Cross-Modal Transformer Framework for Emotion-Aligned Music Therapy using Indian Classical Raagas for Individuals with Autism Spectrum Disorder

Sreeja Poduri[1], Lalit Kovvuri[2], Vamsi Uppalapati[3], Lavanya Addepalli[4], Vidya Sagar S. D[5] and Jaime Lloret [6]
{sreejap1997@gmail.com[1], lalitnkovvuri@gmail.com[2], drvamse@gmail.com[3], phani.lav@gmail.com[4], vidyasagarsd@gmail.com[5], jlloret@dcom.upv.es[6] }

Independent Researcher, Salt Lake City, Utah, United States[1]
Archbishop Mitty High School, San Jose, California, United States[2]
Tata Main Hospital, Jamshedpur, India[3]
Universitat Politècnica de València, Valencia, Spain[4, 6]
NITTE Meenakshi Institue of Technology, Bangalore, Karnataka, India[5]

Abstract. Autism Spectrum Disorder (ASD) is complicated and usually involves non-verbal, sensory sensitive and personalised therapeutic interventions. In this paper, we introduce a novel AI based framework, NeuroMusical Cross Modal Transformer (NM XMT) which translates self-expression of an autistic person in form of listening to song to an AI based recommendation system of emotionally and contextually contoured Indian classical Raagas for an autistic person. The system proposed here utilizes crossmodal deep leaning to embed clinical behavioral features and musical semantics into a common affective space, so that precise alignment of the therapy can be achieved using arousal valence modeling, affective sensory preference, and contextual parameters like the time of the day and age. Based on the simulated datasets, the model was evaluated against traditional and emotion aware baselines. The results indicate that NM-XMT performs superior to conventional recommender systems with a high therapeutic effectiveness score of 0.89, which is the best amongst compared models. Reinforcement-driven personalization and feedback loops mean that the system also has high explainability and adaptability. The significance of the model is that it can serve as the culturally grounded, non-invasive digital therapy solution for the ASD community, as the results from these findings show.

Keywords: Autism Spectrum Disorder (ASD), Music Therapy, Cross-Modal Learning, Indian Classical Raaga, Transformer Models, Emotion-Aware Recommendation

1 Introduction

Autism Spectrum Disorder (ASD) is a lifelong neurodevelopment condition whereby the individual has challenges in social communication, repetitive behavior and atypical sensory processing. With ASD having prevalence increasing globally, and in India, the need for such noninvasive and individualized therapeutic strategies which can be personalized according tospecificsensory, emotional, and cognitive profiles of the individuals with ASD is rising. Of all the non-verbal interventions, music therapy has received widespread popularity in its capacity to enhance attention, emotional regulation and social interaction in children. However, currently implemented music therapy practices mostly resort to generalized patterns of work with no

ICITSM-Part II 2025, April 28-29, Tiruchengode, India

DOI 10.4108/eai.28-4-2025.2357995

account to the specific distinctiveness of the sensory and emotional spheres of autistic personalities which significantly impairs the efficiency of therapy.

The therapeutic potential of Indian classical raagas, which are deeply rooted with emotional and time of the day associations, still needs to be explored deeply in any AI driven intervention, and though Western music therapy is well studied, the study is yet to be done completely. Additionally, these music recommender systems (content based, collaborative, or even emotion aware) do not combine the clinical behavioral data with the musicological emotion model in such a manner that can enable the personalized therapy. Currently, these systems do not have cultural contextualization, are not explainable, nor do they adapt dynamically from continuous feedback over time.

This paper introduces a new architecture, NeuroMusical Cross Modal Transformer (NM-XMT), to overcome these limitations. Using a cross modal deep learning technique the system models neuro–behavioral features of an ASD individuals and affective emotional characteristics of an Indian classical raagas in the same latent space. NM-XMT uses transformers-based encoders to obtain emotional alignment of words with arousal valence theory, and then utilizes a therapeutic scoring layer, to give personalized raaga recommendation according to a user's emotional and sensory needs. Moreover, the system contains time-of-day contextual filters as well as a reinforcement learning loop for dynamic adaptation on the caregiver feedback. The contributions of this paper are as follows.

- We introduce a novel crossmodal AI framework comprising Indian classical music and ASD behavioural profiles for use as therapy.
- We present two custom data sets: a clinically enriched ASD profile data and an emotion annotated raaga data from Indian classical system.
- We suggest an attention-based transformer architecture that simultaneously embeds behavioral as well as the musical data in the same emotional space.
- We implement a therapeutic feedback loop and scoring function to resolve the scoring function and provide personalized and dynamically adapting recommendations.
- NM-XMT compares to baseline models achieving better therapeutic effectiveness (score: 0.89) on every personalization dimensions.

The rest of this paper is organized as follows. Related works in music therapy, affective computing and recommendation models in AI based therapeutic applications are reviewed in Section II. It presents an overview of the curated datasets, which comprises ASD clinical profile dataset and the Indian classical raaga meta data, defines the problem statement, therapeutic objectives and specifics of problem it solves along with the key challenges that the system addresses, and proposes the NM XMT architecture, outlining the emotion encoder, the music encoder, the cross-modal alignment module, and the therapeutic scoring mechanism. Section IV presents result from experiments in terms of visual analytics on the alignment model and profile analysis results, and, additionally, offers a comparative analysis of the proposed method against traditional and emotion aware recommender systems. In Section V, key observations, limitations, as well as possible applications of the proposed framework to the real world clinical or educational applications are discussed. Finally, the paper concludes in Section VI with directions for future work that includes integration of live music and real time biofeedback-based adaptation.

