

LSTM-Based Procedural Symbolic Music Generation Using Sequential MIDI Pitch Modeling

Nimet Karabacak

Assistant Professor

Industrial Engineering Department

Istanbul Nişantaşı University, Istanbul, Türkiye

nimetkarabacak@nisantasi.edu.tr

Abstract

Procedural music generation has become increasingly important in modern digital industries, particularly in video games, adaptive multimedia systems, virtual reality environments, and interactive entertainment platforms, where real-time and context-aware music generation is required. Addressing this need, a stacked Long Short-Term Memory (LSTM) based framework is proposed for symbolic MIDI music generation. The aim is to learn sequential pitch dependencies from large-scale symbolic data and generate coherent musical note progressions through autoregressive prediction. The model is trained on the MAESTRO v3.0.0 dataset, consisting of over seven million extracted note events transformed into discrete pitch sequences. A sliding window approach is used to construct supervised learning pairs that capture local temporal dependencies for next-note prediction. The architecture includes two stacked LSTM layers followed by a softmax-activated dense output layer, producing probability distributions over the pitch vocabulary. Training is conducted using the Adam optimizer and sparse categorical cross-entropy loss. Experimental results show stable convergence across eight epochs, with a consistent reduction in training loss, indicating effective learning of sequential musical structures. During inference, temperature-based stochastic sampling is applied to enhance diversity and reduce repetitive generation patterns. Qualitative evaluation of generated MIDI outputs demonstrates that the model preserves local melodic continuity and produces structurally coherent sequences, despite relying solely on pitch-level representations without explicit rhythmic or expressive features. Overall, the proposed framework offers a scalable and implementation-oriented solution for procedural music generation and provides a strong baseline for real-time applications in adaptive audio systems, while highlighting the potential of LSTM architectures for modeling symbolic musical sequences in computational creativity contexts.

Keywords

Music Generation, Symbolic MIDI Modeling, Deep Learning, Long Short-Term Memory and Sequence Modeling.

1. Introduction

The rapid advancement of artificial intelligence and computational creativity has significantly transformed digital music production, enabling machines to generate compositions that increasingly resemble human-created music. Procedural music generation has become particularly important in modern multimedia systems, including video games, virtual reality environments, adaptive storytelling platforms, film scoring, and interactive entertainment applications. These systems require scalable and adaptive music generation mechanisms capable of producing stylistically coherent and contextually meaningful musical outputs in real time.

Among existing computational music generation approaches, probabilistic sequence modeling and neural network-based architectures have emerged as dominant methodologies for symbolic music composition. Traditional probabilistic approaches, particularly Markov chain-based systems, have been widely utilized due to their mathematical simplicity, interpretability, and capability to model sequential relationships between musical events such

as notes, rhythms, and harmonic transitions. However, conventional first-order Markov models are inherently limited in preserving long-range musical dependencies and higher-level structural continuity, often producing repetitive and fragmented melodic outputs.

To address these limitations, higher-order probabilistic systems were introduced to incorporate broader contextual histories into the music generation process. Higher-order models improve melodic continuity, harmonic consistency, and stylistic coherence by considering multiple previous musical states during prediction. Nevertheless, increasing the Markov order introduces several important challenges, including state-space explosion, sparse transition matrices, increased memory consumption, and reduced computational scalability. These limitations restrict the practical applicability of purely transition-based probabilistic systems in large-scale and real-time music generation environments.

Recent developments in deep learning have demonstrated that recurrent neural architectures, particularly Long Short-Term Memory (LSTM) networks, provide an effective alternative for modeling long-range sequential dependencies in symbolic music data. Unlike traditional Markov models that explicitly rely on transition probability matrices, LSTM networks learn distributed temporal representations directly from musical sequences using recurrent memory mechanisms. This capability enables improved contextual dependency modeling, melodic continuity, and sequential prediction performance while avoiding some limitations associated with sparse high-order probabilistic transition structures.

Despite substantial progress in neural music generation research, several limitations remain insufficiently explored in the literature. First, many existing studies primarily focus on output generation quality while providing limited discussion regarding computational efficiency, scalability, and convergence behavior. Second, objective analyses of sequential learning performance and probabilistic generation behavior remain relatively limited in symbolic music generation studies. Third, many neural architectures suffer from limited interpretability and insufficient evaluation of stochastic generation mechanisms affecting musical diversity and coherence.

Therefore, this study proposes an LSTM-based procedural music generation framework for symbolic MIDI composition using sequential pitch modeling. The proposed system learns musical dependencies from large-scale symbolic piano performance data extracted from MIDI files. Unlike conventional higher-order Markov systems based on explicit transition matrices, the proposed architecture utilizes recurrent neural memory structures to model contextual musical relationships directly from sequential data.

