu, Jiaju Lin, Bilei Zhu, and Qiuqiang Kong. Joint music and language attention models for zero-shot music tagging. In *ICASSP 2024-2024 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 1126–1130. IEEE, 2024.

Zhouyu Fu, Guojun Lu, Kai Ming Ting, and Dengsheng Zhang. A survey of audio-based music classification and annotation. *IEEE Transactions on Multimedia*, 13(2):303–319, 2011. doi: 10.1109/TMM.2010.2098858.

Masataka Goto and Yoichi Muraoka. A beat tracking system for acoustic signals of music. In *Proceedings of the second ACM international conference on Multimedia*, pages 365–372, 1994.

Masataka Goto and Yoichi Muraoka. Musical understanding at the beat level: real-time beat tracking for audio signals. In *Computational auditory scene analysis*, pages 157–176. CRC Press, 2021.

Yun-Ning Hung, Chao-Han Huck Yang, Pin-Yu Chen, and Alexander Lerch. Low-resource music genre classification with cross-modal neural model reprogramming. In *ICASSP 2023 - 2023 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 1–5, 2023. doi: 10.1109/ICASSP49357.2023.10096568.

Jiaming Jiang, Wanlu Cheng, Shengwen Gong, and Jingjing Wang. A deep learning-based data augmentation method for marine mammal call signals. *Frontiers in Marine Science*, 12:1586237, 2025.

Florian Krebs, Sebastian Böck, and Gerhard Widmer. Rhythmic pattern modeling for beat and downbeat tracking in musical audio. In *Ismir*, pages 227–232, 2013.

Vassilis Lyberatos, Spyridon Kantarelis, Edmund Dervakos, and Giorgos Stamou. Challenges and perspectives in interpretable music auto-tagging using perceptual features. *IEEE Access*, 2025.

Atharva Mehta, Shivam Chauhan, and Monojit Choudhury. Exploring adapter design tradeoffs for low resource music generation. *arXiv preprint arXiv:2506.21298*, 2025.

Y. V. S. Murthy, S. G. Koolagudi, and V. G. Swaroop. Vocal and non-vocal segmentation based on the analysis of formant structure. In *2017 Ninth International Conference on Advances in Pattern Recognition (ICAPR)*, pages 1–6. IEEE, 2017. doi: 10.1109/ICAPR.2017.8593164.

Thomas Nuttall, Genís Plaja-Roglans, Lara Pearson, and Xavier Serra. The matrix profile for motif discovery in audio-an example application in carnatic music. In *International Symposium on Computer Music Multidisciplinary Research*, pages 228–237. Springer, 2021.

R Priyadharsini. Nattuvangam-the angam of natyam. *World Wide Journal of Multidisciplinary Research and Development*, 3(1):182–184, 2017. URL https://wwjmrd.com/upload/nattuvangam-the-angam-of-natyam_1674026486.pdf.

D. Anantha Rao. Tala and its significance. *Naad – Nartan Journal of Dance and Music*, 11:5–8, May 2023. ISSN 2349-4654. URL https://naadnartan.in/wp-content/uploads/2023/06/D.-Anantha-Rao-1.pdf.

Wangduk Seo, Sung-Hyun Cho, Paweł Teisseyre, and Jaesung Lee. A short survey and comparison of cnn-based music genre classification using multiple spectral features. *IEEE Access*, 12:245–257, 2023.

Snehalata B Shirude and Satish R Kolhe. Recognizing raga of indian classical songs using regular expressions. In *International Conference on Information Science and Applications*, pages 367–383. Springer, 2023.

Yanjie Sun, Kele Xu, Chaorun Liu, Yong Dou, Huaimin Wang, Bo Ding, and Qinghua Pan. Automated data augmentation for audio classification. *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, 32:2716–2728, 2024.

Matthew Thomas, Y.V. Srinivasa Murthy, and Shashidhar G. Koolagudi. Detection of largest possible repeated patterns in indian audio songs using spectral features. In *2016 IEEE Canadian Conference on Electrical and Computer Engineering (CCECE)*, pages 1–5, 2016. doi: 10.1109/CCECE.2016.7726863.

Ruibin Yuan, Hanfeng Lin, Yi Wang, Zeyue Tian, Shangda Wu, Tianhao Shen, Ge Zhang, Yuhang Wu, Cong Liu, Ziya Zhou, et al. Chatmusician: Understanding and generating music intrinsically with llm. *arXiv preprint arXiv:2402.16153*, 2024.

Shaoxiang Zhang, Peng Lin, Yongchang Ma, and Li Xie. An attention based cnn-lstm hybrid approach for music genre classification. In *2024 7th International Conference on Information Communication and Signal Processing (ICICSP)*, pages 133–137. IEEE, 2024.