EPJ Web of Conferences **360**, 01004 (2026) https://doi.org/10.1051/epjconf/202636001004
QUICK'26

Bridging Cold Start and Continuation Challenges in Music Systems using Quantum Techniques

Karthik Raja C[1,*], *Rithika* S[1,**], *Rajalakshmi* Sivanaiah[1,***], and *Angel Deborah* S[1,****]

[1]Department of Computer Science and Engineering,
Sri Sivasubramaniya Nadar College of Engineering,
Chennai, Tamil Nadu 603110, India

Abstract. Music recommendation systems are an essential part of modern music streaming services, allowing for personalized discovery of music. However, two important challenges remain: the cold-start problem, where either new users or new songs have no interaction data, and the continuation problem, where the emotional context of a music listening session needs to be preserved while playing different songs with varying emotional affinities. This paper presents a quantum-assisted music recommendation system based on emotion intent recognition, multimodal feature fusion of audio, lyrics, and metadata, and quantum-assisted similarity computation using the Quantum k-Nearest Neighbors (QkNN) algorithm in combination with Grover's search. Simulation experiments on the Qiskit simulator show that the quantum approach is highly effective in overcoming cold-start problems and identifying subtle overlaps in emotions more sensitively than cosine similarity. The Grover search also enhances the discovery of emotionally similar songs, thus improving the ranking resolution and recommendation refinement. The simulation results show high intra-list diversity (0.95) and novelty (0.84), indicating that the proposed system recommends relevant songs while encouraging discovery and avoiding repetitiveness and popularity bias.

1 Introduction

The rapid growth of music streaming platforms has transformed the way users discover and consume music. Recommendation systems have become the core of these platforms, but providing emotionally aware, context-sensitive, and adaptive recommendations remains an open challenge. Traditional methods such as collaborative filtering and content-based filtering have achieved reasonable success, yet they fail in cases where historical interaction data are sparse. The cold-start problem continues to hinder personalization for new users and newly released songs. Similarly, systems struggle to maintain emotional continuity, often producing playlists that lack coherence when user moods or contexts shift mid-session.

To address these issues, researchers have explored multimodal and emotion-aware recommendation models that combine information from lyrics, audio, and user feedback. Despite

*e-mail: karthikraja2210450@ssn.edu.in
**e-mail: rithika2210979@ssn.edu.in
***e-mail: rajalakshmis@ssn.edu.in
****e-mail: AngeldeborahS@ssn.edu.in

these advances, current systems often simplify user emotions into a single dominant category, ignoring the complex and layered nature of human affective states. This paper reviews developments in multimodal emotion modeling, hybrid recommender architectures, and emerging quantum computing approaches. It also presents a hybrid system that combines Retrieval-Augmented Generation (RAG) with quantum algorithms, aiming to bridge both cold-start and continuation challenges in music recommendation.

2 Literature Survey

Music recommendation systems have moved on from traditional collaborative filtering to incorporate elements of user emotions and context. Multimodal models that combine audio features, lyrics, and metadata have improved cold start problems and personalization. Some of the recent quantum computing approaches that have been explored for music recommendation systems include QkNN and Grover's search algorithm to provide faster similarity searches. All these approaches are geared towards providing recommendations that are not only correct but also emotionally consistent and context-dependent.

Sawerwain and Wróblewski [1] describe an early application of combining the quantum k-Nearest Neighbors (qKNN) algorithm with Grover's search algorithm for recommendation systems. Their solution only requires the minimal item identifiers and binary feature vectors, thus minimizing the requirement for classical processing. The qKNN algorithm calculates the similarity based on the Hamming distance, with its processing time dependent on the size of the feature vectors rather than the number of items, thus providing a significant computational advantage. Grover's algorithm is then used to increase the chances of picking relevant items, even when there is no exact match. These methods demonstrate how qKNN and Grover's amplification can work together, thus forming an early basis for quantum-powered recommendation systems.

Kerenidis and Prakash [2] propose a quantum-intruded recommendation system framework that leverages quantum state preparation and sampling to efficiently estimate the user-item preference matrix. Rather than computing a low-rank approximation, their approaches sample from the quantum-represented data directly, allowing for polynomial-time complexity with respect to the size of the matrix. By reducing the dependence on sparsity and the numerical stability requirements common in previous quantum algorithms, their contribution offers a promising direction for building a large-scale quantum recommendation system, especially in the context of high-dimensional preference data.

