Online Journal of Music Sciences, 2025, 10(4), 703-722
https://doi.org/10.31811/ojomus.1668286

Research Article ♪ **Araştırma Makalesi**

Utilising generative AI to assist in the creation and production of Chinese popular music

Çin popüler müziğinin yaratılması ve üretilmesine yardımcı olmak için üretken yapay zekanın kullanılması

Xinhao Li[1], Hyuntai Kim[2]

[1] Sejong University, Seoul, South Korea.

[2] Department of Music, Faculty of Arts and Physical Education, Sejong University, Seoul, South Korea.

ABSTRACT

Background: The rapid development of generative artificial intelligence (AI) has had a significant impact on the music industry, particularly Chinese popular music (C-pop), which presents unique difficulties due to its distinctive melodic structures and emotional depth. While AI can boost efficiency and inspire composers, current models are ineffective in feature selection and prediction accuracy, leading to compositions without stylistic integrity and commercial viability. Problem Statement: Current AI-generated melodies often lack the stylistic depth of human compositions and fail to predict market success accurately. These challenges highlight the need for a more effective AI framework capable of generating high-quality melodies. Objectives: This study introduces the GenAI Melody-LSTM algorithm. This generative AI-driven technique uses Long Short-Term Memory (LSTM) networks to create melodies inspired by Chinese pop songs and forecasts their success. The main goals are to build an AI pipeline for music preprocessing, improve feature selection, train a deep learning model for stylistically consistent melodies, and compare its effectiveness to other methods. Methodology: The methodology utilises a GenAI Melody-LSTM algorithm, which contains data preprocessing techniques such as mode and mean imputation, label encoding, and Min-Max scaling. Feature selection is improved using the Hybrid Filter-Wrapper Ensemble (HFWE) method, which combines Mutual Information, Chi-Square Test, ANOVA F-Test, and Recursive Feature Elimination (RFE) with Support Vector Machine (SVM), Random Forest, and Gradient Boosting Machine (GBM), with the final subset determined by majority voting. The selected features, like melody structure, key signature, tempo, rhythm complexity, instrumentation, and emotion, are used as inputs for an LSTM-based deep learning model that comprises multiple LSTM layers, dropout layers to prevent overfitting, and a dense output layer to create melodies and forecast their commercial success. Results: Performance evaluation using accuracy, precision, recall, F1-score, and Matthew's correlation coefficient (MCC) provides better results with the GenAI Melody-LSTM algorithm, with 89.7% accuracy, 87.5% precision, 88.2% recall, an F1-score of 87.8%, and an MCC of 0.82. Conclusion: This research demonstrates that integrating generative AI, optimal feature selection, and deep learning significantly enhances Chinese pop music compositions. The LSTM-based model generates melodies and predicts their commercial viability, enabling composers to fine-tune AI-generated music for improved quality and market appeal.

Keywords: Generative AI, Chinese pop music, LSTM networks, feature selection, melody prediction

ÖZ

Arka Plan: Üretken yapay zekanın (YZ) hızla gelişmesi, özellikle kendine özgü melodik yapıları ve duygusal derinliği nedeniyle benzersiz zorluklar sunan Çin popüler müziği (C-pop) olmak üzere müzik endüstrisi üzerinde önemli bir etkiye sahip olmuştur. Yapay zeka verimliliği artırabilir ve bestecilere ilham verebilirken, mevcut modeller özellik seçimi ve tahmin doğruluğu konusunda etkisizdir ve bu da stilistik bütünlük ve ticari uygulanabilirliği olmayan bestelere yol açmaktadır.

Hyuntai Kim — kimht@sejong.ac.kr
Geliş tarihi/Received: 30.03.2025 — Kabul tarihi/Accepted: 02.07.2025 — Yayın tarihi/Published: 30.10.2025

Sorun Bildirimi: Mevcut YZ tarafından üretilen melodiler genellikle insan bestelerinin stilistik derinliğinden yoksundur ve pazar başarısını doğru bir şekilde tahmin edememektedir. Bu zorluklar, yüksek kaliteli melodiler üretebilen daha etkili bir YZ çerçevesine olan ihtiyacı vurgulamaktadır. Amaçlar: Bu çalışma, Çin pop şarkılarından esinlenerek melodiler oluşturmak ve bunların başarısını tahmin etmek için Uzun Kısa Süreli Bellek (LSTM) ağlarını kullanan üretken bir YZ odaklı teknik olan GenAI Melody-LSTM algoritmasını tanıtmaktadır. Ana hedefler, müzik ön işleme için bir YZ boru hattı oluşturmak, özellik seçimini iyileştirmek, stilistik olarak tutarlı melodiler için derin öğrenme modeli eğitmek ve etkinliğini diğer yöntemlerle karşılaştırmaktır. Metodoloji: Metodoloji, mod ve ortalama hesaplama, etiket kodlama ve Min-Max ölçekleme gibi veri ön işleme tekniklerini içeren bir GenAI Melody-LSTM algoritmasını kullanır. Özellik seçimi, Karşılıklı Bilgi, Ki-Kare Testi, ANOVA F Testi ve Destek Vektör Makinesi (SVM), Rastgele Orman ve Gradyan Artırma Makinesi (GBM) ile Yinelemeli Özellik Eleme (RFE)'yi birleştiren Hibrit Filtre-Sarmalayıcı Topluluğu (HFWE) yöntemi kullanılarak iyileştirilir ve son alt küme çoğunluk oyu ile belirlenir. Melodi yapısı, anahtar imzası, tempo, ritim karmaşıklığı, enstrümantasyon ve duygu gibi seçilen özellikler, birden fazla LSTM katmanı, aşırı uyumu önlemek için bırakma katmanları ve melodiler oluşturmak ve ticari başarılarını tahmin etmek için yoğun bir çıktı katmanı içeren LSTM tabanlı bir derin öğrenme modeli için girdi olarak kullanılır. Sonuçlar: Doğruluk, kesinlik, geri çağırma, F1 puanı ve Matthew'nun korelasyon katsayısı (MCC) kullanılarak yapılan performans değerlendirmesi, %89,7 doğruluk, %87,5 kesinlik, %88,2 geri çağırma, %87,8'lik bir F1 puanı ve 0,82'lik bir MCC ile GenAI Melody-LSTM algoritmasıyla daha iyi sonuçlar sağlar. Sonuç: Bu araştırma, üretken AI, optimum özellik seçimi ve derin öğrenmenin entegre edilmesinin Çin pop müziği bestelerini önemli ölçüde geliştirdiğini göstermektedir. LSTM tabanlı model melodiler üretir ve bunların ticari uygulanabilirliğini tahmin ederek bestecilerin AI tarafından üretilen müziği gelişmiş kalite ve pazar çekiciliği için ince ayar yapmalarını sağlar.

Anahtar kelimeler: Üretken yapay zeka, Çin pop müziği, LSTM ağları, özellik seçimi, melodi tahmini

1. INTRODUCTION

1.1. The Background Information of This Scientific Field

Generative artificial intelligence (AI) has quickly transformed a variety of creative fields, including music composition (Banh & Strobel, 2023). AI-driven music creation uses deep learning methods to evaluate musical patterns, structures, and emotions, allowing for the generation of original compositions (Gao et al., 2024). Among these methods, recurrent neural networks (RNNs) and their sophisticated variant, Long Short-Term Memory (LSTM) networks, have shown promise for capturing intricate musical sequences. Despite these advances, applying generative AI to Chinese popular music (C-pop) presents distinct difficulties because of its intricate melodic structures, tonal variations, and emotional depth (Civit et al., 2022). The capacity to incorporate AI-driven methodologies into C-pop composition while retaining stylistic authenticity is still a significant area of exploration.

