Journal of Quality Measurement and Analysis
e-ISSN: 2600-8602
https://doi.org/10.17576/jqma.2202.2026.17

JQMA 22(2) 2026, 303-328
http://www.ukm.my/jqma

COMPARING EMOTIONAL SIGNATURES IN TAMIL INDIE AND MAINSTREAM SONGS USING NEURO SYMBOLIC AI

(*Perbandingan Corak Emosi dalam Lagu-lagu Indie dan Aliran Umum Tamil Menggunakan AI Neurosimbolik*)

SALINI ANBALAGAN, MOHD AFTAR BIN ABU BAKAR & BERNARD KOK BANG LEE*

ABSTRACT

This study introduces a neurosymbolic AI approach to examining emotional expression in Tamil romance songs, addressing the broader gap in how non-Western musical emotions are represented in computational models. By combining machine learning-based audio pattern recognition with symbolic reasoning rules generated via large language models, the framework bridges statistical learning and cultural interpretation. The approach is applied to 117 independently produced and mainstream commercial Tamil romance tracks. Audio features such as tempo, energy and spectral centroid are extracted and interpreted using emotion classification rules generated by GPT-4o-mini, whose outputs are treated as LLM-consistent symbolic labels for exploratory analysis. These rules serve as a basis for symbolic inference, enabling the system to logically deduce the emotion conveyed by a song segment based on its audio attributes. Three new tools are introduced: the *Emotional Authenticity Index (EAI)*, assessing the richness of emotional expression by aggregating entropy-based diversity and symbolic variety; the *Symbolic Logic Consistency (SLC)*, evaluating the coherence of symbolic audio signatures; and *Temporal Emotion Flow (TEF)* visualisations, mapping emotional evolution across a song. Dimensionality reduction using UMAP on symbolic emotion scores reveals distinct emotional clustering by emotion with no meaningful separation by genre. Emotion categories align with the Geneva Emotional Music Scale (GEMS), a validated framework that captures music-specific emotions such as nostalgia, transcendence and tenderness. A Classification and Regression Tree (CART) decision tree model is used to classify symbolic emotion categories based on three key audio features, revealing which attributes most strongly influence the expression of specific emotions. Preliminary findings suggest that while Tamil indie and mainstream music may differ in symbolic strategies, both appear to exhibit convergent emotional trajectories consistent with a shared cultural grammar of emotion. These insights not only enrich cross-cultural musicology but also demonstrate how interpretable AI may illuminate the cultural logic of affective expression, expanding the scope of emotion-aware computational modelling.

Keywords: neurosymbolic AI; large language models (LLM); Tamil music; computational musicology; cultural analysis in music; extraction of audio features; retrieval of music information; GEMS framework

ABSTRAK

Kajian ini memperkenalkan pendekatan AI neurosimbolik untuk mengkaji ekspresi emosi dalam lagu-lagu romantis Tamil. Ia menggabungkan pengecaman corak audio berasaskan pembelajaran mesin dengan peraturan penaakulan simbolik yang dijana melalui arahan kepada model bahasa besar (LLM). Pendekatan ini diterapkan kepada 117 lagu cinta Tamil yang diterbitkan secara bebas dan aliran utama dagangan. Ciri-ciri audio seperti tempo (kelajuan), tenaga (keamatan) dan centroid spektrum (kedudukan purata frekuensi) diekstrak dan ditafsir menggunakan peraturan klasifikasi emosi yang dijana oleh GPT-4o-mini. Peraturan ini menjadi asas kepada inferens simbolik yang membolehkan sistem membuat deduksi secara logik terhadap emosi yang disampaikan oleh segmen lagu berdasarkan atribut audionya. Tiga alat baharu diperkenalkan: *Indeks Keaslian Emosi (EAI)* yang menilai ketulenan ekspresi emosi; *Skor Konsistensi Logik (SLC)* yang menilai kekoherenan tandatangan audio simbolik; dan *Visualisasi*

Aliran Emosi Temporal (TEF) yang memetakan perkembangan emosi sepanjang sesebuah lagu. Pengurangan dimensi menggunakan UMAP ke atas skor emosi simbolik menunjukkan corak pengelompokan emosi yang berbeza antara genre indie dan aliran utama. Kategori emosi diselaraskan dengan Skala Emosi Muzik Geneva (GEMS), sebuah kerangka yang telah disahkan dan merangkumi emosi khusus dalam muzik seperti nostalgia, keunggulan dan kelembutan. Model pokok keputusan CART digunakan untuk mengelaskan kategori emosi simbolik berdasarkan tiga ciri audio utama, bagi mengenal pasti atribut yang paling mempengaruhi ekspresi emosi tertentu. Dapatan kajian menunjukkan bahawa muzik indie beroperasi berasaskan logik simbolik yang lebih konsisten, di mana tandatangan audio tertentu mempunyai padanan langsung dengan emosi yang jelas serta tidak berbelah bagi. Sebaliknya, muzik arus perdana memanfaatkan pemetaan audio-emosi yang lebih fleksibel dan kurang tegas. Dapatan awal mencadangkan bahawa walaupun muzik indie dan arus perdana Tamil mungkin berbeza dalam strategi simbolik, kedua-duanya nampaknya memperlihatkan trajektori emosi yang konvergen, selaras dengan tatabahasa emosi budaya yang dikongsi bersama. Pendekatan ini menawarkan bukti penerokaan yang memperkayakan muzikologi rentas budaya dan menggambarkan bagaimana AI yang boleh ditafsir berupaya menjelaskan logik budaya di sebalik pengekspresan muzik secara lebih berkesan.

Kata kunci: AI neurosimbolik; model bahasa besar (LLM); muzik Tamil; muzikologi komputasi; analisis budaya dalam muzik; pengekstrakan ciri audio; perolehan maklumat muzik; kerangka GEMS

1. Introduction

Digital platforms such as YouTube and Spotify have transformed the global music ecosystem, reshaping how music is produced, distributed and consumed (Cavalcanti *et al.* 2021; Bourreau *et al.* 2022). Within the Tamil-speaking world, these platforms sustain the dominance of mainstream Tamil film music (Kollywood) while simultaneously enabling the rise of an independent (indie) music scene. Both traditions thrive online but diverge in their emotional grammars, production aesthetics and symbolic vocabularies (Harris *et al.* 2019). Mainstream Tamil romantic songs emphasise universality, polished production and cinematic integration (Lu *et al.* 2019), while indie productions prioritise personal expression, experimentation and cultural specificity, often created outside institutional frameworks (Becker 2004). This tension between commercial standardisation and independent authenticity raises critical questions about how emotion is encoded, symbolised and interpreted across genres.

Despite their cultural importance, these distinctions remain quantitatively underexamined in computational musicology. Only a small number of cross-cultural music emotion recognition (MER) studies, primarily within Chinese (Xie *et al.* 2024) or Hindustani (Wieczorkowska *et al.* 2010) repertoires, have attempted to model non-Western emotional grammars, and none to date have addressed Tamil music specifically. Music Information Retrieval (MIR) has largely focused on Western repertoires, training models on annotated datasets such as DEAM, MIREX and the Million Song Dataset that do not generalise well to non-Western traditions (Hu & Lee 2012). Furthermore, state-of-the-art deep learning systems achieve high predictive accuracy but operate as "black boxes", obscuring the mechanisms by which acoustic features map to emotions (Guidotti *et al.* 2018). This interpretability problem limits their ability to capture the symbolic systems and culturally embedded logics through which music conveys meaning. Parallel advances in music psychology reinforce the complexity of the challenge: emotions arise through mechanisms such as brain stem reflexes, evaluative conditioning and emotional contagion (Juslin & Vastfjall 2008), while musical features such as melody, harmony, rhythm, timbre and dynamics shape affective experience (Koh *et al.* 2022; Zentner *et al.* 2008). Yet these insights are rarely integrated into computational frameworks for culturally diverse repertoires. South Asian musical traditions, including Tamil, rely heavily on *raga*-based modality, metaphorical lyricism and distinct timbral palettes that resist reduction to Western-centric taxonomies (Jiang & Zheng 2024; Li *et al.* 2025). This gap underscores the need for integrative approaches that combine statistical power with symbolic and cultural interpretability.

Similarly, while music-psychological research has identified mechanisms such as brain stem reflexes and emotional contagion (Juslin & Vastfjall 2008), such insights are seldom integrated into computational frameworks for culturally diverse repertoires, appearing in fewer than a dozen studies across major MIR conferences (Eerola & Vuoskoski 2012; Han *et al.* 2021). This scarcity underscores the need for integrative models that unite statistical learning with symbolic interpretability within non-Western contexts.

To address these limitations, this study develops a neurosymbolic framework that unites the feature-learning capacity of neural networks with the reasoning clarity of symbolic AI (Hamilton *et al.* 2022; Cunnington *et al.* 2023; Garcez *et al.* 2019; Garcez & Lamb 2023). Neural components process acoustic descriptors such as tempo, energy and spectral centroid, while symbolic reasoning formalises their mapping to higher-order emotional categories. This hybrid strategy enhances interpretability without sacrificing predictive performance, enabling models that are both transparent and culturally sensitive. Applied to a curated corpus of 117 Tamil romantic songs, comprising 45 indie and 72 mainstream songs sourced from Spotify editorial and algorithmic playlists, the framework exploratorily operationalises authenticity, coherence and diversity as measurable constructs. Symbolic inferences are aligned with the Geneva Emotional Music Scale (GEMS), providing a cross-cultural ontology that situates Tamil music within broader affective categories such as *Sadness*, *Tenderness* and *Vitality* (Zentner *et al.* 2008). This integration aims to ensure comparability while preserving cultural nuance.

2. Literature Review

2.1. *Tamil music in the digital age*

The Tamil film industry has historically dominated regional music, with composers such as Ilaiyaraaja and A. R. Rahman shaping sonic conventions that remain central to Tamil cultural identity (Velayudham 2011; Devdas 2002). Film songs serve a dual role: advancing cinematic narratives and providing emotional registers for audiences. However, the rise of digital platforms has disrupted this dominance, enabling new circuits of production and distribution. Recent industry reports indicate that regional genres now account for over 14% of global on-demand streams, with Tamil releases often surpassing millions of plays within weeks (Ken Research 2024; Spotify 2025). This digital infrastructure has catalysed an independent Tamil music scene where artists such as Arivu, OfRo and Sean Roldan use platforms to assert creative autonomy, experiment with form and engage socio-political themes absent from mainstream film narratives. Unlike Kollywood's polished romantic spectacles, indie songs often foreground minimalist aesthetics and lyrical authenticity, reflecting contemporary social realities. These parallel trajectories, namely mainstream polish versus indie authenticity, define the cultural field in which Tamil romantic music operates.