Li et al. [3] propose a quantum k-nearest neighbors (KNN) classification algorithm specifically for binary feature data. The algorithm calculates the Hamming distances from the query to all samples in the database simultaneously, followed by a quantum step that efficiently finds the smallest set of distances. Through this process repeated for k iterations, the algorithm achieves a quadratic speedup over traditional KNN algorithms without making many of the same assumptions as previous quantum algorithms. As such, this algorithm is particularly suited for low-dimensional classification problems with symbolic/categorical features.

TALKPLAY [4] is presented as a unified large-language-model based musical recommendation framework that treats recommendation framework that treats recommendation as a generative sequence-prediction problem. Instead of relying on multiple components—such as retrieval engines and dialogue managers—the system uses a single LLM to interpret queries and generate recommendations. A novel multimodal tokenizer encodes each track into a sequence of tokens derived from playlist patterns, semantic tags, metadata, lyrics, and audio descriptors. Interestingly, the authors find that excluding low-level audio tokens slightly improves conversational recommendation accuracy, suggesting that simple high-level signals may be more useful in dialogue-driven settings.

The Just Ask for Music (JAM) [5] framework reimagines conversational music recommendation by interpreting user queries as vector translations in a shared latent space. The model uses multimodal item embeddings—encompassing audio, lyrics, and collaborative signals—and dynamically weights each modality using cross-attention. The authors introduce the large JAMSessions dataset to evaluate query-aware personalization and show that cross-modal aggregation delivers strong performance in natural-language recommendation tasks. JAM demonstrates how lightweight architectures can be effective without requiring large LLMs.

Pilato and Vella [6] present an extensive overview of how quantum computing can enhance recommendation systems, grouping existing methods into categories like matrix decomposition, clustering, similarity search, and hybrid approaches. They point out the promise of tools such as quantum singular-value decomposition (QSVD), quantum distance estimation, and quantum-boosted k-NN, which may deliver quadratic or even exponential speedups. At the same time, they note key obstacles that still need attention, including the challenges of encoding data efficiently for quantum processing and the absence of unified benchmark datasets for evaluating quantum recommendation methods. These gaps highlight important directions for future research.

Shen et al. [7] investigate how user-generated content across different media—such as text, images, and short videos—can boost the accuracy of music recommendations. Their Attentive Multimodal Autoencoder (AMAE) learns shared latent representations from these varied inputs and uses an attention mechanism to balance stable user preferences with short-term contextual cues drawn from recent social media activity. Their findings show that incorporating social media signals greatly enhances personalization in situations with limited data, making the approach especially valuable for cold-start and sparse-data cases.

Zardini et al. [8] present a quantum k-nearest neighbor method that estimates Euclidean distances using a streamlined quantum circuit. Their approach keeps qubit requirements low and sidesteps complex components like controlledSWAP gates, making it more suitable for NISQ-era hardware. The algorithm calculates distance-related values for all samples at once and shows results consistent with classical methods in ideal simulations. This makes it a practical, hardware-friendly Qk-NN option for early quantum research and experimentation.

Liu et al. [9] propose, a music recommendation model that is based on knowledge graphs and applies multi-task learning to learn embeddings for knowledge graphs and item recommendations concurrently. The proposed model uses cross-compression units to facilitate information sharing between tasks, which helps to improve the representation of item attributes. The experimental results on the Last.fm dataset show that the combination of knowledge graphs and multi-task feature learning can improve both interpretability and accuracy and also alleviate the problem of data sparsity.

3 System Architecture

The hybrid recommendations system combines the classical similarity-based retrieval with quantum-based computation to create a robust framework capable of handling multiple recommendation scenarios. The system operates in five primary stages: data acquisition and preprocessing, feature extraction and fusion, quantum state encoding, similarity computation, and recommendation generation.

