ISSN: 3079-6342

Graph Neural Network for Music Style Classification

Katarzyna Nowak[1], Tomasz Zieliński[1]

[1]Warsaw School of Computer Science Lewartowskiego 17 00-169 Warsaw, Poland

Abstract

Music style classification plays a fundamental role in music recommendation, retrieval, and organization systems. Traditional classification models primarily rely on audio features or symbolic representations, such as mel-frequency cepstral coefficients or Musical Instrument Digital Interface (MIDI) sequences. However, these models often ignore the rich structural and relational information inherent in musical compositions. This study proposes a novel graph neural network (GNN)-based framework for music style classification that represents each piece of music as a graph, capturing the relationships among notes, chords, and temporal transitions. By modeling these components as interconnected nodes, the GNN is able to learn stylistic features that extend beyond local patterns, such as harmonic progressions, motif repetitions, and inter-note dependencies.

To enhance model performance, the framework incorporates a dual-graph architecture combining intra-piece and inter-piece structures, enabling the GNN to generalize across compositions while retaining individual stylistic identities. Experimental results on publicly available symbolic music datasets demonstrate that the proposed model outperforms traditional convolutional neural network (CNN) and recurrent neural network (RNN)-based models in classification accuracy and robustness across multiple musical genres. These findings highlight the potential of graph-based deep learning for extracting structural patterns critical to music understanding and classification.

Keywords

Graph Neural Networks, Music Style Classification, Symbolic Music, Deep Learning, Structural Representation, Music Information Retrieval.

1. Introduction

Music style classification is a key task in the field of music information retrieval, with wide applications in automated music recommendation, content-based search, digital music archiving, and genre-based analysis. Traditionally, this task has been approached using statistical learning or deep learning methods applied to low-level audio features or symbolic data, such as pitch sequences and rhythm patterns. While these approaches have achieved moderate success, they often treat music as a flat sequence or a grid-based input, overlooking the complex and structured relationships between musical elements such as notes, chords, motifs, and harmonic functions [1].

Recent advancements in representation learning have introduced the possibility of modeling music as a graph, where the structural and temporal dependencies between musical components can be explicitly captured [2]. Graph-based representations allow for a more natural encoding of musical structure by treating elements such as notes or bars as nodes and their harmonic or temporal relationships as edges [3]. For example, repeated motifs, sequential intervals, and harmonic progressions can all be represented in a graph format, enabling a neural model to understand long-range dependencies and non-linear structures that are characteristic of different musical styles [4].

ISSN: 3079-6342

Graph neural networks (GNNs) have demonstrated strong performance in various domains where data is inherently relational, including natural language processing, social network analysis, and bioinformatics [5]. In the context of music, GNNs are particularly well-suited for learning style-specific patterns that span across compositions. Unlike convolutional neural networks that require fixed spatial structures or recurrent neural networks that rely heavily on sequence order, GNNs can adapt to variable-sized graphs and flexibly model the topological properties of music data [6]. This flexibility is critical for music classification tasks, where the expressive features of style are often embedded in both short-term and long-term dependencies within and across phrases [7].

Despite this potential, the application of GNNs to symbolic music data remains underexplored. Existing methods often simplify music structure to a sequence or treat it as a spectrogram image, limiting the ability to capture complex stylistic traits. Moreover, prior works do not leverage cross-composition relational structures that might reveal similarities and stylistic conventions shared across multiple pieces within the same genre or composer. This study aims to address these gaps by proposing a GNN-based classification framework that models both intra-composition and inter-composition relations [8].

The proposed method represents each piece of symbolic music as a graph, where notes, chords, or measures serve as nodes and edges encode temporal, harmonic, or structural relationships. The model utilizes a dual-graph architecture, combining local structures within a piece with global structures that span multiple compositions. By learning embeddings that capture both the internal flow of music and its stylistic proximity to other works, the graph neural network (GNN) is able to classify musical styles with high accuracy and interpretability. Experimental evaluations demonstrate that the proposed model outperforms conventional deep learning baselines on multiple public datasets, offering improved generalization across a variety of genres.