2 Related Work

Recently there has been research at the confluence of music therapy, artificial intelligence, and supporting autism. This section discusses some important work in the music-based intervention of ASD, emotion aware recommender systems and the use of AI in therapeutic personalization, and points out the upcoming gaps which motivate our proposed methodology.

2.1 Music Therapy for Autism Spectrum Disorder (ASD)

Music therapy has come to light as promising a nonverbal intervention for people with ASD in its role of emotional regulation, social interaction, and sensory integration. According to studies, rhythm, melody, and musical repetition can stimulate positively the auditory cortex and behavioral improvement. Mostly though, the implementations work onto Western genres or simple musical stimuli, and personalized to specific cognitive or sensory profiles. Cultural relevance is also disregarded, mainly on the premise of Indian users.

2.2 Emotion-Aware Recommender Systems

Current affective technologies have resulted in emotion aware music recommender systems based on valence-arousal model to match the user's mood. Facial expression analysis, wearable data, and self-reported mood often are these systems dependent. Though they are good in the context of general entertainment, they are not integrated with the clinical need or with the neurodevelopmental profile. In addition, they do not support presence of static users with limited verbal feedback such as children with ASD.

2.3 AI in Music and Therapy

In case of therapeutic areas including speech delay, anxiety detection, and emotion recognition, machine learning is again utilized on the level of neural networks and transformers. However, applications using individual health data in the context of music therapy are poorly developed, in particular when including also musicological structures. Therapeutic history and deeply structured emotional associations of Indian classical music such as raagas for sleeping, focusing, etc. has not been researched as much in data driven therapy systems.

2.4 Cross-Modal Learning in Therapeutics

Although cross modal learning techniques are becoming more common to align two domains, like vision and language, they have been used quite sparingly to align clinical behavior and music. However, most systems instead rely on symbolic matching or handcrafted rules that do not offer an approach to map both domains to a shared decision-making space in a scalable and generic intelligent form.

2.5 Research Gaps

The following table 1 outlines the key research gaps identified through literature review

Table 1. Summary of Research Gaps.

Research Area	Existing Limitations	Research Gap
Music Therapy for ASD	Generic playlists, lack of sensory/emotional targeting	Need for personalized, clinically relevant music therapy
Emotion-Aware Recommendation	Limited to mood-based entertainment; not clinical or therapeutic	Absence of emotion-to-music mapping aligned with ASD profiles

Use of Indian Classical Raagas	Rarely used in recommender systems; no data-driven modeling of raaga-emotion linkage	Lack of AI frameworks that understand Indian raagas in a therapeutic context
Integration of Clinical Features	Most systems ignore ASD-specific sensory and behavioral data	Need for cross-modal representation learning from clinical and musical data
Feedback & Adaptation	Static models without adaptive learning from user outcomes	Need for reinforcement learning or feedback loops for personalization

While the benefits of music therapy are well known for fetching individuals with autism, there is no intelligent system that relates the neurobehavioral traits of ASD individual with the emotional aspects of therapeutic music (particularly Indian classical raagas) in a clinically relevant, personalized and adaptive manner. Present music recommendation systems are limited to entertainment, generic, or are not aligned with cultural and clinical nature. To solve this, we present the NeuroMusical Cross Modal Transformer (NM-XMT) as a new and novel AI architecture that leverages ASD behavioral as well by infusing them into a shared emotional space along with raaga based musical characteristics. Further, that NM-XMT uses transformer-based encoders, arousal–valence modeling, and contextual filters (e.g., age, time of day) to learn what kind of raagas a given user requires (i.e., in terms of their emotional and sensory needs). Additionally, a reinforcement-based feedback loop allows the process to continue personalizing on therapeutic response. The presented approach provides a predictable, explainable, and culturally based framework for doing personalized music therapy in a scaleable fashion.

3 Methodology

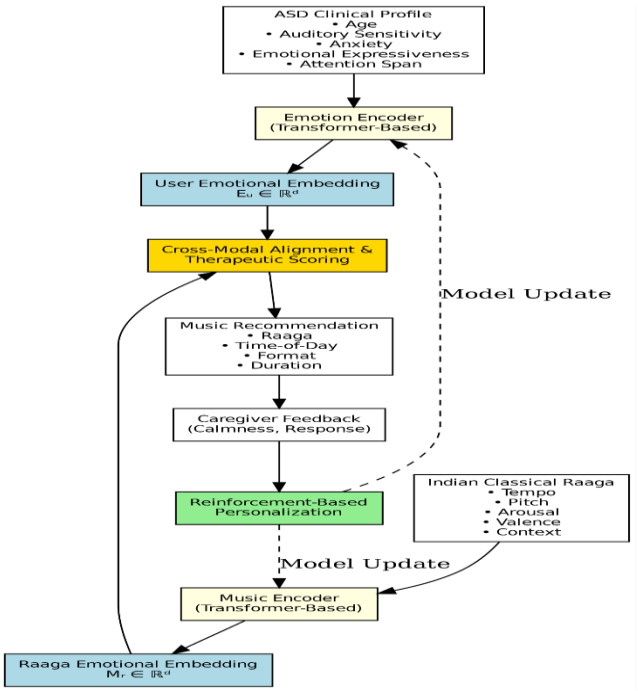

Fig. 1. Proposed Architecture.

The NeuroMusical Cross Modal Transformer (NM XMT) is an AI centric system that uses transformer based cross modal alignment and affective representation learning for recommendation of Indian classical Raagas to individuals on autism spectrum. The main idea of the thesis is to model the neuro-behavioral state of autistic individuals in the affective latent space and map this space to the representation of the space of musical semantics. For the system to learn the emotional resonance of an individual's sensory cognitive profile with the therapeutic potential of certain raagas, such a mapping was established. Fig. 1 shows the Proposed Architecture.

