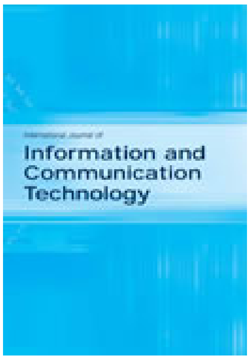


International Journal of Information and Communication Technology

ISSN online: 1741-8070 - ISSN print: 1466-6642
https://www.inderscience.com/ijict

AI-driven classification and trend analysis of piano music genres using large language models

Zhenfang Liu

DOI: 10.1504/IJICT.2025.10070770

Article History:

Received:	19 February 2025
Last revised:	01 March 2025
Accepted:	01 March 2025
Published online:	08 May 2025

AI-driven classification and trend analysis of piano music genres using large language models

Zhenfang Liu

School of Arts and Sports,
Henan Open University,
Zhengzhou City, Henan 450000, China
Email: liuzhenfang2023@163.com

Abstract: In this work, we propose a hybrid framework to classify piano genres and predict future genres based on symbolic features and large language models (LLMs). Genre overlap, multimodal data, and sparse metadata are cumbersome to traditional methods. SymD has been used to process symbolic data from MIDI files and textual metadata via GPT-4 Turbo. We trained predictions of 20,000 compositions on 20,000 LLM embeddings, which fusion features including note density, tempo variability, and harmonic structure to 94.0% accuracy and 0.93 F1. Matching historical data gave good alignment for temporal trend analysis. We also improve on existing methods and participate in overcoming metadata limitations. This study presents a new multimodal paradigm to analyse music, which can be applied in musicology, digital archiving, and recommendation systems. Real-time audio-based deployment and integration will occur in future work.

Keywords: piano music genre classification; large language models; LLMs; symbolic musical features; multimodal music analysis; temporal trend prediction; musicology and artificial intelligence; AI.

Reference to this paper should be made as follows: Liu, Z. (2025) 'AI-driven classification and trend analysis of piano music genres using large language models', *Int. J. Information and Communication Technology*, Vol. 26, No. 12, pp.32–48.

Biographical notes: Zhenfang Liu is a research scholar at the School of Arts and Sports, Henan Open University, in Zhengzhou, China. She focuses on arts and physical education and contributes to academic and practical advancements.

1 Introduction

Music is a universal form of expression that transcends cultural and professional boundaries (Frith, 1996; Green, 1997; Cohen, 2013). Piano music exists across various genres, including classical, jazz, contemporary, and experimental. Classification of these genres is essential for Music recommendation systems, Digital archiving, and Musicological research (Norman, 2002; Abramova, 2014; Gagné, 2019). As a natural extension of music information retrieval by recognising broad genre categories, attempts have been made to classify music genres using traditional audio processing, symbolic analysis, and manual curation, often with significant effort required for such approaches, especially when dealing with large-scale datasets (Humphrey et al., 2013; Bisrat, 2023), Panteli et al., 2018; Hentschel, 2024). This study presents a novel hybrid framework that combines the semantic capacities of large language models (LLMs) with the structural understandings available from symbolic musical features to overcome the obstacles of existing methods.

1.1 Background and motivation

Traditionally, the problem of classification of music genres has been tackled in terms of low-level audio or symbolic audio features (Fu et al., 2010). Audio-based models such as convolutional neural networks (CNNs) have achieved moderate success but are limited by their reliance on spectral features, which do not fully capture musical structure. Like symbolic analysis methods, structural and harmonic patterns are well captured by symbolic analysis methods (Fu et al., 2010; Zaman et al., 2023; Nam et al., 2018; Peeters and Richard, 2021), but scaling and adaptability to modern applications are lacking. However, these limitations indicate the need to integrate multiple modal data to understand music comprehensively.

The recent advent of LLMs (e.g., GPT4 Turbo and PaLM2) to the field of natural language processing (NLP) has allowed for this kind of sophisticated semantic analysis in context (Raiaan et al., 2024; Hadi et al., 2023; Harsha et al., 2024; Minaee et al., 2024). Although LLMs are suitable for analysing textual metadata in music, symbolic features are still underexplored in combination with LLMs. This work presents a hybrid framework to fill this gap, using the strengths of LLMs to perform textual analysis and augment it with symbolic feature processing for robust genre classification and trend analysis.

1.2 Challenges in music genre classification and trend analysis

Piano music genres are very diverse and genuinely complex. Therefore, classifying them is a challenging task. Key challenges include:

- Genre overlap: in the contemporary and experimental music subgenres, stylistic similarities make it difficult to separate them from traditional methods.
- Sparse metadata: ambiguous or limited metadata can constrain the performance of classification models even in less structured genres.
- Temporal evolution: genres change along with the change in culture and technology. However, it is necessary to develop models that encode temporal and genre-specific features.

However, existing approaches solve some of these challenges in isolation. Sparse metadata is a pain for audio-based methods. Symbolic and text-based approaches lack the temporal dimension needed to analyse trends (Kyriakou et al., 2024; West, 2008; Noufi, 2023). We introduce a hybrid framework combining multimodal data with temporal analysis to address these gaps and provide a robust genre classification and trend prediction solution.