2.2. *Evolution of music emotion recognition techniques*

2.2.1. *Pre-deep learning approaches from the early 2000s to mid 2010s*

Early MER research applied statistical learning to hand-crafted features such as tempo, timbre and pitch. Regression models estimated continuous affective dimensions (Friberg *et al.* 2002), while classifiers such as SVMs and decision trees distinguished mood categories (Feng *et al.* 2003). These studies relied on small, Western-dominated datasets, such as MIREX, DEAM and the Million Song Dataset, and reinforced Western-centric taxonomies such as Hevner's adjectives and Russell's circumplex model (Russell 1980; Hu & Lee 2012). While effective for linking acoustic cues to valence-arousal dimensions, such models faced limitations of subjectivity, dataset bias and limited cross-cultural generalisability (Eerola & Vuoskoski 2012; Liyanarachchi *et al.* 2025).

2.2.2. *Deep learning era from the mid 2010s to present*

The mid 2010s saw a shift towards deep learning, with CNNs and RNNs learning directly from spectrograms and replacing manual feature design. Benchmarks on datasets such as AMG1608, DEAM and PMEmo showed that hybrid CNN-LSTM and attention-based architectures outperformed SVM/SVR baselines, achieving superior concordance correlation coefficients for valence-arousal prediction (Ferreira & Whitehead 2021). While predictive accuracy improved, interpretability declined as models became opaque "black boxes" (Guidotti *et al.* 2018).

2.2.3. *Multimodal fusion approaches from 2018 to present*

More recent research has explored multimodal fusion by integrating audio with lyrics, metadata or physiological signals, such as EEG, to reflect the multifaceted nature of emotion. For example, CNN-Transformer models in the MuSe 2020 challenge outperformed audio-only baselines in valence-arousal prediction (Stappen *et al.* 2021). Despite these advances, multimodal models remain limited by data requirements, Western bias and lack of transparency, underscoring the need for interpretable approaches tailored to diverse repertoires (Liyanarachchi *et al.* 2025).

2.3. *Audio-based emotion modelling in music research*

Traditional MER methods emphasised low-level acoustic descriptors such as MFCCs, tempo, pitch and timbral features to train classifiers for categorical or dimensional labels (Kim *et al.* 2010; Yang & Chen 2012). While moderately effective, these approaches offered little interpretability and were ill-suited to culturally embedded affective cues (Eerola & Vuoskoski 2012). Deep learning later enhanced predictive accuracy (Choi *et al.* 2017; Trigeorgis *et al.* 2016) but compounded the interpretability crisis. In response, recent work has revived symbolic and knowledge-driven models that link acoustic features to perceptual constructs such as timbral smoothness, rhythmic complexity or harmonic richness (Han *et al.* 2021). These approaches integrate domain knowledge to yield more transparent, human-readable mappings between features and affective states. Such methods are particularly relevant for underrepresented repertoires such as Tamil music, where emotion is often conveyed through culturally specific melodic, harmonic and rhythmic idioms not captured in Western corpora.

2.4. *The role of neurosymbolic AI in audio emotion modelling*

Neurosymbolic AI has emerged as a promising paradigm that reconciles the predictive strength of neural networks with the interpretability of symbolic reasoning (Garcez *et al.* 2019; Marcus *et al.* 2022). Neural layers capture complex acoustic patterns, while symbolic layers encode domain knowledge through logical rules. Applications in MIR include mapping high tempo and elevated RMS energy to arousal-related emotions while retaining flexibility to detect nonlinearities. Beyond performance gains, this approach enhances cultural adaptability: symbolic abstraction can be tuned to local ontologies rather than confined to Western taxonomies. In repertoires such as Tamil music, where affect is encoded through *raga*-based modality, poetic lyricism and timbral conventions, neurosymbolic frameworks provide an avenue for connecting acoustic features to culturally specific emotional grammars in an interpretable way.

2.5. *Musical semiotics and emotional ontologies*

Understanding music-induced emotion requires engagement with semiotics and the cultural construction of meaning (Meyer 1956; Nattiez 1990). Parameters such as tempo, modality, timbre and rhythm serve as signifiers whose affective associations are socially conditioned (Becker 2004; Clayton 2016). Dimensional models such as Russell's circumplex model, namely valence–arousal, remain influential (Russell 1980) but lack granularity for culturally embedded affect. Tamil romance songs illustrate this limitation: *Vaseegara* (2001) conveys *viraha*, which refers to longing in separation, through lilting tempo and *veena* ornamentation (Vaseegara 2001); *Munbe Vaa* (2006) blends oscillating strings with soft vocals to evoke *anbu migudha*, which refers to intensified love with melancholy (Munbe Vaa 2006); and the indie track *Usurae Pogudhey* (2019) uses acoustic minimalism to evoke *sindhanaiyudan amaidhi*, which refers to meditative calm (Usurae Pogudhey 2019). Such expressions resist reduction to bipolar axes. The Geneva Emotional Music Scale (GEMS) offers a categorical ontology with nine musically grounded categories (Zentner *et al.* 2008), making it suitable for Tamil traditions where *Longing*, *Tenderness* and *Nostalgia* are central. Building on this, the present study introduces symbolic labelling via GPT-4o-mini, which infers Prolog-style rules mapping quantised audio features onto GEMS categories. This methodology combines computational precision with cultural grounding, enabling interpretable analysis of Tamil emotional expression.

2.6. *Gaps in literature and justification for the present study*

Despite advances in MER, three limitations persist. First, high-performing deep models lack interpretability, obscuring which features drive predictions and undermining trust (Guidotti *et al.* 2018). Second, reliance on Western datasets and taxonomies constrains cross-cultural validity (Eerola & Vuoskoski 2012). Third, MER pipelines rarely integrate symbolic reasoning or cultural frameworks, leaving unexamined how emotion is structured in diverse repertoires. Tamil music exemplifies this gap: despite its prominence, no systematic study has modelled its emotional logics or compared mainstream and indie romance genres in terms of diversity, coherence or authenticity.

This study addresses these gaps through three contributions. First, it develops a neurosymbolic pipeline that combines acoustic descriptors with GPT-4o-mini-inferred rules, enhancing accuracy and transparency. Second, it adopts the GEMS ontology to capture culturally resonant categories beyond valence–arousal. Third, it introduces new indices, namely the Emotional Authenticity Index (EAI), Emotional Consistency Score (ECS), Temporal Emotion Flow (TEF) and Symbolic Logic Consistency (SLC), to quantify symbolic-emotional structures. UMAP projections of symbolic emotion scores further reveal clustering across genres. Together, these contributions provide a transparent, reproducible and culturally grounded framework for analysing emotional expression in Tamil music.

Compared to conventional MER baselines, such as SVM classifiers trained on low-level acoustic features (Kim *et al.* 2010; Yang & Chen 2012) and deep learning approaches that optimise predictive accuracy at the expense of interpretability (Choi *et al.* 2017), the proposed neurosymbolic framework offers two distinct advantages. First, it generates human-readable symbolic rules that directly link audio features to emotion categories, thereby enabling cultural interpretation rather than opaque prediction. Second, by grounding its inferences in the GEMS ontology, it extends beyond Western-centric valence–arousal dimensions towards culturally resonant emotional categories. While baseline approaches may achieve higher predictive accuracy on annotated Western datasets, the present framework prioritises transparency and cultural adaptability over raw performance, a trade-off that is particularly relevant for underrepresented repertoires.

3. Methodology

3.1. *Research design*

This study employs a quantitative, cross-sectional design that integrates low-level audio descriptors with symbolic reasoning to analyse emotional signatures in Tamil romantic music.

The framework is grounded in a neurosymbolic paradigm that combines the feature-learning capacity of neural models with the interpretability of symbolic logic. The design proceeds in four stages:

(1) Corpus construction and preprocessing, involving the curation and standardisation of audio files from Spotify and YouTube.
(2) Feature engineering and symbolic abstraction, involving the extraction of acoustic descriptors and their discretisation into interpretable symbolic tags.
(3) Neurosymbolic emotion modelling, involving the prompting of GPT-4o-mini with symbolic inputs to infer emotion labels and rules, followed by CART consolidation.
(4) Evaluation and metric development, involving the assessment of model fidelity and derivation of corpus-level indices, namely EAI, ECS, TEF and SLC.

This design addresses the interpretability gap in music emotion recognition (MER) by generating human-readable rules while retaining quantitative rigour. This study is exploratory in nature and is guided by three working hypotheses. First (H1), low-level acoustic features, namely tempo, RMS energy and spectral centroid, are not expected to differ significantly between Tamil indie and mainstream songs, given the shared cultural context of Tamil romantic music. Second (H2), symbolic-level emotion distributions are expected to exhibit directional differences between genres, with indie songs tending towards melancholic affect and mainstream songs towards vitality. Third (H3), despite any symbolic-level differences, both genres are expected to converge on shared higher-order emotional categories under the GEMS ontology, consistent with the notion of a shared cultural grammar of Tamil romantic expression. Evidence against H3 would be indicated by statistically significant genre separation in UMAP projections or chi-square tests of emotion distributions.

3.2. *Corpus construction and signal processing*

A total of 117 Tamil romantic songs were curated, comprising 72 mainstream (Kollywood) and 45 indie tracks. Songs were sourced from Spotify editorial and algorithmic playlists, namely “Tamil Indie Romance” and “Tamil Romance”, ensuring representation of both expert-curated and algorithmically prominent works. To mitigate duplication, remixes, live versions and unofficial uploads were excluded. Only official releases available on both Spotify and YouTube were retained, ensuring cross-platform consistency.

Audio was obtained from the corresponding YouTube “official video” or “lyric video” using the YouTube Data API v3, filtered by artist and title metadata. Files were standardised as follows:

- Format: WAV, mono.
- Sample rate: 22.05 kHz, satisfying the Nyquist theorem while balancing computational load.
- Bit depth: 16-bit PCM.
- Preprocessing: silence trimming and amplitude normalisation to -1 dBFS.

For consistency, only the first 180 seconds of each track were analysed, capturing the main emotional arc while reducing bias from extended repetitions.

3.3. *Feature engineering and symbolic abstraction*

Three low-level audio descriptors were extracted using `librosa` v0.11.0:

- Tempo, measured in beats per minute (BPM), via `librosa.beat.beat_track()`, representing rhythmic pace.
- Energy, measured using root mean square (RMS), via `librosa.feature.rms()`, representing perceived loudness.

- Spectral centroid, measured in Hz, via `librosa.feature.spectral_centroid()`, representing timbral brightness.

Each descriptor was discretised into three quantile-based bins, namely low, medium and high, as shown in Table 1, to ensure balanced symbolic categories.