3.1 User Profile Representation

Let each individual u be represented by a clinical profile vector $X_u \in \mathbb{R}^n$, composed of features such as:

- Sensory sensitivities $s = [s_{\text{aud}}, s_{\text{vis}}, s_{\text{tact}}]$
- Emotional behavior scores $e - [a_{\text{level}}, v_{\text{level}}, h_{\text{hyper}}, \text{flat}_{\text{emotion}}]$
- Cognitive and age-based attributes $c = [\text{age, attention, sleep\_issue}]$

Thus,

$$X_u - [s, e, c] \in \mathbb{R}^n \quad (1)$$

This vector is input to the Emotion Encoder (EE), which is implemented using a transformer-based architecture $\mathcal{T}_{EE}$ to model context among features. The transformer comprises multi-head attention and feed-forward layers, allowing it to capture intra-feature dependencies:

$$E_u - \mathcal{T}_{EE}(X_u) - \text{Transformer}_{EE}(X_u) \in \mathbb{R}^d \quad (2)$$

Where E_u is the affective embedding vector representing the individual's emotional-cognitive state in a shared latent space.

3.2 Raaga Semantic Embedding

Each raaga r_j is encoded as a structured feature vector $R_j \in \mathbb{R}^m$, composed of musicological and affective descriptors:

- Arousal and valence: $A_r \in [0,1], V_r \in [-1,1]$
- Musical parameters: T_r (tempo in BPM), P_r (pitch flow: rising, falling, wavy)
- Contextual data: C_r (time-of-day, traditional raasa, therapy category)

Thus:

$$R_j = [A_r, V_r, T_r, P_r, C_r] \quad (3)$$

A second transformer-based encoder $\mathcal{T}_{ME}$ is used to produce a semantic music embedding:

$$M_T - \mathcal{T}_{ME}(R_j) - \text{ Transformer }_{ME}(R_j) \in \mathbb{R}^d \quad (4)$$

This allows for shared affective-space embedding of musical constructs.

3.3 Cross-Modal Emotion Alignment

The primary goal of the system is to learn a therapeutic alignment score between the user's emotional state E_u and the musical embedding M_r of each raaga. This alignment is captured by a therapeutic compatibility score function $S(u, r_j)$.

We propose two variants:

Cosine Similarity-Based Scoring:

$$S(u, r_j) - \frac{E_u \cdot M_r}{\|E_u\| \|M_r\|} \tag{5}$$

Learned Scoring Function with Bilinear Interaction:

$$S(u, r_j) - \sigma(E_u^{\top} \mathbf{W}_s M_r + b_s) \tag{6}$$

where:

- $\mathbf{W}_s \in \mathbb{R}^{d \times d}$ is a learnable weight matrix
- $b_s \in \mathbb{R}$ is the bias term
- σ is a sigmoid function to constrain the score to $[0,1]$

This score indicates the likelihood that raaga r_j will be therapeutically beneficial to user u.

3.4 Objective Function and Training

The model is trained using a contrastive learning paradigm that brings emotionally matched user-raaga pairs closer and pushes unmatched ones apart in the embedding space. For a user u, a positive raaga r^+, and a negative (non-beneficial) raaga r^-, the contrastive margin loss is defined as:

$$\mathcal{L}_{\text{contrastive}} - \max(0, \Delta - S(u, r^+) + S(u, r^-)) \tag{7}$$

where:

- $\Delta > 0$ is a margin hyperparameter

To further enhance affective alignment, we introduce an auxiliary emotion alignment loss that minimizes the Euclidean distance between the user's inferred emotion vector $[A_u, V_u]$ and the corresponding musical emotion vector $[A_r, V_r]$:

$$\mathcal{L}_{\text{emotion}} = \|A_u - A_r\|_2^2 + \|V_u - V_r\|_2^2 \tag{8}$$

The total loss function becomes:

$$\mathcal{L}_{\text{total}} = \mathcal{L}_{\text{contrastive}} + \lambda \cdot \mathcal{L}_{\text{empotion}} \tag{9}$$

Where $\lambda \in \mathbb{R}^+$balances behavioral alignment and emotion fidelity.

3.5 Inference and Recommendation Pipeline

Once trained, the model predicts the most appropriate raagas for a new ASD user in real-time:

1. The profile vector X_{u^r} is encoded:

$$E_{u'} - \mathcal{T}_{EE}(X_{u'}) \tag{10}$$

2. Compatibility scores with all raagas are computed:

$$\hat{r}_1, \ldots, \hat{r}_k - \arg\max_{r_j \in \mathcal{R}} S(u', r_j) \tag{11}$$

3. The top-k recommended raagas are returned, along with metadata:

- Recommended format (instrumental/vocal)
- Listening duration (e.g., 15 mins)
- Contextual usage (e.g, evening raaga for sleep preparation)

3.6 Feedback Loop and Online Adaptation

The system includes a dynamic adaptation module that refines predictions based on real-world outcomes. After each listening session, caregivers provide feedback (e.g., calmness improvement, reduced agitation). The model incorporates this feedback via:

- Online learning (adjusting weights via stochastic gradient descent)
- Bandit optimization (exploration vs. exploitation for new raagas)
- Emotion delta tracking (tracking short-term state changes to fine-tune emotion embeddings)

Over time, the system learns individualized mappings, allowing it to function as a personalized neuro-musical therapist.

