SSRG International Journal of Electronics and Communication Engineering
ISSN: 2348-8549/ https://doi.org/10.14445/23488549/IJECE-V11I7P115

Volume 11 Issue 7, 150-162, July 2024

Original Article

Exploring Hybrid GRU-LSTM Networks for Enhanced Music Generation

Suman Maria Tony[1], S. Sasikumar[2]

[1,2]*Department of Electronics and Communication Engineering, Hindustan Institute of Technology and Science, Tamilnadu, India.*

[1]*Corresponding Author : sumantony27@gmail.com*

Received: 11 May 2024 Revised: 10 June 2024 Accepted: 10 July 2024 Published: 27 July 2024

Abstract *- The use of deep learning for creating music has been receiving a lot of interest nowadays due to its capacity for innovation and originality. This paper investigates how well hybrid networks that combine Gated Recurrent Units (GRU) and Long Short-Term Memory (LSTM) units perform music-generating tasks. Improving the model's ability to identify long-term relationships and maintain context while generating musical sequences is the aim of the proposed hybrid architecture. It accomplishes this by fusing the benefits of LSTM and GRU units. Comprehensive experiments are conducted on multiple music datasets to evaluate the performance of the hybrid GRU-LSTM networks in generating musical compositions. The quality of created music sequences is evaluated using performance measures like overall musicality, harmonic consistency, and melody coherence. Expert musicians also conduct qualitative assessments to offer insights into the creative and artistic elements of the developed compositions. When compared to current LSTM-based models, the results show how well the hybrid GRU-LSTM networks perform in generating high-quality music sequences with better coherence, consistency, and inventiveness. Furthermore, the study investigates the effects of various coaching strategies and design features on the performance of hybrid networks. All things considered, by exploring novel architectures and strategies for applying deep learning techniques to enhance the uniqueness and quality of generated music, this research enhances the field of music creation. The findings shed light on how hybrid GRU-LSTM networks might encourage innovation in the sector and raise the bar for music production skills.*

Keywords *- Creativity, Deep Learning, Gated Recurrent Units, Hybrid networks, Harmonic consistency, Long Short-Term Memory, Music composition, Melody coherence, Quality assessment.*

1. Introduction

The synthesis of music is a fascinating new area in artificial intelligence that has significant ramifications for human-machine partnerships and the creative industries. By exploring the hybridization of GRU and LSTM networks, this research aims to improve the capabilities of automated music creation systems and explore the complexities of music composition [1]. The aim of creating more intricate and illuminating musical compositions is the main motivation behind this study. RNNs that are now in use, such as GRU and LSTM networks, are excellent at capturing temporal dependencies, which makes them suitable for jobs involving sequential input. Every architecture, though, has advantages and disadvantages of its own. In order to create a hybrid model that takes advantage of the memory retention powers of LSTMs for lengthier musical structures and the efficiency of GRUs for learning short-term dependencies, it is necessary to combine the complementing qualities of these two types [2]. The ability to write engaging tunes programmatically becomes increasingly important as the demand for varied and high-quality music keeps growing. The aim of this study is to investigate the unique contributions made by GRU and LSTM components in the hybrid architecture. This will provide insight into how their synergistic interaction can improve the quality of musical sequences that are generated [3].

In order to generate music automatically, this study sets out on an exciting journey to maximise the synergies of hybrid GRU-LSTM networks. Seek to push the limits of AI-generated music by examining their collaboration potential, which offers more originality as well as transparency in the creation process. The results of this research have the potential to add to the growing body of knowledge on AI-augmented artistic expression as they move through this multidisciplinary terrain [4]. In the age of digital devices, music is one of the most widely used forms of pleasure. Music is seen as a product of human ingenuity that conveys thoughts and feelings via the use of melodies, harmonies, and rhythmic patterns. Numerous genres of music exist, including folk, blues, jazz, rock, pop, etc. Smartphones have characteristics that allow music to be

played both offline and online; consuming music has become simpler in the age of technology.