Table 1: Symbolic mapping of acoustic features

Feature	Low	Medium	High
Tempo (BPM)	< 104.57	104.57 to 123.05	> 123.05
Energy (RMS)	< 0.17	0.17 to 0.21	> 0.21
Spectral centroid (Hz)	< 1929.30	1929.30 to 2222.80	> 2222.80

The resulting triplets formed symbolic signatures for each song:

$$\textit{Signature} = (\textit{tempo}_{\text{tag}}, \textit{energy}_{\text{tag}}, \textit{centroid}_{\text{tag}}), \quad \text{where each tag} \in \{\textit{low}, \textit{medium}, \textit{high}\}.$$

3.4. *Neurosymbolic emotion modelling*

3.4.1. *Symbolic inference with GPT-4o-mini*

Each symbolic signature was supplied to GPT-4o-mini using a structured, schema-constrained prompt rather than free-form natural language. Acoustic features were encoded as explicit key–value mappings, such as `tempo_tag: high`, `energy_tag: low` and `centroid_tag: medium`, and presented in a fixed JSON-like format. GPT-4o-mini was instructed to:

(1) assign a single dominant emotion label from a predefined symbolic taxonomy;
(2) generate a corresponding Prolog-style logical rule linking the symbolic feature configuration to the inferred emotion.

This structured prompt design constrained the model's output space and minimised linguistic variability across runs.

Accordingly, all findings derived from GPT-4o-mini inferences should be interpreted as LLM-consistent symbolic patterns rather than claims about the ground truth emotional content of the music. Three complementary validation mechanisms were employed: (i) deterministic reproducibility ensured by executing all prompts with a fixed temperature of 0.0 and a constrained output schema; (ii) rule-level consistency assessed by consolidating distributed per-song symbolic rules into a global CART decision tree; and (iii) Symbolic Logic Consistency (SLC), which quantified agreement between window-level GPT-4o-mini inferences and the global symbolic model.

3.4.2. *CART consolidation*

While GPT-4o-mini produced per-song rules, a CART decision tree using `scikit-learn` v1.6.1 was trained on symbolic signatures and GPT-4o-mini emotion labels to derive a global decision structure. This post hoc model was not used for classification but for interpretability:

- Input features: $(tempo_q, energy_q, centroid_q)$.
- Target: GPT-4o-mini emotion label.
- Output: global decision tree and feature importance scores.

Performance was evaluated using five-fold stratified cross-validation, with metrics reported as mean $\pm$ SD:

$$\textit{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN}, \qquad F_1 = \frac{2 \cdot \textit{Precision} \cdot \textit{Recall}}{\textit{Precision} + \textit{Recall}}.$$

Bootstrap resampling ($B = 1000$) provided 95% confidence intervals (CI) for accuracy and macro-F_1.

3.4.3. *Corpus-level indices*

To quantify symbolic-emotional structures, four indices were developed.

(1) *Emotional Authenticity Index (EAI)*
This index captures emotional diversity through entropy and rarity:

$$EAI = H_{\text{emotion}} + H_{\text{tags}} + \frac{N_{\text{unique}}}{10} + F_{\text{rare}}, \tag{1}$$

$$H_{\text{emotion}} = -\sum_i p(e_i)\log_2 p(e_i), \tag{2}$$

$$H_{\text{tags}} = -\sum_j p(t_j)\log_2 p(t_j), \tag{3}$$

$$F_{\text{rare}} = \sum_{e_i : p(e_i) \le 0.05} p(e_i). \tag{4}$$

Here, H_{emotion} denotes the Shannon entropy of the emotion distribution, H_{tags} denotes the entropy of symbolic signatures, N_{unique} denotes the number of distinct symbolic triplets, namely tempo, energy and centroid, and F_{rare} denotes the proportion of songs with emotion frequency $\le 5\%$. The term $N_{\text{unique}}/10$ serves as a normalisation factor to align the magnitude of symbolic-diversity values with the entropy-based components of the index. Pilot analyses indicated that the number of unique symbolic signatures typically ranged between 15 and 25 per genre, which would otherwise dominate the composite score. Dividing by 10 therefore provides an empirical calibration that balances the contribution of all subcomponents while preserving relative sensitivity to corpus-level diversity.

(2) *Emotional Consistency Score (ECS)*
This score measures the proportion of windows matching the song's overall emotion, where each song, analysed over the first 180 seconds, is segmented into N_w non-overlapping windows of fixed duration $L = 10$ seconds, yielding a maximum of 18 windows per track:

$$ECS_{\text{song}} = \frac{1}{N_w}\sum_{i=1}^{N_w} \mathbb{I}(e_{w_i} = e_{\text{song}}).$$

(3) *Temporal Emotion Flow (TEF)*
This index quantifies the frequency of changes between adjacent windows:

$$TEF_{\text{song}} = \frac{\sum_{i=2}^{N_w} \mathbb{I}\big(e_{w_i} \ne e_{w_{i-1}}\big)}{N_w - 1}.$$

(4) *Symbolic Logic Consistency (SLC)*
This index evaluates the agreement between GPT-4o-mini labels and CART global rules:

- $\mathrm{SLC}_{\mathbf{1}}$, CART versus GPT-4o-mini windows: accuracy of a global model in predicting window-level emotions.
- $\mathrm{SLC}_{\mathbf{2}}$, deterministic triplets: share of symbolic triplets that deterministically map to a single emotion.

3.5. *Reproducibility*

All analyses in this study were conducted using Python version 3.10 within the Google Colab environment to ensure transparency and reproducibility. The workflow encompassed data acquisition and preprocessing from Spotify and YouTube, including format conversion, downsampling and silence trimming; feature extraction and symbolic quantisation of tempo, energy and spectral centroid; neurosymbolic modelling through GPT-4o-mini prompts and Prolog-style rule generation; CART consolidation; and computation of corpus-level indices, including the Emotional Authenticity Index (EAI), Emotional Consistency Score (ECS), Temporal Emotion Flow (TEF) and Symbolic Logic Consistency (SLC). The complete set of scripts, annotated notebooks and step-by-step documentation is publicly available at `https://tinyurl.com/rkpyh53c`.

4. Results

4.1. *Tempo*

Tempo distributions showed no statistically significant differences between Tamil indie and mainstream songs. Mainstream songs had a mean tempo of 114.44 BPM (SD $=$ 19.15, $\text{Median} = 112.35$, IQR $=$ 99.38–129.20, $n = 72$), while indie songs averaged 118.89 BPM (SD $=$ 19.91, $\text{Median} = 117.45$, IQR $=$ 103.36–129.20, $n = 45$). A Mann–Whitney U test indicated that this difference was not statistically significant ($U = 1833.0$, $p = 0.232$), although the effect size was small to moderate ($r = 0.13$). The 95% confidence interval (CI) for the median difference, mainstream minus indie, was $[-16.85, 5.11]$ BPM.

Figure 1 presents a boxplot of tempo distributions. Indie songs exhibited a slightly higher mean tempo and included slower outliers (< 90 BPM), suggesting a broader stylistic range at the lower end. In contrast, mainstream tracks showed slightly more spread within the mid-tempo range ($\text{IQR} \approx 30$ BPM), but lacked extreme slow-tempo values, resulting in a more uniform overall pacing profile. These subtle, non-significant distinctions nonetheless provide the symbolic foundation for subsequent neurosymbolic reasoning, where tempo is discretised into symbolic tags, namely *low*, *medium* and *high*, that directly inform GPT-4o-mini rule-based emotional inferences.

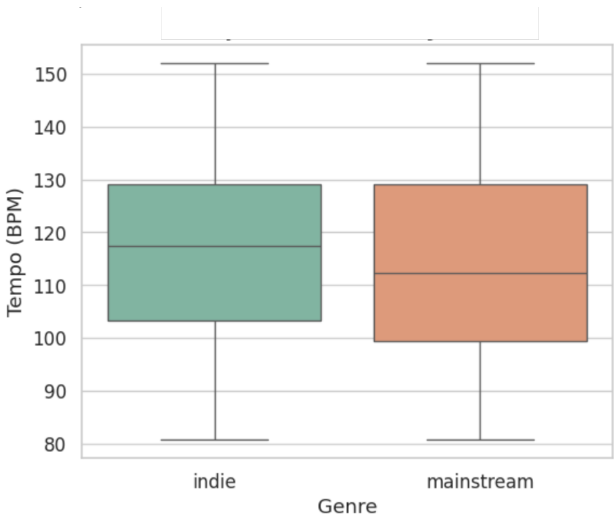

Figure 1: Boxplot of tempo distributions for Tamil indie and mainstream songs. Indie tracks show slightly higher mean tempo and broader spread at the lower end, while mainstream tracks display a more uniform pacing profile

4.2. *Energy (RMS)*

Mainstream songs showed a mean RMS of 0.1910 (SD $= 0.0468$, $\text{Median} = 0.1842$, IQR $=$ 0.1576–0.2174, $n = 72$), compared with 0.1874 for indie songs (SD $= 0.0347$, $\text{Median} =$ 0.1869, IQR $= 0.1629$–0.2094, $n = 45$). The Mann–Whitney U test was not significant ($U = 1587.0$, $p = 0.856$, $r = 0.02$), with a 95% CI for the median difference of [-0.0227, 0.0173].

Figure 2 visualises RMS distributions. Although the differences were not statistically significant, mainstream tracks exhibited slightly higher average loudness and a wider spread of RMS values, including both lower and higher extremes. Indie tracks showed a slightly tighter central distribution. These patterns may reflect greater dynamic range in mainstream mastering, which sometimes raises both floor and ceiling levels due to compression and limiting, whereas indie productions may remain closer to natural acoustic levels. RMS-based symbolic tags, namely *low*, *medium* and *high*, play a critical role in the neurosymbolic reasoning pipeline, helping distinguish high-arousal states, such as *Excitement*, from low-arousal states, such as *Melancholy*.

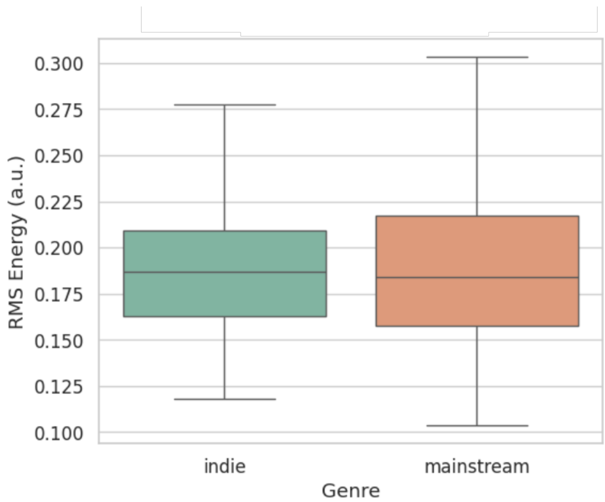

Figure 2: RMS energy distributions for Tamil indie and mainstream songs. Mainstream tracks show slightly wider spread, including higher extremes, potentially reflecting compression and limiting in commercial mastering

4.3. *Spectral centroid*

Mainstream tracks had a mean centroid of 2093.43 Hz (SD $= 353.90$, $\text{Median} = 2123.45$, IQR $= 1866.63$–2358.55, $n = 72$), while indie tracks averaged 2060.38 Hz (SD $= 342.25$, $\text{Median} = 2039.46$, IQR $= 1793.69$–2289.30, $n = 45$). The Mann–Whitney U test indicated no significant difference ($U = 1475.0$, $p = 0.418$, $r = 0.09$), with a 95% CI for the median difference of [-57.16, 274.32] Hz.