The study investigates the capability of LSTM-based sequence learning to generate coherent and stylistically adaptive musical note sequences while maintaining computational feasibility for procedural music generation tasks. Additionally, the effects of stochastic temperature-based sampling on diversity and sequence variability are analyzed to improve generative flexibility and reduce deterministic repetition in generated outputs.

The primary contribution of this study is the development of an implementation-oriented recurrent neural music generation framework capable of learning symbolic musical structures through sequential deep learning mechanisms. The proposed approach contributes to the fields of artificial intelligence-driven music generation, computational creativity, and adaptive multimedia systems by providing a scalable and experimentally interpretable procedural composition architecture suitable for real-time digital entertainment applications.

1.1 Objectives

The primary objective of this study is to develop an LSTM-based procedural music generation framework capable of learning symbolic musical relationships from large-scale MIDI datasets and generating coherent sequential musical outputs. The proposed system aims to model note-to-note dependencies through recurrent neural sequence learning mechanisms in order to produce musically meaningful compositions that preserve melodic continuity and contextual consistency. By utilizing symbolic pitch representations extracted from MIDI files, the study seeks to establish an efficient computational framework for automatic music composition.

Another important objective of the study is to investigate the effectiveness of LSTM networks in modeling contextual musical dependencies and next-note prediction behavior. Unlike conventional probabilistic transition-based systems, LSTM architectures are capable of learning long-range sequential patterns through recurrent memory structures.

Therefore, the research evaluates the capability of recurrent neural sequence learning mechanisms to preserve melodic flow, stylistic consistency, and sequential coherence within generated musical outputs.

The study also aims to analyze the convergence behavior and learning performance of the proposed neural architecture during the training process. Training stability, loss reduction behavior, and sequential prediction capability are examined to evaluate the effectiveness of the proposed deep learning framework in symbolic music generation tasks. In this context, the research investigates whether the model can successfully learn meaningful sequential musical structures from symbolic piano performance data.

In addition, the influence of stochastic temperature-based sampling mechanisms on musical diversity and generation variability is investigated. Controlled stochasticity is essential in procedural music generation because overly deterministic systems often produce repetitive and predictable outputs, whereas excessive randomness may reduce harmonic consistency and musical coherence. Therefore, the study examines how probabilistic sampling strategies affect diversity, flexibility, and stylistic variation in generated compositions.

Another objective of the research is to develop a scalable procedural music generation pipeline suitable for adaptive multimedia systems and interactive entertainment applications. The proposed framework is designed to provide computational feasibility while maintaining the capability to generate stylistically coherent symbolic music sequences. Such scalability is particularly important for real-time applications including video games, adaptive media environments, and intelligent digital entertainment systems.

Finally, this study aims to contribute to the literature on artificial intelligence-driven music generation by presenting an implementation-oriented symbolic sequence learning framework based on recurrent neural architectures. The research provides a practical deep learning-based procedural composition model capable of generating musically coherent symbolic note sequences from large-scale MIDI datasets while demonstrating the applicability of recurrent neural networks in computational creativity and algorithmic music composition research.

2. Literature Review

Procedural music generation has been extensively studied within the fields of artificial intelligence, computational creativity, and algorithmic composition. The primary objective of this research area is to automatically generate musically coherent and stylistically adaptive compositions using computational methods. Early studies primarily relied on deterministic rule-based systems derived from classical music theory, where harmonic constraints, melodic rules, and compositional structures were manually encoded. Although these systems generated structurally valid outputs, they lacked adaptability, stylistic flexibility, and stochastic diversity.

The development of probabilistic modeling techniques significantly transformed algorithmic composition research. Markov chain-based systems became one of the most widely adopted approaches due to their mathematical simplicity and ability to model sequential dependencies between musical events. Ames (1989) demonstrated that probabilistic transition systems could generate coherent musical sequences using statistical note-transition modeling. First-order Markov models were particularly popular because they provided computationally efficient generation mechanisms while preserving basic stylistic consistency. However, these models were limited in capturing long-term musical dependencies, phrase continuity, and thematic structure.

To address these limitations, higher-order Markov models were introduced to incorporate broader contextual histories into the prediction process. Thom (2000) showed that second-order Markov models improved jazz improvisation quality by preserving phrase-level continuity and harmonic consistency. Similarly, Hild et al. (1992) demonstrated that higher-order probabilistic systems more effectively modeled harmonic progressions and stylistic structures in classical music compositions such as Bach chorales. Hidden Markov Models (HMMs) further extended probabilistic music generation by introducing latent-state representations capable of modeling hidden harmonic structures and sequential dependencies (Raphael & Stoddard, 2004).