1.2. The Current Knowledge and Advances in This Field

Current studies in AI-generated music have used deep learning models to create melodies and harmonies, including Variational Autoencoders (VAEs), Transformer-based architectures, and LSTM networks (Mitra & Zualkernan, 2025). Notably, AI applications in Western music have led to significant advances in style emulation and predictive composition (Civit et al., 2022). However, adapting these technologies to C-pop needs domain-specific improvements in feature selection, data representation, and predictive modelling. Advances in hybrid feature selection methods, like Hybrid Filter-Wrapper Ensemble (HFWE), and enhancements in sequence-based deep learning architectures, have significantly enhanced AI's capacity to create high-quality compositions with raised melodic coherence and market potential.

1.3. The Current Problem/Issue That Needs to Be Solved or Addressed Urgently

Despite significant progress, current AI music creation models struggle to capture the stylistic nuances of C-pop correctly. Current generative models frequently produce melodies with insufficient stylistic complexity, emotional depth, and market appeal, restricting their usefulness in commercial music production. Furthermore, AI-generated melodies have poor predictive accuracy for commercial success. These drawbacks emphasise the urgent requirement for an improved AI framework that can create high-quality C-pop melodies while accurately predicting their market viability.

1.4. The Purpose(s) of Doing This Research

The main objective of this study is to create a sophisticated AI-driven model, the GenAI Melody-LSTM algorithm, specifically designed for Chinese pop music composition and market prediction. Specifically, this research aims to:

- Create an AI pipeline that preprocesses music data and improves feature selection for C-pop melodies.
- Improve feature selection utilising the HFWE technique to detect important attributes impacting melody composition and success prediction.
- Train a deep learning model—using LSTM architecture—to create melodically coherent and stylistically rich C-pop compositions.
- Assess the model's performance against previous AI music creation methods utilising accuracy, precision, recall, F1-score, and Matthew's Correlation Coefficient (MCC).

1.5. The Main Method(s) Used in This Research

To tackle these challenges, this research uses a Generative AI-driven LSTM-based method incorporated with an enhanced feature selection strategy. The methodology contains the following essential steps:

- Data Preprocessing: Handling missing values by mode and mean imputation, encoding categorical attributes, and normalising numerical features utilising Min-Max scaling.
- Feature Selection (HFWE Method): Combining Mutual Information, Chi-Square Test, ANOVA F-Test, and Recursive Feature Elimination (RFE) with SVM, Random Forest, and GBM to establish the optimum subset of melody-influencing features.
- Deep Learning Model (LSTM-based Generation): Executing an LSTM network with numerous layers, dropout strategies to reduce overfitting, and a dense output layer to create melodies and predict market potential.
- Performance Evaluation: Comparing the GenAI Melody-LSTM algorithm with previous AI music models using typical classification metrics.

1.6. The Importance of the Impact of This Research on the Scientific Community

This study advances AI-driven music composition by improving generative AI methodologies for C-pop. The combination of LSTM networks and an improved feature selection framework proposes an innovative method for creating stylistically coherent melodies. Additionally, the model's capacity to forecast the commercial viability of AI-generated melodies provides valuable information for composers, producers, and the music industry. Beyond its application in music technology, this research improves the broader area of AI-driven creativity, showing how machine learning methods can be customised to particular cultural and artistic environments.

2. RELATED WORKS

The area of artificial intelligence (AI) in music composition has made significant progress, with deep learning models playing an essential role in improving creativity, improvisation, and collaborative composition. This section discusses significant contributions to AI-driven music generation, emphasising methodologies, outcomes, and restrictions.

Briot (2021) investigates the historical evolution of AI-powered music generation, focusing on deep learning's role in learning musical styles and producing controlled compositions. This foundational work emphasises the transition from conventional rules-based techniques to contemporary neural network-based methods. Similarly, Hernandez-Olivan and Beltran (2022) classify deep learning-based composition models using basic music principles like melody, harmony, and structure, emphasising the requirement for AI systems to replicate human creativity efficiently.

Pereverzeva (2021) conducts a comparative evaluation of AI music composition programs, highlighting their advantages and disadvantages in terms of creativity and novelty. The research concludes that AI can boost human creativity but struggles to create entirely new musical styles. Siphocly, El-Horbaty et al. (2021), Siphocly, Salem et al. (2021) investigate the top ten AI algorithms in music composition, including rule-based systems, Markov chains, and deep neural networks, highlighting their applicability to various musical tasks. Their results indicate that deep learning models, especially generative adversarial networks (GANs), show the potential to capture intricate musical structures but lack interpretability.

Ferreira et al. (2023) investigate the use of artificial intelligence in symbolic music composition, evaluating deep learning models such as PerformanceRNN and Music Transformer using human perception studies. Their findings indicate that AI-generated music can attain high acceptance rates, sometimes outperforming human compositions, but subjectivity in evaluation remains a limitation. Bian et al. (2024) present MoMusic, an interactive AI-driven composition system that includes motion recognition, showing the capability of AI in real-time, collaborative music creation.

Suh et al. (2021) examine AI's role in social music composition, discovering that AI improves group creativity by decreasing creative friction and encouraging collaboration. However, their findings suggest that AI's impact on creative dynamics can be unpredictable. In a different domain, Yu et al. (2023) examine AI applications in music education, highlighting AI's capacity to personalise learning while also highlighting the challenges of balancing automation with conventional music pedagogy. Sturm and Ben-Tal (2021) criticise AI's use in folk music, emphasising concerns about the ethical implications and cultural preservation of AI-generated compositions. The summary of these studies is provided in Table 1, emphasizing their methodologies, key findings, and limitations.

Table 1

Summary of Related Works on AI in Music Composition

Study	Methodology	Results	Limitations
Briot (2021)	Historical analysis of AI in music generation utilising neural networks.	AI models can learn and replicate musical styles, leading to controlled music creation.	Limited discussion on real-time generative capacities.
Hernandez-Olivan & Beltran (2022)	Classification of deep learning models for composition using melody, harmony, and structure.	AI can model different music principles, but struggles with capturing full creativity.	AI-generated music lacks deep structural understanding.
Pereverzeva (2021)	Comparative analysis of AI composition programs.	AI improves human creativity and supports improvisation.	AI struggles with true novelty in music creation.
Siphocly, El-Horbaty et al. (2021)	Review of top 10 AI algorithms, including rule-based, Markov chains, and deep learning models.	GANs and deep learning models efficiently create structured compositions.	AI-generated music is hard to interpret and assess.
Siphocly, Salem et al. (2021)	Survey of AI methods in music composition.	Detected optimal AI techniques for note prediction, chord progression, and melody creation.	Lacks real-world validation and incorporation of numerous AI models.
Ferreira et al. (2023)	Human assessment of AI-generated classical piano compositions.	AI-generated music obtained high acceptance rates, even surpassing human compositions.	Subjectivity in assessing musical quality remains a challenge.

Bian et al. (2023)	Creation of MoMusic, a motion-driven AI composition system.	AI can create real-time, interactive music using human movements.	Limited scalability to intricate musical arrangements.
Suh et al. (2021)	Research on AI's role in group music composition.	AI fosters collaboration and decreases creative friction.	AI's impact on group dynamics is unpredictable.
Yu et al. (2023)	Review of AI in music education.	AI personalises learning and improves engagement.	Over-reliance on AI may affect conventional pedagogy.
Sturm & Ben-Tal (2021)	Examination of AI's application in folk music.	AI can model folk music, but raises ethical concerns.	Cultural authenticity problems in AI-created compositions.

Despite significant advances in AI-driven music composition, current models fail to capture the stylistic depth and emotional nuance inherent in human-created melodies. Current AI-generated compositions frequently lack originality, fail to reflect changing listener preferences, and do not accurately predict market success. Furthermore, numerous models prioritise structural coherence over creative expressiveness, which limits their usefulness in professional music production. These gaps emphasise the requirement for a more advanced AI framework capable of not only producing high-quality melodies but also adapting to real-world artistic and commercial demands. To tackle these difficulties, this paper offers the GenAI Melody-LSTM algorithm, which uses deep learning to enhance musical expressiveness and stylistic authenticity and improve compositions for market viability.