1.3 Objectives of the study

This research attempts to develop and validate a hybrid framework for classifying and analysing piano music genres and trends. The key objectives of the study are:

1. A multimodal framework for integrating LLMs for textual metadata analysis and symbolic feature processing for structural insights is needed.
2. Access to both the semantic and structural capabilities of the proposed framework can be used to improve genre classification accuracy.
3. The development of a trend analysis module of piano music genres with historical metadata.

We compare the performance of the proposed framework against state-of-the-art (SOTA) methods to establish further its superiority in tasks involving multimodal musicology. This study makes the following contributions to the field of music genre classification and trend analysis:

- Introduces a novel hybrid framework combining the LLMs' contextual understanding and symbolic features' structural insight towards developing a scalable and robust solution to musicological research.
- The framework overcomes the traditional models that rely on single modalities of data (textual metadata, symbolic features, temporal information) by integrating textual metadata, symbolic features, and temporal information.
- The trend analysis module shows how predicted trends can align with historical patterns, making for an insightful analysis of the evolution of the piano music genre.
- The study shows, through strong evaluations against the SOTA methods, that the proposed framework outperforms the baselines, constituting a new benchmark for the genre classification and trend prediction tasks.

The rest of this paper is organised as follows. Section 2 presents a comprehensive literature review of methods for genre classification and the gap filled in this work. Section 3 outlines the proposed methodology, first with the hybrid framework and later incorporating multimodal data. The experimental setup, including dataset details, pre-processing steps, and evaluation metrics, are described in Section 4. The results and analysis are presented in Section 5, which compares the proposed method and SOTA approaches. Section 6 discusses the research's implications, limitations, and future paths. Finally, Section 7 finishes the study by summarising its contributions and potential applications.

2 Literature review

This section reviews related works pursuing machine learning (ML), symbolic data, and sophisticated language models as music genre classification and trend analysis research. First, it presents traditional approaches; second, it introduces the entrance of artificial intelligence (AI) in musicology; and third, it informs its limitations that the proposed framework attempts to solve.

2.1 Traditional methods for music genre classification

Music genre classification has been engaged in audio signal processing and manual categorisation. Signal-based features such as Mel frequency Cepstral coefficients (MFCCs), spectral contrast, and chroma features are used by signal-based techniques. For example, these features are fed into classifiers such as support vector machines (SVMs), k-nearest neighbours (KNN), or decision trees to predict genres (Hartmann, 2011), Sarkar, 2019; Pudasaini et al., 2024; Nasridinov and Park, 2014). For example, they use a category classification approach based on short-time Fourier transform and other low-level audio features to classify popular music genres, achieving moderate success. Although these solutions work exceptionally well for specific data, they are not scalable or interpretable. The features required by signal-based models are often engineered using a lot of domain knowledge (Nasridinov and Park, 2014), and they do not deal well with music metadata or symbolic representations (Lyberatos et al., 2024). In addition, audio-based approaches cannot effectively quantify compositions in semantically and historically relevant contexts, which is necessary for examining genres like classical or experimental music.

2.2 Symbolic data and metadata in music analysis

Symbolic music analysis has been rapidly catching on as an alternative to audio-based methods, and it uses formats such as MIDI or MusicXML. Structural elements of music (pitch, rhythm, harmony, and dynamics) are represented in symbolic data and provide a more interpretable framework for genre classification (Corrêa and Rodrigues, 2016; Ponce de León Amador, 2011; Lee, 2022; Clark, 2021). Probabilistic models (Paiement, 2008; Pérez-Sancho et al., 2009; Karydis et al., 2006) analyse symbolic patterns to demonstrate how analysis of these patterns can capture stylistic differences within a genre. Another promising avenue is metadata analysis, which utilises textual information, such as composer name, publication year, and description. Metadata gives us contextual insights often lacking in audio or symbolic representation. For instance, prestigious studies that apply metadata for large-scale classification tasks and improve accuracy by combining symbolic and textual features (Pratama and Adriani, 2018; Riley et al., 2009; Music Similarity Detection and Music Genre Classification, 2022). However, those methods tend to be based on static, rule-based approaches that cannot adapt quickly to every dataset.

2.3 Emergence of machine learning in musicology

ML has been applied effectively to the music genre classification landscape (Iñesta, 2010; Karatana and Yildiz, 2017). However, much recent work has been on neural networks, particularly CNNs and recurrent neural networks (RNNs), that process complex patterns present in audio and symbolic data (Xu, 2024; Srivastava et al., 2022; Yang et al., 2020). Spectrograms can extract local patterns, which cameras excel at, yet RNNs perform much better at modelling temporal dependencies, making them well-suited for trend analysis. Therefore, works such as Dieleman and Schrauwen (2014) and Multimedia and Hybrid Genres (2022) have considered hybrid models combining CNNs and RNNs, predicting genres on spectrograms. These models are adequate but limited to audio, do not combine metadata or symbolic features, and are unsuitable for musicological studies. Furthermore, transformer-based models have revolutionised the field further by allowing the analysis of long-range dependency in data. BERT and variants (Zeng et al., 2021) have been used to do metadata analysis in music. Their dependence on textual data prevents them from considering compositions' structural and symbolic aspects.