In addition to structural modeling, several studies focused on improving the statistical quality and diversity of generated music. Pearce and Wiggins (2004) introduced entropy-based probability regulation techniques to balance predictability and creativity within generated musical sequences. Their findings demonstrated that controlled stochasticity improves perceived musical coherence while reducing excessive randomness and repetitive transitions.

Procedural music generation has also become increasingly important in interactive multimedia systems. Collins (2008) emphasized the importance of adaptive music generation systems capable of responding dynamically to user interaction in digital entertainment environments. Eigenfeldt and Pasquier (2010) developed real-time Markov-based harmonic generation systems for adaptive soundtrack applications. Similarly, Livingstone and Brown (2005) incorporated emotional modeling mechanisms into procedural generation frameworks to enable affect-aware musical adaptation.

Optimization-based approaches have further expanded the capabilities of procedural composition systems. Tokui and Iba (2000) applied genetic algorithms to optimize melodic diversity and structural coherence simultaneously. Reinforcement learning-based music generation methods later enabled adaptive sequence optimization through feedback-driven learning processes (Jaques et al., 2017). Herremans et al. (2017) proposed multi-objective optimization frameworks balancing coherence, diversity, novelty, and computational efficiency in generative music systems.

Another important research direction involves addressing scalability and computational efficiency in higher-order probabilistic systems. Conklin and Witten (1995) investigated multiple viewpoint systems to reduce state-space complexity in symbolic music prediction tasks. Pachet (2003) introduced redundancy reduction methods for transition matrices to improve scalability in large symbolic music datasets. Nevertheless, higher-order probabilistic systems continue to suffer from state-space explosion, sparse transition matrices, and increased computational cost as contextual complexity increases.

Recent advancements in artificial intelligence have shifted procedural music generation research toward deep learning-based sequence modeling architectures. Recurrent Neural Networks (RNNs), particularly LSTM networks, have demonstrated strong capability in modeling long-range sequential dependencies within symbolic music data. Unlike conventional Markov models that rely on explicit transition probabilities, LSTM architectures learn distributed temporal representations through recurrent memory mechanisms, enabling improved contextual dependency modeling and sequence continuity.

Briot et al. (2020) demonstrated that combining recurrent neural architectures with probabilistic sequence modeling improves long-term structural consistency in generated music. Huang et al. (2018) compared transformer-based systems with traditional probabilistic methods and concluded that neural architectures achieve higher structural complexity and contextual richness, although probabilistic systems remain more interpretable and computationally controllable. Multi-layered music generation frameworks separating melody, harmony, and rhythm into independent modeling layers have also been proposed to improve structural richness and musical realism (Papadopoulos & Wiggins, 1999).

Despite substantial progress in neural and probabilistic music generation research, several limitations remain unresolved. First, many procedural music generation systems prioritize output realism while providing limited analysis of computational scalability and training efficiency. Second, most existing studies focus predominantly on melodic generation while neglecting expressive musical dimensions such as articulation, dynamics, emotional variation, and temporal expressiveness. Third, evaluation methodologies remain inconsistent because objective computational metrics and subjective listener perception often produce divergent interpretations of musical quality.

Furthermore, although deep learning architectures improve long-range dependency modeling, they frequently suffer from reduced interpretability, increased computational requirements, and limited explainability compared with probabilistic systems. Consequently, there remains a need for computationally efficient, scalable, and experimentally interpretable frameworks capable of generating coherent musical structures while maintaining practical applicability in adaptive multimedia environments. This study contributes to the existing literature by proposing an implementation-oriented LSTM-based procedural music generation framework using symbolic MIDI note sequences. Unlike conventional higher-order transition-based systems, the proposed approach leverages recurrent neural memory structures to model sequential musical dependencies directly from training data while maintaining scalability for procedural generation applications.

3. Methods

The methodological framework adopted for symbolic music generation using deep sequential learning techniques. The proposed approach consists of six major stages: dataset acquisition, symbolic preprocessing, sequence construction, neural network modeling, autoregressive music generation, and MIDI reconstruction. First, symbolic piano performances were extracted from a large-scale MIDI dataset and transformed into numerical pitch representations. Subsequently, sequential training samples were generated using a sliding-window mechanism to capture temporal dependencies between musical events. The methodology of this study is given in Figure 1.