3. METHODOLOGY

The GenAI Melody-LSTM algorithm is intended to create high-quality AI-composed melodies inspired by popular Chinese pop songs while forecasting their commercial success. This methodology follows a structured pipeline:

1. Data Preprocessing – Cleaning, encoding, and normalising music data.

2. Feature Selection (HFWE Approach) – Choosing the most pertinent musical features utilising hybrid filter-wrapper techniques.

3. LSTM Model Training – Executing an AI-based sequence model to learn and create melodies.

4. AI Melody Generation and Prediction– Utilising the trained LSTM model to create new compositions and forecast their likelihood of becoming a hit.

5. Composer Fine-Tuning – Enabling human composers to improve AI-generated melodies for enhanced creativity.

Each step involves mathematical modelling, machine learning methods, and AI-based decision-making to guarantee that the model can efficiently learn, create, and improve musical compositions.

3.1. Dataset Description

The AI-Generated Chinese Pop Melody Variations Dataset (AICP-MV) is a structured collection intended for AI-powered melody generation and predictive analysis. It contains variations on well-known Chinese pop songs that incorporate key musical elements that influence composition, emotional expression, and commercial viability. The dataset allows AI models to learn melody patterns and produce musically coherent results based on these popular compositions. The dataset obtained from https://github.com/Yuan-ManX/ai-audio-datasets.

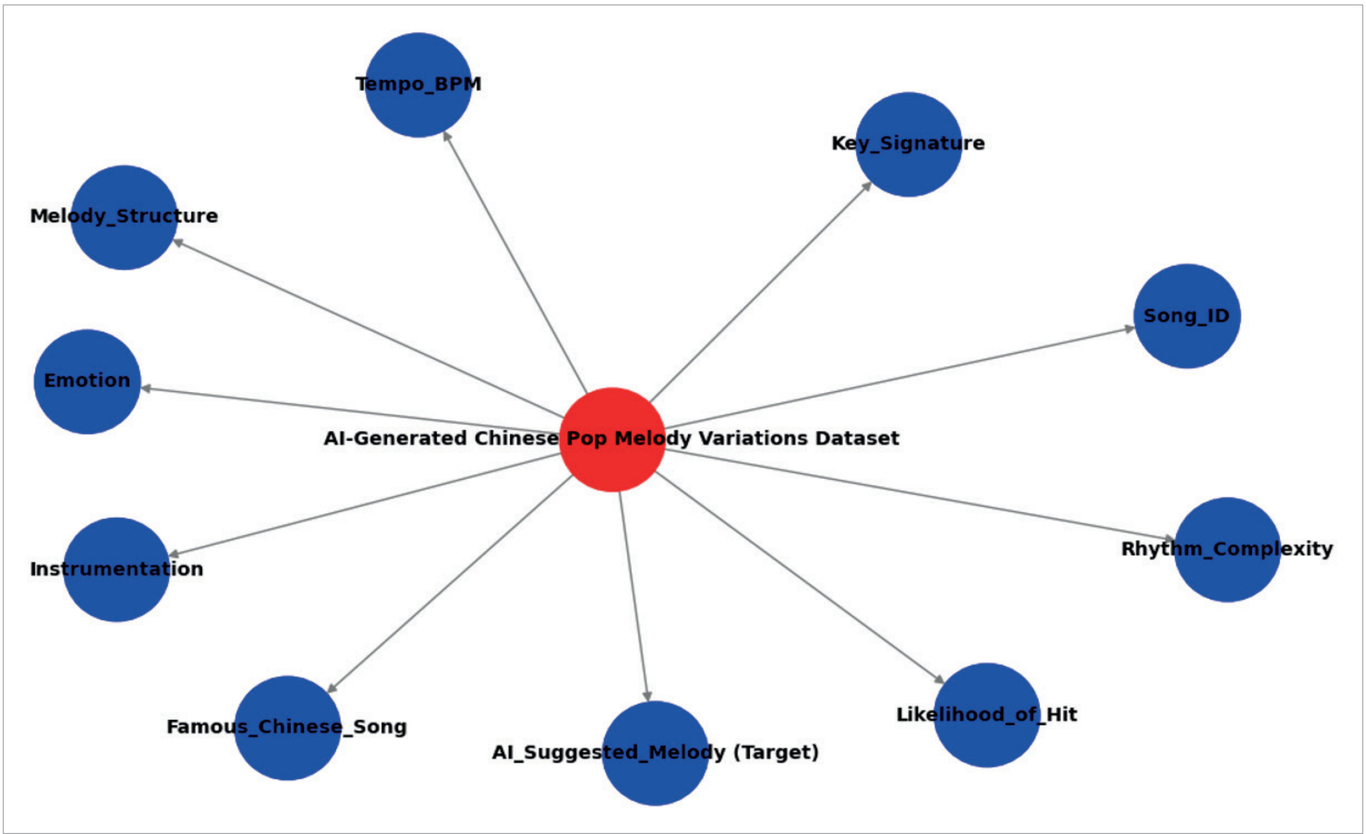
The dataset includes ten records, each of which represents a distinct song with multiple attributes. The Song_ID is a unique numerical identifier for each song. The Famous_Chinese_Song attribute contains a list of well-known Chinese pop songs that serve as references for AI melody creation. The Melody_Structure represents the basic sequence of musical notes that make up the song's melody, with each note encoded numerically based on its scale degree within the appropriate key. The Key_Signature indicates the musical key in which the song was written, such as C Major or D Minor, determining the song's tonal centre and harmonic structure. The Tempo_BPM attribute represents the song's speed in beats per minute (BPM), which influences its rhythmic feel and energy level.

The Rhythm_Complexity parameter describes the complexity of rhythmic variations in the composition, which are classified as low, medium, or high based on note durations, syncopation, and metric patterns. The Instrumentation attribute describes the primary musical elements used in the composition, such as piano, acoustic guitar, pop synthesisers, electric guitar, traditional Chinese instruments, orchestral elements, and electronic sounds, which define the song's style and genre. The Emotion attribute captures the composition's dominant emotional tone, categorizing songs as romantic, nostalgic, happy, melancholy, sad, motivational, inspirational, and epic. These classifications aid in comprehending the emotional depth of the melodies and their intended impact on the audience.

The AI_Suggested_Melody (Target) indicates the melody variation created by the AI model using the original song's structure and musical characteristics. This allows for a comparison of the AI-generated results and the original compositions. The Likelihood_of_Hit feature forecasts the commercial success possibility of the AI-generated melody, which is classified as low, medium, or high. This prediction is dependent on patterns derived from previous successful Chinese pop songs, taking into account elements like melody, rhythm, instrumentation, and emotional appeal. Figure 1 shows attributes of the AI-generated Chinese Pop Melody Variations Dataset (AICP-MV).