Figure 3 shows the distributions. Mainstream songs tended towards slightly brighter timbres, indicated by higher centroids, while indie songs were more associated with darker tonal palettes. Although not statistically significant, these timbral tendencies align with genre conventions. Symbolic tags derived from centroid values, namely *low*, *medium* and *high*, provide essential cues in the neurosymbolic reasoning pipeline.

The three low-level descriptors, namely tempo, RMS energy and spectral centroid, did not reveal statistically significant genre differences in this dataset. However, subtle distributional contrasts were observed, where indie songs trended towards greater tempo variability, while mainstream tracks showed slightly higher RMS and centroid values. These findings underscore the limits of low-level acoustic features in capturing genre distinctions and motivate the

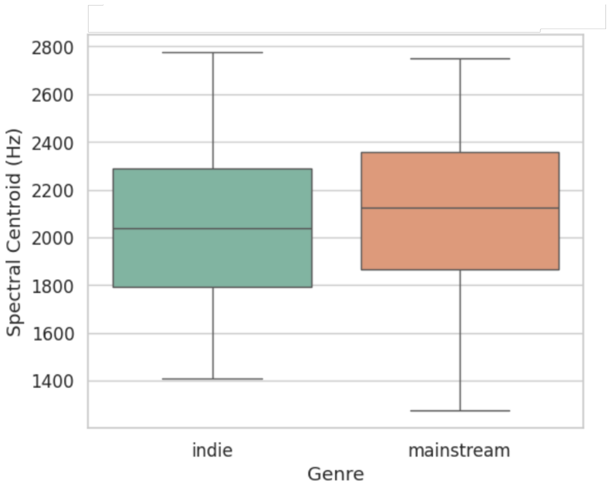

Figure 3: Spectral centroid distributions for Tamil indie and mainstream songs. Mainstream songs trend towards slightly brighter timbres, while indie tracks lean towards darker tonal palettes, although the differences were not statistically significant

integration of higher-order neurosymbolic reasoning in subsequent sections.

Across all three descriptors, group differences were small and statistically non-significant with the current feature estimation settings. A consolidated statistical summary is reported in Table 2.

Given the exploratory nature of this study and the small number of comparisons ($k = 3$), no multiple testing correction was applied. However, all differences remained non-significant at the uncorrected $\alpha = 0.05$ level. Taken together, these results indicate that low-level acoustic descriptors alone offer limited discriminative power between Tamil indie and mainstream romance songs.

Although subtle directional tendencies are observable, such as greater tempo variability in indie tracks and slightly higher RMS and spectral centroid values in mainstream productions, these differences are small and statistically non-significant. This suggests that genre distinctions in Tamil romantic music are not primarily encoded at the level of isolated acoustic features, thereby motivating the need for higher-order symbolic abstraction.

4.4. *Statistical comparison of emotion distributions by genre*

To examine whether the distribution of GPT-4o-mini-inferred emotions differed significantly between Tamil indie and mainstream songs, song genre, namely indie versus mainstream, was cross-tabulated with three dominant emotional categories: *Excitement*, *Melancholy* and *Nostalgia*. These GPT-4o-mini-inferred emotions are treated here as LLM-consistent symbolic labels rather than verified emotional ground truth. Categories with fewer than ten cases were excluded to satisfy chi-square assumptions. Table 3 presents the raw counts and proportions. Indie songs were more frequently classified as *Melancholy* ($n = 23$, 51.1%) relative to mainstream songs ($n = 25$, 34.7%), whereas mainstream songs showed higher proportions of *Excitement* ($n = 33$, 45.8%) compared to indie songs ($n = 17$, 37.8%). *Nostalgia* was less common overall but more pronounced in mainstream tracks ($n = 14$, 19.4%) than in indie tracks ($n = 5$, 11.1%).

Although the chi-square test did not reach statistical significance, the observed distribution reveals modest directional tendencies that should be interpreted cautiously. Indie songs show a higher concentration of *Melancholy*, whereas mainstream songs more frequently express *Excitement* and *Nostalgia*. These trends, reflected in the standardised residuals and effect size, suggest differences in emotional emphasis rather than categorical separation.

Table 2: Consolidated statistical summary of low-level acoustic descriptors for indie and mainstream tracks

Feature	**Group**	n	**Mean ± SD**	**Median [IQR]**	**Mann–Whitney** U	p	**Effect size** (r)	**95% CI (Δ median)**
Tempo (BPM)	Indie	45	118.89 ± 19.91	117.45 [103.36–129.20]	1833.0	0.232	0.13	$[-16.85, 5.11]$
	Mainstream	72	114.44 ± 19.15	112.35 [99.38–129.20]				
RMS (a.u.)	Indie	45	0.1874 ± 0.0347	0.1869 [0.1629–0.2094]	1587.0	0.856	0.02	$[-0.0227, 0.0173]$
	Mainstream	72	0.1910 ± 0.0468	0.1842 [0.1576–0.2174]				
Centroid (Hz)	Indie	45	2060.38 ± 342.25	2039.46 [1793.69–2289.30]	1475.0	0.418	0.09	$[-57.16, 274.32]$
	Mainstream	72	2093.43 ± 353.90	2123.45 [1866.63–2358.55]				

Table 3: Distribution of GPT-4o-mini-inferred emotions by genre

Emotion category	Indie (n, %)	Mainstream (n, %)	Total
Excitement	17 (37.8%)	33 (45.8%)	50
Melancholy	23 (51.1%)	25 (34.7%)	48
Nostalgia	5 (11.1%)	14 (19.4%)	19
Total	**45 (100%)**	**72 (100%)**	**117**

A chi-square test of independence did not reach statistical significance, $\chi^2(2, N = 117) = 3.42, p = 0.181$, indicating that the overall distribution of emotions was not strongly dependent on genre. However, the effect size, Cramer's $V = 0.171$, suggests a weak association, implying that larger or more balanced samples may be required to detect robust genre differences. Inspection of standardised residuals, not shown, indicated that indie songs exceeded expected frequencies for *Melancholy*, whereas mainstream songs exceeded expectations for *Excitement* and *Nostalgia*.

These results suggest that while emotional profiles overlap substantially across genres, there are subtle divergences which, although not statistically significant and supported only by weak effect size trends, warrant further exploration in larger datasets. Indie music leans towards reflective or melancholic affect, whereas mainstream productions emphasise vitality and nostalgic tones, consistent with a commercial emphasis on energetic and audience-oriented soundscapes.

4.5. *Symbolic rule extraction and CART consolidation*

To interpret GPT-4o-mini symbolic outputs at the corpus level, a CART decision tree classifier was trained post hoc on GPT-4o-mini predicted emotion labels. GPT-4o-mini first produced per-song Prolog rules that captured local symbolic reasoning. CART then acted as a meta-level consolidation step, learning from these label patterns to assemble a single, global decision hierarchy. This approach transformed GPT-4o-mini distributed, rule-based inferences into a coherent corpus-wide structure, enabling interpretable, hierarchical comparisons of emotion across song types.

Figure 4 presents the decision tree learned from the symbolic tags, namely `tempo_tag`, `energy_tag` and `centroid_tag`. The first split was on `energy_tag_high`, with a threshold of 0.5, highlighting its dominant role in genre-wide emotion differentiation. Subsequent splits revealed pathways such as:

(1) Rule 1: if energy is low and centroid is low, then *Melancholy*, with support of $n = 30$ songs.
(2) Rule 2: if energy is high and centroid is low, then *Excitement*, with support of $n = 26$ songs.
(3) Rule 3: if centroid is high and tempo is medium, then *Nostalgia*, with support of $n = 19$ songs.

These rules demonstrate that energy serves as the primary determinant of emotional state, with centroid, which represents brightness, and tempo, which represents pacing, refining the classification boundaries. To further quantify the structure, feature importance scores are shown in Figure 5. Energy contributed the most variance, at 65.5%, followed by spectral centroid at 23.0% and tempo at 11.5%. Other symbolic features, namely low and medium bins, contributed negligibly. Overall, CART achieved 87.2% accuracy, with a weighted F_1 score of 0.87, in reproducing GPT-4o-mini symbolic labels, confirming that the LLM's distributed reasoning can be summarised in a compact, interpretable global decision structure.

The dominance of energy as the primary splitting variable indicates that perceived intensity plays a central role in symbolic emotion differentiation across Tamil romantic music. Spectral centroid and tempo function as secondary modifiers, refining emotional categorisation rather

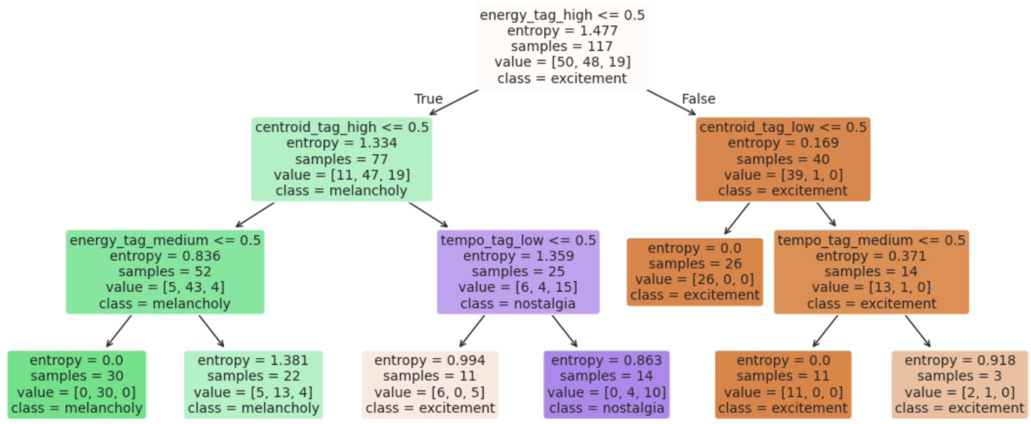

Figure 4: CART decision tree consolidating GPT-4o-mini symbolic rules. The primary split on energy tag, with a threshold of 0.5, highlights perceived intensity as the dominant determinant of emotional categorisation, with spectral centroid and tempo serving as secondary modifiers

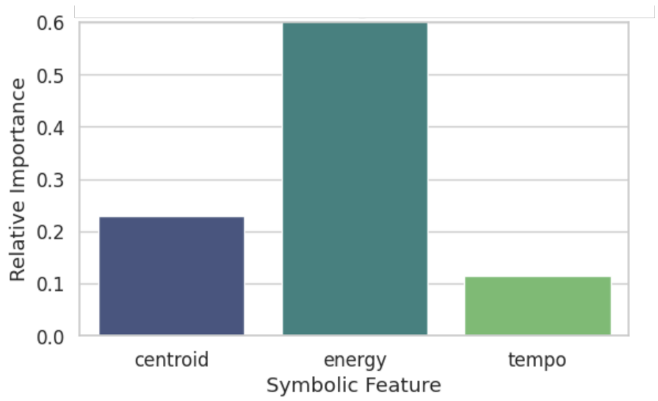

Figure 5: Relative feature importance scores derived from CART consolidation. Energy accounts for the largest proportion of variance, at 65.5%, followed by spectral centroid at 23.0% and tempo at 11.5%

than driving it independently. This hierarchical structure suggests that emotional meaning is constructed through interaction among features rather than isolated acoustic cues.