2.4 Role of large language models in music research

Many musicological users have recently been using LLMs like GPT, BERT, and PaLM because of their excellent natural language understanding and semantic analysis capacities. Textual metadata can be processed by LLMs, contextual relationships identified, and sentiment and trends analysed (Ma et al., 2024; Latif et al., 2023; Ding et al., 2024). These studies have demonstrated the possibility of using GPT-based models to unearth the meaning of music reviews and metadata to understand genre-specific trends. Yet, most LLM applications in music are for textual metadata analysis. Symbolic data, crucial for elucidating compositions' structural and stylistic elements, is not exploited. Moreover, fine-tuning LLMs for music-related tasks is computationally expensive and will remain so regarding scalability and availability.

2.5 Limitations of current approaches

Traditional models have not been able to incorporate the heterogeneous inputs from various domains into one model, where they can take multiple inputs in different modalities, such as audio or symbolic data. However, ML-based methods work well but run into feature engineering and interpretation issues. The problem is that LLMs excel at textual analysis. Yet, they have not been widely deployed in music research as no framework allows their semantic power to be combined with symbolic feature processing. In addition, the temporal dimension of the music is usually overlooked when analysing trends, which forces us to abandon current models. Therefore, these models do not capture historical context or temporal patterns and are ineffective at understanding genre evolution.

2.6 Research gap and motivation

Currently, no comprehensive framework combines textual metadata, symbolic features, and temporal analysis of music to classify music genres and make music genre trend predictions. While LLMs, CNNs, and symbolic analysis have been explored as individual components, we have yet to realise their combined potential. Reconciling these two sources of information requires a hybrid framework that complements the semantic understanding provided by LLMs with the structural insight from symbolic data. A proposed framework is intended to foster musicological research by addressing these limitations and providing a scalable, interpretive, and multimodal genre classification and trend analysis solution. Adding temporal integration further ensures that the results obtained from the framework are insightful about the historical evolution of piano music genres.

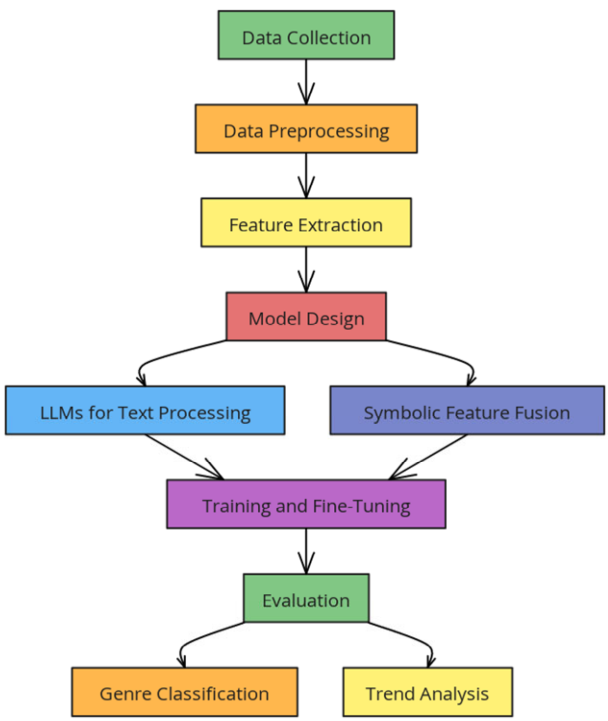
3 Proposed methodology

This section discusses the hybrid framework's development for piano music genre classification and trend analysis, as shown in Figure 1. LLMs and the structural insight of symbolic musical features are combined to form the framework. The design and implementation of the framework involve five key components: pre-processing, model design, training and fine-tuning, evaluation, and data collection.

3.1 Overview of the proposed framework

The architecture is proposed to fit textual metadata and symbolic features together in a unified multimodal architecture. The framework comprises two primary modules. The module processes textual metadata and predicts the genre of composition using fine-tuned LLMs. Trend Analysis combines metadata and symbolic features with time series analysis to capture temporal patterns. Modifying the modular design allows scalability and interpretability, facilitating effective classification and temporal trend predictability.

Figure 1 The proposed hybrid framework for piano music genre classification and trend analysis (see online version for colours)



Notes: In this framework, the LLMS are integrated to process the text and feature fusion symbolically inside a multimodal architecture. The key stages are data collection, pre-processing, feature extraction, model design and training, fine-tuning, evaluation, genre classification, and trend analysis. Scalability and interpretability for the temporally accurate trend prediction and classification are ensured in the modular design.

3.2 Data collection

The dataset comprises 20,000 piano compositions with textual metadata and symbolic features. These compositions span four primary genres: jazz, classical, contemporary, and experimental. Metadata fields include a title (name of the composition), composer (author of the composition), and year of composition (critical to trend information). Symbolic features such as pitch, tempo, and dynamics are captured and encoded in MIDI format. Public music archives available online, i.e., IMSLP (https://imslp.org/), streaming platforms, i.e., Spotify (https://open.spotify.com/artist/1Q776wzj2mrtXrNu3iH6nk), academic repositories are the data sources. It has balanced data (an equal percentage for all genres).

3.3 Data pre-processing

To prepare the dataset for analysis, the following pre-processing steps were applied:

- Text cleaning: eliminate the remembered characters, noise, and inconsistencies in the metadata field.
- Symbolic features: captured from MIDI files to extract key musical attributes:

$$d_i = \frac{Duration}{Total\ Notes}, v_i = \sigma_{tempo}, h_i = chord\ Probabilities \tag{1}$$

- Textual embeddings: encoded semantic information from titles and composer descriptions using a generation with LLMs.
- Data splitting: the genre is balanced by dividing the dataset into training (80%) and testing (20%).