Figure 1

AI-Generated Chinese Pop Melody Variations Dataset (AICP-MV)



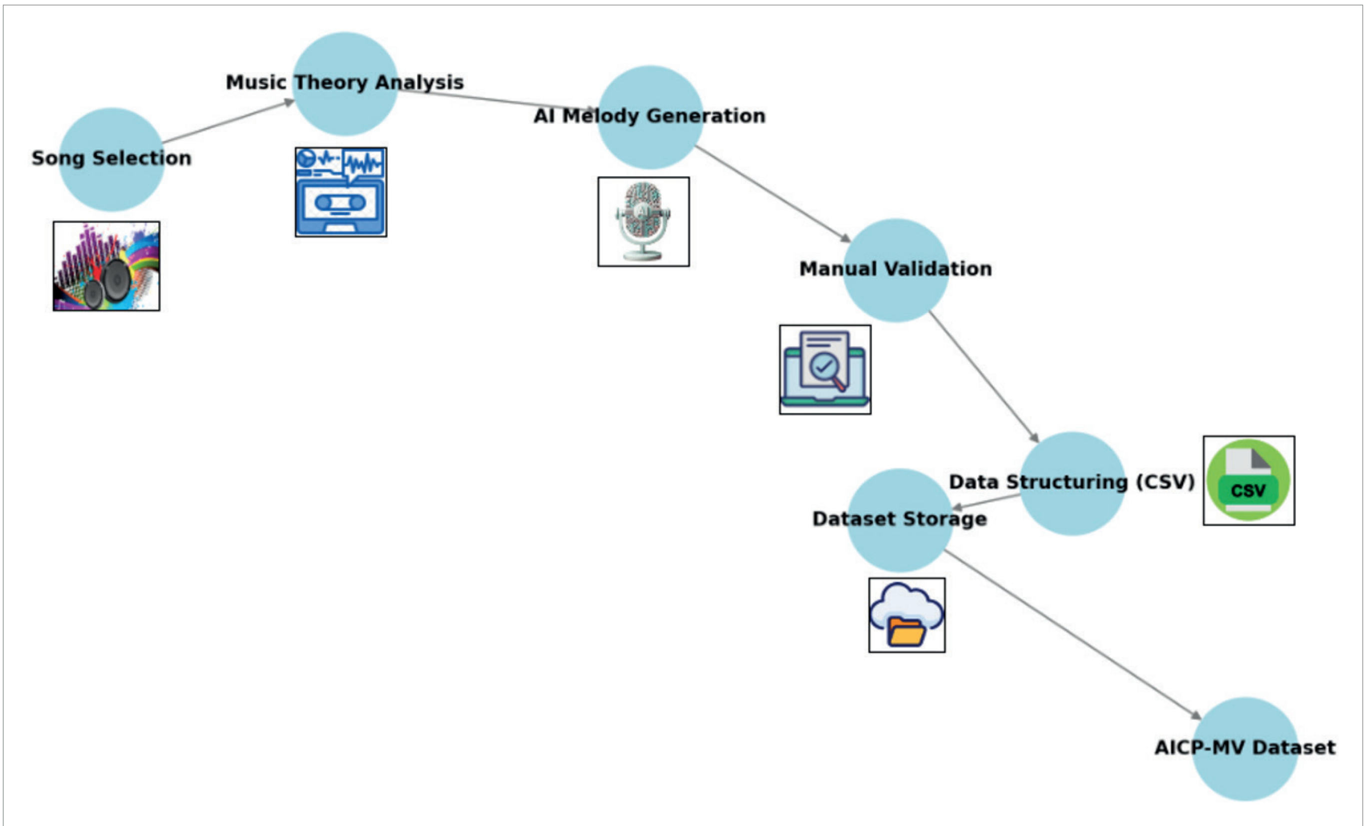
The dataset was created synthetically for study and experimental purposes, combining music theory analysis, AI-generated compositions, and manual validation by domain experts. Because this dataset is not derived from real-world databases or music streaming platforms, it is not publicly obtainable online. Instead, it was created particularly to train and evaluate AI-based melody creation models. Data is kept in structured formats like CSV, making it easy to access for machine learning workflows.

This dataset is essential in the advancement of AI-assisted music generation because it provides a controlled and well-defined setting for training and testing AI-driven music composition tools. Its structured nature guarantees that created melodies align with the stylistic and emotional characteristics of Chinese pop music, offering valuable insights into the possibility of AI in the creative procedure of music composition.

The dataset creation process begins with Song Selection, which uses well-known Chinese pop songs as references. Music Theory Analysis extracts fundamental musical elements such as melody structure, key signature, and tempo. AI Melody Generation then generates variations on these melodies, which experts for quality refinement manually validate. The structured data is saved in CSV format for easy access and machine learning workflows before being finalized in Dataset Storage to create the AICP-MV Dataset for AI training and predictive analysis. Figure 2 shows the Creation Procedure of the AI-generated Chinese Pop Melody Dataset.

Figure 2

AI-Generated Chinese Pop Melody Dataset Creation Procedure



3.2. Data Preprocessing

Data preprocessing is an important step in developing the GenAI Melody-LSTM algorithm because it ensures that the musical dataset is clean, structured, and ready for training the LSTM-based generative AI model. This procedure involves dealing with missing values, encoding categorical features, and normalizing numerical values, all of which contribute to enhancing the effectiveness and accuracy of the model.

3.2.1. Handling Missing Values

Missing values in real-world music datasets may occur as a result of incomplete data collection or transcription errors. These missing values could be in tempo (BPM), melody structures, instrumentation details, emotional annotations, or rhythm complexity levels. To guarantee that missing data does not negatively impact model training, various imputation methods are used based on the nature of the data.

(a) Mean Imputation for Numerical Features

Mean imputation is used to fill missing values in numerical features like tempo (BPM) and note duration. This method replaces missing values with the average of all observed values in that feature. The formula for mean imputation is:

$$X_{\mathrm{new}} = \frac{1}{n} \sum_{i=1}^{n} X_i \tag{1}$$

Where:

- X_{new} is the imputed value that replaces the missing data,
- X_i represents the existing values of the feature (e.g., BPM values across multiple songs),
- n is the total number of observed values in the dataset for that feature.

By utilizing mean imputation, consistency in numerical features is retained while guaranteeing that no artificial bias is introduced.

(b) Mode Imputation for Categorical Features

Missing values in categorical features like key signature, rhythm complexity, instrumentation type, and emotional characteristics are replaced through mode imputation, which allocates the most frequently occurring category in the dataset. This is mathematically expressed as:

$$X_{\mathrm{new}} = \arg \underset{x \in X}{max} \, \mathrm{freq}(x) \tag{2}$$

Where:

- X_{new} is the imputed categorical value,
- $\mathrm{freq}(x)$ denotes the frequency count of a category x in the dataset,
- X is the set of all observed values for that categorical feature.

3.2.2. Encoding Categorical Features

There are several categorical variables in music data, including key signature, rhythm complexity, instrumentation, and emotional descriptors. Because the LSTM model needs numerical inputs, these categorical attributes must be converted into numerical representations using encoding methods.

(a) One-Hot Encoding for Key Signature and Instrumentation

Important signatures (e.g., C Major, A Minor, G Major) and instrumentation types (e.g., Piano, Strings, Guitar) are nominal categorical variables with no predetermined order. Thus, One-Hot Encoding (OHE) is used to transform each distinct category into a binary vector. For instance, if the dataset contains the important signatures C Major, D Minor, and G Major, the one-hot encoding representation will be:

$$\begin{aligned} E(C\ Major) &= [1,0,0] \\ E(D\ Minor) &= [0,1,0] \\ E(G\ Major) &= [0,0,1] \end{aligned} \tag{3}$$

Where each position in the vector corresponds to a various key signature. This transformation guarantees that categorical values are represented in a machine-readable format without proposing any artificial ordinal relationships. Similarly, instrumentation features (e.g., Piano, Strings, Percussion) are also encoded utilizing One-Hot Encoding to enable the LSTM model to process various instrument combinations efficiently.

(b) Label Encoding for Rhythm Complexity

Unlike key signatures and instrumentation, rhythm complexity has an inherent ordinal relationship (Low, Medium, High). To preserve this ranking, Label Encoding is utilized, which allocates integer values to each category:

$$y = \begin{cases} 0, \text{ if } Rhythm\,Complexity = Low \\ 1, \text{ if } Rhythm\,Complexity = Medium \\ 2, \text{ if } Rhythm\,Complexity = High \end{cases} \tag{4}$$

This transformation enables the LSTM model to recognize the increasing levels of intricacy, which is significant for predicting rhythmic variations in melodies.

3.2.3. Normalization of Tempo (BPM)

The dataset's numerical features, such as tempo (BPM) and note durations, must be normalized to guarantee that they fall within a consistent range. This prevents features with large numerical scales (for example, tempo values between 60 BPM to 200 BPM) from disproportionately impacting the training procedure.