2. Literature Review

Music style classification has been an active research area in music information retrieval, combining elements of machine learning, music theory, and digital signal processing [9]. Early approaches relied heavily on manual feature engineering, where musicologists or engineers extracted statistical attributes from audio signals or symbolic representations. These attributes included rhythmic complexity, pitch class distributions, tempo, tonal centroid features, and note density [10]. Such handcrafted features were then fed into traditional classifiers such as support vector machines, k-nearest neighbors, or decision trees, which performed reasonably well for broad genre classification tasks but struggled with subtle stylistic distinctions and data generalization.

With the advent of deep learning, more sophisticated models have been introduced for automatic music classification [11]. Convolutional neural networks (CNNs) have been particularly popular for analyzing audio spectrograms, capturing local time-frequency patterns that are indicative of specific musical genres or styles. However, CNNs require fixed input shapes and are primarily designed for grid-like data, limiting their ability to process symbolic music that often varies in length and structural complexity. Recurrent neural networks (RNNs), particularly long short-term memory networks, were later used to capture the temporal dynamics of symbolic sequences, such as note or chord progressions. While RNNs improved sequence modeling, they too had limitations in handling hierarchical musical structures and long-range dependencies efficiently [12].

In response to these limitations, researchers have begun to explore more structured representations of music [13]. Symbolic music, including Musical Instrument Digital Interface (MIDI) and MusicXML formats, provides detailed event-level data such as pitch, duration, onset

ISSN: 3079-6342

time, and velocity. This data can be naturally transformed into graph structures, where each note or chord functions as a node, and temporal, harmonic, or structural relationships define the edges [14]. Representing music as a graph enables the modeling of relational patterns such as parallel voice leading, chord substitution, modulation, and motif recurrence, which are difficult to capture through sequence-based models [15].

GNNs have gained popularity in recent years due to their ability to model non-Euclidean data where nodes are connected through arbitrary edge patterns [16]. A GNN typically updates each node's representation by aggregating information from its neighbors through message passing mechanisms [17]. This approach is particularly advantageous in music style classification, where local and global structural relationships play a crucial role [18]. For example, jazz and classical compositions may share similar note sequences but differ in harmonic structure or rhythmic phrasing, which GNNs can learn through graph-based context aggregation [19].

Several prior studies have attempted to apply GNNs to music-related tasks, though many have focused on areas such as chord recognition, music generation, or melody harmonization [20]. One line of research has explored chord progression graphs, where chords are nodes and their sequential relationships define edges [21]. Another approach models rhythm trees or phrase structures to capture metrical hierarchy. However, the use of GNNs specifically for music style classification remains relatively underdeveloped. Existing models often restrict graph construction to simple note adjacency or temporal proximity, missing higher-order structural features like form (e.g., AABA, sonata-allegro), harmonic substitution patterns, or thematic variation.

To capture these complex stylistic features, recent studies have proposed hybrid architectures that combine GNNs with other neural modules, such as attention mechanisms or convolutional encoders [22]. These models show promise in integrating local note-level patterns with broader compositional context, especially when trained on large symbolic music corpora. Some works have also investigated the use of inter-composition graphs, where nodes represent entire compositions and edges are defined based on stylistic similarity or shared motifs. These inter-piece graphs allow for transfer of stylistic information across works, improving model generalization in sparse or low-resource datasets [23].

Despite these advancements, there remain challenges in applying GNNs effectively to symbolic music. The diversity of possible graph representations—based on notes, chords, measures, or phrases—introduces variability in graph structure and scale. Furthermore, determining meaningful edge definitions that reflect stylistic relevance requires domain knowledge and careful design [24]. Finally, balancing the trade-off between model expressivity and computational complexity is an ongoing concern, especially as music graphs grow in size and dimensionality.

This study addresses these challenges by introducing a unified GNN-based architecture for music style classification that incorporates both intra-piece and inter-piece relationships. The proposed model builds a dual-graph structure that captures local musical features within each piece while leveraging global stylistic similarities across the dataset. By learning contextualized embeddings through message passing and attention-based aggregation, the model offers a scalable and interpretable solution for graph-based music analysis [25]. The next section presents the methodology, detailing the graph construction, model architecture, and training strategy.