In contrast to earlier times, there is an abundance of digital music available today, making it time-consuming and exhausting to go through it all [5]. Music recommendations are an attribute of music streaming services like Spotify and Pandora. These characteristics can assist in obtaining a list of suitable songs from well-known music collections based on already-heard music. The system of recommendation is crucial to sustaining the digital streaming music industry. Music suggestions are made by comparing similarities between songs or by favouring an individual over others [6].

To create customized suggestions for the requirements of various audiences, changes are required. As a result, the personalized recommender system for music is more intricate than the standard recommender network. To extract the music characteristics, it is essential to take into account customer requirements in their entirety and integrate music feature identification and sound processing techniques. The goal of this study is to put into action a personalized recommended framework, which is important both in terms of application and academic value. A study strategy is centered on studying how comparable the elements of an audio signal are to achieving the goal. This method can be referred to as content-based music recommendations because it bases its suggestions on the consumer's prior listening history. An understanding of the resemblance metric, which is utilized to compare audio data, is necessary for this strategy [7].

There are different ways to express musical information, including sheet music, audio signals, and symbolic representation. This work has adopted a symbolic representation of music. The first contributory work represents the music in alpha-numeric symbols and then converts them into a wave file using a synthesizer. The music in the subsequent two contributed compositions is in the Musical Instrument Digital Interface (MIDI) format [8].

1.1. Research Gap

In the field of music generation using deep learning techniques, the Research Gap lies in the following areas:

Hybrid Architectures: There is a lack of exploration and development of hybrid deep learning architectures that effectively combine different models, such as GRU and LSTM networks, to enhance the quality and creativity of generated music compositions.

Long-Term Dependency Capture: Existing deep learning models may struggle to effectively capture long-term dependencies in music sequences, leading to limitations in coherence and continuity in generated compositions.

Context and Coherence Maintenance: There is a need to improve the capability of deep learning models to maintain context and coherence in music generation tasks, ensuring that generated compositions flow naturally and maintain musical structure and integrity.

Quality and Creativity Evaluation: Current evaluation metrics and techniques may not accurately assess the quality, creativity, and musicality of generated compositions, particularly in capturing aesthetic and creative aspects.

User Experience Enhancement: There is a gap in focusing on enhancing the user experience by generating personalized and customized music compositions tailored to individual preferences, thereby increasing user engagement and satisfaction.

Complex Musical Structures: Limited exploration and synthesis of complex musical structures, including melodies, harmonies, rhythms, and arrangements, using deep learning models beyond existing composition methods.

Collaboration and Co-creation Facilitation: There is a need to facilitate collaboration and co-creation among musicians, composers, and AI systems, providing a platform for interaction, experimentation, and interdisciplinary collaboration in music composition.

Pushing Boundaries of Artistic Expression: The exploration of new artistic concepts, ideas, and forms of musical expression using deep learning models is limited, with opportunities to push the boundaries of artistic expression in music composition remaining largely unexplored.

Advancing Research and Innovation: There is a gap in contributing to advancing research and innovation in the music-generating industry by developing novel deep-learning techniques, architectures, and evaluation methodologies that improve the quality, creativity, and coherence of generated music compositions.

2. Related Works

Enhanced music generation has its roots in the early experiments with computer-generated music, dating back to the mid-20th century. Early pioneers like Alan Turing and Max Mathews explored the potential of computers to create musical compositions, laying the foundation for algorithmic music generation. Over the decades, advancements in computer science, artificial intelligence, and digital signal processing have propelled the field forward. The introduction of neural networks and deep learning in the late 20th and early 21st centuries marked a significant milestone, allowing for more sophisticated and realistic music generation. Techniques such as Recurrent Neural Networks (RNNs), Long Short-Term Memory (LSTM) networks, and, more recently, Generative Adversarial Networks (GANs) have been employed to model and generate complex musical patterns, capturing the intricate structures of rhythm, melody, and harmony. The recent surge