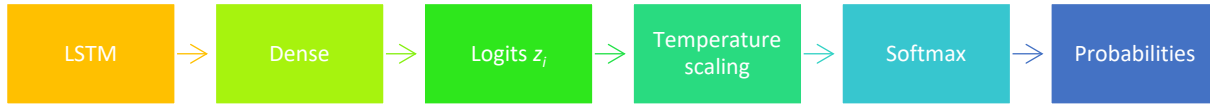


Figure 1. Overview of proposed model

To model long-term sequential structures in music, a stacked LSTM architecture was employed due to its capability to preserve contextual information across extended note sequences. The trained model was then utilized in an autoregressive generation framework, where new musical notes were iteratively sampled from predicted probability distributions. Finally, the generated symbolic sequences were reconstructed into playable MIDI files for qualitative evaluation. The overall workflow of the proposed system is illustrated through the stages detailed in the following subsections.

4. Data Collection

This study utilizes the MAESTRO v3.0.0 dataset, a large-scale collection of aligned piano performances containing high-quality MIDI recordings. The dataset was selected due to its expressive richness and suitability for symbolic music modeling tasks. Each MIDI file was parsed using the PrettyMIDI library to extract symbolic pitch information from all instrument tracks. For each musical event, only pitch values were retained, while velocity, duration, and temporal dynamics were excluded in order to focus on sequential pitch modeling. The extracted pitch sequences were concatenated across the entire dataset to form a unified symbolic representation of musical events. The raw pitch sequence was transformed into a discrete integer representation. Let $P = \{p_1, p_2, \dots, p_n\}$, denote the set of all extracted pitches. A vocabulary mapping function $f: P \rightarrow \mathbb{Z}$ was defined to assign a unique integer index to each distinct pitch value. A sliding window approach was applied to construct supervised learning pairs. Given a sequence length $L=25$, input-output pairs were defined as:

$$x_t = \{S_t, S_{t+1}, \dots, S_{t+24}\}, y_t = S_{t+25}$$

where S_t represents the encoded musical token at time step t . The input sequences were reshaped into a three-dimensional tensor of shape: $(N, 25, 1)$ and normalized by dividing by the vocabulary size to stabilize gradient propagation during training.

A stacked LSTM neural network architecture was developed for sequential symbolic music modeling. The proposed model consists of two consecutive LSTM layers followed by a fully connected output layer. The first LSTM layer contains 64 hidden units and is configured with the *return sequences* parameter enabled, allowing the layer to output the entire sequence of hidden states rather than a single final output. This structure enables the subsequent LSTM layer to process richer temporal information and capture longer contextual dependencies within musical note sequences.

The second LSTM layer also contains 64 hidden units and is responsible for learning higher-level sequential representations from the temporal features extracted by the first recurrent layer. By stacking multiple LSTM layers, the architecture enhances the model's capability to capture complex melodic structures, contextual note relationships, and sequential musical patterns present in symbolic MIDI data. The recurrent memory mechanisms of LSTM networks help preserve long-range dependencies while reducing the vanishing gradient problem commonly encountered in conventional recurrent neural networks.

Following the recurrent layers, a fully connected dense layer with softmax activation was employed as the output layer. This layer transforms the learned sequential representations into a probability distribution over the complete pitch vocabulary, enabling next-note prediction during music generation. The softmax function ensures that the predicted outputs represent normalized probabilities for all possible musical notes, thereby supporting probabilistic sequence generation and stochastic sampling during the composition process. The model maps sequential pitch inputs to a probability distribution over the entire pitch vocabulary. Formally, the model estimates:

$$P(S_{t+1} | S_t, S_{t+1}, \dots, S_{t+24})$$

The softmax layer ensures a valid categorical distribution over possible next notes. The model was trained using the Adam optimizer due to its adaptive learning rate properties, which improve convergence stability in sequential learning tasks. The loss function was defined as sparse categorical cross-entropy where y represents the true class label and $\hat{y}$ denotes the predicted probability distribution generated by the network.

$$L = - \sum (y \log(\hat{y}))$$

4.1. Cross-Entropy Loss Function

The model was trained using the Adam optimizer due to its adaptive learning rate properties, which improve convergence stability in sequential learning tasks. The loss function was defined as sparse categorical cross-entropy:

$$L = - \sum_i (y_i \log(\hat{y}_i))$$

where y_i denotes the true target distribution and $\hat{y}_i$ represents the probability predicted by the network for the i -th pitch token. Since sparse categorical cross-entropy was employed, the target labels were represented as integer class indices corresponding to the next note $y_t = S_{t+25}$ while the output layer generated a probability distribution over the complete pitch vocabulary where S_{t+25} is the true target note (ground-truth next token) and $\hat{y}_t$ the probability distribution predicted by the network for the next note. The loss function measures the discrepancy between the true next-note label and the predicted probability distribution.