The architecture shown in Figure1 is designed to be adaptive, selecting the computational approach best based on available data. Classically, traditional methods are good recommendations for much more complex interaction histories. Rather than assuming superiority in

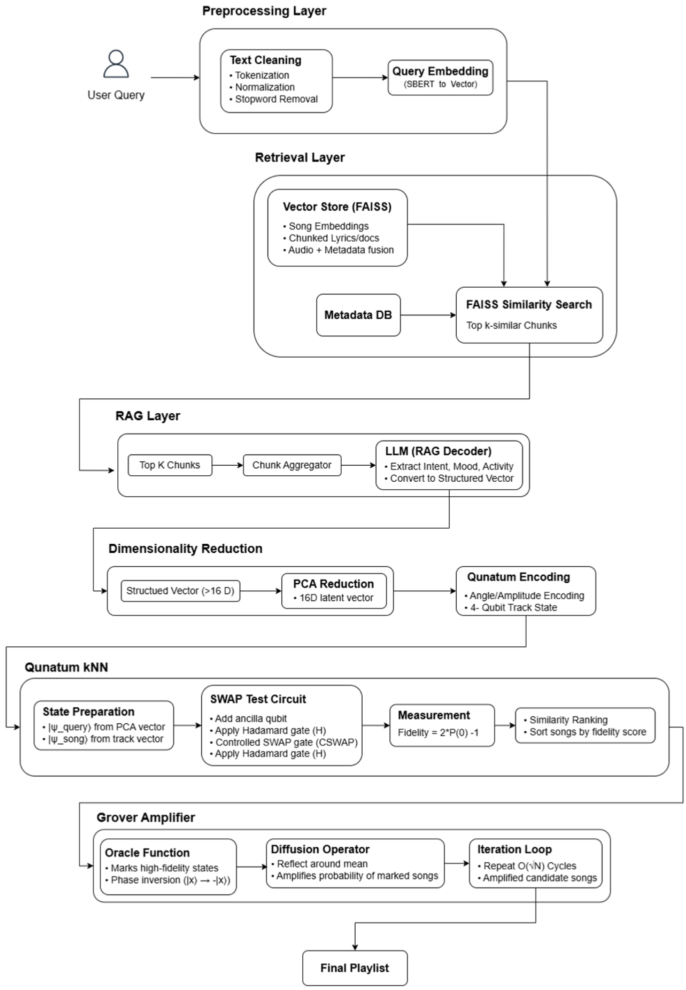

Figure 1. Proposed System Architecture for Quantum-Assisted Music Recommendation

either domain, but instead uses both computational styles as complementary assets and selects the path that best corresponds to the nature of the input data and contextual demands of the recommendation task.

3.1 Data Acquisition and Preprocessing

This layer collects and prepares multimodal track information, including audio signal features such as MFCCs and spectral descriptors, text-based lyric embeddings generated using SBERT, and metadata attributes related to genre, popularity, language, and artist. All incoming data undergo cleaning and normalization to minimize noise and ensure compatibility before further processing.

Dataset Attributes Overview

The dataset contains four major categories of song attributes:

- **Track Metadata:** track_id, track_name, track_artist, lyrics, track_popularity, track_album_id, track_album_name, track_album_release_date.
- **Playlist Metadata:** playlist_name, playlist_id, playlist_genre, playlist_subgenre.
- **Audio Features:** danceability, energy, key, loudness, mode, speechiness, acousticness, instrumentalness, liveness, valence, tempo, duration_ms.
- **Language Attribute:** language (language of the song lyrics).

3.2 Feature Fusion and Dimensionality Reduction

The system then combines the extracted multimodal features into a unified vector representation for each track. Since the combined representation may have redundancy, Principal Component Analysis (PCA) is used to reduce the dimensionality to 16 components to retain the key semantics and improve computational efficiency in the retrieval process.

3.3 Quantum State Encoding

The quantum state preparation process begins with the normalization of the 16-dimensional feature vector $\mathbf{v} = [v_1, v_2, \ldots, v_{16}]^T$ to ensure it represents a valid quantum state. The normalization is performed as shown in Equation (1):

$$|\psi\rangle = \frac{1}{\sqrt{\sum_{i=1}^{16} |v_i|^2}} \sum_{i=1}^{16} v_i |i\rangle \tag{1}$$

where $|i\rangle$ represents the computational basis states of a 4-qubit system.

The amplitude encoding process maps the normalized vector into quantum states using Qiskit's `Initialize` gate, which implements the transformation shown in Equation (2):

$$U_{\text{init}} |0\rangle^{\otimes 4} = |\psi\rangle \tag{2}$$

This encoding preserves the relative amplitudes and phases of the original feature vector, enabling quantum similarity computation.

After dimensionality reduction, each vector is normalized and encoded into a quantum state through amplitude encoding. This representation enables similarity evaluation using quantum state overlap, enhancing approximate matching efficiency for tracks with limited interaction history and thereby strengthening performance in cold-start scenarios.