Although GPT-4o-mini symbolic inferences were deterministic in format, they are not entirely free from uncertainty arising from representational ambiguity and probabilistic reasoning. To minimise such effects, prompts were executed with a fixed temperature of 0 and a constrained schema, namely JSON-like key–value mapping, ensuring consistent symbolic outputs across runs. Residual uncertainty, such as borderline cases between *Melancholy* and *Nostalgia*, may contribute minor variability in the observed distributions. However, cross-validation and rule-consistency checks indicated that these effects did not materially alter the overall emotion proportions or genre-level trends.

4.6. *Model performance evaluation*

The reliability of the neurosymbolic pipeline was assessed in terms of fidelity to GPT-4o-mini symbolic reasoning rather than accuracy against human-annotated labels, since no gold-standard emotional annotations were available for the Tamil music corpus. Accordingly, the evaluation focused on whether GPT-4o-mini distributed per-song symbolic inferences could be consolidated into a global explanatory model.

As shown in Table 4, the CART model achieved an overall fidelity accuracy of 0.838, with a weighted F_1 score of 0.832. The macro F_1 score was 0.786, with a 95% CI of [0.687, 0.865], estimated via bootstrapping with 1000 iterations. The lower macro F_1 score reflects the extreme class imbalance in GPT-4o-mini symbolic labels. Dominant categories such as *Excitement* ($n = 50$) and *Melancholy* ($n = 48$) were reproduced strongly, whereas *Nostalgia* ($n = 19$) showed moderate performance. In contrast, *Peace* ($n = 0$) and *Longing* ($n = 0$) provided insufficient support for reliable per-class estimates.

Table 4: Performance metrics of the GPT-4o-mini neurosymbolic model by emotion class based on CART fidelity

Emotion	**Precision**	**Recall**	**F_1 score**	**Support**
Excitement	0.88	0.90	0.89	50
Melancholy	0.83	0.90	0.86	48
Peace	0.00	0.00	0.00	0
Nostalgia	0.71	0.53	0.61	19
Longing	0.00	0.00	0.00	0

Overall fidelity: accuracy = 0.84, macro F_1 = 0.79, weighted F_1 = 0.83
95% CI for macro F_1 based on bootstrap resampling with $B = 1000$: [0.687, 0.865]

The categories *Longing* and *Peace* had no instances in GPT-4o-mini symbolic outputs. These absences are not due to explicit exclusion but likely reflect GPT-4o-mini difficulty in confidently mapping subtle or low-arousal affective states into discrete symbolic rules, given the limited availability of musical cues. As such, metrics for these categories are not interpretable. The weighted F_1 score is therefore reported alongside the macro F_1 score to account for class imbalance.

While the current evaluation prioritises fidelity to GPT-4o-mini internal symbolic reasoning, future validation will incorporate expert-annotated emotional labels to ensure cultural and affective authenticity. A panel of Tamil musicologists and affective-music researchers will rate a subset of songs using the same five-emotion taxonomy, enabling quantitative comparison between model-inferred and human-judged emotions. This triangulation would provide an empirical benchmark for the neurosymbolic model's cultural validity and interpretive reliability.

The framework demonstrates strong reproducibility for dominant categories, although its robustness across the full emotional taxonomy is constrained by class imbalance and systematically underrepresented states. Nonetheless, the high weighted F_1 score and tight bootstrap confidence intervals confirm that GPT-4o-mini symbolic reasoning is reproducible and structurally coherent, establishing a reliable foundation for higher-level indices such as the Emotional Authenticity Index (EAI).

4.7. *UMAP visualisation of symbolic emotion space*

To explore the structural geometry of symbolic features, the discretised tempo, energy and spectral centroid tags were projected into a two-dimensional manifold using Uniform Manifold Approximation and Projection (UMAP). Figure 6 shows each track plotted by symbolic emotion label, represented by colour, and genre, represented by shape.

The UMAP projection further reinforces this interpretation. Symbolic emotion labels exhibit modest but interpretable clustering, whereas genre labels show no meaningful separation. This indicates that emotional organisation is more structurally salient than genre classification in the symbolic feature space.

Each point in Figure 6 represents a song, colour-coded by GPT-4o-mini-inferred emotion, namely *Excitement*, *Melancholy*, *Nostalgia* and *Peace*, and shaped by genre, namely *Indie* and *Mainstream*. Cluster separability was quantified using internal validation metrics. The UMAP

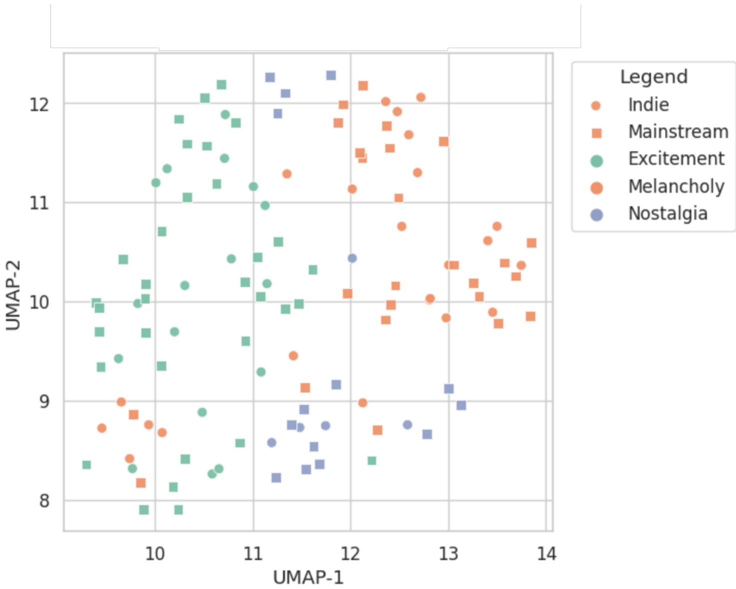

Figure 6: UMAP projection of the symbolic emotion space derived from discretised tempo, energy and spectral centroid tags, with `n_neighbors` = 15, `min_dist` = 0.15 and `metric` = `euclidean`. Points are colour-coded by GPT-4o-mini-inferred emotion and shaped by genre. The silhouette score is 0.167 for emotional clustering and -0.015 for genre clustering

embedding achieved a silhouette score of 0.167 and a Davies–Bouldin index of 2.56 for emotional clustering, indicating weak but interpretable grouping, with *Excitement* and *Melancholy* forming the most distinct clusters and *Nostalgia* remaining more diffuse. This diffuseness may reflect the cultural and musical context of Tamil romantic traditions, where nostalgic or peaceful affective states are often intertwined with other emotions rather than expressed in isolation, resulting in softer cluster boundaries. In contrast, clustering by genre produced a silhouette score of -0.015 and a Davies–Bouldin index of 43.50, confirming no meaningful separation between indie and mainstream tracks. These results suggest that the symbolic abstraction layer captures broad emotional geometries in Tamil music, although with modest separation strength. The UMAP embedding therefore provides a visual bridge between per-song symbolic rules and corpus-level indices such as the Emotional Authenticity Index (EAI).

4.8. *Emotional Authenticity Index (EAI)*

To quantify corpus-level expressive diversity, an Emotional Authenticity Index (EAI) was computed as the weighted average of four components: (i) emotion entropy, representing the spread of GPT-4o-mini labels; (ii) symbolic tag entropy, representing the diversity of tempo, energy and centroid triplets; (iii) unique signature count, representing distinct symbolic combinations; and (iv) rare-emotion frequency, representing the proportion of categories with frequency $\leq 5\%$. Bootstrap resampling with $B = 1000$ was used to estimate 95% confidence intervals (CI) for each component and for the composite score.

Mainstream tracks exhibited a higher composite EAI score of 7.748, with a 95% CI of [7.198, 8.194], than indie tracks, which recorded an EAI score of 7.087, with a 95% CI of [6.386, 7.702]. This difference was driven by consistently higher emotion entropy, 1.484 versus 1.341, and symbolic tag entropy, 4.114 versus 3.913, as well as a larger unique signature count, 21.4 versus 18.3. Rare-emotion frequency was negligible for both groups and lower in mainstream tracks, 0.000 versus 0.004. Overall, mainstream tracks exhibited directionally higher EAI scores, although the overlapping confidence intervals suggest only modest separation under this bootstrap formulation.

These results suggest that, within this corpus, mainstream tracks display broader emotional and symbolic diversity at the corpus level than indie tracks. Importantly, this corpus-level diversity complements earlier findings: although low-level descriptors showed only subtle distributional contrasts, the combination of symbolic abstractions and GPT-4o-mini emotion inferences yields a richer space of expressive configurations. Thus, EAI provides a compact, statistically grounded summary of how neurosymbolic representations scale from per-song rules to corpus-level variety, as shown in Tables 5–7.

At the component level, mainstream songs exhibited slightly higher emotional entropy than indie tracks, indicating a marginally broader emotional palette, whereas indie songs appeared more concentrated on a smaller set of core emotions. In parallel, symbolic tag entropy was greater for mainstream tracks, reflecting more diverse combinations of tempo, energy and timbral tags. Consistent with this pattern, mainstream tracks also produced a larger number of unique symbolic signatures, suggesting richer layering and production diversity. Rare-emotion frequency was negligible in both genres, with indie showing only a marginally higher value.