in enhanced music generation can be attributed to the integration of these advanced neural network architectures with large-scale datasets and powerful computational resources. Researchers have developed models that not only mimic human composers but also innovate by blending styles, instruments, and genres in novel ways. Notable projects like OpenAI's MuseNet and Google's Magenta have demonstrated the potential of AI to generate high-quality music that can be indistinguishable from human-created compositions. The focus has also shifted towards creating more interactive and user-friendly tools, enabling musicians and producers to collaborate with AI in real time. This symbiotic relationship between human creativity and artificial intelligence is driving the evolution of music, offering new possibilities for artistic expression and reshaping the landscape of music production and consumption.

A software program and algorithm known as a recommender system makes suggestions for products that are likely to capture the audience's attention. Recommendations are connected to various types of usage, such as purchasing products, listening to music, or recently reading news [9]. However, after Apple acquired Beats Music, the production of music commodities shifted. Over the past few years, the music industry has transitioned from commodity sales to a subscription and streaming-based revenue model. This new business approach has made electronic music more accessible than ever before [10].

Recurrent Neural Networks (RNNs) are a part of the feed-forward neural network family, differing from traditional feed-forward networks by transmitting data in time increments. This capability allows the network to learn from both recent and past data [11], making it particularly adept at handling sequences, especially temporal series. RNNs, however, faced a learning issue known as the vanishing gradient problem due to the difficulties in estimating gradients with Back-Propagation Through Time (BPTT). This problem could lead to gradients becoming excessively reduced or amplified. BPTT is a back-propagation method used to estimate the gradients for each phase on an unrolled graph [12]. In RNNs, the outputs from the hidden layer (ht+1) re-enter as inputs for calculating subsequent values, allowing the system to utilize both past and present information and learn from sequences, even periodic ones. The Digital Audio Workstation (DAW) is the industry-preferred method for recording, modifying, learning, and mixing in contemporary audio production. Musical instruments or synthesizers connected to a DAW can be played individually [13].

Computational compositional techniques include generative grammars, Markov models, artificial neural networks, and transition networks. RNNs have historically struggled with learning issues due to the challenges of estimating gradients via BPTT, leading to problems such as vanishing gradients caused by overly minimizing or amplifying consequences. The BPTT method uses reverse propagation to determine the gradient for each step in an unfolded graph. The initial LSTM unit has undergone several minor adjustments since its introduction, and the version currently in use includes block input, a single cell with the Constant Error Carousel, a single gate (input, hidden, and output), an output activation function, and peephole connectors [14]. Music can effectively communicate emotions through the imaginative use of sound and timing. Composing music involves creatively assembling the parts or elements of music to produce an original piece. Music varies in style and convention across different cultures, with Indian classical music and Western music being two notable categories, each with unique histories, traditions, and musical characteristics. The literature identifies two commonly used approaches for generating artificial music: data-based learning methods and metaheuristic methods. Data-based methods are often employed for music imitation, transcription, and generating similar musical sequences, effectively capturing the characteristics and patterns in the training datasets. Metaheuristics, on the other hand, are used for creative exploration of the musical search space and the generation of new musical compositions, adhering to specific rules or guidelines. This work proposes methods for music composition that merge the qualities of both approaches, considering the exploration of the musical search space, adherence to musical guidelines, and capturing the characteristics of musical data.

3. Proposed Methodology

One of three emotions tension, grief, or tenderness was entrusted to each of the 1,160 songs in the dataset used in this study's investigation. The dataset was taken straight from the Diptanu repository. By applying the Adam optimizer and sparse categorical volatility on the validation set, these models were evaluated after being trained on the training set. The data about musical styles greatly influence the model's capacity to pick up particular styles. The following portion introduces the music generation model SCTG. Figure 1 depicts the SCTG model's general architecture. The ultimate goal was to choose which structure, based on its testing set performance and precision, would be evaluated on the testing set. Figure 2 displays the dataset label transmission. As the internal principle of LSTM and GRU models is an entryway, use them. The data that is kept and that is discarded can be managed by these entrances. After comparing the methods, it is determined that the hybrid model outperforms the SVM model, which gains a precision of 60.8% for text classification, with an accuracy of 82.6%. Since this decline pertains to a sample with a single label, employ sparse categorical entropy for the calculations. In this instance, don't utilize many labels—just one, as the Adam optimizer lowers the cost of the loss function compared to other optimizers and has a quicker computation time.