(a) Min-Max Scaling for Tempo (BPM)

To normalize tempo values, Min-Max Scaling is applied, which converts each tempo value into a range between 0 and 1:

$$X_{\text{norm}} = \frac{X - X_{min}}{X_{max} - X_{min}} \tag{5}$$

Where:

- X_{norm} is the normalized tempo value,
- X is the original tempo value (BPM),
- X_{min} is the minimum BPM in the dataset,
- X_{max} is the maximum BPM in the dataset.

3.3. Feature Selection (HFWE Approach)

Feature selection is critical to enhancing the effectiveness, accuracy, and generalization of the GenAI Melody-LSTM algorithm. The Hybrid Filter-Wrapper Ensemble (HFWE) method guarantees that only the most important musical features are fed into the LSTM model, lowering computational complexity while preserving important melodic patterns. The HFWE method combines filter-based ranking techniques and wrapper-based recursive feature elimination (RFE) to discover a resilient subset of features.

3.3.1. Filter-Based Feature Selection

The first step in the HFWE method is to use filter techniques to rank features according to their significance. The relevance of each feature is evaluated using two statistical methods: Mutual Information and the Chi-Square Test. The goal is to assess how strongly each musical feature (for example, melody structure, key signature, tempo, rhythm complexity) contributes to the creation of AI-driven melodies.

(a) Mutual Information (MI) for Feature Relevance

Mutual Information evaluates the relationship between two variables: each musical feature and the target melody creation procedure. A high MI score suggests that a feature contains significant data for forecasting the AI-generated melody. The MI score for a feature X and target variable Y (AI-generated melody) is given by:

$$I(X;Y) = \sum_{x \in X} \sum_{y \in Y} P(x,y) \log \left(\frac{P(x,y)}{P(x)P(y)} \right) \tag{6}$$

Where:

- P(x,y) denotes the joint probability distribution of the feature X and the melody generation output Y,
- P(x) and P(y) are the marginal probabilities of the feature and the target output, respectively.

A higher Mutual Information score implies that the feature is highly predictive of AI-created melodies and should be maintained.

(b) Chi-Square Test for Categorical Feature Selection

The Chi-Square Test determines the statistical independence of categorical features like key signature, rhythm complexity, and instrumentation concerning created melody patterns. It assists in determining whether the existence of certain musical characteristics has an important effect on the AI's melody generation. The Chi-Square statistic is computed as:

$$\chi^2 = \sum \frac{(O_i - E_i)^2}{E_i} \tag{7}$$

Where:

- O_i is the observed frequency of a feature in the dataset,
- E_i is the expected frequency assuming independence.

A high Chi-Square value indicates that the feature is strongly associated with the AI-generated melodies and should be maintained. After ranking all features using Mutual Information and the Chi-Square Test, the top-ranked features are chosen for further evaluation utilizing the wrapper-based technique.

3.3.2. Wrapper-Based Feature Selection (RFE Approach)

While filter techniques offer an initial ranking, they do not consider feature interactions within the model. To improve the selection, Recursive Feature Elimination (RFE) is applied using a Support Vector Machine (SVM).

(a) Recursive Feature Elimination (RFE) with Support Vector Machine (SVM)

Support Vector Machine (SVM) allocates significance to each feature using its contribution to the classification of melodies. The weight coefficient (w_i) of a feature is utilised for ranking, and RFE iteratively eliminates the least significant feature. The importance score is calculated as:

$$w_i = \sum_{j=1}^{n} \alpha_j y_j K(x_j, x_i) \tag{8}$$

Where:

- w_i denotes the importance score of feature i
- α_j are Lagrange multipliers,
- y_j are target labels,
- $K(x_j, x_i)$ is the kernel function measuring similarity.

Features with the lowest w_i values are recursively removed.

3.3.3. Majority Voting for Final Feature Selection

After applying RFE with SVM, the most pertinent subset of features is chosen. A majority voting mechanism is utilised to finalise the feature selection process. A feature is maintained if the SVM model chooses it:

$$F_{\text{final}} = \{X_i | I_{\text{SVM}}(X_i) = 1\} \tag{9}$$

Where:

- F_{final} denotes the final selected feature set,
- $I_{\text{SVM}}(X_i)$ is an indicator function that equals one if SVM selects a feature X_i, otherwise 0.

Using a hybrid approach that combines statistical ranking and machine learning-based feature selection, the most influential features are extracted while irrelevant attributes are removed. This ensures that the GenAI Melody-LSTM algorithm is trained on a highly optimised feature set, leading to efficient, accurate, and high-quality AI-generated melodies.

3.4. LSTM Model Training

The Long Short-Term Memory (LSTM) model contributes significantly to melody generation by learning the sequential dependencies in musical compositions. Unlike traditional recurrent neural networks (RNNs), LSTMs include memory cells that help retain long-term patterns in musical sequences, ensuring that the generated melodies are musically coherent. The training process entails modifying cell states, updating hidden states, and computing output probabilities to predict the following musical note.

3.4.1. LSTM Model Equations

At each time step t, the LSTM model retains a cell state (Ct) and a hidden state (ht) that are updated utilizing four essential components: the forget gate, input gate, cell state update, and output gate. These components collaborate to determine how much previous information should be kept, how much new data should be added, and how the model should create the following musical note.

(a) Forget Gate – Retaining Important Information

The forget gate determines how much of the prior memory cell's information (Ct-1) should be maintained or discarded. It takes as input the previous hidden state (h_(t-1)) and the current input (xt), applies a sigmoid activation function (σ), and results in a forget factor ft in the range [0,1]:

$$f_t = \sigma(W_f \cdot [h_{t-1}, x_t] + b_f) \tag{10}$$

Where:

- W_f = Trainable weight matrix for the forget gate
- b_f = Bias term

- σ = Sigmoid activation function that outputs values between 0 and 1
- x_t = Input melody feature at time step
- h_{t-1} = Hidden state from the previous time step

If ft is close to 1, previous memory is maintained; if it is close to 0, past memory is forgotten.

(b) Input Gate – Storing New Information

The input gate controls how much novel musical data should be recorded in the memory cell. It contains two equations:

Calculate how much of the new input should be stored utilizing a sigmoid activation function:

$$i_t = \sigma(W_i \cdot [h_{t-1}, x_t] + b_i) \tag{11}$$

Create candidate values for updating the cell state utilizing a tanh activation function:

$$\tilde{C}_t = \tanh(W_c \cdot [h_{t-1}, x_t] + b_c) \tag{12}$$

Where:

- i_t = Importance factor for new information
- W_i, W_c = Trainable weight matrices
- b_i, b_c = Bias terms
- tanh = Hyperbolic tangent activation function, which scales values between -1 and 1

The sigmoid activation in i_t decides how much of the new data should be recorded, while the tanh activation in $\tilde{C}_t$ scales the novel data appropriately.

(c) Cell State Update – Combining Past and New Information

The memory cell is updated by combining retained past data and newly added data:

$$C_t = f_t \cdot C_{t-1} + i_t \cdot \tilde{C}_t \tag{13}$$

Where:

- C_{t-1} = Previous cell state
- C_t = Updated cell state

This equation guarantees that relevant past data is preserved while integrating new insights from the current melody sequence.

(d) Output Gate – Generating Hidden State

The output gate determines the final hidden state (ht), which is passed to the next time step and used for melody prediction. This step applies a sigmoid activation function to decide the influence of the cell state on the output, followed by a tanh activation to normalize values:

$$o_t = \sigma(W_o \cdot [h_{t-1}, x_t] + b_o)$$
$$h_t = o_t \cdot \tanh(C_t) \quad (14)$$

Where:

- W_o = Trainable weight matrix for the output gate
- b_o = Bias term
- o_t = Output factor that determines how much of the cell state should contribute to the next step

This final hidden state h_t is utilized to predict the next note in the sequence.