These results nuance the perception of indie authenticity, suggesting that mainstream songs in this corpus display broader computational diversity, at least as measured by entropy-based metrics. From a computational and symbolic perspective, mainstream songs exhibited higher EAI scores, reflecting greater structural richness and emotional diversity. This suggests that authenticity, when modelled through symbolic features, does not equate to minimalism or rawness, but rather emerges from the complexity and diversity of expressive pathways.

By situating EAI within the neurosymbolic framework, this study shows that authenticity can be measured transparently and reproducibly through a fusion of entropy-based metrics and symbolic abstraction, thus bridging computational rigour with cultural interpretation.

Table 5: Emotional Authenticity Index (EAI), part 1: entropy measures

Category	Emotional entropy	95% CI	Tag entropy	95% CI
Indie	1.341	[1.126, 1.506]	3.913	[3.592, 4.175]
Mainstream	1.484	[1.347, 1.569]	4.114	[3.879, 4.322]

Table 6: Emotional Authenticity Index (EAI), part 2: symbolic signatures

Category	Unique signatures	95% CI
Indie	18.3	[15.0, 21.0]
Mainstream	21.4	[18.0, 24.0]

Table 7: Emotional Authenticity Index (EAI), part 3: rare emotions and composite score

Category	Rare emotions	95% CI	EAI score	95% CI
Indie	0.004	[0.000, 0.044]	7.087	[6.386, 7.702]
Mainstream	0.000	[0.000, 0.000]	7.748	[7.198, 8.194]

4.9. *Emotional Consistency Score (ECS)*

The Emotional Consistency Score (ECS) measures the degree to which symbolic features within a song converge on the same emotional inference across time windows. Higher ECS values indicate that tempo, energy and spectral centroid tags reinforce a stable emotional category, while lower values reflect fluctuations or contradictory cues within a track.

Mainstream songs recorded an average ECS of $\mathrm{Mean} = 0.581$, SD $= 0.219$, $\mathrm{Median} = 0.597$, $\mathrm{IQR} = [0.416, 0.755]$ and $n = 72$, while indie songs averaged $\mathrm{Mean} = 0.506$, SD $= 0.210$, $\mathrm{Median} = 0.500$, $\mathrm{IQR} = [0.359, 0.692]$ and $n = 45$. A Mann–Whitney U test indicated that the difference was not statistically significant ($U = 1959.0$, $p = 0.058$, $r = 0.209$), with a 95% CI for the median difference of [-0.028, 0.184].

Figure 7 presents a boxplot comparing ECS distributions across indie and mainstream songs. Although not statistically significant, the trend suggests that mainstream tracks may more often sustain a single emotional trajectory. Indie tracks, in contrast, tended towards lower ECS values and greater variability, tentatively reflecting more frequent emotional fluctuations and hybridity, such as transitions between *Melancholy* and *Nostalgia* within a single piece. From a neurosymbolic standpoint, higher ECS values in mainstream songs suggest that symbolic rules apply more consistently within tracks, whereas indie songs reveal richer intra-track variability that contributes to their perceived emotional nuance.

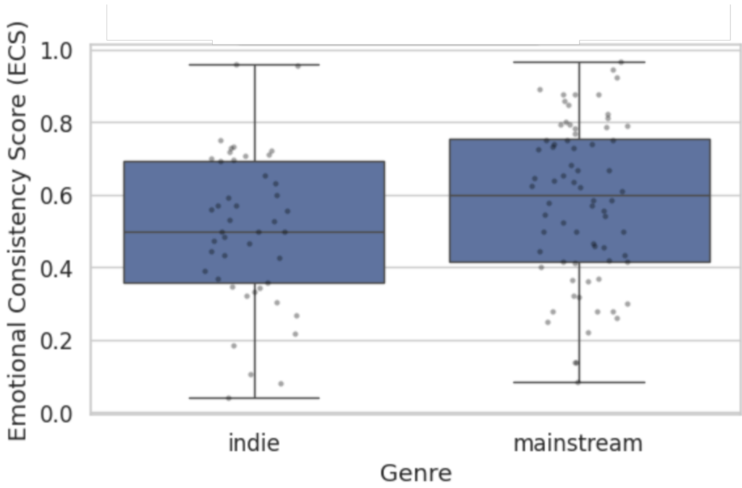

Figure 7: Emotional Consistency Score (ECS) distributions by genre. Mainstream tracks trend towards higher ECS values, suggesting more stable single-emotion trajectories, although the difference was not statistically significant ($U = 1959.0, p = 0.058$)

4.10. *Temporal Emotion Flow (TEF)*

The Temporal Emotion Flow (TEF) captures the extent to which a song exhibits emotional shifts across time windows. A higher TEF value indicates more frequent changes between emotional categories within a track, reflecting dynamic transitions and hybrid affective structures, while lower TEF values indicate smoother and more stable emotional trajectories.

For indie songs, TEF values averaged $M = 0.360$, with a 95% CI of [0.323, 0.395] and $n = 45$, while mainstream songs averaged $M = 0.346$, with a 95% CI of [0.310, 0.381] and $n = 72$. Although indie tracks showed slightly higher TEF, tentatively suggesting greater emotional variability within songs, the confidence intervals overlap substantially, suggesting no strong genre-level differences.

Figure 8 presents a boxplot comparing TEF distributions between indie and mainstream songs. Although the distributions largely overlap, visual inspection suggests that indie tracks

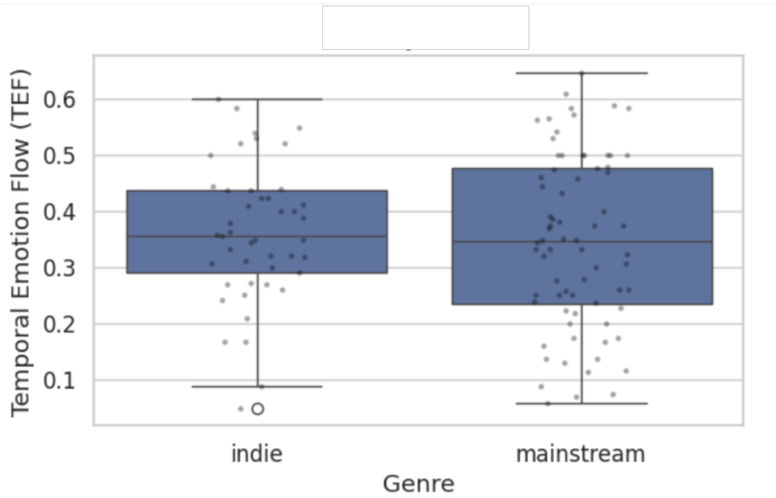

Figure 8: Temporal Emotion Flow (TEF) distributions by genre. Indie tracks show slightly higher TEF values, tentatively suggesting greater within-song emotional variability, although confidence intervals overlap substantially

may exhibit slightly more within-song variability, while mainstream productions maintain more stable emotional arcs, often dominated by *Excitement* or *Melancholy*.

From a neurosymbolic perspective, TEF quantifies the temporal stability of symbolic rules. The results suggest that indie music, consistent with its aesthetic of experimentation and hybridity, embraces emotional dynamism, whereas mainstream music maintains continuity, aligning with production strategies geared towards accessibility and broad audience appeal.

4.11. *Symbolic Logic Consistency (SLC)*

The Symbolic Logic Consistency (SLC) metric evaluates the degree to which GPT-4o-mini per-window symbolic inferences align with the global decision structure derived from CART. Two complementary measures were used:

- $\mathbf{SLC_1}$, CART versus GPT-4o-mini windows: the CART model, trained on song-level symbolic tags, was applied to window-level tags. Agreement between CART predictions and GPT-4o-mini labels yielded an overall SLC_1 accuracy of 0.711. When broken down by genre, indie songs achieved higher consistency ($M = 0.758$, 95% CI [0.717, 0.801], $n = 45$) compared to mainstream songs ($M = 0.684$, 95% CI [0.633, 0.732], $n = 72$).
- $\mathbf{SLC_2}$, deterministic triplets: for each triplet, namely `tempo_tag`, `energy_tag` and `centroid_tag`, the analysis tested whether it mapped to a single emotion across the corpus. The proportion of deterministic triplets was 0.889, indicating that nearly 90% of symbolic feature configurations consistently implied the same emotion.

Figure 9 presents a boxplot of SLC_1 per-song values by genre. The results show that indie songs are more internally consistent in applying symbolic rules across windows, whereas mainstream songs exhibit slightly more variability. At the same time, the high SLC_2 score confirms that the symbolic feature space is largely deterministic, with only a small fraction of triplets supporting multiple possible emotions.

This finding complements earlier observations on ECS and TEF. Indie songs, while showing greater emotional fluctuation across windows, reflected by lower ECS, nonetheless maintain stronger logical alignment with the global symbolic rules, reflected by higher SLC_1. This suggests that indie music embraces temporal emotional variability, such as shifting between *Melancholy* and *Nostalgia*, but does so within a rule-consistent framework. Mainstream songs,

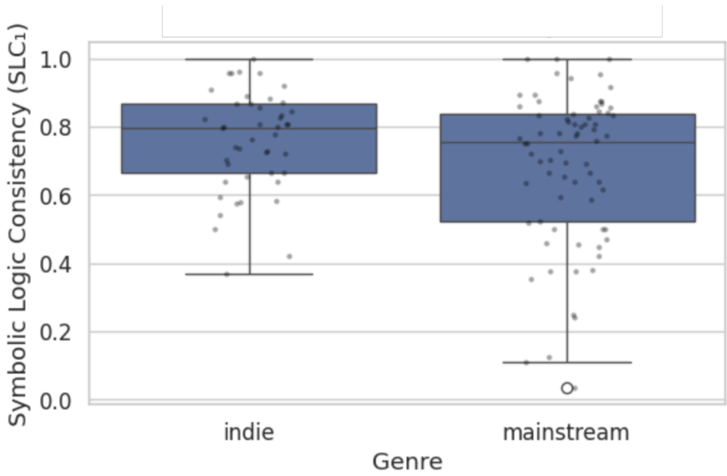

Figure 9: Symbolic Logic Consistency (SLC_1) distributions by genre. Indie songs exhibit higher alignment between window-level GPT-4o-mini inferences and the global CART structure ($M = 0.758$) compared to mainstream tracks ($M = 0.684$)

in contrast, sustain more stable emotional trajectories, reflected by higher ECS, but allow occasional symbolic deviations from the global CART structure, perhaps reflecting production techniques that prioritise affective immediacy over strict logical coherence.

These results demonstrate that symbolic consistency and emotional consistency capture distinct dimensions of neurosymbolic behaviour: indie tracks tend to be emotionally dynamic yet logically structured, while mainstream tracks are emotionally stable but symbolically looser in their rule alignment.