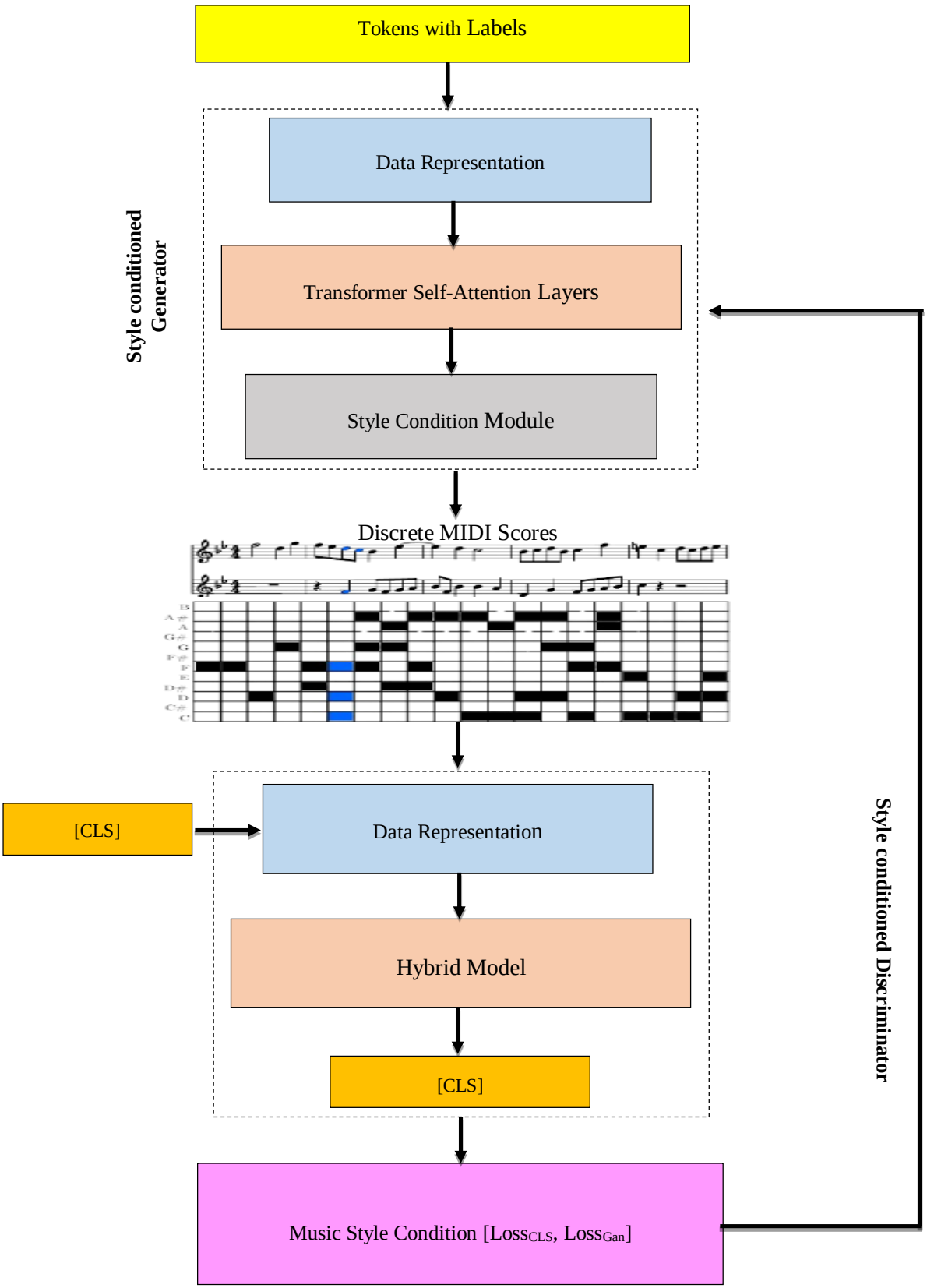


Fig. 1 Proposed methodology

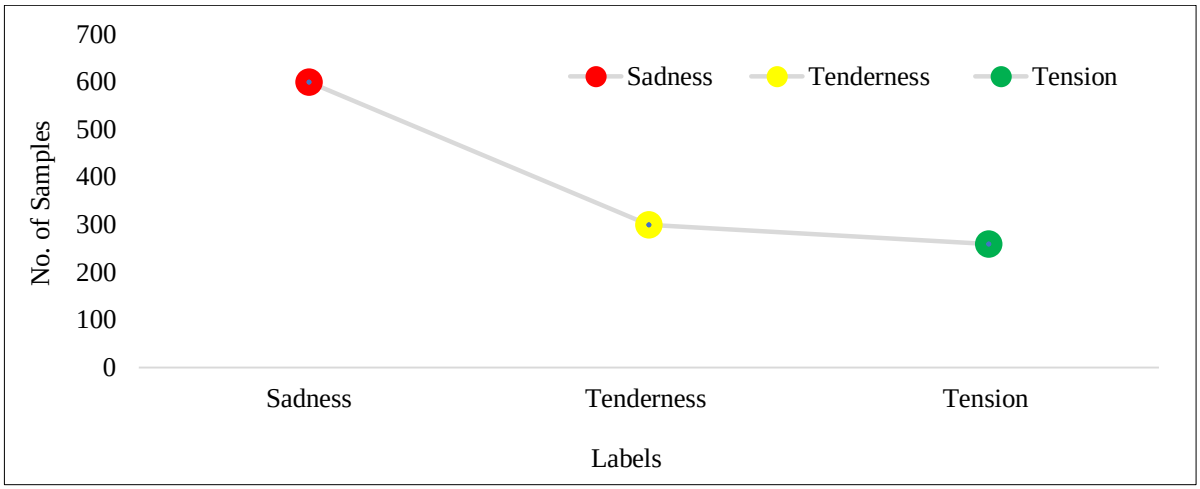


Fig. 2 Distribution of dataset

Table 1. Dataset description

Data Source	Description	Examples
MIDI Databases	Collections of MIDI files that provide detailed information on musical compositions, including notes, timing, and dynamics.	Lakh MIDI Dataset, Classical Archives
Audio Recordings	High-quality audio recordings of music are used for training models to understand timbre, instrumentation, and production techniques.	YouTube Music Dataset, Free Music Archive
Sheet Music	Digitized versions of musical scores that provide structured information on musical pieces, including melody, harmony, and arrangement.	MuseScore, MusicXML datasets
Lyric Databases	Collections of song lyrics that help in understanding the relationship between textual content and musical composition.	MetroLyrics, LyricWiki
Genre-Specific Datasets	Curated collections of music from specific genres are used to train models on the characteristics and nuances of different musical styles.	GTZAN Genre Collection, Million Song Dataset
Performance Data	Data capturing live performances, including tempo variations, expressive timing, and dynamics, are used to add realism to generated music.	MAESTRO Dataset (piano performances), Live Music Archive
Cultural Music Archives	Collections of traditional and folk music from various cultures to ensure diversity and inclusion in music generation models.	Smithsonian Folkways, Global Music Archive
User Interaction Data	Data from user interactions with music platforms, including playlists, likes, and skips, are used to personalize and adapt music generation models.	Spotify Million Playlist Dataset, Last.fm dataset
Music Theory Databases	Structured datasets that include information on music theory, such as chord progressions, scales, and harmonic analysis.	Hooktheory Trends, TonalHarmonyAnalysis

3.1. Data Sources

Table 1 describes the data sources collectively providing a rich and diverse foundation for training and enhancing music generation models, enabling the creation of complex and contextually rich musical compositions.