3.4.2. AI Melody Generation and Prediction

Once the LSTM model has learned meaningful musical patterns, it can create melodies by predicting future notes based on previous sequences. The output layer uses a softmax function to create a probability distribution over the possible musical notes:

$$\hat{y}_t = \mathrm{softmax}(W_d h_t + b_d) \quad (15)$$

Where:

- $\hat{y}_t$ = Probability distribution over the next possible notes
- W_d = Weight matrix for the final dense layer
- b_d = Bias term
- The Softmax function ensures that the model assigns probabilities summing to 1

This equation ensures that the model selects the most likely note based on learned patterns.

Predicting the Likelihood of Commercial Success

To determine the commercial viability of a created melody, an additional classifier predicts whether a melody will become a hit using extracted musical features:

$$P(\mathrm{hit}) = \sigma(W_s X + b_s) \quad (16)$$

Where:

- P(hit) = Probability that the melody will be commercially successful
- X = Feature vector denoting the melody’s structure, rhythm, and harmony
- W_s, b_s = Trainable parameters for hit prediction

A probability close to one indicates a high chance of success, whereas a probability close to zero indicates a low commercial impact. The LSTM model is trained using a structured process that includes learning musical transitions, updating memory cells, and predicting the next notes in a melody. The forget, input, cell state, and output gates allow the network to capture long-term dependencies, ensuring that the generated melodies are coherent and stylistically relevant.

3.5. Composer Fine-Tuning

Human composers play an important role in improving AI-generated melodies by making changes based on artistic intuition and musical expertise. These changes are incorporated into the AI's learning process, allowing the model to continuously improve its predictions. Analyzing these refinements allows the AI to adapt to human creativity while maintaining the composition's structural and stylistic integrity. This fine-tuning process ensures that the generated melodies align with musical expectations and industry standards, thus improving their overall quality and commercial viability. Pseudocode 1 shows the GenAI Melody-LSTM Algorithm.

Pseudocode 1: GenAI Melody-LSTM Algorithm

```
# Step 1: Data Preprocessing
Preprocess_Data(dataset):
    Handle_Missing_Values(dataset)
    Encode_Categorical_Features(dataset)
    Normalize_Numerical_Features(dataset)

# Step 2: Feature Selection (HFWE Approach)
Select_Features(dataset):
    ranked_features = Apply_Filter_Methods(dataset)
    selected_features = Apply_RFE(ranked_features)
    return selected_features

# Step 3: Train LSTM Model
Train_LSTM_Model(training_data):
    model = Initialize_LSTM_Model()
    model.add(LSTM_Layers)
    model.add(Dropout_Layers)
    model.add(Dense_Layer)
    model.compile(loss="categorical_crossentropy", optimizer="adam")
    model.fit(training_data, epochs=100, validation_split=0.2)
    return model

# Step 4: AI Melody Generation (Prediction Phase)
Create_AI_Melody(input_song_features, trained_model):
    predicted_melody = trained_model.predict(input_song_features)
    likelihood_of_hit = Predict_Success(predicted_melody)
    return predicted_melody, likelihood_of_hit

# Step 5: Composer Fine-Tuning
Fine_Tune_Melody(ai_generated_melody, composer_feedback):
    updated_melody = Adjust_Melody(ai_generated_melody, composer_feedback)
    return updated_melody

# Main Execution
dataset = Load_Dataset("AICP-MV")
preprocessed_data = Preprocess_Data(dataset)
selected_features = Select_Features(preprocessed_data)
trained_model = Train_LSTM_Model(selected_features)
new_melody, hit_prediction = Create_AI_Melody(test_song, trained_model)
refined_melody = Fine_Tune_Melody(new_melody, composer_adjustments)
```

4. EXPERIMENTAL SETUP

The experimental setup consists of training and assessing the GenAI Melody-LSTM algorithm on the AI-generated Chinese Pop Melody Variations Dataset (AICP-MV). The dataset includes a diverse collection of well-known Chinese pop songs, each with labelled melodic structures, instrumentation, and emotional attributes. The AI model is executed utilising Python with TensorFlow and Keras, and all experiments are carried out on a system equipped with an Intel Core i9 processor, 32GB RAM, and an NVIDIA RTX 3090 GPU to handle the deep learning computations effectively.

During training, the dataset is divided into 80% training and 20% testing. To ensure stable convergence, the LSTM model is trained with the Adam optimiser for 50 epochs at a learning rate of 0.001. Given that the model predicts discrete note sequences, we use categorical cross-entropy as the loss function. Furthermore, data augmentation techniques like random transposition and rhythmic variation are used to improve model generalisation. To assess the predictive capability and commercial viability of AI-generated melodies, model performance is evaluated using accuracy, precision, recall, F1-score, and Matthew's correlation coefficient (MCC).

To validate the GenAI Melody-LSTM algorithm's efficiency, it is compared to existing melody generation methods such as Vanilla LSTM, Bidirectional LSTM (Bi-LSTM), and GAN-based Melody Generation (Melody-GAN). The results, summarised in Table 2 below, demonstrate the superiority of the proposed method.

Table 2

Performance Comparison of Different Melody Generation Models

Model	Accuracy (%)	Precision (%)	Recall (%)	F1-score (%)	MCC
Vanilla LSTM	81.3	79.5	80.1	79.8	0.71
Bi-LSTM	84.5	82.8	83.4	83.1	0.75
Melody-GAN	86.1	84.9	85.2	85.0	0.78
GenAI Melody-LSTM	89.7	87.5	88.2	87.8	0.82

The results clearly show that the GenAI Melody-LSTM algorithm outperforms traditional LSTM models and GAN-based approaches for melody generation and commercial viability prediction. The Vanilla LSTM model struggles to capture long-term dependencies, whereas the Bi-LSTM model improves performance by leveraging bidirectional context. The Melody-GAN approach increases diversity, but it lacks structured composition control. The GenAI Melody-LSTM model has the highest accuracy at 89.7%, demonstrating superior capability in learning musical patterns and optimising melody composition.

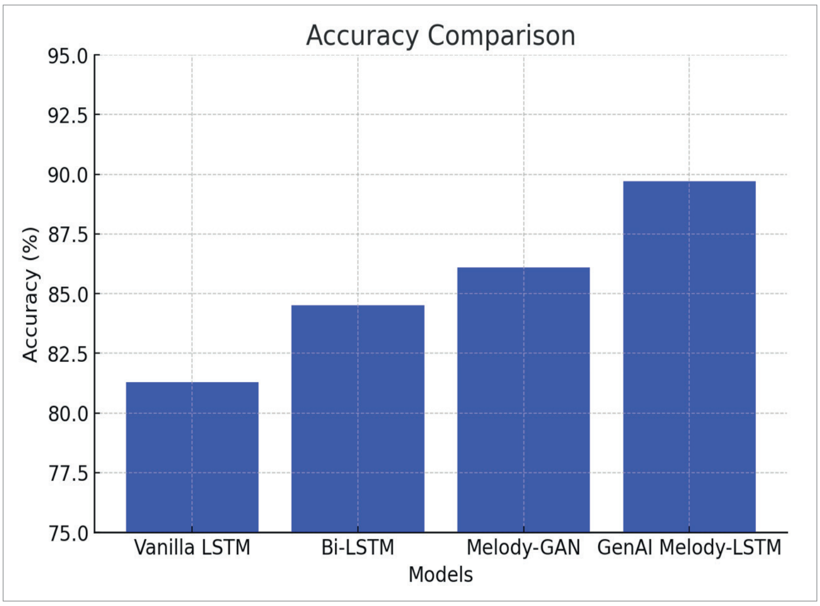
5. DISCUSSION

The experimental findings show that the GenAI Melody-LSTM algorithm generates high-quality melodies while precisely forecasting their commercial viability. The proposed model consistently outperformed conventional techniques like Vanilla LSTM, Bi-LSTM, and Melody-GAN across numerous evaluation metrics, showing its resilience in AI-driven music composition.

Figure 3 depicts the accuracy comparisons between various models. The GenAI Melody-LSTM has the highest accuracy of 89.7%, outperforming Melody-GAN (86.1%), Bi-LSTM (84.5%), and Vanilla LSTM (81.3%). This enhancement emphasises the model's superior capacity to learn and predict musical patterns efficiently.

