

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

Ankitha Sudarshan*
 Atharva Vikas Jadhav*
 Rohini Srihari

ASUDARSH@BUFFALO.EDU
 AJADHAV9@BUFFALO.EDU
 ROHINI@BUFFALO.EDU

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. [GitHub-https://github.com/04rookie/nattuvangam_dataset](https://github.com/04rookie/nattuvangam_dataset)

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

* Equal contribution

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 (*sol kattus*) within *Nattuvangam* performances. We settled on a set of salient attributes and went on to annotate.

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 *sol kattus*. 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 annotations—particularly of the vocalized rhythmic syllables (*sol kattus*)—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 profiles, 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 *an-gas*—*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, *khandasra*=5, *misra*=7, *sankirna*=9) to form 35 derived *talas*. For instance, Tripura *tala* with *tisra jati* becomes *Tisra Jati Tripura Tala*. These derived *talas* define the rhythmic framework for compositions. Table 1 lists their nomenclature.

<i>Tala</i>	<i>Tisra</i> (3)	<i>Chatusra</i> (4)	<i>Khandasra</i> (5)	<i>Misra</i> (7)	<i>Sankirna</i> (9)
<i>Dhruva</i>	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
<i>Matya</i>	L ₃ D L	L ₄ D L	L ₅ D L	L ₇ D L	L ₉ D L
<i>Rupaka</i>	D L ₃	D L ₄	D L ₅	D L ₇	D L ₉
<i>Jhampa</i>	L ₃ A D	L ₄ A D	L ₅ A D	L ₇ A D	L ₉ A D
<i>Tripura</i>	L ₃ D D	L ₄ D D	L ₅ D D	L ₇ D D	L ₉ D D
<i>Ata</i>	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
<i>Eka</i>	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 (Dineen, 2023) 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.

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., [ɖ]) 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 diṇ gi ṇa tom
6	ta ri ki ṭa ta ka	ta ki ṭa ta ki ṭa, ta diṇ • 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 • diṇ • gi • ṇa • tom
8	ta ka di mi ta ka jo ṇu	ta diṇ • gi • ṇa • tom
9	ta ka di mi ta ka ta ki ṭa	ta ka di ku ta diṇ gi ṇa tom, ta • diṇ • gi • ṇa • tom

Table 2: Example *Solkattu* counting patterns for 1–9 beats: left column shows primary syllables, right column lists alternate combinations.

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 *solkattu* 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 *sol kattu* recordings and downloaded audio from the Hugging Face repository ¹ for our experiments. 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. The distribution overview is present in Appendix C.

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. For details on data availability and licensing, please refer to Appendix A.2.

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

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

4.3. LLM-based Annotation and Verification

The dataset was annotated for *tala* type, *sol kattu* transcription, *aksharas*, *angas*, and tempo (*temp_bpm*) to capture the hierarchical structure of *Nattuvangam*, guided by prior research, tutorials, and performance analysis. Initial annotations used zero-shot prompting with a large language model (Gemini) on textual and acoustic cues, followed by verification and refinement by one of the authors formally trained in Carnatic music and a native

1. https://huggingface.co/datasets/vibhuti16/bharatnatyam_adavus

Hindi–Marathi speaker, due to phonetic similarities. The prompt is provided in Appendix A. A listening-based verification step, detailed in Appendix A.1, compared ground-truth audio with LLM-generated syllables to correct mismatches, demonstrating that LLMs can support structured, musically valid annotations for low-resource datasets.

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 *Tripura*) 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 for *tāla* classification using full-sequence MFCCs processed with ROCKET (RandOm Convolutional KERNel Transform) (Dempster et al., 2020), a state-of-the-art time-series feature extractor. Features were classified with XGBoost (Chen and Guestrin, 2016) 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
Tripura	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 for predicting *tāla* types “*Eka*” and “*Tripura*”

The baseline achieves 69.8% accuracy and a weighted F1-score of 0.66. The model successfully identifies the majority *Tripura* 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 + 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**) (Wu et al., 2022). To isolate the effect of adding CLAP embeddings, we trained a Random Forest classifier both on MFCC-only features and on the hybrid MFCC + CLAP features. The Random Forest with 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 shown in Table 6. This indicates that the observed performance gain is attributable to the addition of CLAP embeddings rather than the classifier choice alone.