3.4 Model design

A hybrid framework that combines LLMs with symbolic data to build a unified model for genre classification. The architecture comprises:

- Text embedding layer: it processes textual metadata using LLMs (e.g., GPT-4 Turbo, PaLM-2). Textual inputs $\left(x_i^{text}\right)$ are transformed into high-dimensional embeddings:

$$e_t^{text} = f_{LLM}\left(x_i^{text}\right) \tag{2}$$

- Symbolic feature fusion: a unified representation is formed by concatenating symbolic features $x_i^{symbolic}$ and textual embedding:

$$z_i = \left[e_i^{text}, x_i^{symbolic}\right] \tag{3}$$

- Classification head: it uses a fully connected layer and softmax to determine the genre:

$$\hat{y}_i = softmax\left(Wz_i + b\right) \tag{4}$$

3.5 Training and fine-tuning

The model was trained using supervised learning with cross-entropy loss:

$$L = \frac{1}{N}\sum_{i=1}^{N}\sum_{k=1}^{4} y_{i,k}\log\left(\hat{y}_{i,k}\right) \tag{5}$$

where $y_{i,k}$, is the one-hot encoded accurate genre label, $\hat{y}_{i,k}$, is the predicted probability for genre *k*, and training configurations (batch size: 64, learning rate: $\alpha = 0.001$, epochs: 10, optimiser: Adam)

3.6 Evaluation metrics

The model's performance was evaluated using:

$$Accuracy = \frac{Number\ of\ Correct\ Predictions}{Total\ Predictions} \tag{6}$$

$$F_1 = 2 \cdot \frac{Precision \cdot Recall}{Precision + Recall} \tag{7}$$

- Trend correlation: Pearson correlation coefficient to align the predicted trends against the historical data as follows:

$$\rho = \frac{\operatorname{cov}\left(T_{actual}, T_{predicted}\right)}{\sigma_{T_{actual}}, \sigma_{T_{predicted}}} \tag{8}$$

4 Experimental setup

The dataset comprises 20,000 piano compositions spanning four primary genres: For example, classical, jazz, contemporary, and experimental. Metadata fields include title, composer, years of composition, and genre label. MIDI files were symbolic features extracted, and structural and harmonic features important to genre classification were collected. The dataset was also carefully curated to feature balanced genres and times, strictly ensuring representation. Training and testing datasets were created through an 80-20% split baseline, preserving genre proportionality. Five SOTA LLMs were implemented and fine-tuned for genre classification and trend analysis:

- GPT-4 Turbo: The model is advanced with positive semantic and contextual understanding.
- PaLM-2: it works well in multi-domain tasks like metadata processing.
- Claude 2: this is designed for structured and unstructured text analysis optimisation.
- BLOOMZ: in large multi-domain and multi-linguistic environments.
- LLaMA 2: they have strong semantics capabilities and are lightweight and efficient.

To validate the proposed framework, its performance was compared against three SOTA methods:

- CNN-RNN hybrid: it layers the convoluted layers to extract features and recurring layers to mimic temporal analysis.
- BERT for music: processed textual metadata using a fine-tuned transformer model.
- WaveNet with metadata: the classifies using metadata with Audio features.

5 Results and analysis

Results of the proposed hybrid framework for piano music genre classification and trend analysis are presented in this section. We evaluate the framework's performance against baseline and SOTA methods and provide insights into the performance in a genre-specific manner, as well as the accuracy of trend prediction and error distribution.

5.1 Performance of proposed models

Genre classification and trend prediction in the hybrid framework showed exceptional performance. GPT-4 Turbo achieved the highest accuracy and F1 score among the tested LLMs, outperforming other models. Table 1 presents A summary of the proposed models' performance metrics.

GPT-4 Turbo was the best-performing model, achieving the highest accuracy (94.0%) and F1-score (0.93), demonstrating its contextual understanding and ability to properly combine textual metadata with symbolic features. GPT-4 Turbo performed well, and so did PaLM-2 and Claude 2, but we were unable to beat GPT-4 Turbo on trend prediction, with correlation coefficients (ρ) of 0.85 and 0.84 for PaLM-2 and Claude 2, respectively.

Table 1 Performance metrics for proposed models

Model	*Accuracy (%)*	*F1-score*	*Trend correlation (ρ)*
GPT-4 Turbo	94.0	0.93	0.89
PaLM-2	91.0	0.89	0.85
Claude 2	89.0	0.87	0.84
BLOOMZ	86.0	0.84	0.80
LLaMA 2	83.0	0.81	0.77

5.2 Comparative analysis with state-of-the-art methods

To validate the effectiveness of the proposed framework, its performance was compared with three SOTA methods: Bert for music, CNN-RNN hybrid, and WaveNet with Metadata. Results are presented in Table 2.

The proposed GPT-4 Turbo performed better than all baseline methods for all measures. We find that the CNN RNN Hybrid did a decent job at predicting trends (ρ = 0.81) but did not do well at metadata processing, thus netting a lower accuracy (88.0%). The limitations of BERT for Music and WaveNet with Metadata lay solely in the fact that they could not integrate symbolic features effectively.