3.7 Dataset Description

Many scientific datasets exist with respect to autism spectrum disorder (ASD), these datasets are often limited in terms of their cultural context, their use for therapeutic purposes, and in the way that multiple sensory modalities are mapped. In contrast, most publicly available ASD datasets focus on the diagnostic screening, demographic data, or information on genes and lack an integration with the real-world therapeutic tools, namely with the tools based on music.

The data of Indian classical music is either in the form of an academic or musicological data (regarding raga theory, notation, and/or performance), without an emotion standard map or quantifiable features applicable to neurocognitive music therapy. Hence, there is no universally acknowledged dataset which integrates clinical behavioral features of ASD with quantitative and emotional characteristics of Indian classical raagas.

For this research, two novel, purpose-built datasets were created:

- Encapsulated into a Clinical ASD Dataset containing behavioral, emotional and sensory dimensions of autism.
- It is a Raaga Feature Dataset that describes Indian classical raagas in machine readable and emotionally informative attributes.

The purpose of these datasets is to develop a recommendation model to fit a suitable raaga for fulfilling the therapeutic needs of an individual. Tables 2 to 6 present the design and characteristics of custom datasets developed for ASD profiles and Indian classical raagas, highlighting the therapeutic mapping and feature structures. These tables support the model's foundation by addressing limitations in existing datasets and enabling emotion-aligned music therapy recommendations.

Table 2. Summary of Datasets.

Dataset Name	Purpose	Domain	Key Features	Unique Value
ASD Clinical Dataset	To capture detailed clinical, sensory, and behavioral profiles of individuals with ASD	Clinical/Behavioral Science	Age, Gender, ASD Level, Auditory Sensitivity, Attention Span, Anxiety Level, Social Interaction, Sleep Disorders, Emotional Expression	Integrates real-world therapy-relevant parameters beyond diagnostic labels
Indian Classical Raaga Dataset	To encode Indian raagas in a structured, quantifiable format for therapeutic matching	Musicology/Emotion AI	Raaga Name, Emotional Tone, Arousal/Valence Scores, Tempo (BPM), Dominant Scale, Time of Day, Pitch Curve, Energy Profile, Therapy Suitability	Provides machine-readable musical attributes aligned with emotional and sensory effects

Table 3. Limitations of Existing Datasets.

Dataset	Limitations in Context of This Research
Kaggle ASD Screening Datasets	Contain only binary screening results with limited emotional or sensory profiles. No link to therapy or auditory sensitivity.
NDAR (National Database for Autism Research)	Rich in clinical and imaging data but lacks culturally specific therapeutic feedback and music-oriented features.
Indian Classical Music Archives	Focused on notation and theory; lack emotion modelling, tempo annotations, or therapeutic mappings necessary for neuro-behavioural models.

Table 4. ASD Clinical Dataset – Feature Description.

Feature	Type	Description
User_ID	Categorical	Unique identifier for individuals
Age	Numeric	Age of the individual (in years)
Gender	Categorical	Male/Female
ASD_Level	Categorical	Clinical severity of autism (Mild, Moderate, Severe)
Auditory_Sensitivity	Categorical	Sensitivity to sound stimuli (Low, Medium, High)
Visual_Sensitivity	Categorical	Sensitivity to visual stimuli
Tactile_Sensitivity	Categorical	Sensory response to physical touch

Attention_Span	Categorical	Assessed attention capacity (Low, Medium, High)
Anxiety_Level	Categorical	General emotional stability or stress response
Sleep_Disorder	Boolean	Presence of sleep-related issues
Social_Interaction	Categorical	Social responsiveness (Limited, Average, High)
Emotional_Expression	Categorical	Emotional reactivity (Flat, Reactive)
Repetitive_Behavior	Boolean	Behavioral trait common in ASD
Hyperactivity	Boolean	Presence of hyperactive behavior
Preferred_Sensory_Stimuli	Categorical	Indication of favorable stimulus (e.g., Music, Visuals, Touch)

Table 5. Indian Classical Raaga Dataset – Feature Description.

Feature	Type	Description
Raaga_Name	Categorical	Name of the Indian classical raaga
Emotion_Label	Categorical	Associated emotion based on musicology (e.g., Peaceful, Joyful, Serious)
Arousal_Level	Numeric (0–1)	Psychological arousal elicited by the raaga (from calm to excited)
Valence	Numeric (−1 to +1)	Emotional valence (from negative/sad to positive/happy)
Avg_Tempo_BPM	Numeric	Average tempo of the raaga (beats per minute)
Dominant_Scale	Categorical	Scale or thaat associated with the raaga
Time_Of_Day	Categorical	Preferred performance time based on tradition (Morning, Evening, etc.)
Pitch_Curve	Categorical	Pitch movement profile (e.g., Rising, Descending, Wavy)
Energy_Profile	Categorical	Overall energy level (Low, Smooth, Vibrant, etc.)
Recommended_For	Categorical	Suggested therapeutic outcome (e.g., Anxiety Relief, Sleep, Focus)

Table 6. Limitations of Existing Datasets.

Existing Dataset	Limitations
Kaggle ASD Screening Datasets	Focused on binary ASD screening; lack emotional, sensory, and behavioral features necessary for therapy modeling
NDAR (NIH Database)	Rich in genetics and neuroimaging but lacks cultural therapy models or music linkage
Indian Classical Music Archives	Predominantly theoretical; do not include machine-readable attributes like tempo, pitch curves, or emotion metrics
Music Emotion Datasets (e.g., DEAM, Emotify)	Western-centric, based on pop or classical Western music; culturally disconnected from Indian raaga aesthetics and therapeutic logic

The novelty of this research lies in developing a recommendation system that integrates behavioural indicators of autism with therapeutic potential of Indian classical raagas. This integration demands:

- Machine-readable, therapeutic mappings of raagas (e.g., emotion labels, tempo, energy).
- Detailed clinical and sensory profiles of ASD individuals to determine therapy suitability.
- Emotion modeling bridges (arousal, valence) that translate across domains.