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 1a shows that hybrid MFCC + CLAP embeddings separate the frequent tālas (*Eka* and *Triputa*) more distinctly, while Figure 1b indicates that ROCKET features capture structure across all tāla classes, including the underrepresented ones. This suggests that both feature types encode discriminative rhythmic patterns useful for few-shot recognition of rarer tālas.

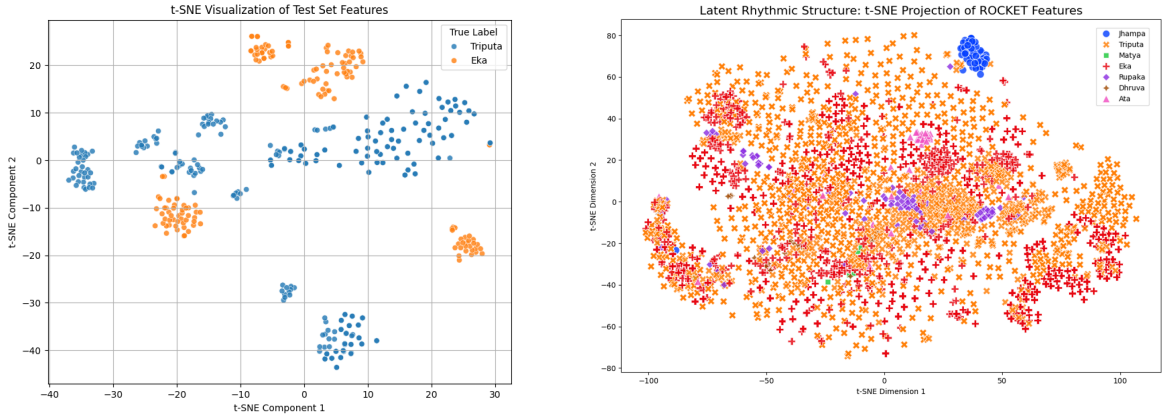


Figure 1: t-SNE visualizations of rhythmic feature spaces. (a) Hybrid test set features (MFCC statistics + CLAP embeddings) showing clear separation between the two most frequent tālas, Triputa and Eka. (b) ROCKET feature projections reveal latent rhythmic structure across all seven tāla classes, with even underrepresented classes (*Jhampa*, *Rupaka*, *Ata*) forming distinct clusters.

Models trained on frequent *tālas* capture transferable rhythmic structures, enabling low-resource learning for rarer forms. Combining statistical and learned audio features improves

detection of minority classes and reveals generalizable rhythmic patterns, as reflected in t-SNE projections. To validate this, we performed binary classification between the most frequent *tāla* (Eka, N=1005) and the rarest *tālas* (Ata, N=36; Dhruva, N=28), using Stratified 5-Fold Cross-Validation (80/20 split) with hybrid sampling to address extreme class imbalance (30:1). Despite limited training data, the pipeline achieved a Recall of 0.71 for Dhruva and 0.56 for Ata, with the Confusion Matrix confirming effective separation of minority patterns. Full metrics and the Confusion Matrix are provided in Appendix B. These results show that the dataset preserves distinct, learnable rhythmic signatures even for very low-resource classes, highlighting its potential for few-shot rhythm modeling.

5.3. Latent Structure and Clustering of Rhythmic Features

To examine rhythmic variability in our dataset, we analyzed audio clips using mean inter-onset interval (IOI), spectral flux, and tempo. PCA projections with K-means clustering and t-SNE embeddings reveal both global cluster structure and local relationships (Fig. 2).