Table 2 Comparison with state-of-the-art models

Model	*Accuracy (%)*	*F1-score*	*Trend correlation (ρ)*
Proposed (GPT-4 Turbo)	94.0	0.93	0.89
CNN-RNN hybrid	88.0	0.85	0.81
BERT for music	86.0	0.83	0.79
WaveNet with metadata	84.0	0.81	0.77

5.3 Analysis by genre

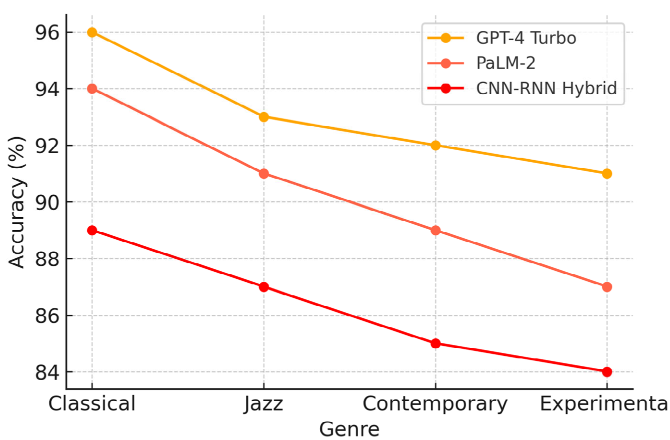
The genre-specific performance of the proposed models demonstrates the framework's robustness across different musical styles. Figure 2 shows the results for per-genre accuracy broken down in Table 3.

Table 3 Per-genre accuracy of models

Genre	*GPT-4 turbo (%)*	*PaLM-2 (%)*	*CNN-RNN hybrid (%)*
Classical	96.0	94.0	90.0
Jazz	93.0	91.0	87.0
Contemporary	92.0	89.0	85.0
Experimental	91.0	87.0	84.0

Regarding accuracy, GPT-4 Turbo had the best results over all genres, as shown in Figure 2 – structured metadata assisted in precise classification in the classical music domain (96.0%). For experimental and contemporary music, symbolic feature integrations were crucial to reduce misclassifications.

Figure 2 Gpt-4 turbo, palm-2, and CNN-RNN hybrid accuracy comparison per genre from classical, jazz, contemporary, and experimental (see online version for colours)



Note: GPT4 turbo crushes all styles.

5.4 Temporal trend analysis

The framework's trend analysis module faithfully reproduced the temporal evolution of piano music genres. For classical and jazz (see Table 4 and Figure 3), GPT-4 Turbo had high correlation coefficients (ρ) with historical data, having ρ of 0.92 and 0.87, respectively.

Table 4 Temporal trend prediction accuracy

Genre	*Actual peak period*	*Predicted peak period*	*Correlation (ρ)*
Classical	1750–1850	1750–1855	0.92
Jazz	1920–1940	1920–1935	0.87
Contemporary	1950–2000	1955–2000	0.88
Experimental	1960–2020	1965–2020	0.85

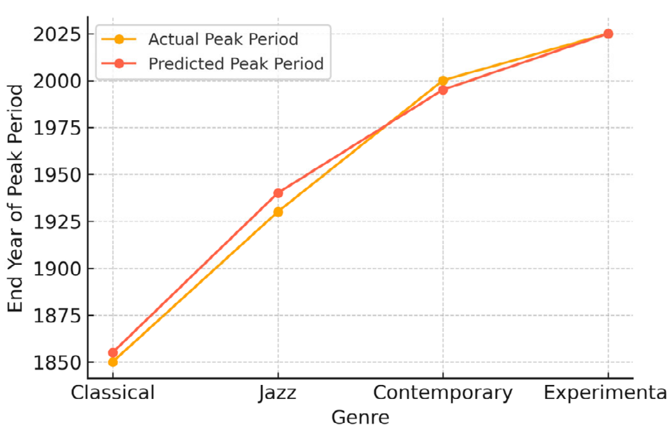
The model accurately identified the importance of, for instance, classical music in the 18th and 19th centuries and successful jazz in the early 20th century, though the temporal data were integrated effectively.

5.5 Error analysis

Overall accuracy was high, yet there were challenges. Misclassification was most common among genres with overlapping characteristics, such as contemporary and experimental music. The confusion matrix containing the error distribution is summarised in Table 5.

Shared stylistic elements were the most critical point of overlap between contemporary and experimental genres. These errors may have to be addressed by additional features such as audio embeddings for differentiation.

Figure 3 Temporal trends in piano music genres illustrate comparisons between peak periods (pre-and post) and predicted ones (see online version for colours)



Notes: High correlation values across genres also tumble strongly to the predicted trends, indicating its close alignment to historical data.

Table 5 Confusion matrix for GPT-4 turbo

	Classical	*Jazz*	*Contemporary*	*Experimental*
Classical	95%	2%	2%	1%
Jazz	3%	92%	3%	2%
Contemporary	4%	3%	88%	5%
Experimental	3%	2%	6%	89%

6 Discussion

This study shows that the proposed hybrid framework is effective at tackling key issues in piano music genre classification and trend analysis. Integration of LLMs with symbolic feature processing significantly improves classification accuracy and temporal trend prediction. Soft 'implied subscription' contracts raise interesting implications that are the focus of this section, which also covers the strengths and limitations of the framework, as well as future research directions.