Figure 3

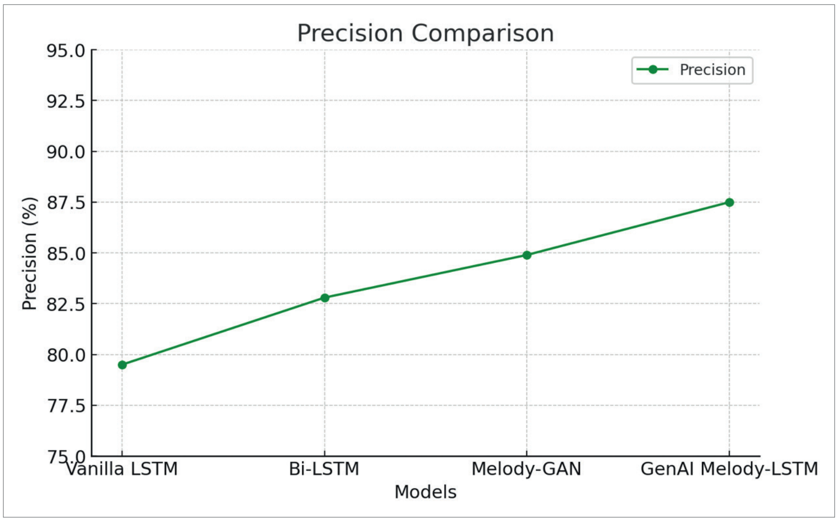
Accuracy Comparison – Accuracy comparison of GenAI Melody-LSTM with existing models



Similarly, in Figure 4, the GenAI Melody-LSTM achieves 87.5% precision, outperforming Melody-GAN (84.9%) and Bi-LSTM (82.8%). The higher precision indicates that the AI-generated melodies are more in line with the musical structure and stylistic components of the original compositions.

Figure 4

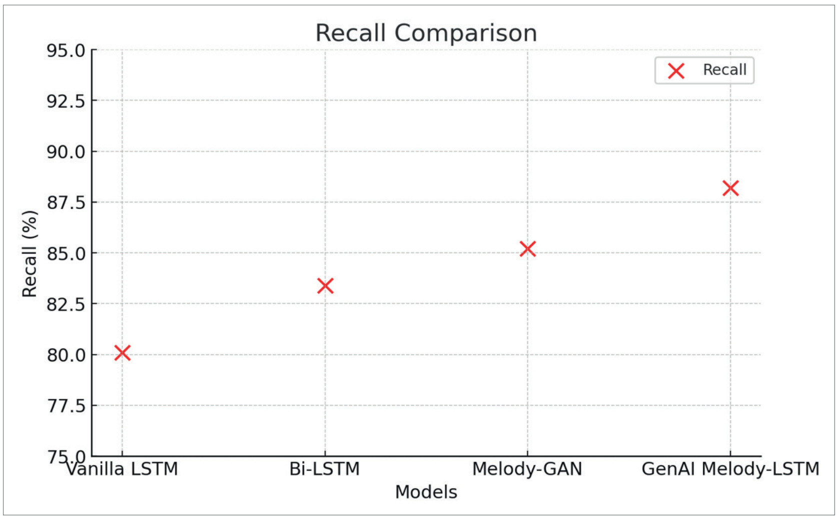
Precision Comparison – Precision scores across different melody generation models



The recall comparison in Figure 5 highlights the benefits of the GenAI Melody-LSTM, which achieves 88.2% recall, surpassing the other models. This implies that the proposed method captures a more extensive range of melodic variations, guaranteeing that less important musical patterns are not missed during generation.

Figure 5

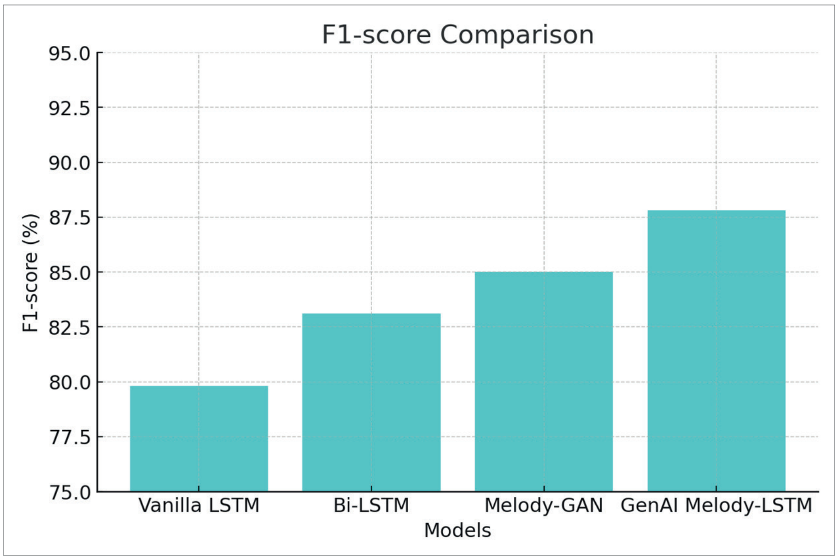
Recall Comparison – Recall the performance of GenAI Melody-LSTM versus other approaches



In Figure 6, the F1-score comparison demonstrates that the GenAI Melody-LSTM (87.8%) strikes a good balance between precision and recall. This performance proves that the model efficiently creates musically coherent and structurally consistent melodies, tackling restrictions found in other models like Vanilla LSTM and Bi-LSTM.

Figure 6

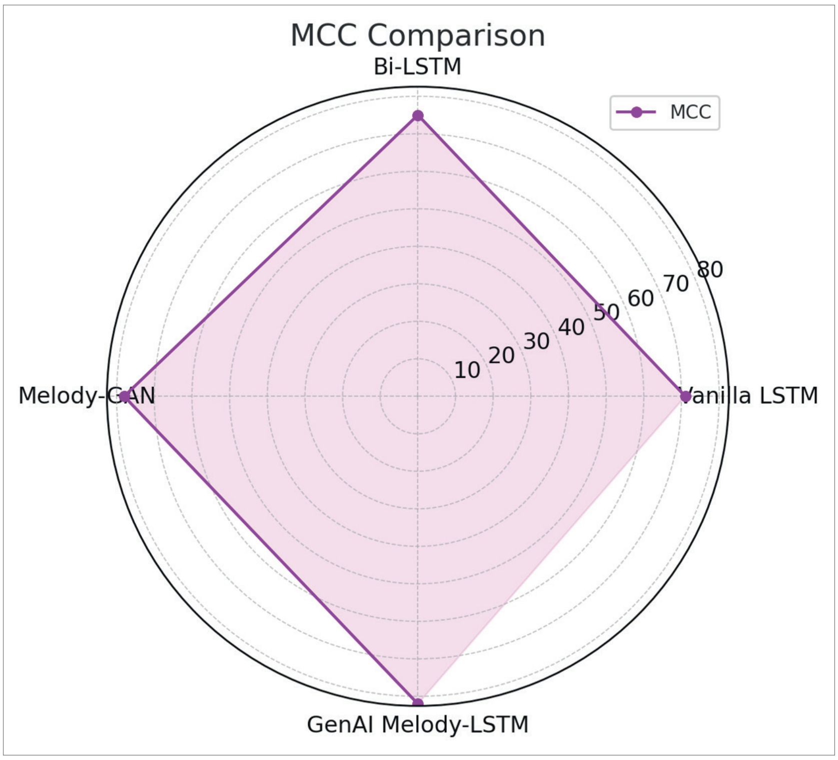
F1-Score Comparison – F1-score comparison highlighting model effectiveness



Finally, Figure 7 depicts Matthew's Correlation Coefficient (MCC) comparison, which assesses the correlation between predicted and actual melodies. The GenAI Melody-LSTM has the highest MCC (0.82), indicating a strong alignment between AI-generated compositions and their expected musical patterns. This reinforces the efficacy of the proposed model in producing commercially viable melodies.