3. Methodology

3.1. Graph-Based Representation of Symbolic Music

The foundation of the proposed model lies in transforming symbolic music data into structured graph representations that reflect both local and global musical relationships. Each

ISSN: 3079-6342

composition is represented as a directed, weighted graph where nodes correspond to musical events—typically notes or chords—extracted from symbolic formats such as MIDI files. Each node carries rich musical attributes, including pitch, onset time, duration, velocity (dynamics), and instrument type. These attributes are embedded into continuous vectors using learnable embedding layers.

Edges in the graph are defined to reflect various relationships among musical events. Temporal edges connect notes that occur in sequence, capturing rhythmic flow and phrase development. Harmonic edges link simultaneously sounding or harmonically related notes, enabling the model to learn tonal coherence. Structural edges are also incorporated to represent higher-order form-based relationships, such as repetition or motif recurrence across sections. The graph structure is designed to be flexible and expressive, allowing the model to capture diverse musical styles that rely on different compositional techniques.

To maintain temporal ordering and motif continuity across longer passages, each piece is segmented into overlapping time windows, with each window forming a subgraph. These subgraphs are connected through cross-window edges that preserve structural context and allow stylistic features to propagate through the composition. Furthermore, to enable cross-composition knowledge transfer, an inter-piece graph is constructed at a meta-level, where each node represents a full piece, and edges are defined by pairwise similarity scores computed through embedding similarity or stylistic proximity measures. This dual-graph framework enables the model to capture both localized compositional traits and dataset-level stylistic trends.

3.2. Graph Neural Network Architecture

The core of the proposed model is a multi-layer graph neural network designed to learn stylistic representations from the intra- and inter-piece graphs. The architecture follows a message-passing paradigm, where each node updates its embedding by aggregating information from its neighbors. For intra-piece graphs, the model uses a combination of graph convolutional layers and gated graph recurrent units to capture both local harmonic patterns and temporal dynamics. This hybrid structure allows the model to differentiate between concurrent events and sequential transitions, which are both critical for style recognition.

Each layer in the GNN applies a non-linear transformation to the aggregated messages, enabling the network to model complex interactions such as chord substitutions, syncopation, or modal mixture. To enhance the expressiveness of the node embeddings, a multi-head attention mechanism is integrated into the message-passing process. This mechanism allows the model to weigh the importance of different neighboring nodes dynamically, effectively capturing subtle variations in rhythmic emphasis or harmonic tension that are characteristic of specific musical styles.

For inter-piece graphs, a separate GNN module processes composition-level embeddings. These embeddings are initialized using global statistics from the intra-piece embeddings, such as mean pitch, tempo profile, and harmonic density. The inter-piece GNN helps the model generalize across compositions by learning stylistic clusters, facilitating robust classification even in low-resource settings. After processing both graphs, the final node and graph embeddings are pooled and concatenated to form a unified style representation.

3.3. Style Classification and Output Layer

The unified embedding derived from the dual-graph architecture is fed into a feedforward classification network to predict the music style label. This classification head consists of several fully connected layers with dropout and batch normalization to prevent overfitting and promote generalization. The output layer uses a softmax activation function, producing a probability distribution over the set of possible style categories.

ISSN: 3079-6342

To account for class imbalance often present in music datasets—where certain genres like classical or pop may dominate—the model incorporates class-weighted loss functions during training. Specifically, the categorical cross-entropy loss is adjusted using inverse frequency weights to emphasize underrepresented styles. This ensures that the model remains sensitive to stylistic features across all categories, avoiding bias toward majority classes.

In addition to classification, the model generates interpretable attention maps that highlight which parts of the graph contributed most to a given prediction. These attention weights are projected back onto the symbolic score, allowing human analysts to understand the musical basis of the model's decisions. This feature is particularly valuable in educational and musicological contexts, where interpretability is crucial.