4.12. *GEMS ontology mapping*

To situate the reduced symbolic taxonomy within an established affective framework, the five symbolic categories, namely *Excitement*, *Peace*, *Melancholy*, *Nostalgia* and *Longing*, were mapped onto higher-order clusters from the Geneva Emotional Music Scale (GEMS). Following mapping conventions in affective science (Juslin & Vastfjall 2008; Eerola & Vuoskoski 2012), the symbolic categories were collapsed into three principal GEMS clusters: *Sadness*, absorbing *Melancholy*; *Tenderness*, absorbing *Peace*, *Nostalgia* and *Longing*; and *Vitality*, absorbing *Excitement*. Other GEMS dimensions, such as *Transcendence*, *Wonder* and *Joyful Activation*, were not represented in this dataset, as they were not invoked by the symbolic tagging stage.

Figure 10 presents the proportional distribution of songs across these three clusters by genre. Indie music exhibited a strong tendency towards *Sadness*, representing 51.1% of songs, while mainstream tracks leaned more heavily towards *Vitality*, representing 45.8%, and *Tenderness*, representing 19.4%. Indie music contained fewer *Tenderness*-oriented songs, representing 11.1%, compared with mainstream tracks.

A chi-square test of independence indicated that these differences were not statistically significant, $\chi^2(2, N = 117) = 3.42$, $p = 0.181$, although clear genre-level tendencies were observed. Expected frequencies, namely *Sadness*: 18.5 indie versus 29.5 mainstream; *Tenderness*: 7.3 indie versus 11.7 mainstream; and *Vitality*: 19.2 indie versus 30.8 mainstream, suggested that mainstream songs slightly exceeded expectations for *Vitality* and *Tenderness*, while indie songs exceeded expectations for *Sadness*.

These results suggest that both genres occupy a shared emotional space when viewed through the GEMS ontology, yet differ subtly in emphasis. Indie music skews towards low-valence affective states, consistent with its aesthetic emphasis on *Intimacy* and *Melancholy*, whereas

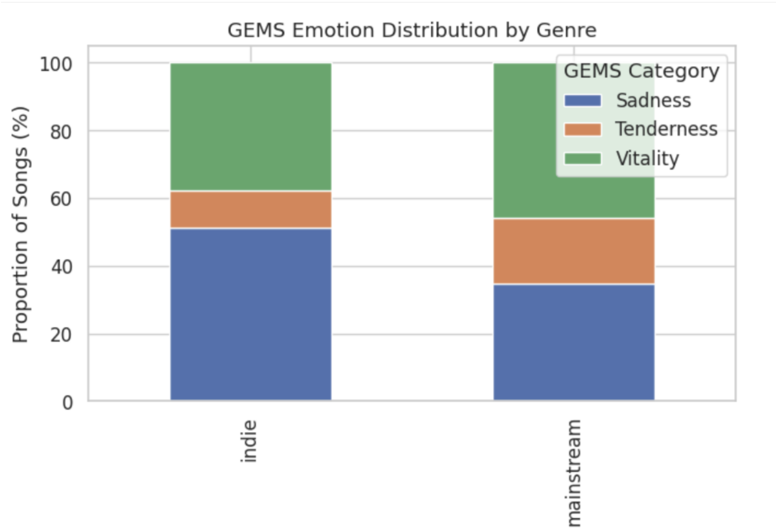

Figure 10: Proportional distribution of GEMS emotional clusters by genre. Indie music exhibits a stronger tendency towards *Sadness* (51.1%), whereas mainstream tracks are more associated with *Vitality* (45.8%) and *Tenderness* (19.4%). However, these differences are not statistically significant ($\chi^2(2) = 3.42$, $p = 0.181$, Cramer's $V = 0.171$)

mainstream music distributes more evenly across *Vitality* and *Tenderness*, aligning with its orientation towards energetic and accessible production.

5. Discussion

5.1. *The sonic divide: production philosophy versus emotional palette*

At the acoustic level, Tamil indie and mainstream romance songs differ more in their variability than in their absolute averages. Mean differences in tempo, RMS energy and spectral centroid were modest and not statistically significant. However, variance analyses revealed narrower interquartile ranges in mainstream tracks, consistent with standardised production practices, while indie songs displayed broader spreads, including markedly slower and darker pieces. This pattern may suggest that genre differentiation in Tamil romance music is expressed primarily through controlled variability rather than shifts in mean acoustic properties, a finding that contrasts with Western-centric MER studies where genre effects are often captured through average feature differences (Eerola & Vuoskoski 2012).

This divergence suggests two contrasting aesthetic logics. Mainstream productions, with their tighter clustering, appear optimised for accessibility and recognisability, aligning with commercial imperatives. Indie music, in contrast, embraces heterogeneity and experimentation, using broader sonic palettes to foster intimacy and stylistic individuality. From a MIR perspective, this aligns with prior observations that production constraints can compress acoustic feature distributions without necessarily altering perceptual emotion outcomes (Serra *et al.* 2012).

Although the framework demonstrates structural coherence, its validation currently depends on model-internal consistency rather than human-expert judgement. Incorporating expert annotation in future iterations will be essential to evaluate the cultural and emotional alignment of GPT-4o-mini symbolic reasoning and to refine its interpretive sensitivity to Tamil affective idioms.

5.2. *The paradox of authenticity: coherence versus complexity*

The neurosymbolic analysis highlights a paradox in how authenticity manifests across Tamil indie and mainstream music. Symbolic Logic Consistency was higher in indie tracks, with GPT-4o-mini local symbolic rules aligning more closely with the global CART structure. Nearly 90% of symbolic triplets mapped deterministically to a single emotion, and indie songs adhered to these mappings with greater reliability. By contrast, mainstream tracks showed slightly lower symbolic alignment but achieved higher Emotional Consistency Scores, maintaining stable single-emotion trajectories. This contrast suggests that symbolic coherence and emotional stability may capture distinct dimensions of affective organisation, supporting prior findings that musical emotion cannot be reduced to a single dominant mechanism (Juslin & Vastfjall 2008). At the corpus level, mainstream tracks also scored higher on the Emotional Authenticity Index, reflecting broader symbolic diversity, while indie songs displayed greater within-song variability, as captured by the Temporal Emotion Flow metric.

These findings suggest that authenticity in Tamil music may operate on two complementary registers. Indie authenticity appears grounded in symbolic coherence: emotional variability is permitted, but it unfolds within a logically stable framework. Mainstream authenticity appears grounded in emotional stability: songs sustain consistent affective trajectories while allowing occasional symbolic deviations. Although the observed effects were modest and, in some cases, descriptive rather than statistically robust, they tentatively point to authenticity as a multidimensional construct in which both coherence and complexity may function as valid markers.

5.3. *Convergence in emotional storytelling: a shared narrative grammar*

Despite the divergences observed in production styles and emotional emphasis, both indie and mainstream Tamil romance songs appear to converge on a common emotional grammar. UMAP visualisations demonstrated that symbolic emotions separated more distinctly than genres, with clustering indices confirming that emotional groupings, namely silhouette = 0.171 and Davies–Bouldin index = 2.48, were more coherent than genre groupings, namely silhouette = -0.015 and Davies–Bouldin index = 43.5. Similarly, chi-square analysis indicated no statistically significant differences in emotion distribution across genres, $\chi^2(2, N = 117) = 3.42$, $p = 0.181$, Cramer's $V = 0.17$, suggesting substantial overlap.

Both genres appear to foreground *Melancholy*, *Vitality* and *Tenderness* as dominant categories, with *Nostalgia* serving as a shared secondary thread. These findings tentatively suggest that Tamil romantic music may draw from a culturally grounded symbolic-emotional repertoire that transcends industrial categorisation. Genre boundaries thus appear as shifts in emphasis rather than categorical divides. In contrast to Western genre-classification studies that report stronger emotion–genre coupling, these results suggest that emotional overlap may be more pronounced in culturally specific repertoires (Eerola & Vuoskoski 2012).

However, these results matter beyond genre comparison. Theoretically, they suggest that emotional authenticity and coherence in Tamil music could be computationally formalised without erasing cultural nuance, representing a step towards inclusive and interpretable models of musical affect. Methodologically, this validates neurosymbolic AI as a bridge between statistical learning and symbolic reasoning, illustrating how LLM-based inference can yield culturally grounded explanations rather than opaque predictions. Practically, such frameworks could inform more equitable music-recommendation systems by recognising non-Western emotional logics and supporting culturally diverse content visibility. Collectively, these implications extend the scope of computational musicology towards an ethically and epistemologically plural vision of AI in culture.

6. Conclusion

This study compared the emotional signatures of Tamil indie and mainstream romantic songs through a neurosymbolic framework that integrates machine learning inference with symbolic reasoning. By discretising core audio descriptors into symbolic tags and applying GPT-4o-mini to infer emotions and generate human-readable rules, the analysis suggests that emotion in Tamil music may be not only measurable but also interpretable through culturally grounded symbolic logics.

At the acoustic level, mainstream tracks clustered more tightly around standardised production values, while indie songs displayed greater variability and experimentation. At the symbolic-emotional level, indie songs appeared to reveal stronger logical coherence amid emotional fluctuation, whereas mainstream songs showed greater emotional stability and broader corpus-level diversity. Despite these differences, both repertoires appear to converge on a shared symbolic-emotional ontology, tentatively underscoring the cultural continuity of Tamil romantic music.

The contributions of this study are threefold. Theoretically, it suggests that authenticity in Tamil music may be multidimensional, potentially realised either through symbolic coherence, as observed in indie music, or through emotional stability, as observed in mainstream music. Methodologically, it illustrates how neurosymbolic AI may bridge predictive performance with interpretability by producing transparent symbolic rules that scale from individual tracks to corpus-level indices. Practically, it highlights how indie and mainstream genres cultivate complementary strategies of audience engagement, with implications for cultural production and music recommendation systems.

These contributions must be understood within the limits of a modest, audio-only dataset. Future research should extend this approach to multimodal corpora that integrate lyrics, video aesthetics and audience reception, while also testing its generalisability across other languages and musical traditions. Scaling neurosymbolic models to larger datasets will also provide a crucial test of whether interpretability can be preserved alongside predictive performance.

In conclusion, this study offers both a comparative account of Tamil indie and mainstream romantic songs and a methodological template for interpretable, culturally sensitive AI. By situating machine reasoning within symbolic-emotional logics, it contributes to a cross-disciplinary agenda that connects musicology, affective computing and cultural studies. Beyond the Tamil corpus, the proposed neurosymbolic pipeline offers a scalable framework for analysing emotional and symbolic dynamics across culturally situated repertoires. By integrating interpretable AI reasoning with domain-specific musical ontologies, this approach may be adapted to other linguistic or regional traditions, extending the scope of computational musicology towards a more inclusive, cross-cultural understanding of affect in sound.