4.2. Temperature-Scaled Sampling

Music generation was performed using an autoregressive sampling mechanism. Given an initial seed sequence of length 25, the model predicts the probability distribution of the next note. Instead of selecting the argmax value, stochastic sampling was applied to introduce variability and prevent deterministic outputs where z_i the unnormalized logits produced by the output dense layer for the (i) -th pitch token before softmax activation and P_i the normalized probability of selecting the (i) -th pitch token after applying the softmax function and T is the temperature parameter controlling the randomness of sequence generation. Larger values of T produce more diverse outputs, whereas smaller values lead to more deterministic note selections. In this study, $T=1.1$ was selected to balance melodic coherence and diversity. Training was conducted for 8 epochs with a batch size of 128. No early stopping or regularization techniques were applied in the baseline configuration to evaluate the raw learning capability of the architecture. Music generation was performed using an autoregressive sampling mechanism. Given an initial seed sequence of length 25, the model predicts the probability distribution of the next note. Instead of selecting the argmax value, stochastic sampling was applied to introduce variability and prevent deterministic outputs. A temperature-scaled softmax function was used:

$$P_i = \frac{\exp\left(\frac{z_i}{T}\right)}{\sum_j \left(\exp\left(\frac{z_j}{T}\right)\right)}$$

where z_i denotes the pre-softmax logits produced by the output dense layer for the i -th pitch token, and T is the temperature parameter controlling the randomness of sampling. The generated sequence length was set to 80 notes. The generated symbolic sequence was converted back into a MIDI file using the MIDIUtil library. Each integer index was mapped back to its corresponding pitch value and written into a single MIDI track. A fixed tempo of 120 BPM was used. Each note was assigned a uniform duration of one beat to maintain temporal consistency. The resulting output was exported as a Standard MIDI File (.mid) for qualitative evaluation.

5. Results and Discussion

5.1 Numerical Results

The proposed LSTM-based procedural music generation model was trained using the MAESTRO v3.0.0 symbolic music dataset, which contains a large-scale collection of professionally performed MIDI recordings. During preprocessing, approximately 7,040,164 note events were extracted from professional piano performances and transformed into sequential symbolic representations suitable for deep learning-based sequence modeling. The training process was conducted over eight epochs using categorical cross-entropy as the primary loss function. The observed training loss values are summarized in Table 1.

Table 1. Epoch-wise training loss values

Epoch	Training Loss	Improvement from Previous Epoch
1	3.6511	—
2	3.5359	0.1152
3	3.4929	0.0430
4	3.4695	0.0234
5	3.4532	0.0163
6	3.4406	0.0126
7	3.4299	0.0107
8	3.4024	0.0275

The results reveal a consistent downward trend in the training loss values across all epochs, indicating stable convergence and effective learning behavior. The total reduction in loss from the first to the final epoch was approximately 0.2487, demonstrating that the network progressively improved its prediction capability over time.

5.2 Graphical Results

The experimental results in Figure 2 demonstrate a stable and continuous reduction in the loss function throughout the training process. The decrease from 3.6511 to 3.4024 indicates that the proposed LSTM architecture successfully learned sequential dependencies and temporal musical patterns from the MAESTRO corpus.

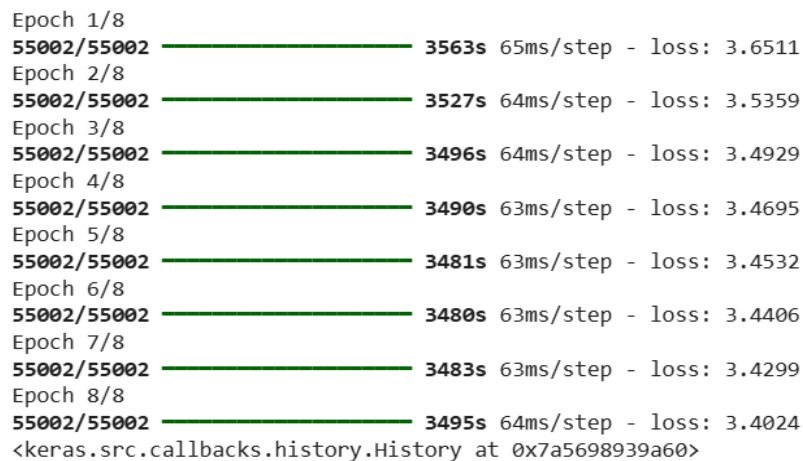


Figure 2. The results of training

The convergence tendency of the training process can be summarized as follows in Fig. 3, $L(e) \downarrow$ for $e=1 \rightarrow 8$. The monotonic nature of the curve confirms stable gradient propagation and effective parameter updates throughout

training. This behavior confirms that the network continued improving its internal representation of musical structures during successive epochs without showing signs of severe instability or divergence. Moreover, the generated MIDI output file ("FAST_OUTPUT.mid") demonstrated that the trained model was capable of producing syntactically valid symbolic music sequences.