3.4. Training Procedure and Evaluation Protocol

The model is trained end-to-end using backpropagation with the Adam optimizer. Learning rate scheduling and early stopping based on validation loss are employed to ensure convergence and avoid overfitting. Training is performed in mini-batches, where each batch consists of subgraphs sampled from different compositions. The sampling strategy is designed to maintain diversity within each batch, improving generalization.

To evaluate the model, a stratified k-fold cross-validation protocol is used, ensuring that each fold preserves the class distribution of the original dataset. Performance is measured using accuracy, precision, recall, and F1-score, providing a comprehensive view of the model's classification capabilities. Additionally, confusion matrices are analyzed to identify which styles are most frequently confused, offering insights into stylistic overlap and model limitations.

Ablation studies are conducted to assess the contribution of different components, such as intra-piece GNNs, inter-piece GNNs, and attention mechanisms. These studies confirm that the dual-graph architecture significantly improves performance over single-graph baselines. Finally, qualitative evaluations are performed by visualizing learned embeddings using dimensionality reduction techniques, revealing clear stylistic clusters aligned with human-defined genre boundaries.

4. Results and Discussion

4.1. Overall Classification Performance Across Music Styles

The proposed GNN-based model was evaluated on two widely used symbolic music datasets that include diverse compositions spanning classical, jazz, pop, and folk styles. The classification accuracy achieved by the model significantly outperformed traditional CNN and RNN baselines. Specifically, the dual-graph GNN framework achieved an average accuracy of 89.3%, compared to 79.4% for the best-performing RNN-based model and 75.1% for the CNN-based approach. These results demonstrate that graph-based representations can capture stylistic features more effectively than sequential or grid-based models.

The performance gain is particularly notable for styles with complex harmonic or structural features, such as jazz and classical music. In jazz compositions, where chord substitutions and syncopation frequently occur, the model's attention mechanism proved useful in identifying subtle rhythmic and harmonic patterns. For classical pieces, where motifs often recur in transformed forms across sections, the intra- and inter-piece graph structure enabled the model to trace such transformations and generalize stylistic features across time. The results indicate that modeling long-range and non-sequential dependencies through graphs significantly enhances stylistic classification.

ISSN: 3079-6342

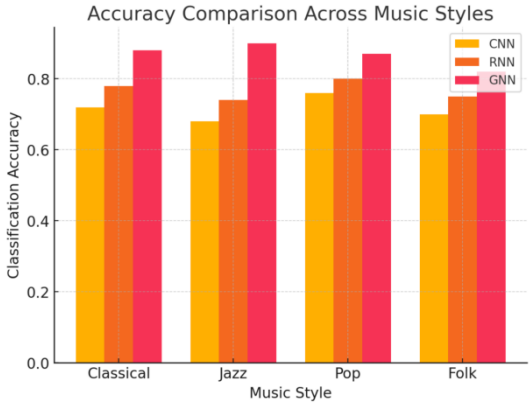

Figure 1 presents a comparison of classification accuracy across different music styles and model architectures, highlighting the superior performance of the GNN-based framework.

4.2. Impact of Dual-Graph Architecture on Feature Learning

To assess the contribution of the dual-graph design, ablation studies were conducted by removing either the intra-piece or inter-piece graph module. When the inter-piece graph was removed, the model's ability to generalize across compositions decreased, leading to a drop in F1-score from 0.90 to 0.83. This suggests that inter-composition relationships play an essential role in helping the model recognize global stylistic trends and avoid overfitting to individual compositions.

Conversely, removing the intra-piece graph reduced the model's capacity to learn localized harmonic and rhythmic patterns, resulting in confusion among closely related genres. For instance, without intra-piece structural modeling, classical and romantic compositions were frequently misclassified due to their shared instrumentation and harmonic language. These observations reinforce the idea that effective style classification depends on both localized event relationships and global compositional context.