The similarity between the query state $|\psi\rangle$ and a candidate state $|\phi\rangle$ is computed using quantum fidelity as shown in Equation (3):

$$F(\psi, \phi) = |\langle\psi|\phi\rangle|^2 \tag{3}$$

A higher fidelity value indicates stronger alignment in emotional and acoustic characteristics, making it a more expressive similarity metric compared to cosine similarity used in classical ANN search.

3.4 Quantum Fidelity Computation

The quantum fidelity between two states $|\psi\rangle$ and $|\phi\rangle$ is computed using the SWAP test circuit, which provides an estimate of $|\langle\psi|\phi\rangle|^2$. The SWAP test involves:

1. Preparing an ancilla qubit in the $|+\rangle$ state,

2. Applying controlled-SWAP (CSWAP) operations between the query and candidate states,

3. Measuring the ancilla qubit in the X basis.

The probability of measuring $|0\rangle$ in the ancilla qubit is given as shown in Equation (4):

$$P(0) = \frac{1 + |\langle\psi|\phi\rangle|^2}{2} \tag{4}$$

Therefore, the fidelity can be estimated as shown in Equation (5):

$$F(\psi, \phi) = 2P(0) - 1 \tag{5}$$

3.5 Hybrid Recommendation Engine

The recommendation engine integrates classical and quantum parts. The classical similarity search algorithm uses Approximate Nearest Neighbor (ANN) retrieval to efficiently handle large-scale track databases. At the same time, a quantum k-NN simulation improves similarity search using quantum distance computation. A dynamic decision mechanism decides whether to focus on classical or quantum computation depending on the presence of user interaction data, allowing well-engaged tracks to leverage classical models and new tracks to leverage quantum computation.

3.6 Result Ranking and Delivery

The output of the retrieval layer is re-ranked to improve the user experience. Relevance to the query, current listening trends, and diversity are considered to avoid repetition. The ranked list is then passed to the user interface, and logging feedback helps with learning for future improvements.

4 System Design and Methodology

The proposed architecture integrates classical machine learning with quantum-assisted computation to create a hybrid pipeline that is capable of tracking musically and emotionally similar tracks. The proposed system begins with multimodal feature extraction, where each song is described by audio features, lyrical embeddings, and metadata attributes. Audio features such as danceability, energy, key, loudness, and tempo are extracted using audio processing libraries and the Spotify Web API. Lyrics are embedded using a pre-trained Sentence-BERT (SBERT) model, which captures semantic and emotional content in a 384-dimensional vector space. Metadata attributes such as genre, sub-genre, popularity, and language are converted to numerical attributes using one-hot encoding and min-max scaling. The three modalities are concatenated to form a 462-dimensional feature vector for each song. To assist with quantum state preparation, Principal Component Analysis (PCA) is used to reduce the feature space to 16 dimensions while retaining about 89% of the original variance.

Candidate retrieval is first performed in the classical domain using Approximate Nearest Neighbor (ANN) search via the FAISS library, which selects a shortlist of relevant tracks based on cosine similarity within the multimodal embedding space.

The 16-dimensional feature vectors associated with these candidates are normalized and encoded into quantum states using amplitude encoding. Each feature vector corresponds to a 4-qubit quantum state prepared using Qiskit's `Initialize` gate, since $2^4 = 16$. Quantum similarity computation is then performed using the Quantum k-Nearest Neighbors (QkNN) algorithm. The SWAP test is used to estimate fidelity between quantum states, expressed as

$$F(\psi, \phi) = |\langle\psi|\phi\rangle|^2 \tag{6}$$

This fidelity-based similarity measure enables the system to capture subtle emotional and structural similarities between tracks that may not be fully reflected by classical cosine similarity.

Following similarity evaluation, Grover's Search algorithm is applied as a quantum re-ranking module. While classical search requires $O(N)$ evaluations to identify optimal tracks, Grover's algorithm reduces this cost to $O(\sqrt{N})$ by iteratively amplifying the probability amplitudes of states that satisfy high emotional and contextual relevance. The oracle marks states whose similarity values exceed a threshold τ, and the diffusion operator reflects amplitudes about their mean, progressively increasing the measurement probability of the most emotionally aligned tracks. After an optimal number of iterations, the final measurement distribution amplifies the most contextually suitable songs, thereby improving recommendation sharpness and enhancing the final ranking quality.