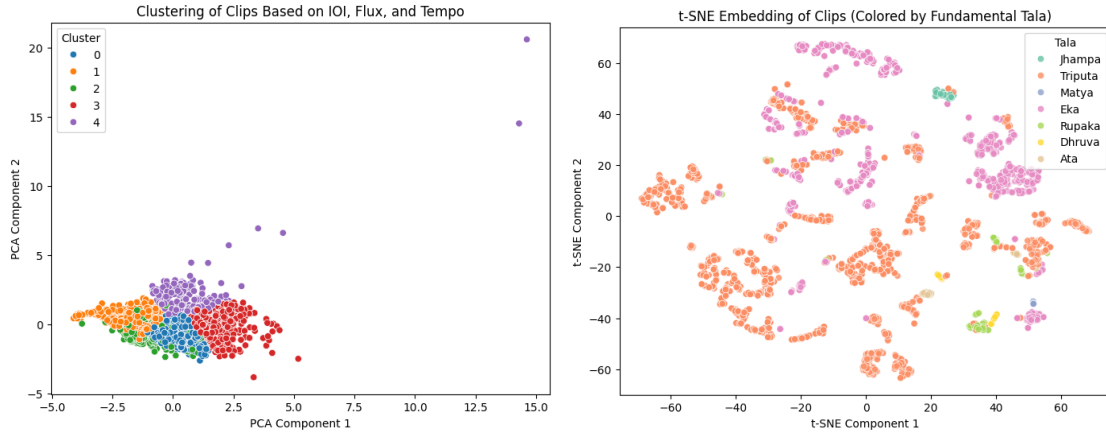


Figure 2: Left: PCA projection of K-means clustered clips based on mean IOI, spectral flux, and tempo (colors indicate cluster membership). Right: t-SNE embedding of the same clips colored by fundamental *Tāla*, highlighting intra- and inter-*Tāla* rhythmic variability.

PCA and K-means highlight distinct rhythmic groupings for each *Tāla*. *Tripura* spans multiple clusters, capturing expressive diversity, while *Jhampa* forms compact clusters with consistent timing and energy. Sparse clusters identify outliers with extreme tempo or flux, and *Dhruva* and *Eka* occupy slower, irregular regions. These patterns align with IOI and spectral flux trends (Figs. 6, 5), showing that onset timing and energy variation are strong rhythmic discriminators.

t-SNE embeddings emphasize intra- and inter-*Tāla* variability: dense *Tripura* regions indicate expressive variation, compact *Eka* and *Matya* clusters reflect uniformity, partial overlaps among *Jhampa*, *Rupaka*, and *Āta* reveal shared temporal traits, and isolated *Dhruva* points highlight unique accentuation and pacing.

These analyses demonstrate that the dataset encodes structured, learnable rhythmic signatures across both frequent and rare *Tālas*. By capturing canonical patterns and expressive

variation, it supports classification, synthesis, and pedagogical applications, highlighting its potential for few-shot learning and generalizable rhythmic modeling.

6. Analysis and Discussion

6.1. Tempo Distribution Across Tāla Types

Figure 3 shows standardized tempo (BPM) across canonical *Sapta Tālas*. *Chaturasra Jāti Rūpaka* and *Khanda Jāti Eka* exhibit higher medians and broader variance, reflecting flexible pacing, while *Tisra* and *Khanda Āta* have lower, tightly constrained tempos. *Chaturasra Jāti Tripuṭa* shows moderate variability, indicating stylistic diversity. Derived tālas like *Miśra Chāpu* were excluded for clarity. These patterns highlight the dataset’s util-

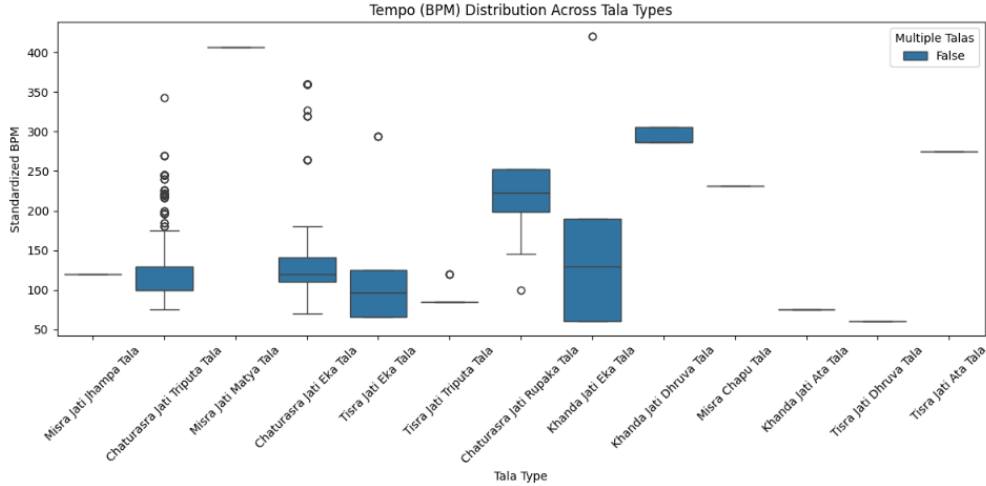


Figure 3: Distribution of standardized tempo (BPM) across canonical *Sapta Tālas*.

ity for rhythmic modeling. Frequent tālas capture diverse temporal structures, enabling transfer learning to rarer classes. Tempo variability complements features like MFCCs and CLAP embeddings, helping classifiers distinguish underrepresented rhythms and supporting few-shot learning of complex tāla patterns.