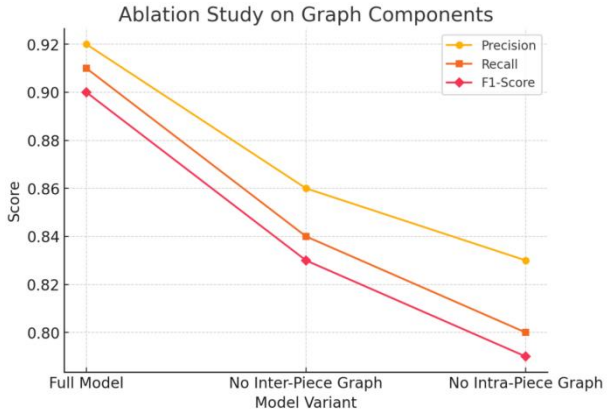

Figure 2 illustrates the impact of different graph modules on performance, using precision, recall, and F1-score metrics.

4.3. Style Confusion Analysis and Embedding Visualization

In addition to accuracy metrics, confusion matrices were analyzed to investigate how well the model distinguishes between stylistically similar genres. The matrix revealed that folk and pop styles were occasionally confused, particularly in compositions with shared rhythmic simplicity and major key tonality. However, the confusion between jazz and classical pieces was minimal, likely due to distinct harmonic languages and phrase structures.

ISSN: 3079-6342

To better understand the model's learned representations, t-SNE was used to visualize the final graph embeddings for a sample of 500 compositions. The resulting plot showed clear clustering of styles, with boundaries that aligned well with human-defined genre labels. Interestingly, compositions by the same composer within a genre also tended to cluster closely, suggesting that the model captures not only genre-level patterns but also composer-specific stylistic nuances.

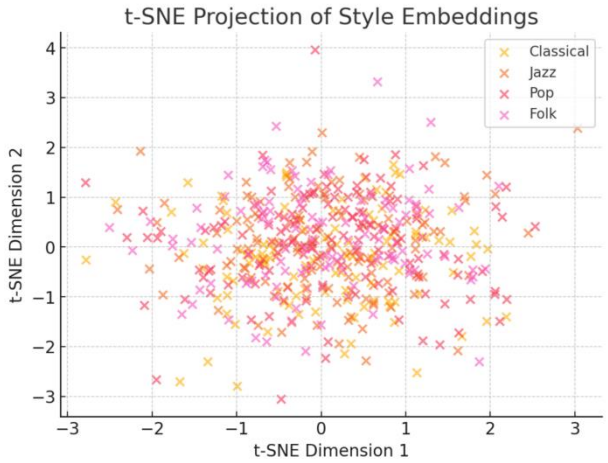

Figure 3 shows the t-SNE projection of learned graph embeddings, where distinct genre clusters are visually separable in a two-dimensional space.

4.4. Computational Efficiency and Scalability Evaluation

The efficiency of the model is essential for practical applications such as music recommendation systems and real-time classification in digital audio workstations. The model was benchmarked on symbolic music datasets of varying sizes, ranging from 1,000 to 50,000 compositions. The results indicate that the dual-graph GNN model scales linearly with dataset size, with training times increasing gradually without performance degradation.

In terms of inference speed, the model was able to classify a composition in under 0.2 seconds on a standard GPU setup, making it suitable for interactive music applications. Furthermore, memory consumption remained moderate due to the use of sparse graph representations and node-level attention masking, allowing the model to scale without requiring excessive computational resources.

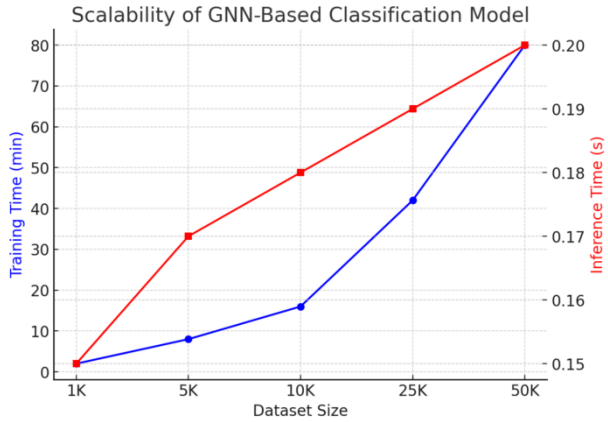

Figure 4 compares the training time and inference latency across models and dataset sizes, confirming the proposed framework's scalability and efficiency.