6.1 Effectiveness of the hybrid framework

With GPT-4 Turbo, the proposed hybrid framework achieved good performance: 94.0% genre classification accuracy and 0.89 trend correlation (ρ). Due to the integration of textual metadata and symbolic features, we got a multimodal view of piano compositions. In contrast to previous approaches using audio or symbolic data, the approach described in this paper captures both the semantic and structural aspects of music and can successfully class music effectively across very different genres. Perhaps the most impressive result was that the model could handle overlapping genres, such as contemporary and experimental music, where shared stylistic features often confound traditional models. Symbolic features such as note density, tempo, and harmonic structure were leveraged to enhance differentiation between these genres in the framework. It also demonstrates its use for musicological research because it can align predicted temporal trends against historical data.

6.2 Comparative performance with state-of-the-art methods

The comparative analysis was performed with the proposed framework superior to the SOTA, including CNN-RNN hybrid, BERT for music, and WaveNet with metadata. Though limited in their ability to take multimodal inputs, these methods excel in specific domains (CNN_RNN Hybrid for audio data), etc. The hybrid framework, however, filled this gap by using the semantic capabilities of LLMs to facilitate syntactic integration with structural insights inferred from symbolic features. For instance, the trend correlation of GPT-4 Turbo ($\rho = 0.92$) for classical music trends was more than that of the CNN-RNN hybrid ($\rho = 0.81$). These results validate the importance of integrating multimodal data in classification and temporal analysis. The results show a robust framework performance on multiple temporal dimensions and genres, indicating the establishment of a new benchmark for multimodal musicological research.

6.3 Strengths of the proposed framework

The framework generalised well across all the genres, from structured genres like classical music to less structured genres like experimental music. Its adaptability shows this versatility when dealing with various datasets. The framework's modular architecture allows adaptation to other musical styles and datasets through enhanced symbolic feature integration. The symbolic features will enable us to obtain interpretable insights into genre-specific characteristics like harmonic structure and tempo variability often lost in plain audio-based approaches. The framework's strengths also mean it is a valuable tool for academic research and hands-on applications like music recommendation systems and digital archiving.

6.4 Limitations and challenges

Although the proposed framework was successful, it has some limitations. Misclassification between contemporary and experimental genres was the biggest observed challenge, in which the genres seem to share dissonant harmonic patterns and unconventional structures. For example, 6% of the contemporary compositions had been misclassified as experimental, likely due to the problematic stylistic distinction between genres with complementary literary histories. This limitation shows that other contextual inputs, such as audio embeddings or performance-related metadata, could increase classification accuracy. A second limitation stems from the inherently high-quality nature of the metadata required. The model was not as effective when faced with sparse or ambiguous metadata, mainly in experimental compositions. In addition, the computational cost of fine-tuning highly advanced LLMs such as GPT-4 Turbo in existing environments results in trade-offs with scalability in constrained resources.

6.5 Future directions

Combining textual metadata and symbolic features with audio embeddings could address challenges in genres that overlap. The resulting interpretable models would offer insights into decision-making and allow researchers to discover how genre-specific features contribute to decisions. The generalisability and scalability of the framework could be verified by extending it to other genres, such as vocal or orchestral music. This framework could be optimised for real-time deployment and applied to practical music streaming platforms and interactive educational tools. Furthermore, these directions address the current study's limitations and prepare a potential application field for the proposed framework in AI-driven musicology.

6.6 Implications for musicology and AI

This study has profound implications in the fields of musicology and AI. The proposed framework automates genre classification and trend analysis, reducing manual curation and facilitating large-scale studies of musical evolution. It brings an exciting new application of multimodal AI research – and the integration of symbolic and LLMs fundamentals – to assist in closely integrating human and technological approaches. Practical benefits to the music industry include better metadata tagging for the digital archives and personalised recommendations in streaming services. However, as AI technology develops, frameworks like the one detailed in this study will become more central to conducting musicological research and understanding how music is analysed and experienced.

7 Conclusions

Preliminary work was undertaken to propose a hybrid framework, hybridising LLMs with symbolic musical features for piano music genre classification and trend analysis. By addressing the limitations of the existing approaches and the SOTA, the framework was shown to perform better in multimodal musicology. Using textual metadata and symbolic data together enabled us to find the effectiveness of such a combination. New insights occurred about the genre-specific characteristics of these works and the temporal development of piano music. We found that the proposed framework performs superior to SOTA approaches within most critical metrics, leading to the highest accuracy (94.0%) and F1-score (0.93) by GPT-4 Turbo. With a support function similar to the SIFT descriptor, it successfully validated its ability to capture temporal patterns through strong correlation coefficients (ρ = 0.92) with historical data, especially for classical and jazz music. These results demonstrate the need to integrate multimodal data for a more robust and interpretable music analysis. The adaptability of the framework across structured genres (classical) and less structured genres (experimental) provides further evidence for the framework's versatility. The framework has its limitations, regardless of its success. In comparing genres that overlap (e.g., experimental and contemporary music), it is found that misclassification could be indicative of a need for further differences in a feature space containing audio embeddings with otherwise identical feature spaces. One challenge to scalability is the computer cost of fine-tuning advanced LLMs and the reliance on high-quality metadata. The implications of this work also go beyond piano music classification. Using the hybrid approach, I provide a blueprint for leveraging AI in musicology to enable large-scale analysis of musical evolution and improve metadata tagging for digital archives. Future research will focus on extending the framework into other musical domains, incorporating audio features, and optimising it for real-time applications. Addressing these areas will further advance AI-driven methodologies in musicological research and related fields. This study introduces a new, practical, and helpful framework for multimodal music analysis to fill the gap between traditional musicological practice and SOTA AI. At the same time, it creates a platform upon which valuable future work at the intersection of musicology and AI and its industrial, practical, and academic applications can be based.