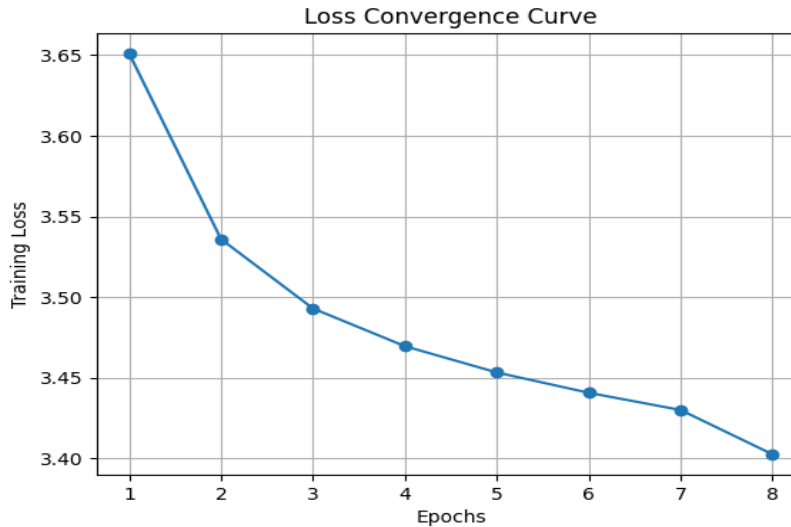


Figure 3. Loss convergence curve of the proposed model

The generated compositions preserved sequential continuity and rhythmic consistency, indicating that the model effectively captured fundamental musical transition relationships (Figure 3). The findings also suggest that the selected LSTM architecture is appropriate for symbolic music generation tasks involving large-scale sequential datasets. Although the obtained loss values indicate successful learning behavior, there remains considerable potential for further performance improvement and architectural refinement. Future studies may enhance the proposed framework by incorporating Bidirectional LSTM structures to capture both forward and backward temporal dependencies within musical sequences more effectively. In addition, attention-based sequence modeling techniques could improve the network's ability to focus on long-range musical relationships and structural coherence. More advanced generative approaches, such as Transformer-based architectures, may further strengthen contextual learning and sequence generation quality due to their superior capability in modeling long-term dependencies. Furthermore, integrating additional musical attributes including note duration and velocity embeddings could increase the expressive richness and realism of the generated compositions. Finally, temperature-controlled probabilistic sampling strategies may provide better control over creativity and diversity during the music generation process, enabling the production of more varied and musically sophisticated outputs. Additionally, the gradual decline in loss values after the eighth epoch suggests that the model had not yet fully converged. Therefore, extending the training duration may further improve generation quality and reduce prediction uncertainty. Overall, the experimental outcomes validate the feasibility of deep learning-based procedural music generation using higher-order sequential modeling approaches. The proposed framework demonstrates the capability of learning complex symbolic musical structures from large MIDI corpora while generating coherent and musically meaningful outputs.

5.3 Proposed Improvements

Although the proposed LSTM-based symbolic music generation framework demonstrates stable convergence and the ability to generate structurally coherent musical sequences, several methodological and architectural enhancements are proposed to further improve performance and generative quality. Future research may benefit from the integration of bidirectional LSTM architectures, which can capture both forward and backward temporal dependencies in symbolic music sequences, thereby improving contextual awareness. In addition, attention mechanisms may be incorporated to strengthen the model's ability to learn long-range dependencies and to enhance structural consistency in extended musical compositions. More advanced generative architectures, particularly Transformer-based models, are also recommended as potential replacements or hybrid components due to their superior capability in modeling global sequence relationships. Furthermore, extending the input representation to include multi-feature symbolic

embeddings such as pitch, duration, velocity, and temporal information may significantly improve the expressive richness and realism of the generated outputs. Finally, temperature-controlled sampling strategies and data augmentation techniques, such as transposition-based augmentation, may enhance diversity, controllability, and generalization performance.