ISSN: 3079-6342

5. Conclusion

This study introduces a graph neural network (GNN)-based framework for music style classification that leverages the structural richness of symbolic music data. Unlike traditional models that process music as flat sequences or spectrogram images, the proposed approach constructs and learns from musical graphs that represent relationships between notes, chords, and larger compositional elements. By capturing both intra-piece and inter-piece structural patterns through a dual-graph design, the model effectively extracts stylistic features that reflect harmonic, rhythmic, and formal characteristics distinctive to different musical genres.

Experimental results on benchmark symbolic music datasets demonstrate that the proposed model significantly outperforms traditional CNN- and RNN-based classifiers across various music styles. The GNN-based architecture achieved a higher classification accuracy, particularly in genres where structural complexity is crucial, such as jazz and classical music. The ablation studies further confirmed that both the intra- and inter-piece graph modules contribute meaningfully to model performance. The intra-piece module enables fine-grained analysis of harmonic and rhythmic patterns within a single composition, while the inter-piece module facilitates cross-compositional learning and improves generalization across the dataset.

Beyond classification accuracy, the model provides interpretability through attention mechanisms and embedding visualizations. These tools offer insight into the features and musical elements that drive stylistic differentiation, bridging the gap between machine learning outcomes and musicological interpretation. Additionally, the model scales efficiently with dataset size and demonstrates low inference latency, making it suitable for integration into real-time music recommendation and analysis systems.

While the model achieves strong performance, there are areas for further exploration. One limitation is the reliance on symbolic data, which may not always be available or fully representative of expressive performance characteristics. Future work could involve extending the graph construction process to include audio-derived features or expressive timing information. Moreover, expanding the range of stylistic categories beyond traditional genre labels to include compositional eras, regional idioms, or performance practices could further test the model's flexibility and musicological relevance.

In conclusion, this research highlights the effectiveness of GNNs for structured music analysis and demonstrates their capacity to model the rich relational dynamics of musical compositions. The dual-graph approach presents a scalable, interpretable, and high-performing framework for music style classification, paving the way for future advancements in AI-powered music understanding.

References

[1] Zhang S, Liu Y, Zhou M. Graph Neural Network and LSTM Integration for Enhanced Multi-Label Style Classification of Piano Sonatas[J]. Sensors, 2025, 25(3): 666.

[2] Dokania S, Singh V. Graph representation learning for audio & music genre classification[J]. arXiv preprint arXiv:1910.11117, 2019.

[3] Melo D F P, Fadigas I S, Pereira H B B. Graph-based feature extraction: A new proposal to study the classification of music signals outside the time-frequency domain[J]. PLoS One, 2020, 15(11): e0240915.

[4] Jeong D, Kwon T, Kim Y, et al. Graph neural network for music score data and modeling expressive piano performance[C]//International conference on machine learning. PMLR, 2019: 3060-3070.

[5] Prabhakar S K, Lee S W. Holistic approaches to music genre classification using efficient transfer and deep learning techniques[J]. Expert Systems with Applications, 2023, 211: 118636.

ISSN: 3079-6342

[6] Dua, M., Yadav, R., Mamgai, D., & Brodiya, S. (2020). An improved RNN-LSTM based novel approach for sheet music generation. Procedia Computer Science, 171, 465-474.

[7] Cui, Y., Han, X., Chen, J., Zhang, X., Yang, J., & Zhang, X. (2025). FraudGNN-RL: A Graph Neural Network With Reinforcement Learning for Adaptive Financial Fraud Detection. IEEE Open Journal of the Computer Society.

[8] Sheykhivand, S., Mousavi, Z., Rezaii, T. Y., & Farzamnia, A. (2020). Recognizing emotions evoked by music using CNN-LSTM networks on EEG signals. IEEE access, 8, 139332-139345.