6.2. Temporal Coverage and Structural Consistency of the First Solkattu

Figures 4a and 4b show the temporal structure of the first *solkattu*. The coverage plot reveals a bimodal distribution: most clips capture the full first *solkattu*, reflecting its role as a foundational template, while partial clips correspond to stylistic variations or improvisation. The length ratio distribution reinforces this, with a primary peak near 1 for full-cycle renditions, a secondary peak around 0.5 for partial cycles, and rare cases above 1.25 likely due to expressive elongation or annotation inconsistencies. These patterns inform modeling: full *solkattus* support learning internal rhythmic structure, partial cycles highlight boundary variability for generalization, and the bimodality provides a supervisory signal for cycle alignment. Together, they link musicological structure with computational modeling,

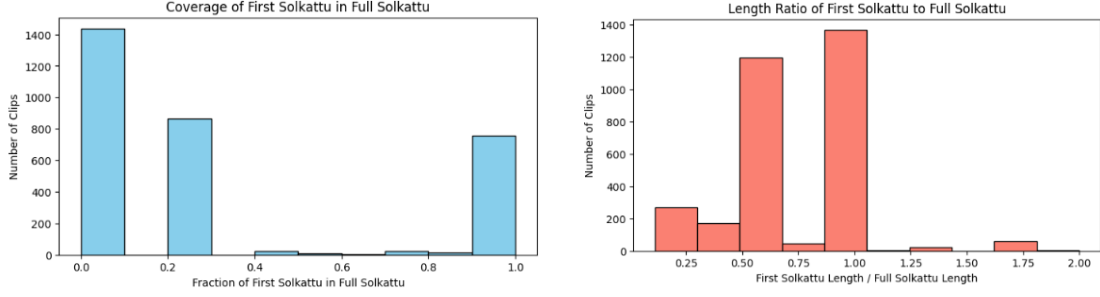


Figure 4: (a) Coverage of first Solkattu across the beat cycle. (b) Distribution of first Solkattu's length ratio relative to the full cycle. The two plots together illustrate both completeness and variability in initial rhythmic phrases.

benefiting tasks like *tāla* classification, *solkattu* synthesis, and transfer learning to rarer rhythmic forms.

6.3. Spectral Flux at Onsets Across Tālas

Figure 5 shows the distribution of mean spectral flux at onsets across the seven fundamental *Tālas*, capturing micro-dynamics and timbral intensity. *Triputa* and *Eka* exhibit higher median flux with greater variability, reflecting pronounced accents and expressive emphasis, while *Jhampa* and *Aṭa* show lower, tightly clustered flux, indicating consistent energy transitions.

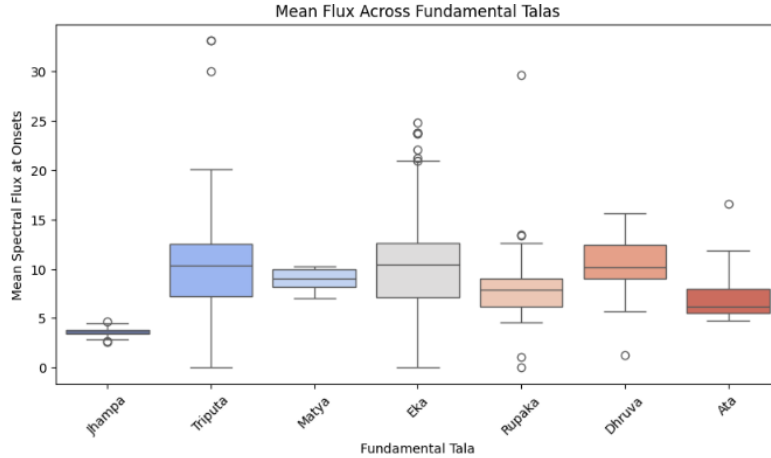


Figure 5: Mean spectral flux at onsets across seven *Tālas*, highlighting differences in rhythmic intensity and micro-dynamic variation.