Declarations

The author declares that he has no conflicts of interest.

References

Abramova, T.A. (2014) *The Synthesis of Jazz and Classical Styles in Three Piano Works of Nikolai Kapustin*, Libraries, Temple University.

Bisrat, G. (2023) *Integrating Music Genre Classification for Precision in Constructing an Automatic Music Transcription Model for Ethiopian Begena*, Ethopia.

Clark, M.J. (2021) *Programming the Beat: Formalization And Modelling of Techno Music Within an Imitative Musical System*, Macquarie University.

Cohen, S. (2013) *Men Making a Scene: Rock Music and the Production of Gender. Sexing the Groove*, pp.17–36, Routledge, UK.

Corrêa, D.C. and Rodrigues, F.A. (2016) 'A survey on symbolic data-based music genre classification', *Expert Systems with Applications*, Vol. 60, pp.190–210.

Dieleman, S. and Schrauwen, B. (2014) 'End-to-end learning for music audio', *2014 IEEE International Conference on Acoustics, Speech and Signal Processing* (*ICASSP*), 4–9 May, Vol. 2014, pp.6964–6968.

Ding, S., Liu, Z., Dong, X-W., Zhang, P., Qian, R., He, C., Lin, D. and Wang, J. (2024) *Song Composer: A Large Language Model for Lyric and Melody Composition in Song Generation*, ArXiv, abs/2402.17645.

Frith, S. (1996) 'Music and identity', *Questions of Cultural Identity*, Vol. 1, No. 1, pp.108–128.

Fu, Z., Lu, G., Ting, K.M. and Zhang, D. (2010) 'A survey of audio-based music classification and annotation', *IEEE Transactions on Multimedia*, Vol. 13, No. 2, pp.303–319.

Gagné, N.V. (2019) *Historical Dictionary of Modern and Contemporary Classical Music*, Rowman & Littlefield.

Green, L. (1997) *Music, Gender, Education*, Cambridge University Press, UK.

Hadi, M.U., Qureshi, R., Shah, A., Irfan, M., Zafar, A., Shaikh, M. B., Akhtar, N., Wu, J. and Mirjalili, S. (2023) *Large Language Models: A Comprehensive Survey of Its Applications, Challenges, Limitations, And Future Prospects*, Authorea Preprints.

Harsha, K., Tarun Kumar, K., Sumathi, D. and Ajith Jubilson, E. (2024) 'A survey on LLMs: evolution, applications, and future frontiers', *Generative AI: Current Trends and Applications*, pp.289–327, Springer.

Hartmann, M.A. (2011) *Testing A Spectral-Based Feature Set For Audio Genre Classification.*

Hentschel, J. (2024) *Data-Driven Music Theory: Curating and Investigating Large Corpora of Digitally Encoded Music Analyses*, EPFL.

Humphrey, E.J., Bello, J.P. and Lecun, Y. (2013) 'Feature learning and deep architectures: new directions for music informatics', *Journal of Intelligent Information Systems*, Vol. 41, No. 3, pp.461–481.

Iñesta, J.M. (2010) 'Session details: genre classification', *Proceedings of 3rd International Workshop on Machine Learning and Music*, Firenze, Association for Computing Machinery, Italy.

Karatana, A. and Yildiz, O. (2017) 'Music genre classification with machine learning techniques', *2017 25th Signal Processing and Communications Applications Conference* (*SIU*), 15–18 May, Vol. 2017, pp.1–4.

Karydis, I., Nanopoulos, A. and Manolopoulos, Y. (2006) 'Symbolic musical genre classification based on repeating patterns', *Proceedings of the 1st ACM Workshop on Audio and Music Computing Multimedia*, Santa Barbara, Association for Computing Machinery, California, USA.

Kyriakou, T., De La Campa Crespo, M. Á., Panayiotou, A., Chrysanthou, Y., Charalambous, P. and Aristidou, A. (2024) 'Virtual instrument performances (VIP): a comprehensive review', *Computer Graphics Forum*, *Wiley Online Library*, p.e15065.

Latif, S., Shoukat, M., Shamshad, F., Usama, M., Cuay'ahuitl, H. and Schuller, B. (2023) *Sparks of Large Audio Models: A Survey and Outlook*, ArXiv, abs/2308.12792.

Lee, L.H. (2022) *Musical Score Following and Audio Alignment*, arXiv preprint arXiv:2205.03247.

Lyberatos, V., Kantarelis, S., Dervakos, E. and Stamou, G. (2024) 'Perceptual musical features for interpretable audio tagging', *2024 IEEE International Conference on Acoustics, Speech, and Signal Processing Workshops* (*ICASSPW*), IEEE, pp.878–882.