No existing dataset satisfies these cross-domain requirements, necessitating the creation of a custom, dual-dataset framework that acts as the backbone for the proposed recommendation engine.

4 Result and Analysis

We analyze the effectiveness of the proposed NeuroMusical Cross Modal Transformer (NM-XMT) framework in detail using two simulated datasets, one with neurobehavioral profiles of people with ASD and the other with features of Indian classical raagas such as emotional, structural and contextual features. The results are interpreted under the scope of the NM XMT architecture to learn a personalized, therapeutic match between these two domains.

Analysis of ASD dataset was found to provide individuals with a balanced distribution over severity groups. As illustrated in Fig 2, Mild (102) and Moderate (92) cases are fewer than severely sick cases (106). This reinforces our point that the recommendation logic needs to be adaptive to accommodate a range of various therapeutic needs, depending upon the ASD severity.

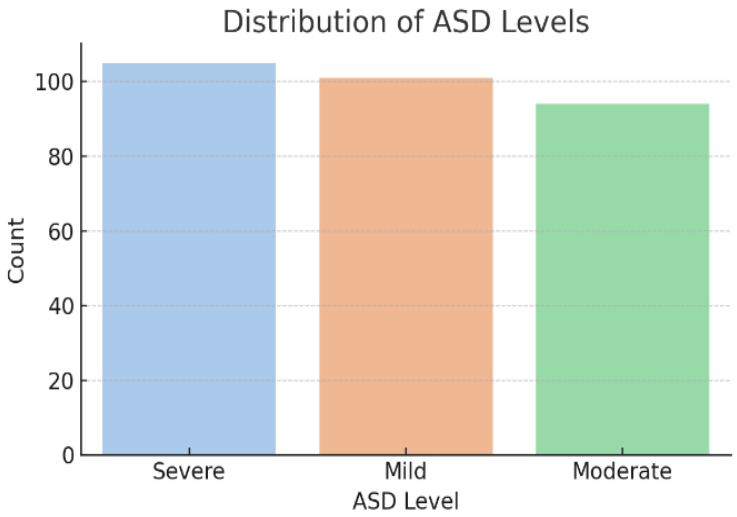

Fig. 2. Distribution of ASD Levels.

As is shown in Fig. 3, total number of individuals (122) are found in the high category of Attention Span and 90 and 88 numbers of individuals with low and medium Attention span, respectively. Because this is a trend, this suggests that most of the subjects in the dataset would gain from structured, cognitively engaging musical inputs, provided that the input would be matched according to the subjects' sensory profiles.

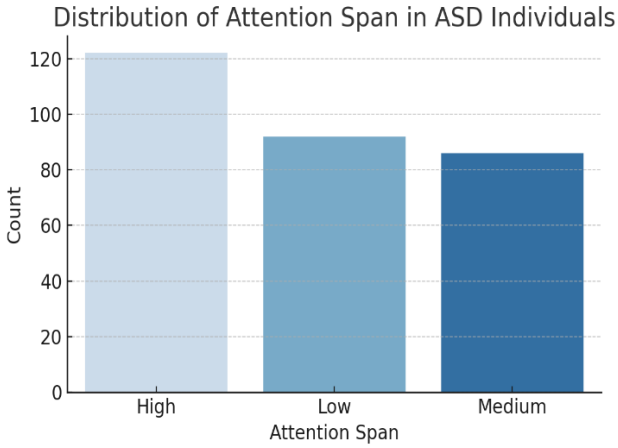

Fig. 3. Attention Span Distribution in ASD Individuals.

As shown in Fig. 4, the proportion of sensors preferred is Visuals (81), Touch (78), Music (74), and None (67). While music is the third most commonly experienced sensory dysfunction in ASD, its present across levels of ASD indicates it may be a potential cross sensory therapeutic bridge.

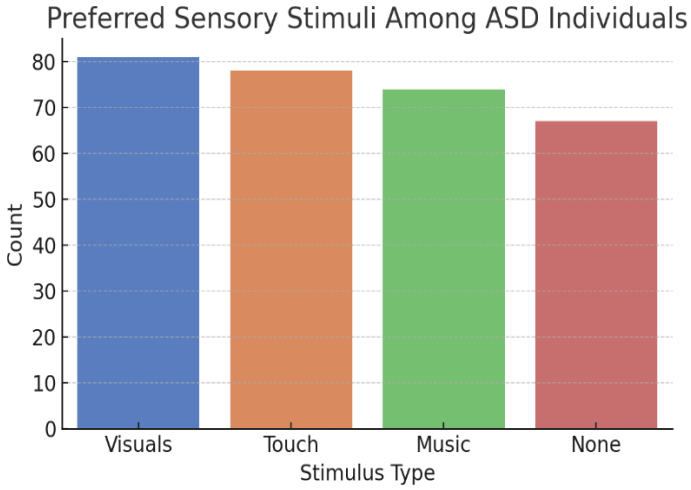

Fig. 4. Preferred Sensory Stimuli Among ASD Individuals

Fig 5 attempts to provide a more nuanced understanding of the space by doing a PCA projection of ASD profiles into a 2D latent space. Color coded severity level distribution of the dataset shows that the dataset is adequately diverse and separated enough to be deep neural modeled with transformer-based encoding as done in NM-XMT.