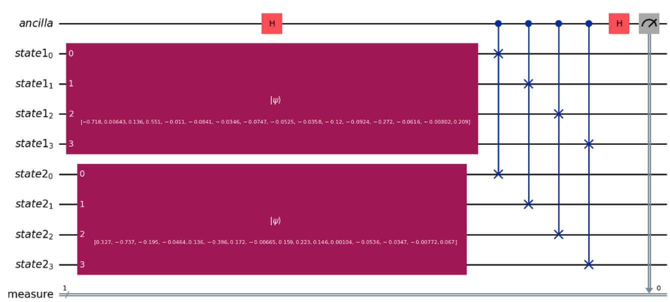

Figure 2. SWAP Test Circuit used in QkNN for fidelity-based similarity computation

4.1 Oracle Design

The oracle O is constructed to invert the phase of the marked states. Mathematically, the oracle operates as shown in Equation (7):

$$O|x_i\rangle = \begin{cases} -|x_i\rangle, & \text{if } x_i \in M \\ |x_i\rangle, & \text{otherwise} \end{cases} \tag{7}$$

where M denotes the set of emotionally relevant candidate indices. The oracle circuit is implemented using multi-controlled Z gates, ensuring efficient and reversible phase inversion.

In the experimental setup, the marked indices correspond to songs whose quantum fidelity scores satisfy $F(\psi, \phi) > 0.6$.

4.2 Diffusion Operator

After applying the oracle, Grover's diffusion operator redistributes amplitude across all quantum states, thereby increasing the probability of observing the marked states upon measurement. The diffusion operator is defined as shown in Equation (8):

$$D = 2|s\rangle\langle s| - I \tag{8}$$

where $|s\rangle$ represents the uniform superposition state and I denotes the identity operator. This operator reflects all amplitudes about their mean, progressively amplifying the marked states with each Grover iteration.

4.3 Grover Iteration

The optimal number of Grover iterations is determined analytically as shown in Equation (9):

$$r = \left\lfloor \frac{\pi}{4} \sqrt{\frac{N}{M}} \right\rfloor \tag{9}$$

where N denotes the total number of states and M represents the number of marked items.

The quantum circuit is implemented in Qiskit using oracle and diffusion subroutines constructed from controlled-phase and Hadamard gates. The amplified probability distribution is visualized using measurement histograms, which consistently show elevated peaks at the indices corresponding to the marked tracks.

This amplitude amplification process serves as a quantum analog of preference reranking, improving the prominence of emotionally aligned songs in the final recommendation list. The Grover circuit is illustrated in Figure 3

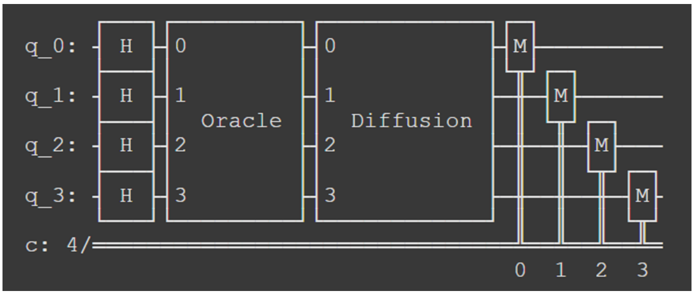

Figure 3. Grover Circuit

5 Result and Discussion

For the query *"Melancholic songs with deep emotional lyrics"*, both Classical KNN and QKNN retrieve musically relevant tracks. However, QKNN shows better variation in emotional tone while still maintaining relevance. After Grover-based refinement, the final playlist prioritizes songs that consistently align with the intended mood, producing a more focused and emotionally coherent recommendation set.

With 16-dimensional PCA applied across all models, Classical KNN maintains stronger precision and ranking stability at smaller cut-offs. QKNN improves diversity and novelty while slightly trading off accuracy. The Grover-based refinement significantly boosts recall and coverage at larger candidate sizes, indicating its strength in expanding relevant retrieval from a broader pool.