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 6 shows the distribution of mean inter-onset intervals (IOIs) across seven fundamental *Tālas*, capturing structural and expressive timing. *Dhruva* and *Tripura* exhibit higher median IOIs with broader variability and outliers, reflecting expressive flexibility, whereas *Jhampa*, *Aṭa*, and *Mātya* have tightly clustered IOIs, indicating metrically stable execution.

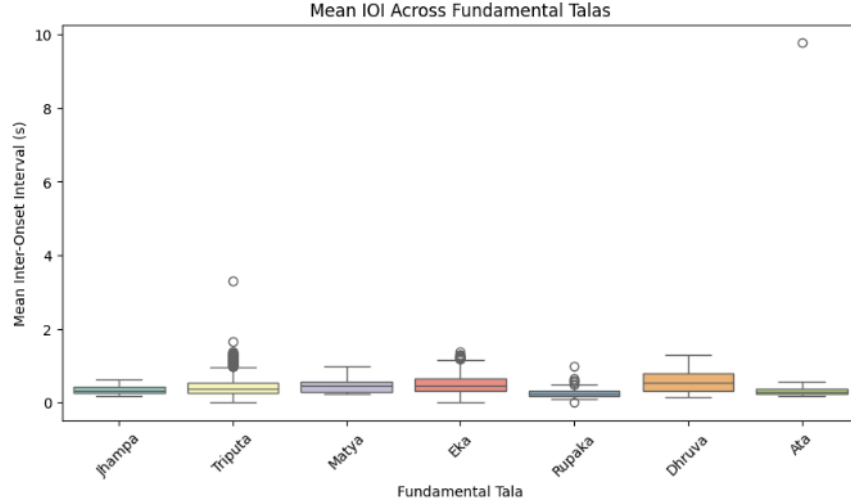


Figure 6: Distribution of mean inter-onset intervals (IOIs) across seven *Tālas*, highlighting differences in rhythmic pacing and temporal variability.

These patterns have both musical and computational significance. Musically, they show which *Tālas* allow expressive timing versus those requiring strict adherence to the cycle. Computationally, IOI distributions provide a discriminative, rhythm-sensitive feature that supports solkattu segmentation, rhythmic synthesis, and transfer learning from frequent to rare *Tālas*. By connecting temporal variability with musical function, this analysis highlights IOI as a key feature for modeling rhythmic structure.

7. Conclusion

The annotated *Nattuvangam* dataset captures hierarchical rhythmic structures through *solkattu* patterns, providing a valuable resource for AI-driven analysis of low-resource musical traditions. It enables models to learn and generalize rhythmic patterns, supporting applications in music education, performance analysis, computational ethnomusicology, and few-shot learning for underrepresented rhythms. While current limitations include dataset size, performer variability, and absence of multimodal features, future expansions incorporating motion, gesture, or expressive timing could enhance modeling of complex rhythmic interactions. Overall, this dataset facilitates structured, interpretable analyses of traditional rhythms, demonstrating its utility for both research and practical musical applications.

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.

A.1. Listening-Based Transcription Evaluation

Ground truth	Annotation by LLM
Dhit Tam Thei Tat Tei Tham Dhit Tam Thei Tat Tei Tham	Dhi Tham Tei Ta Tei Tham Dhitham Thithathei Tham
Dhi Mi Ta Ki Ta Ta Ka Dhi Mi	Ta Ka Dhi Mi Ta Ki Ta
Tai Yum Dhat Tat Tai Yum Tam	Tai yum tat ta tai yum tam
Tat Ta Tei Tei Tat Dhit Tei Tei Tat Ta	Ta Tei Tei Ta
Tai Hat Tai Hi	Tai Ha Ta Ha

Table 7: Ground truth transcription produced by a human annotator through direct listening and verification of the LLM-generated *solkattu* syllables.

The ground truth was annotated by a human judge who is a native speaker of Hindi and Marathi, which aligns well with the phonetic structure of the solkattu syllables. As

shown in Table 7, LLMs performed reliably when translations were processed in smaller batches and shorter segments. Because longer videos exhibit structural variation, chunking the input yielded noticeably better outputs. For future work, we plan to standardize the syllable vocabulary for more consistent human verification and to expand the set of human-annotated samples to better identify where LLMs struggle most.

A.2. Data Availability and Licensing

The audio data utilized in this research was sourced from open-access platforms for non-commercial purposes. Although the original audio content remains the intellectual property of its respective creators, the accompanying author-generated annotations, including time-aligned solkattu transcripts and rhythm labels, are made available under a Creative Commons Attribution (CC-BY) license. This license aims to facilitate future research and ensure the reproducibility of the findings within the community.