3.2. Pre-Processing

3.2.1. Tokenization and Normalization

For text-based data like lyrics and music theory:

Tokenization: Splitting text into tokens (words or symbols).

$$\text{Tokens}=\text{Tokenize}(\text{Text}) \quad (1)$$

Normalization: Converting tokens to a standard form (e.g., lowercasing, removing punctuation).

$$\text{Normalized Tokens}=\text{Normalize}(\text{Tokens}) \quad (2)$$

3.2.2. Audio Feature Extraction

For audio recordings:

Mel-Frequency Cepstral Coefficients (MFCC): Common audio features extracted from recordings.

$$MFCC(t) = \sum_{n=1}^{N} \log\big(X(n)\big) \cos\left[\frac{t(n-0.5)\pi}{N}\right] \quad t = 1,2,\ldots k \quad (3)$$

Where X(n) is the magnitude spectrum of the audio signal, N is the number of metal bands, and K is the number of MFCC coefficients.

3.2.3. MIDI Data Processing

For MIDI files: Note Representation: Extracting note information (pitch, velocity, duration).

$$Note=(Pitch, Velocity, Start\ Time, End\ Time) \quad (4)$$

3.3. Data Integration

3.3.1. Combining Different Modalities

For integrating MIDI, audio, and text data, need to align them on a common timeline or framework:

Alignment**:** Synchronize MIDI events and audio features.

$$Aligned\ Data = Align(MIDI\ Data, Audio\ Features) \quad (5)$$

Concatenation**:** Combine feature vectors from different sources.

$$Combined\ Features = Concat(MIDI\ Features, Audio\ Features, Text\ Features). \quad (6)$$

3.3.2. Feature Vector Construction

Creating a unified feature vector for model input:

$$FV = [MIDI\ Features, MFCC, Lyrics\ Embeddings] \quad (7)$$

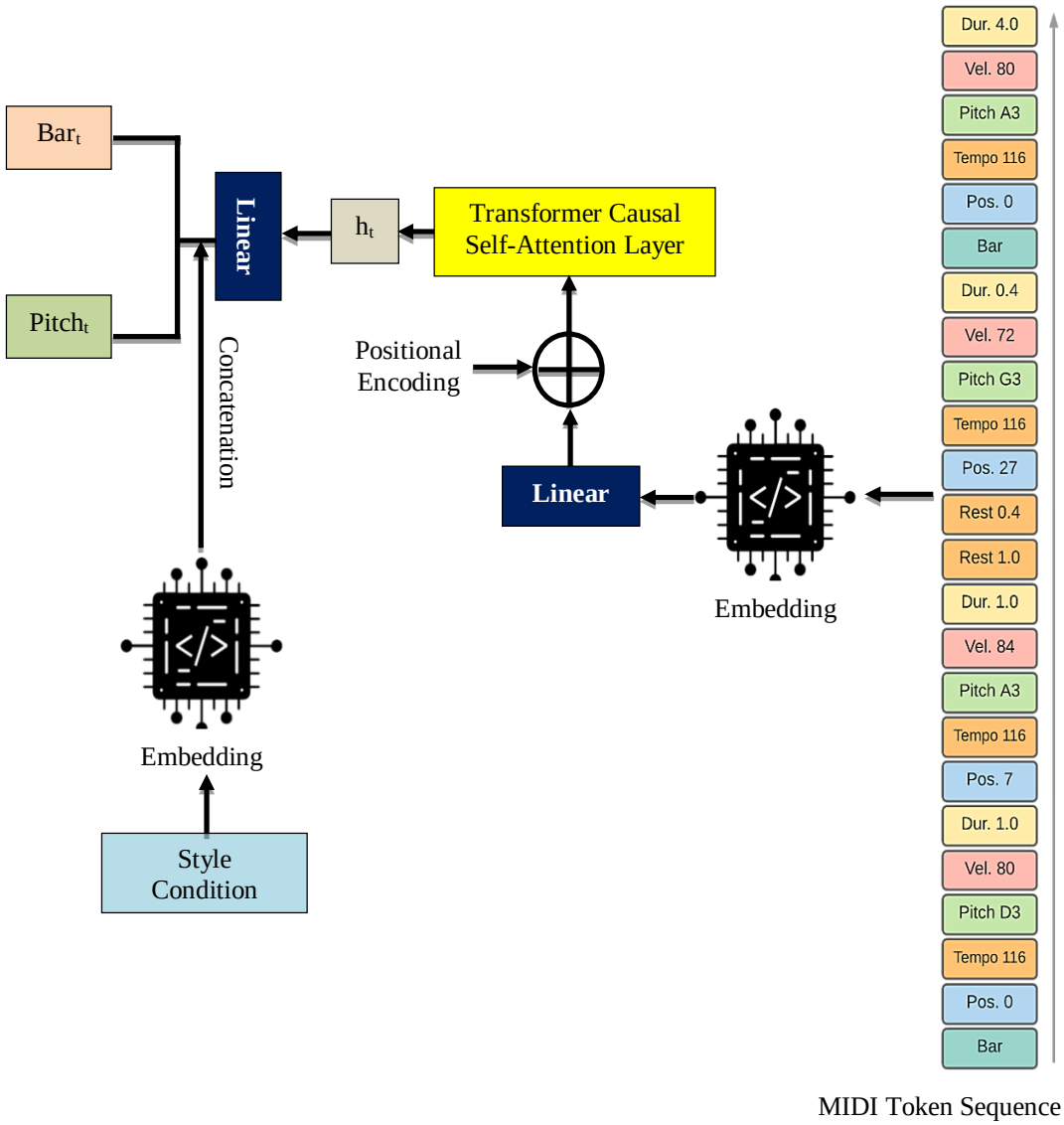


Fig. 3 Architecture of style conditioned linear transformer

3.4. Hybrid Model

An investigation strategy that included training, examination, and validation of datasets was used in the initiation and pre-processing of the data to generate sliding window data. Moreover, the investigation's structure and hyperparameters are utilized for the dataset's forecasting procedure. The methodology served as the foundation for both the structure of the model and the hyperparameter configurations. Two layers of convolution, two pooling layers, two GRU layers, one LSTM layer, one smoothing layer, three dense layers, and two layers for dropouts make up the model employed in this work. There are 192 neurons in the GRU layer and 128 in the LSTM layer. The quantity of layers in the present investigation is one of the enhanced hyper parameters. The lay process of the layer acts as the catalyst mechanism. The layer's function of activation is linear. Describe the actions and modifications that every component makes as it processes the input information in order to produce a mathematical illustration of this design. Here's a summary of it: proportion of the input components to be changed during training, assisting in preventing over-fitting. The dropout operation is a technique used in training; it is not strictly mathematical. Furthermore, the framework created using this analytical approach was applied to forecast data from the test dataset.

The investigation of automatic music generation focuses on the model's interpretive capability. The objective is to create a model that exhibits interpretive capacity and can appropriately represent the sequence information. As an idea, include style data in the final product component relying on the linear Transformer. The style description function, when combined with the model's output attributes, has the ability to alter the pattern's overall output. Figure 3 displays the composition of the simulation. Maximum Likelihood Estimation (MLE) is then used to restrict the procedure. It is not possible to optimize this training method performs badly and exhibits exposure bias when dealing with lengthy sequences like music sequences. The biased goal of a GAN can be introduced to successfully ease this problem. Both the interpretive capacity and the structural knowledge concerning the music-generating model are improved by the discriminator. In Figure 4, the discriminator's structure is displayed.