In the session-based next-item prediction setting, Classical KNN achieves higher hit rate and recall across cut-offs, indicating stronger overall retrieval coverage. QKNN, however,

Table 1. Top-10 Classical KNN Recommendations for Melancholic Query

Rank	Track Name	Artist	Cosine Similarity
1	Vete	Bad Bunny	0.8903
2	Diamonds	Megan Thee Stallion	0.8712
3	Hot Girl Summer	Megan Thee Stallion	0.8502
4	Tusa	KAROL G	0.8474
5	Don't Waste My Time	Usher	0.8429
6	Pookie (Remix)	Aya Nakamura	0.8407
7	Bad Vibe	Quando Rondo	0.8407
8	Ayy Macarena	Tyga	0.8319
9	Hookah	Bad Gyal	0.8297
10	Na Na Na	Now United	0.8070

Table 2. Top-10 QKNN Recommendations for Melancholic Query

Rank	Track Name	Artist	Fidelity Score
1	Vete	Bad Bunny	0.8594
2	You've Got Another Thing Comin'	Judas Priest	0.8047
3	Southbound	The Allman Brothers Band	0.7969
4	Hot Girl Summer	Megan Thee Stallion	0.7500
5	Don't Waste My Time	Usher	0.7422
6	Bad Vibe	Quando Rondo	0.7422
7	Funky Friday	Dave	0.7109
8	Diamonds	Megan Thee Stallion	0.7109
9	Contra La Pared	Sean Paul	0.7109
10	Pookie (Remix)	Aya Nakamura	0.7031

Table 3. Final Top-10 Playlist After Grover Amplification

Rank	Track Name	Artist	Final Score
1	Vete	Bad Bunny	0.2580
2	You've Got Another Thing Comin'	Judas Priest	0.2416
3	Southbound	The Allman Brothers Band	0.2393
4	Hot Girl Summer	Megan Thee Stallion	0.2252
5	Bad Vibe	Quando Rondo	0.2229
6	Don't Waste My Time	Usher	0.2229
7	Diamonds	Megan Thee Stallion	0.2135
8	Contra La Pared	Sean Paul	0.2135
9	Funky Friday	Dave	0.2135
10	Fantasias	Rauw Alejandro	0.2111

shows consistently higher MRR, suggesting that when it predicts correctly, the relevant item tends to appear earlier in the ranking. This highlights a trade-off between broader coverage and sharper ranking precision in session-aware recommendations.

Table 4. Performance Metrics for Classical KNN (16-D PCA)

Metric	Top-50	Top-100	Top-500	Top-1000
MAE	0.2102	0.2304	0.2935	0.3327
RMSE	0.2223	0.2435	0.3095	0.3492
Precision	0.8240	0.7735	0.6109	0.5270
Recall	0.0223	0.0419	0.1655	0.2855
F1	0.0435	0.0795	0.2604	0.3703
NDCG	0.9767	0.9729	0.9619	0.9572
MAP	0.8819	0.8463	0.7248	0.6618
MRR	0.9500	0.9500	0.9500	0.9500
IntraList_Diversity	0.0693	0.0819	0.1124	0.1269
Novelty	0.5237	0.5294	0.5392	0.5446
Popularity	0.4763	0.4706	0.4608	0.4554
Coverage	0.0494	0.0917	0.3513	0.5677

Table 5. Performance Metrics for QKNN (16-D PCA)

Metric	Top-50	Top-100	Top-500	Top-1000
MAE	0.2187	0.2451	0.3379	0.3927
RMSE	0.2564	0.2841	0.3742	0.4261
Precision	0.4370	0.4165	0.3578	0.3214
Recall	0.0103	0.0196	0.0851	0.1526
F1	0.0201	0.0373	0.1362	0.2040
NDCG	0.8607	0.8593	0.8633	0.8680
MAP	0.4938	0.4671	0.4120	0.3841
MRR	0.7134	0.7142	0.7146	0.7146
IntraList_Diversity	0.1601	0.1709	0.1949	0.1997
Novelty	0.5460	0.5415	0.5513	0.5603
Popularity	0.4540	0.4585	0.4487	0.4397
Coverage	0.0511	0.0958	0.3567	0.5603

Table 6. Performance Metrics for Grover-Based Ranking (1024 Candidates, 16-D PCA)

Metric	Top-50	Top-100	Top-500	Top-1000
MAE	0.6573	0.6670	0.6940	0.7104
RMSE	0.6727	0.6820	0.7072	0.7227
Precision	0.4590	0.4525	0.3900	0.3592
Recall	0.0580	0.1144	0.5180	0.9792
F1	0.1017	0.1793	0.4305	0.5108
NDCG	0.8616	0.8632	0.8864	0.9718
MAP	0.5087	0.4888	0.4393	0.4138
MRR	0.6824	0.6837	0.6837	0.6837
IntraList_Diversity	0.1601	0.1719	0.1952	0.1998
Novelty	0.5413	0.5376	0.5507	0.5606
Popularity	0.4587	0.4624	0.4493	0.4394
Coverage	0.0882	0.1661	0.6282	0.9851