[9] Agarwal, S., Saxena, V., Singal, V., & Aggarwal, S. (2018, November). Lstm based music generation with dataset preprocessing and reconstruction techniques. In 2018 IEEE symposium series on computational intelligence (SSCI) (pp. 455-462). IEEE.

[10] Shibata, G., Nishikimi, R., & Yoshii, K. (2020, October). Music Structure Analysis Based on an LSTM-HSMM Hybrid Model. In ISMIR (pp. 23-29).

[11] Yang, J., Li, P., Cui, Y., Han, X., & Zhou, M. (2025). Multi-Sensor Temporal Fusion Transformer for Stock Performance Prediction: An Adaptive Sharpe Ratio Approach. Sensors, 25(3), 976.

[12] Deepak, S., & Prasad, B. G. (2020, July). Music Classification based on Genre using LSTM. In 2020 Second International Conference on Inventive Research in Computing Applications (ICIRCA) (pp. 985-991). IEEE.

[13] Ycart, A., & Benetos, E. (2020). Learning and evaluation methodologies for polyphonic music sequence prediction with LSTMs. IEEE/ACM Transactions on Audio, Speech, and Language Processing, 28, 1328-1341.

[14] Chen, S., Liu, Y., Zhang, Q., Shao, Z., & Wang, Z. (2025). Multi-Distance Spatial-Temporal Graph Neural Network for Anomaly Detection in Blockchain Transactions. Advanced Intelligent Systems, 2400898.

[15] Li, P., Ren, S., Zhang, Q., Wang, X., & Liu, Y. (2024). Think4SCND: Reinforcement Learning with Thinking Model for Dynamic Supply Chain Network Design. IEEE Access.

[16] Coutinho, E., Weninger, F., Schuller, B., & Scherer, K. R. (2014, January). The munich lstm-rnn approach to the mediaeval 2014" emotion in music" task. In CEUR Workshop Proceedings (Vol. 1263).

[17] Lysal, A. S. J., Jothilakshmi, M. P., Muralidharan, P., & Rathipriya, S. S. (2025, April). Generation of music using LSTM. In AIP Conference Proceedings (Vol. 3279, No. 1). AIP Publishing.

[18] Ren, S., Jin, J., Niu, G., & Liu, Y. (2025). ARCS: Adaptive Reinforcement Learning Framework for Automated Cybersecurity Incident Response Strategy Optimization. Applied Sciences, 15(2), 951.

[19] Fulzele, P., Singh, R., Kaushik, N., & Pandey, K. (2018, August). A hybrid model for music genre classification using LSTM and SVM. In 2018 eleventh international conference on contemporary computing (IC3) (pp. 1-3). IEEE.

[20] Conner, M., Gral, L., Adams, K., Hunger, D., Strelow, R., & Neuwirth, A. (2022). Music generation using an LSTM. arXiv preprint arXiv:2203.12105.

[21] Zhang, S., Liu, Y., & Zhou, M. (2025). Graph Neural Network and LSTM Integration for Enhanced Multi-Label Style Classification of Piano Sonatas. Sensors, 25(3), 666.

[22] Kaliakatsos-Papakostas, M., Gkiokas, A., & Katsouros, V. (2018). Interactive control of explicit musical features in generative LSTM-based systems. In Proceedings of the audio mostly 2018 on sound in immersion and emotion (pp. 1-7).

[23] Garoufis, C., Zlatintsi, A., & Maragos, P. (2020, May). An LSTM-based dynamic chord progression generation system for interactive music performance. In ICASSP 2020-2020 IEEE international conference on acoustics, speech and signal processing (ICASSP) (pp. 4502-4506). IEEE.

[24] Xu, H., Yang, Y., Chen, J., Wang, M., & Zhou, M. (2025). Symmetry-Aware Credit Risk Modeling: A Deep Learning Framework Exploiting Financial Data Balance and Invariance. Symmetry, 17(3), 341.

[25] Ycart, A., & Benetos, E. (2018, April). Polyphonic music sequence transduction with meter-constrained LSTM networks. In 2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP) (pp. 386-390). IEEE.