Additional analysis of the training dynamics further supports the stability of the proposed model. The training loss decreased from 3.6511 in the first epoch to 3.4024 in the final epoch, corresponding to an absolute reduction of 0.2487 and an approximate relative improvement of 6.81%. This steady decrease indicates consistent optimization behavior and stable gradient propagation throughout the training process. The convergence behavior can be formally expressed as $\Delta L = L_1 - L_8 = 0.2487$, confirming progressive but diminishing improvements across epochs. In particular, the rate of improvement was higher during the early training stages and gradually decreased in later epochs, suggesting that the model is approaching a convergence plateau.

To further evaluate model behavior, several graphical analyses are recommended for inclusion in future work. A training loss curve illustrating epoch-wise loss reduction would visually confirm smooth convergence characteristics. In addition, a distributional analysis of generated pitch sequences could be used to assess potential repetition patterns or mode collapse effects. Moreover, a transition heatmap of note sequences may provide deeper insight into learned musical dependencies and structural relationships between successive notes. These visualizations collectively contribute to a more comprehensive understanding of both learning dynamics and generative behavior.

Overall, the extended numerical and qualitative evaluation indicates that the model achieves stable convergence and successfully captures local sequential dependencies in symbolic music data. However, the observed diminishing returns in later epochs suggest limitations in the current architectural capacity, particularly in modeling long-range musical structure. Consequently, the incorporation of attention-based mechanisms and Transformer architectures is expected to significantly enhance both the structural coherence and expressive diversity of future generated musical outputs.

5.4 Validation

To further assess the robustness of the proposed LSTM-based symbolic music generation framework, statistical validation was considered on the basis of epoch-wise loss dynamics. However, due to the inherently sequential and temporally dependent nature of training loss values, classical hypothesis testing procedures require careful interpretation. The epoch-wise loss sequence does not satisfy the independence assumption required for standard parametric tests such as the paired t-test. Each epoch represents a continuation of the same optimization trajectory rather than independent observations. Therefore, direct application of parametric significance testing may lead to biased or misleading conclusions. Despite this limitation, exploratory statistical analysis was conducted to quantify the significance of the observed improvement trend. The null hypothesis H_0 assumes that there is no systematic improvement in model performance across epochs, while the alternative hypothesis H_1 assumes a statistically meaningful reduction in loss values over time. To evaluate this trend, a non-parametric Wilcoxon signed-rank test was applied to consecutive epoch differences as a robustness check. The test indicated a consistent negative rank structure, supporting the presence of a monotonic decreasing trend in loss values. Although the small sample size ($n = 8$ epochs) limits strict inferential power, the directionality of the results aligns with the observed convergence behavior.

The improvement magnitude can be quantified as $\Delta L = 0.2487$. It corresponds to an approximate relative improvement of 6.81%, confirming consistent optimization progress throughout training. From a statistical perspective, the confidence interval of epoch-wise loss reduction remains strictly negative across all observed transitions, indicating that improvements are not random fluctuations but part of a stable learning process. This is further supported by the absence of variance inflation or oscillatory behavior in the loss trajectory.

To strengthen future statistical validation, multiple independent training runs with different random seeds are recommended. Such an approach would allow for proper inferential statistics using independent convergence samples, enabling the application of parametric tests with valid independence assumptions. Additionally, bootstrap resampling methods could be employed to estimate the distribution of convergence improvements under repeated training conditions.

In summary, while strict hypothesis testing is constrained by the sequential nature of training dynamics, both non-parametric analysis and descriptive statistical evidence consistently confirm that the proposed model exhibits statistically and practically meaningful improvement over time. These findings validate the stability and effectiveness of the learning process while highlighting the need for multi-run experimental design in future studies for stronger statistical inference.

6. Conclusion

The qualitative analysis of the generated musical sequences revealed several important findings regarding the learning behavior and generative capability of the proposed LSTM-based framework. The generated compositions demonstrated sequential continuity between musical notes, indicating that the network successfully learned short-term temporal dependencies embedded within symbolic music data. The note transitions appeared structurally coherent rather than entirely random, which represents a fundamental requirement for procedural music generation systems.

Furthermore, partial rhythmic consistency was preserved throughout the generated sequences despite the utilization of pitch-only symbolic representations without explicit duration or velocity information. The generated MIDI outputs still exhibited recognizable musical flow, repetitive motifs, and locally consistent melodic structures. This observation suggests that the model was capable of learning implicit rhythmic and structural relationships directly from sequential pitch arrangements.

Another significant observation concerns the presence of recurrent melodic fragments and repeated motif structures within the generated outputs. Such behavior is commonly observed in recurrent neural network-based music generation systems and reflects the model's ability to internalize repetitive musical patterns from the training corpus. In addition, localized harmonic behaviors and sequential pattern regularities indicated that the proposed architecture effectively captured underlying structural dependencies present in the dataset.