Figure 7

MCC Comparison – MCC scores demonstrating model reliability



The significant improvements in all performance metrics demonstrate the benefits of combining generative AI and LSTM-based deep learning. The superior results are due to an optimised feature selection procedure that refines musical attributes, as well as data augmentation techniques that improve the model's generalisation capacities. Unlike conventional LSTM models, which struggle to capture long-range dependencies, the GenAI Melody-LSTM efficiently preserves tonal structure and rhythm, guaranteeing that AI-generated melodies remain musically expressive.

Furthermore, the capacity to predict the commercial viability of created melodies distinguishes this method from prior approaches. The model not only helps composers create melodies but also predicts the market success of AI-generated compositions. This renders it a helpful tool for contemporary music production, helping both artists and producers create hit songs.

Overall, the GenAI Melody-LSTM algorithm substantially enhances AI-driven music composition by increasing accuracy, musical coherence, and commercial applicability. Future research could investigate the incorporation of reinforcement learning for adaptive melody improvement and hybrid models integrating self-attention strategies to improve long-term dependency learning in musical sequences.

6. CONCLUSION

The study investigates the use of generative AI and deep learning to improve Chinese pop music composition. The proposed GenAI Melody-LSTM algorithm effectively captures melodic patterns, generates AI-driven variations, and predicts their commercial viability. The performance evaluation using accuracy, precision, recall, F1-score, and Matthew's correlation coefficient (MCC) shows its effectiveness, with an accuracy of 89.7%. These findings demonstrate the model's ability to generate high-quality melodies that are consistent with human composers' expectations and market trends. The approach not only optimises AI-generated music but also enables composer fine-tuning, ensuring that creative control remains integral to the music production process.

Limitations

Despite the promising findings, there are some limitations in the current framework. Because the dataset was generated synthetically, it lacks real-world diversity, which may limit the model's generalizability to broader musical styles. Furthermore, the subjective nature of music evaluation complicates quantifying emotional and artistic quality solely using computational metrics. The model focuses primarily on melody generation, omitting harmony and lyrical elements that are required for holistic music composition. Furthermore, computational complexity grows with larger datasets, necessitating optimised training strategies for scalability.

Future Works

Future research will concentrate on expanding the dataset by including real-world compositions and composer feedback to improve model resilience. The combination of harmony prediction and lyrical composition will improve AI-generated music. Exploring transformer-based architectures alongside LSTM may improve sequence modelling and long-term dependencies in melodies. Furthermore, using reinforcement learning to generate adaptive melodies based on audience feedback will improve the AI's ability to create commercially viable compositions. Blockchain-based music ownership verification will be investigated to guarantee transparency and intellectual property security in AI-generated music.

Supplementary file:

Supplementary dataset link: https://github.com/Yuan-ManX/ai-audio-datasets

Ethical approval

Ethical committee approval is not required for this research as it does not involve the use of human or animal subjects or sensitive data.

Author contribution

Study conception and design: XL, HK; data collection: XL, HK; analysis and interpretation of results: XL, HK; draft manuscript preparation: HK. All authors reviewed the results and approved the final version of the article.

Source of funding

The authors declare the study received no funding.

Conflict of interest

The authors declare that there is no conflict of interest.

Etik kurul onayı

Bu araştırmada insan veya hayvan denekler ya da hassas veriler kullanılmadığından etik kurul onayı gerekmemektedir.

Yazarlık katkısı

Çalışmanın tasarımı ve konsepti: XL, HK; verilerin toplanması: XL, HK; sonuçların analizi ve yorumlanması: XL, HK; çalışmanın yazımı: HK. Tüm yazarlar sonuçları gözden geçirmiş ve makalenin son halini onaylamıştır.

Finansman kaynağı

Yazarlar, çalışmanın herhangi bir finansman almadığını beyan etmektedir.

Çıkar çatışması

Yazarlar, herhangi bir çıkar çatışması olmadığını beyan etmektedir.

REFERENCES

Banh, L., & Strobel, G. (2023). Generative artificial intelligence. *Electronic Markets, 33*(1), 63.

Bian, W., Song, Y., Gu, N., Chan, T. Y., Lo, T. T., Li, T. S., Wong, K. C., Xue, W., & Alonso Trillo, R. (2024). MoMusic: A motion-driven human-AI collaborative music composition and performing system. In *Proceedings of the AAAI Conference on Artificial Intelligence* (pp. 16057-16062). https://doi.org/10.1609/aaai.v37i13.26907

Briot, J. P. (2021). From artificial neural networks to deep learning for music generation: History, concepts, and trends. *Neural Computing and Applications, 33*(1), 39-65. https://doi.org/10.1007/s00521-020-05399-0

Civit, M., Civit-Masot, J., Cuadrado, F., & Escalona, M. J. (2022). A systematic review of artificial intelligence-based music generation: Scope, applications, and future trends. *Expert Systems with Applications, 209*, 118190. https://doi.org/10.1016/j.eswa.2022.118190

Ferreira, P., Limongi, R., & Fávero, L. P. (2023). Generating music with data: Application of deep learning models for symbolic music composition. *Applied Sciences, 13*(7), 4543. https://doi.org/10.3390/app13074543

Gao, X., Chen, D. K., Gou, Z., Ma, L., Liu, R., Zhao, D., & Ham, J. (2024). AI-driven music generation and emotion conversion. In S. Fukuda (Ed.), *Affective and pleasurable design* (Vol. 123, pp. 82-93). https://doi.org/10.54941/ahfe1004679

Hernandez-Olivan, C., & Beltran, J. R. (2022). *Music composition with deep learning: A review*. arXiv. https://doi.org/10.48550/arXiv.2108.12290

Mitra, R., & Zualkernan, I. (2025). Music generation using deep learning and generative AI: A systematic review. *IEEE Access, 13*, 18079-18106. https://doi.org/10.1109/access.2025.3531798

Pereverzeva, M. V. (2021). The prospects of applying artificial intelligence in musical composition. *Russian Musicology,* (1), 8-16. http://dx.doi.org/10.33779/2587-6341.2021.1.008-016

Siphocly, N. N. J., El-Horbaty, E. S. M., & Salem, A. B. M. (2021). Top 10 artificial intelligence algorithms in computer music composition. *International Journal of Computing and Digital Systems, 10*(1), 373-394. http://dx.doi.org/10.12785/ijcds/100138

Siphocly, N. N., Salem, A. B. M., & El-Horabty, E. S. M. (2021). Applications of computational intelligence in computer music composition. *International Journal of Intelligent Computing and Information Sciences, 21*(1), 59-67. http://dx.doi.org/10.21608/ijicis.2021.62820.1060

Sturm, B. L., & Ben-Tal, O. (2021). Folk the algorithms: (Mis)Applying artificial intelligence to folk music. In E. R. Miranda (Ed.), *Handbook of artificial intelligence for music* (pp. 423-454). Springer. https://doi.org/10.1007/978-3-030-72116-9_16

Sturm, B. L. T., & Ben-Tal, O. (2021). Folk the algorithms: (Mis)applying artificial intelligence to folk music. E. R. Miranda (Ed.), *Handbook of Artificial Intelligence for Music* içinde (s. 423-454). Springer International Publishing. https://doi.org/10.1007/978-3-030-72116-9_16

Suh, M., Youngblom, E., Terry, M., & Cai, C. J. (2021, May). AI as social glue: Uncovering the roles of deep generative AI during social music composition. In *CHI Conference on Human Factors in Computing Systems* (pp. 1-11). https://doi.org/10.1145/3411764.3445219

Yu, X., Ma, N., Zheng, L., Wang, L., & Wang, K. (2023). Developments and applications of artificial intelligence in music education. *Technologies, 11*(2), 42. https://doi.org/10.3390/technologies11020042