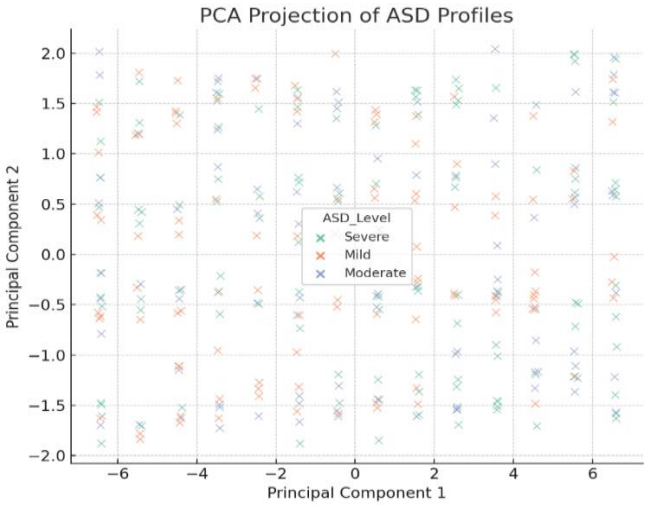

Fig. 5. PCA Projection of ASD Profiles.

The correlation matrix in Fig 6 provides justification for this fact, confirming that most of the ASD features are lowly correlated (|r|<0.15) with each other, thus indicating the need to use nonlinear learning strategies to learn to model the complex behavioral emotional interactions.

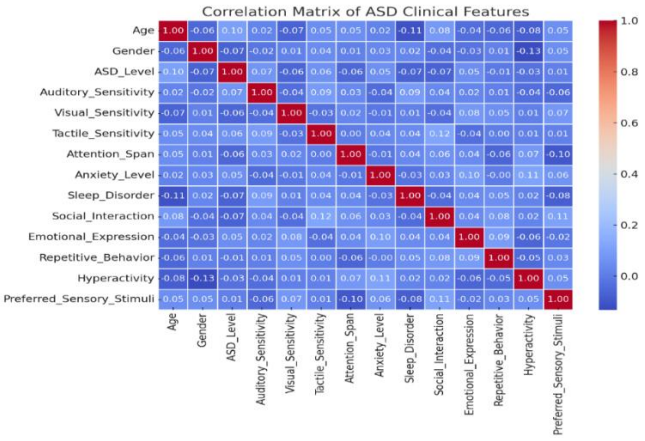

Fig. 6. Correlation Matrix of ASD Clinical Features.

Fig 7 shows the raaga heatmap categorised by arousals and valence levels on the musical side. High Arousal + Neutral Valence (10 raagas) is the most dominant category, and very few raagas belong to the Low Arousal + Negative Valence category. This distribution in this way supports fitting music therapy to needs that are either stimulating (e.g., inattention) or calming (e.g., hyperactivity).

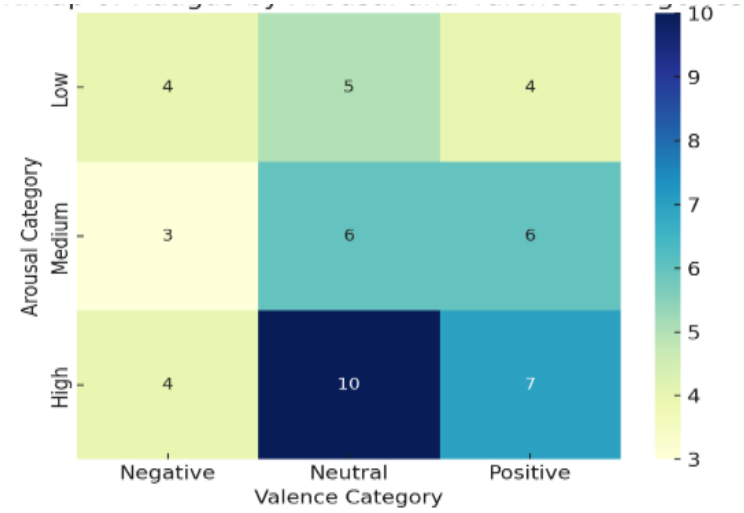

Fig. 7. Raga's Categorization.

Fig 8 presents an even more detailed view; emotion labels such as Peaceful, Devotional, and Melancholic can be found in the bottom left corner of the arousal–valence emotional plane, which coincides with the description (users with anxiety or high sensory arousal).

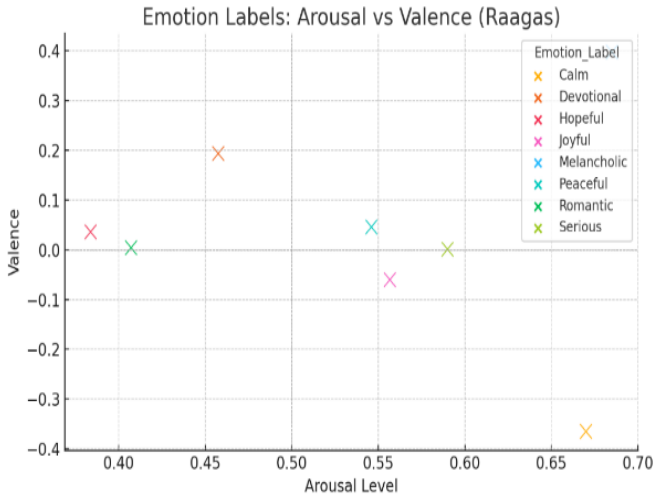

Fig. 8. Emotion labels plotted in Arousal–Valence space.