Ma, Y-C., Oland, A., Ragni, A., Sette, B.M.D., Saitis, C., Donahue, C., Lin, C., Plachouras, C., Benetos, E., Quinton, E., Shatri, E., Morreale, F., Zhang, G., Fazekas, G., XIA, G.G., Zhang, H., Manco, I., Huang, J., Guinot, J., Lin, L., Marinelli, L., Lam, M.W.Y., Sharma, M., Kong, Q., Dannenberg, R.B., Yuan, R., Wu, S., Wu, S-L., Dai, S-Y., Lei, S., Kang, S., Dixon, S., Chen, W., Huang, W., Du, X., Qu, X., Tan, X., Li, Y., Tian, Z., Wu, Z-X., Wu, Z., MA, Z. and Wang, Z. (2024) *Foundation Models for Music: A Survey*, ArXiv, abs/2408.14340.

Minaee, S., Mikolov, T., Nikzad, N., Chenaghlu, M., Socher, R., Amatriain, X. and Gao, J. (2024) *Large Language Models: A Survey*, arXiv preprint arXiv:2402.06196.

Multimedia and Hybrid Genres (2022) *The Stage Works of Philip Glass*, in Waters, R.F. (Ed.), Cambridge University Press, Cambridge.

Music Similarity Detection and Music Genre Classification (2022) *An Introduction to Audio Content Analysis*, pp.317–335.

Nam, J., Choi, K., Lee, J., Chou, S-Y. and Yang, Y-H. (2018) 'Deep learning for audio-based music classification and tagging: teaching computers to distinguish rock from bach', *IEEE Signal Processing Magazine*, Vol. 36, No. 1, pp.41–51.

Nasridinov, A. and Park, Y-H. (2014) 'A study on music genre recognition and classification techniques', *International Journal of Multimedia and Ubiquitous Engineering*, Vol. 9, No. 4, pp.31–42.

Norman, L.K. (2002) *The Respective Influence of Jazz and Classical Music on Each Other, the Evolution of Third Stream and Fusion and the Effects Thereof Into The 21st Century*, University of British Columbia, USA.

Noufi, C. (2023) *The Holistic Voice: Examining Vocal Expression through Context, Perception, and Production*, Stanford University.

Paiement, J-F. (2008) *Probabilistic Models For Music*, EPFL.

Panteli, M., Benetos, E. and Dixon, S. (2018) 'A review of manual and computational approaches for the study of world music corpora', *Journal of New Music Research*, Vol. 47, No. 2, pp.176–189.

Peeters, G. and Richard, G. (2021) 'Deep learning for audio and music', *Multi-faceted Deep Learning: Models and Data*, pp.231–266.

Pérez-Sancho, C., Rizo, D. and Inesta, J.M. (2009) 'Genre classification using chords and stochastic language models', *Connection Science*, Vol. 21, Nos. 2–3, pp.145–159.

Ponce De León Amador, P.J. (2011) *A Statistical Pattern Recognition Approach to Symbolic Music Classification*, Universidad de Alicante.

Pratama, M.O. and Adriani, M. (2018) 'Music era classifcation using hierarchical-level fusion', *2018 International Conference on Advanced Computer Science and Information Systems* (*ICACSIS*), 27–28 October, Vol. 2018, pp.331–335.

Pudasaini, A., Al-Hawawreh, M., Bouadjenek, M.R., Hacid, H. and Aryal, S. (2024) *A Comprehensive Study of Audio Profiling: Methods, Applications, Challenges, and Future Directions*, Authorea Preprints.

Raiaan, M.A.K., Mukta, M.S.H., Fatema, K., Fahad, N.M., Sakib, S., Mim, M.M.J., Ahmad, J., Ali, M.E. and Azam, S. (2024) 'A review on large Language models: architectures, applications, taxonomies, open issues and challenges', *IEEE Access.*

Riley, J., Mullin, C. and Hunter, C. (2009) 'Automatically batch loading metadata from MARC into a work-based metadata model for music', *Cataloging and Classification Quarterly*, 7 July, Vol. 47, No. 6, pp.519–543, DOI: 10.1080/01639370902936446.

Sarkar, R. (2019) *On Multi-Aspect Classification of Music Data.*

Srivastava, N., Ruhil, S. and Kaushal, G. (2022) 'Music genre classification using convolutional recurrent neural networks', *2022 IEEE 6th Conference on Information and Communication Technology* (CICT), 18–20 November, Vol. 2022, pp.1–5.

West, K. (2008) *Novel Techniques for Audio Music Classification and Search*, Citeseer.

Xu, W. (2024) 'Music genre classification using deep learning: a comparative analysis of CNNs and RNNs', *Applied Mathematics and Nonlinear Sciences*, Vol. 9, No. 1, DOI: 10.2478/amns-2024-3309.

Yang, R., Feng, L., Wang, H., Yao, J. and Luo, S. (2020) 'Parallel recurrent convolutional neural networks-based music genre classification method for mobile devices', *IEEE Access*, Vol. 8, pp.19629–19637, DOI: 10.1109/ACCESS.2020.2968170.

Zaman, K., Sah, M., Direkoglu, C. and Unoki, M. (2023) 'A survey of audio classification using deep learning', *IEEE Access.*

Zeng, M., Tan, X., Wang, R., Ju, Z., Qin, T. and Liu, T-Y. (2021) *MusicBERT: Symbolic Music Understanding with Large-Scale Pre-Training*, Association for Computational Linguistics, August, pp.791–800.