Appendix B. Confusion Matrix of Frequent vs Rare Tala and Classification Report

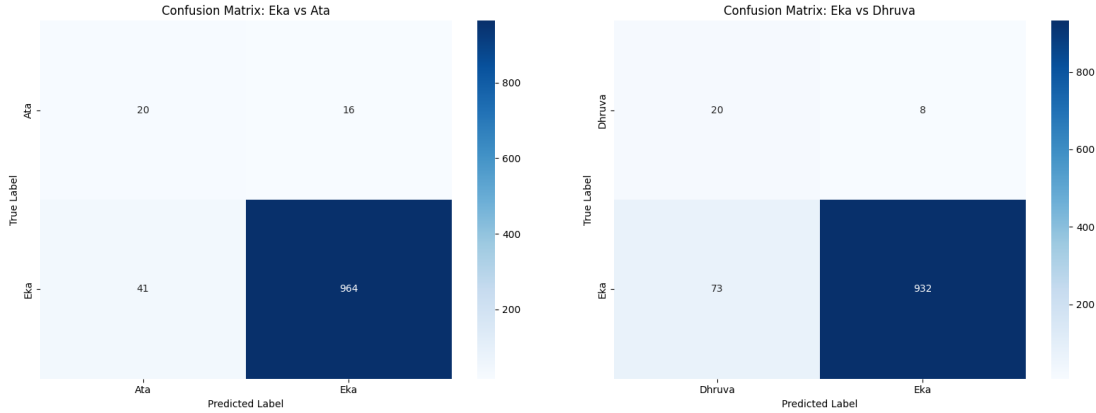


Figure 7: Confusion matrix of binary classification between the frequent class (Eka) and each rare class (Dhruva, Ata).

Appendix C. Dataset Distribution

Table 8 summarizes the number of annotated samples for each Tala class in the dataset. Triputa and Eka account for the majority of the data, with 1915 and 1033 samples respectively. The remaining talas, namely Rupaka, Jhampa, Ata, Dhruva, and Matya, are represented by far fewer samples, indicating a noticeable class imbalance across tala categories.

Tala Name	N Samples
Tripata	1915
Eka	1033
Rupaka	90
Jhampa	48
Ata	36
Dhruva	28
Matya	6

Table 8: Number of samples for each Tala class.

Rare Class	Class	Precision	Recall	F1-score
Dhruva	Dhruva	0.22	0.71	0.33
	Eka	0.99	0.93	0.96
	Accuracy	0.76 (Weighted Avg: P=0.97, R=0.92, F1=0.94)		
Ata	Ata	0.33	0.56	0.41
	Eka	0.98	0.96	0.97
	Accuracy	0.76 (Weighted Avg: P=0.96, R=0.95, F1=0.95)		

Table 9: Classification metrics for binary classification between the frequent class (Eka) and each rare class (Dhruva, Ata).

References

- 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_8NF01bZw.
- 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.
- Tianqi Chen and Carlos Guestrin. Xgboost: A scalable tree boosting system. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, KDD '16, page 785–794, New York, NY, USA, 2016. Association for Computing Machinery. ISBN 9781450342322. doi: 10.1145/2939672.2939785. URL <https://doi.org/10.1145/2939672.2939785>.
- 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.
- Angus Dempster, François Petitjean, and Geoffrey I. Webb. Rocket: exceptionally fast and accurate time series classification using random convolutional kernels. *Data Mining and Knowledge Discovery*, 34(5):1454–1495, July 2020. ISSN 1573-756X. doi: 10.1007/s10618-020-00701-z. URL <http://dx.doi.org/10.1007/s10618-020-00701-z>.
- Douglass Fugan Dineen. Speaking time, being time: Solkaṭṭu in south indian performing arts. https://digitalcollections.wesleyan.edu/_flysystem/fedora/2023-03/22103-Original%20File.pdf, 2023. [Accessed: Dec. 7, 2025].
- 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 trade-offs 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://wwjmr.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.
- 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.
- Yusong Wu, Ke Chen, Tianyu Zhang, Yuchen Hui, Taylor Berg-Kirkpatrick, and Shlomo Dubnov. Large-scale contrastive language-audio pretraining with feature fusion and keyword-to-caption augmentation, 2022. URL <https://arxiv.org/abs/2211.06687>.

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.