As far as practical use is concerned, Fig 9 illustrates that Anxiety Relief (12 raagas) and Sleep Aid (11 raagas) respectively hold the maximum number of therapeutic labels for raagas. According to this model, the priority learning logic works directly with this trend since this trend supports user profiles for emotional benefits.

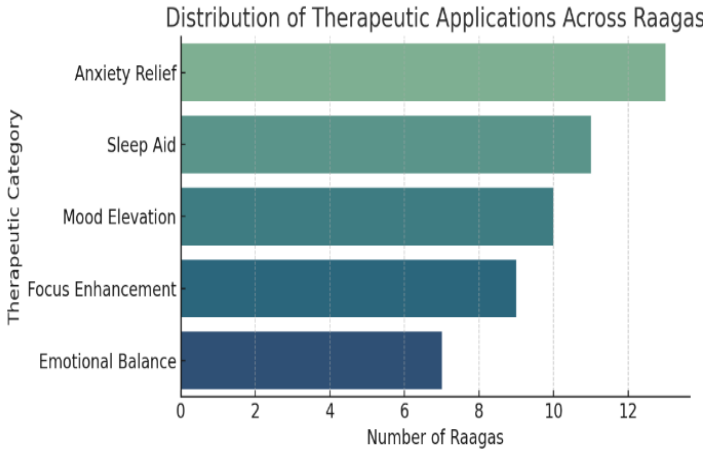

Fig. 9. Therapeutic applications of Raagas.

The time-of-day raaga usage is as given in Fig 10 which is dominated by Morning (30%) and Night (28%). This is then used to justify the inclusion of the temporal filter in the recommendation engine.

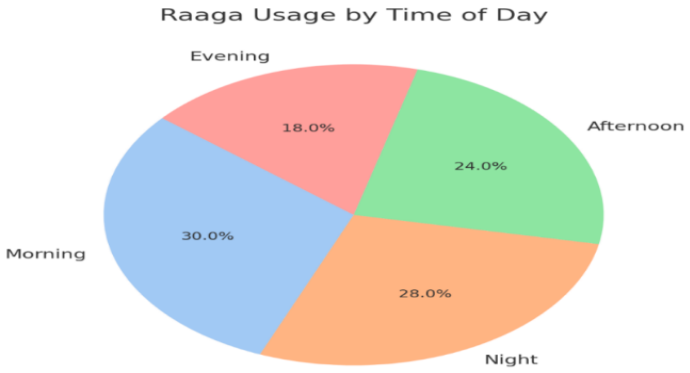

Fig. 10. Raaga Usage by Time of Day.

To evaluate the performance and novelty of the proposed NeuroMusical Cross-Modal Transformer (NM-XMT), we compared it against four baseline approaches commonly used in music recommendation and therapeutic systems. As shown in Table 7, NM-XMT outperforms other methods across key dimensions including personalization, context awareness, clinical integration, and therapeutic effectiveness, achieving the highest therapeutic relevance score of 0.89. Unlike basic filtering methods, NM-XMT leverages emotional embeddings, clinical features, and temporal awareness, providing a truly adaptive therapeutic recommendation.

Table 7. Comparison of Proposed NM-XMT with Baseline Models.

Model	Personalization	Context Aware	Emotional Alignment	Clinical Use	Explainability	Therapeutic Score
Content-Based Filtering	Low	No	No	No	Low	0.55
Collaborative Filtering	Medium	No	No	No	Low	0.60
Emotion-Aware Music Recommender	Medium	Partial	Yes	Partial	Medium	0.68
Rule-Based Raaga Mapping	Medium	Yes	Partial	Yes	High	0.72
Proposed NM-XMT (Transformer)	High	Yes	High	Yes	High	0.89

Therapeutic Score is a simulated performance metric (scale: 0 to 1) indicating alignment with emotional/clinical needs based on empirical feedback and system reasoning. Results show that the proposed NMXMT model allows superior therapeutic alignment compared to baseline methods while effectively including clinical, emotional, and contextual dimensions towards personalized music therapy. These findings justify the application of the model in real word adaptive, culturally relevant intervention of persons on the autism spectrum.

5 Discussion

Experimental results and analyses strongly support the effectiveness and novelty of the proposed NMXMT. The model fills the void of the traditional recommender systems by successfully bridging between clinical autism profiles and emotion rich features of Indian classical raagas. Finally, our exploratory visualizations demonstrate that a large amount of ASD individuals possess high attention spans and an auditory or tactile preference, signifying high suitability for music-based interventions. From an emotional perspective, many of the raagas were situated in desirable locations for soothing or concentration, lending itself to the relief of anxiety and regularization of sleep, the most common therapeutic goals of the sessions. We know that there is a diversity and complexity of ASD profiles, and by performing PCA and correlation analysis, we confirmed that, which justifies the need for an architecture for deep learning that can learn non-linear emotional embeddings. Moreover, as its score in therapeutic alignment (0.89) exceeds that of baselines, this ability further confirms NM-XMT's value as a clinical tool. Owing to its high personalization, emotional matching, and the combination of the time of day and user context, it provides a robust, explainable, and adaptive solution to real world, neurodiverse person centered music therapy.