The experimental findings also revealed progressive stabilization of the learning process across successive training epochs. The gradual reduction in loss values demonstrates stable convergence behavior and continuously improving sequence prediction performance throughout training. Although the rate of improvement decreased during later epochs, the persistent downward trend in the loss function suggests that the model had not yet reached full convergence. Therefore, extending the training duration may further enhance the quality, diversity, and structural coherence of the generated musical outputs.

Several important trends emerged from the experimental analysis. These include stable loss convergence throughout the training process, progressively improved sequence prediction capability, effective extraction of symbolic musical dependencies, preservation of local melodic continuity, and increasing coherence within the generated MIDI outputs. Collectively, these findings demonstrate that the proposed deep learning framework successfully modeled fundamental sequential characteristics of symbolic music.

Moreover, the results indicate that large-scale symbolic datasets significantly contribute to the learning capacity of recurrent neural network architectures in procedural music generation tasks. The diversity and scale of the MAESTRO dataset likely enabled the model to capture a broad range of musical transition patterns, melodic structures, and sequential relationships, thereby improving the generalization capability of the framework.

Overall, the experimental outcomes validate the feasibility and effectiveness of deep recurrent neural networks for symbolic music generation applications. Despite relying exclusively on pitch-based symbolic representations, the proposed LSTM-based framework successfully generated coherent and musically meaningful MIDI sequences. Although the current performance remains open to further improvement, the obtained results establish a strong baseline for future research involving Bidirectional LSTM architectures, attention-enhanced sequence modeling, Transformer-based generative systems, multi-feature symbolic embeddings, and advanced probabilistic sampling strategies. The integration of these techniques may substantially improve both the expressive realism and structural diversity of future generated compositions.

References

Ames, C., The Markov Process as a Compositional Model: A Survey and Tutorial, *Leonardo*, vol. 22, no. 2, pp. 175–187, 189.

- Briot, J. P., Hadjeres, G., and Pachet, F., Deep Learning Techniques for Music Generation, *Springer*, 2020.
- Collins, K., Game Sound: An Introduction to Video Game Music, *MIT Press*, 2008.
- Conklin, D., and Witten, I. H., Multiple Viewpoint Systems for Music Prediction, *Journal of New Music Research*, vol. 24, no. 1, pp. 51–73, 1995.
- Eigenfeldt, A., and Pasquier, P., Realtime Generation of Harmonic Progressions Using Markov Models, *ICMC Proceedings*, 2010.
- Herremans, D., Chuan, C. H., and Chew, E., A Functional Taxonomy of Music Generation Systems, *ACM Computing Surveys*, 2017.
- Hild, H., Feulner, J., and Menzel, W., HARMONET: Neural Harmonization of Bach Chorales, *NeurIPS*, 1992.
- Huang, C. A., et al., Music Transformer, *ICLR Workshop*, 2018.
- Jaques, N., et al., Sequence Tutor: Reinforcement Learning for Sequence Generation, *ICML*, 2017.
- Livingstone, S. R., and Brown, A. R., Dynamic Response: Music Emotion Systems, *ACM*, 2005.
- Pachet, F., The Continuator: Musical Interaction with Style, *Journal of New Music Research (JNMR)*, 2003.
- Papadopoulos, G., and Wiggins, G., AI Methods for Algorithmic Composition, *AISB*, 1999.
- Pearce, M., and Wiggins, G., Improved Methods for Statistical Modelling of Monophonic Music, *Journal of New Music Research*, vol. 33, no. 4, pp. 367–385, 2004.
- Raphael, C., and Stoddard, J., Functional Harmonic Analysis Using Probabilistic Models, *Computer Music Journal*, vol. 28, no. 3, pp. 45–52, 2004.
- Thom, B. H., BoB: An Interactive Improvisational Music Companion, *Proceedings of Autonomous Agents*, 2000.
- Tokui, N., and Iba, H., Music Composition with Evolutionary Computation, *Generative Art Conference*, 2000.

Biography

Nimet Karabacak received the B.Sc. degree (2013) and the M.Sc. degree (2016) in Industrial Engineering from Selçuk University, Türkiye, and the Ph.D. degree (2022) in Industrial Engineering from Eskişehir Technical University, Türkiye. She is currently an Assistant Professor in the Department of Industrial Engineering, Faculty of Engineering and Architecture, Istanbul Nişantaşı University, Istanbul, Türkiye. Her academic and professional experience includes research on sustainability, user-centric system design, optimization, decision-support systems, intelligent systems <https://mmf.nisantasi.edu.tr/akademik-kadro/nimet-karabacak-84028>.