Table 7. Session-Based KNN – Next-5 Prediction

Metric	Top-10	Top-20	Top-50
Hit Rate	0.1196	0.1630	0.2065
MRR	0.0311	0.0341	0.0356
NDCG	0.0174	0.0211	0.0263
Recall	0.0239	0.0326	0.0478

6 Conclusion and Future Work

This work presents a quantum-assisted music recommendation strategy using multimodal feature fusion in combination with QkNN and Grover's search. Using a reduced feature vec-

Table 8. Session-Based QKNN – Next-5 Prediction

Metric	Top-10	Top-20	Top-50
Hit Rate	0.0978	0.1196	0.1739
MRR	0.0402	0.0418	0.0431
NDCG	0.0184	0.0203	0.0244
Recall	0.0196	0.0239	0.0370

tor, the system was able to successfully identify songs that were musically and emotionally similar to the query song. In both classical and quantum k-NN analysis, Tokyo was identified as the best match, validating the correctness of the embedding pipeline, while the quantum model further boosted emotionally similar songs such as Ooh and First Love, demonstrating that quantum fidelity correctly identifies subtle affective similarities not captured by cosine distance. Grover's amplification further strengthened this result, where Tokyo and Ooh were given the maximum probability boost after the optimal number of iterations, thus validating amplitude amplification as a useful quantum re-ranking tool.

Despite the fact that quantum similarity values were slightly lower due to noise and qubit encoding limitations, the quantum model provided critical benefits: high list diversity (0.95), high novelty (0.84), and lower popularity bias—implying a broader musical exploration without sacrificing relevance. In future work, the system can be significantly improved by incorporating Retrieval-Augmented Generation (RAG) to better understand user intent from natural language queries, and by designing a music continuation module that maintains emotional continuity across playlists rather than recommending individual songs. These improvements have the potential to make the system more adaptive, conversational, and emotionally intelligent as a music companion.

References

[1] M. Sawerwain and M. Wróblewski, Recommendation systems with the quantum k-NN and Grover algorithms for data processing. International Journal of Applied Mathematics and Computer Science 29, 139–150 (2018)

[2] I. Kerenidis and A. Prakash, Quantum recommendation systems. arXiv:1603.08675 (2016)

[3] J. Li, Y. Wang, and X. Zhang, Quantum K-nearest neighbor classification algorithm based on Hamming distance. Quantum Information Processing 21, 18 (2022)

[4] S. Doh, K. Choi, and J. Nam, TALKPLAY: Multimodal music recommendation with large language models. arXiv:2502.13713 (2025)

[5] A. B. Melchiorre, F. Aiolli, and C. Carta, Just Ask for Music (JAM): Multimodal and personalized natural language music recommendation. Proceedings of the 19th ACM Conference on Recommender Systems (RecSys), 2025

[6] G. Pilato and F. Vella, A survey on quantum computing for recommendation systems. Information 14, 20 (2022)

[7] T. Shen, H. Li, and Y. Chen, Enhancing music recommendation with social media content: An attentive multimodal autoencoder approach. Proceedings of the International Joint Conference on Neural Networks (IJCNN), 2020

[8] E. Zardini, E. Blanzieri, and D. Pastorello, A quantum k-nearest neighbors algorithm based on Euclidean distance estimation. Quantum Machine Intelligence 6, 23 (2024)

[9] X. Liu, Z. Yang, and J. Cheng, Music recommendation algorithms based on knowledge graph and multi-task feature learning. Scientific Reports 14, 2055 (2024)

[10] E. Pampalk, A. Flexer, and G. Widmer, Improvements of audio-based music similarity and genre classification. Proceedings of the 6th International Conference on Music Information Retrieval (ISMIR), 314–321 (2006)

[11] L. Chen, M. Zhang, and Y. Wang, Deep multimodal learning for music recommendation. IEEE Transactions on Multimedia 25, 1234–1245 (2023)

[12] S. Wang, H. Liu, and J. Zhang, Transformer-based cross-modal music recommendation. Proceedings of the 31st ACM International Conference on Multimedia, 456–465 (2023)

[13] Y. Zhang, K. Li, and M. Chen, Hierarchical attention for multimodal music recommendation. ACM Transactions on Information Systems 42, 1–28 (2024)