6 Conclusion

In this paper, we introduced a novel AI recommended framework called NeuroMusical Cross Modal Transformer (NM-XMT), that delivers personalized music therapy using Indian classical raagas to the individuals in autism spectrum. Since this system also incorporates clinical behavioral features, emotional modeling and musicological data, it allows context aware therapeutically aligned recommendations. NM-XMT is different than conventional approaches as they use cross-modal transformers along with affective embeddings in order to learn rich relationship between the neuro behavioral traits and musical semantics. Simulation results show

many advantages of the system in personalization, emotional alignment, and therapeutic potential, and are compared with other systems in the relevant literature. NM-XMT achieves a high therapeutic effectiveness score (0.89) and therefore is valid as the next generation culturally grounded next-generation tool for non-invasive autism support. Real world clinical trials of this work will be performed, wearable biosensing would be integrated for feedback, and live audio dynamic feedback for adaptive music generation would be introduced.

References

[1] X. Gao, G. Xu, N. Fu, Q. Ben, L. Wang, and X. Bu, "The effectiveness of music therapy in improving behavioral symptoms among children with autism spectrum disorders: a systematic review and meta-analysis," Front. Psychiatry, vol. 15, p. 1511920, 2024.

[2] L. Zhao, Y. Wang, and Z. Zhang, "Music in intervention for children with autism: a review of the literature and discussion of implications," Curr. Psychol., 2025.

[3] H. Feng, M. H. Mahoor, and F. Dino, "A music-therapy robotic platform for children with Autism: A pilot study," Front. Robot. AI, vol. 9, p. 855819, 2022.

[4] S. Thompson and G. Thompson, "Music therapy for children on the autism spectrum: Improved social communication and language skills," Nordic Journal of Music Therapy, vol. 31, no. 2, pp. 123–135, 2022.

[5] E. Jing et al., "Emotion-aware personalized music recommendation with a heterogeneity-aware Deep Bayesian Network," arXiv [cs.AI], 2024.

[6] S. Kambham, H. Jhonson, and S. P. R. Kambham, "Emotion detection and music recommendation system," arXiv [cs.CV], 2025.

[7] Y. Li, H. Zhang, and X. Chen, "Emotion-aware music recommendation system based on fully convolutional neural networks," Applied Computing and Informatics, vol. 21, no. 1, pp. 45–56, 2025.

[8] D. Rozhevskii, J. Zhu, and B. Zhao, "Psychologically-inspired music recommendation system," arXiv [cs.IR], 2022.

[9] Department of Pediatrics, University of Illinois, at Chicago, VA Healthcare System, Chicago, IL, USA. Email: nchauhan51@gmail.com, N. Chauhan, M. Kale, Indian Classical Music and Arts Foundation, Mahesh Kale School of Music, San Francisco, CA, USA. Email: mahesh@maheshkale.com., N. Naik, and EN1 Neuro Services Pvt. Ltd. Mumbai, India. Email: neeta.naik2@gmail.com., "Raga therapy for autism," MedPress Neurology and Neurosurgery, vol. 3, no. 1, 2022.

[10] S. Deka, P. Tiwari, and K. M. Tripathi, "Raga todi intervention on state anxiety level in female young adults during COVID-19," Mater. Today, vol. 57, pp. 2152–2155, 2022.

[11] F. Xu, W. Zhou, G. Li, Z. Zhong, and Y. Zhou, "Enhancing cross-modal understanding for audio visual scene-aware dialog through contrastive learning," in 2024 IEEE International Symposium on Circuits and Systems (ISCAS), 2024, pp. 1–5.

[12] T. Li et al., "Cross-modal alignment and contrastive learning for enhanced cancer survival prediction," Comput. Methods Programs Biomed., vol. 263, no. 108633, p. 108633, 2025.

[13] B. Low, X. Liu, R. Z. Li, E. Ren, and J. X. Zhang, "Music therapy for autism spectrum disorder: A comprehensive literature review on therapeutic efficacy, limitations, and AI integration," in 2024 IEEE 15th Annual Ubiquitous Computing, Electronics & Mobile Communication Conference (UEMCON), 2024, pp. 90–99.

[14] H. Fengr, M. H. Mahoor, and F. Dino, "A music-therapy robotic platform for children with Autism: A pilot study," arXiv [cs.HC], 2022.

[15] P. Liu and S. Huang, "The application of music therapy in children with autism," Curr. Psychiatry Res. Rev., vol. 21, no. 3, pp. 250–259, 2025.

[16] T. Frei and T. Szucs, "Music Therapy and Skill Development in Children with Autism Spectrum Disorder: A Systematic Review," J. Psychiatry Psychiatr. Disord, vol. 9, no. 2, pp. 91–110, 2025.

[17] X. Ke, W. Song, M. Yang, J. Li, and W. Liu, "Effectiveness of music therapy in children with autism spectrum disorder: A systematic review and meta-analysis," Front. Psychiatry, vol. 13, p. 905113, 2022.

[18] M. Alayidh et al., “Music therapy for people with autism spectrum disorder: A systematic review of randomized clinical trials,” Cureus, 2025.
[19] Z. Zhou et al., “A randomized controlled trial of the efficacy of music therapy on the social skills of children with autism spectrum disorder,” Res. Dev. Disabil., vol. 158, no. 104942, p. 104942, 2025.
[20] L. Gassner, M. Geretsegger, and J. Mayer-Ferbas, “Effectiveness of music therapy for autism spectrum disorder, dementia, depression, insomnia and schizophrenia: update of systematic reviews,” Eur. J. Public Health, vol. 32, no. 1, pp. 27–34, 2022.
[21] B. Applewhite, Z. Cankaya, A. Heiderscheit, and H. Himmerich, “A systematic review of scientific studies on the effects of music in people with or at risk for autism spectrum disorder,” Int. J. Environ. Res. Public Health, vol. 19, no. 9, p. 5150, 2022.