Acknowledgement

The authors would like to thank all individuals and platforms that supported the preparation of the dataset and the development of the computational workflow used in this study.

Appendix A. Reproducibility

A.1 Windowing parameters

Each song was analysed over its first 180 seconds and segmented into non-overlapping windows of 10 seconds each, with WINDOW_SEC = 10 and HOP_SEC = 10, yielding up to 18 windows per track. Feature extraction was performed using `librosa` v0.11.0 within Python 3.10 on Google Colab.

A.2 `gpt-4o-mini` prompt design

All symbolic inferences were generated using the `gpt-4o-mini` model with $\mathrm{temperature} = 0.0$, $\mathrm{top\_p} = 1.0$ and $\mathrm{max\_tokens} = 256$. The `response_format` was set to a JSON object to enforce structured outputs.
The system instruction was:

```
‘‘You are an AI researcher specialising in symbolic reasoning
and music emotion analysis.’’
```

The user prompt template was:

```
‘‘You are an expert in symbolic AI and music emotion modelling.
A Tamil song has the following symbolic audio features: Tempo:
{tempo_tag}, Energy: {energy_tag}, Spectral Centroid: {centroid_tag}.
Choose ONE emotion from this list: [excitement, peace, melancholy,
nostalgia, longing]. Return STRICT JSON ONLY with keys: emotion
(one of the listed emotions), rule (Prolog-style: emotion(<song_id>,
<emotion>) :- tempo(<tag>), energy(<tag>), centroid(<tag>).),
rationale (one sentence). IMPORTANT: JSON only (no prose). Use
lowercase for emotion and tags.’’
```

A.3 CART hyperparameters

The CART decision tree was trained using `scikit-learn` v1.6.1 with the following settings: $\mathrm{criterion} =$ entropy, $\mathrm{max\_depth} = 3$ and $\mathrm{random\_state} = 42$. One-hot encoding was applied to the three symbolic tags, namely `tempo_tag`, `energy_tag` and `centroid_tag`, prior to model fitting.

A.4 UMAP hyperparameters

Dimensionality reduction was performed using UMAP with $n_{\text{neighbors}} = 15$, $\mathrm{min\_dist} = 0.15$ and Euclidean distance as the metric. A fixed random seed was applied to ensure reproducibility.

A.5 Complete song list and code

The complete list of 117 songs, including their Spotify and YouTube identifiers, annotated notebooks and step-by-step documentation, is publicly available at `https://tinyurl.com/rkpyh53c`.

References

Becker J. 2004. *Deep Listeners: Music, Emotion, and Trancing*. Bloomington: Indiana University Press.

Bourreau M., Moreau F. & Wikstrom P. 2022. Does digitization lead to the homogenization of cultural content? *Economic Inquiry* **60**(1): 427–453.

Cavalcanti R.C.T., Souza-Leão A.L.M.D. & Moura B.M. 2021. Fan affirmation: Alethurgy on an indie music fandom. *Revista de Administração Contemporânea* **25**(5): e190395.

Choi K., Fazekas G., Sandler M. & Cho K. 2017. Convolutional recurrent neural networks for music classification. *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 2392–2396. New Orleans, USA: IEEE.

Clayton M. 2016. *The Oxford Handbook of Music Psychology*, 2nd Ed., In cross-cultural perspective, pp. 47–62. Oxford University Press.

Cunnington D., Law M., Lobo J. & Russo A. 2023. Neurosymbolic learning of answer set programs from raw data. *Proceedings of the Thirty-Second International Joint Conference on Artificial Intelligence (IJCAI-23)*, pp. 3586–3596. Macao, China: IJCAI.

Devdas 2002. Devdas (2002 hindi film). Wikipedia, available at: `https://en.wikipedia.org/wiki/Devdas_(2002_Hindi_film)`.

Eerola T. & Vuoskoski J.K. 2012. A review of music and emotion studies: Approaches, emotion models, and stimuli. *Music Perception: An Interdisciplinary Journal* **30**(3): 307–340.

Feng Y., Zhuang Y. & Pan Y. 2003. Popular music retrieval by detecting mood. *Proceedings of the 26th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, pp. 375–376. Toronto, Canada: ACM.

Ferreira L.N. & Whitehead J. 2021. Learning to generate music with sentiment. arXiv preprint arXiv:2103.06125.

Friberg A., Schoonderwaldt E., Juslin P.N. & Bresin R. 2002. Automatic real-time extraction of musical expression. *Proceedings of the International Computer Music Conference (ICMC 2002)*, pp. 365–367. Gothenburg, Sweden: ICMA.

Garcez A.D.A., Gori M., Lamb L.C., Serafini L., Spranger M. & Tran S.N. 2019. Neural-symbolic computing: An effective methodology for principled integration of machine learning and reasoning. arXiv preprint arXiv:1905.06088.

Garcez A.D.A. & Lamb L.C. 2023. Neurosymbolic AI: the 3rd wave. *Artificial Intelligence Review* **56**: 12387–12406.

Guidotti R., Monreale A., Ruggieri S., Turini F., Giannotti F. & Pedreschi D. 2018. A survey of methods for explaining black box models. *ACM Computing Surveys* **51**(5): 1–42.

Hamilton K., Nayak A., Bozic B. & Longo L. 2022. Is neurosymbolic AI meeting its promise in natural language processing? a structured review. *Semantic Web* **15**(4): 1265–1306.

Han Y., Chen J., Wang Y. *et al.* 2021. Symbolic and knowledge-driven music emotion recognition. *Proceedings of the International Society for Music Information Retrieval Conference* .

Harris C.B., Baird A., Harris S.A. & Thompson W.F. 2019. "They're playing our song": Couple-defining songs in intimate relationships. *Journal of Social and Personal Relationships* **37**(1): 163–179.

Hu X. & Lee J.H. 2012. A cross-cultural study of music mood perception between American and Chinese listeners. *Proceedings of the 13th International Society for Music Information Retrieval Conference (ISMIR 2012)*, pp. 535–540.

Jiang Y. & Zheng M. 2024. The inductive effect of musical mode types on emotions and physiological responses: Chinese pentatonic scales versus Western major/minor modes. *Frontiers in Psychology* **15**: 1414014.

Juslin P.N. & Vastfjall D. 2008. Emotional responses to music: The need to consider underlying mechanisms. *Behavioral and Brain Sciences* **31**(5): 559–575.

Ken Research 2024. Tamil music market report. Ken Research.

Kim Y.E., Schmidt E.M., Migneco R., Morton B.G., Richardson P., Scott J., Speck J.A. & Turnbull D. 2010. Music emotion recognition: A state of the art review. *Proceedings of the International Society for Music Information Retrieval Conference (ISMIR)*, pp. 937–952. Utrecht, Netherlands: ISMIR.

Koh E.Y., Cheuk K.W., Heung K.Y., Agres K.R. & Herremans D. 2022. MERP: A music dataset with emotion ratings and raters' profile information. *Sensors* **23**(1): 382.

Li M.G., Olsen K.N. & Thompson W.F. 2025. Cross-cultural biases of emotion perception in music. *Brain Sciences* **15**(5): 477.

Liyanarachchi R., Joshi A. & Meijering E. 2025. A survey on multimodal music emotion recognition. arXiv preprint arXiv:2504.18799.

Lu Z., Huyang S., Wu D., Chen Y., Jin Z. & Li Q. 2019. Impairment of romantic emotion and empathy in schizophrenics: A preliminary study. Preprint.

Marcus G., Bommasani R., Davis E. & ten Oever S. 2022. Neuro-symbolic AI: The third wave. *Communications of the ACM* **65**(9): 54–63.

Meyer L.B. 1956. *Emotion and Meaning in Music*. Chicago: University of Chicago Press.

Munbe Vaa 2006. Munbe vaa. Spotify audio track, available at: `https://open.spotify.com/track/6R05n6BoyHIRhsOEUav5q5`.

Nattiez J.J. 1990. *Music and Discourse: Toward a Semiology of Music*. Princeton: Princeton University Press, translated by C. Abbate.

Russell J.A. 1980. A circumplex model of affect. *Journal of Personality and Social Psychology* **39**(6): 1161–1178.

Serra J., Corral A., Boguna M., Haro M. & Arcos J.L. 2012. Measuring the evolution of contemporary western popular music. *Scientific Reports* **2**(1): 1–6.

Spotify 2025. Regional streaming trends. Spotify Newsroom.

Stappen L., Schumann L., Batliner A. & Schuller B.W. 2021. Embracing and exploiting annotator emotional subjectivity: An affective rater ensemble model. *2021 9th International Conference on Affective Computing and Intelligent Interaction Workshops and Demos (ACIIW)*, pp. 1–8. IEEE.

Trigeorgis G., Ringeval F., Brueckner R., Marchi E., Nicolaou M.A., Schuller B. & Zafeiriou S. 2016. Adieu features? End-to-end speech emotion recognition using a deep convolutional recurrent network. *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 5200–5204. Shanghai, China: IEEE.

Usurae Pogudhey 2019. Usurae pogudhey. Spotify audio track, available at: `https://open.spotify.com/track/5jOi67yXzjpMdl6KFNWSIe`.

Vaseegara 2001. Vaseegara. Spotify audio track, available at: `https://open.spotify.com/track/34IgZueFcAjJ1xwVxH1ANy`.

Velayudham 2011. Velayudham. Wikipedia, available at: `https://en.wikipedia.org/wiki/Velayudham`.

Wieczorkowska A., Datta A.K., Sengupta R., Dey N. & Mukherjee B. 2010. *On Search for Emotion in Hindusthani Vocal Music*. Polish-Japanese Institute of Information Technology & ITC Sangeet Research Academy.

Xie L., Wang Y. & Gao Y. 2024. Acoustical feature analysis and optimization for aesthetic recognition of Chinese traditional music. *Journal of Audio, Speech and Music Processing* **7**: 1.

Yang Y.H. & Chen H.H. 2012. Machine recognition of music emotion: A review. *ACM Transactions on Intelligent Systems and Technology* **3**(3): 1–30.

Zentner M., Grandjean D. & Scherer K.R. 2008. Emotions evoked by the sound of music: Characterization, classification, and measurement. *Emotion* **8**(4): 494–521.

Department of Mathematical Sciences
Faculty of Science and Technology
Universiti Kebangsaan Malaysia
43600 UKM Bangi
Selangor, MALAYSIA
*E-mail: p160285@ukm.edu.my, aftar@ukm.edu.my, bernardlkb@ukm.edu.my**

Received: 4 October 2025
Accepted: 23 May 2026

*Corresponding author