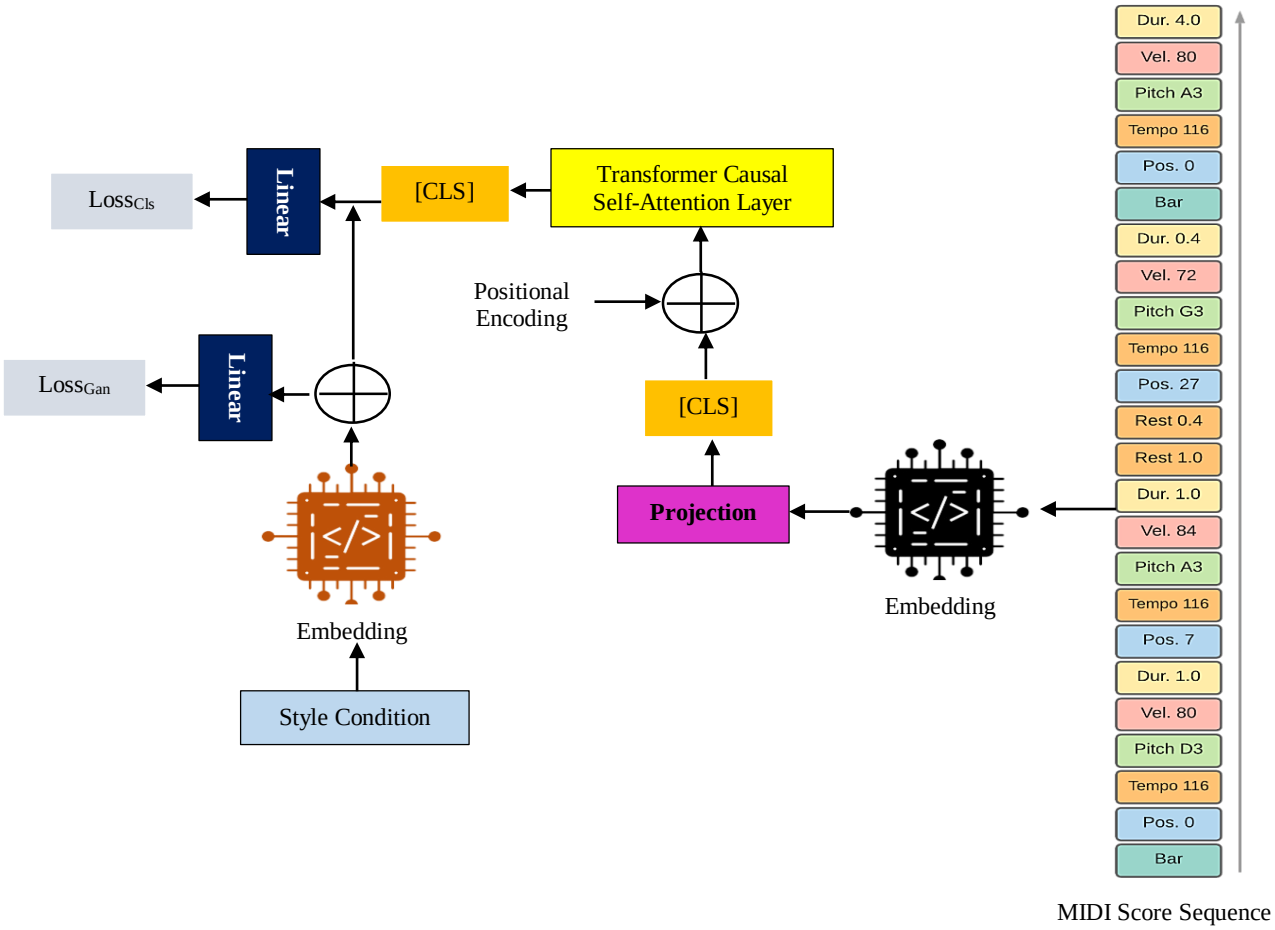


Fig. 4 Architecture of style-conditioned patch discriminator

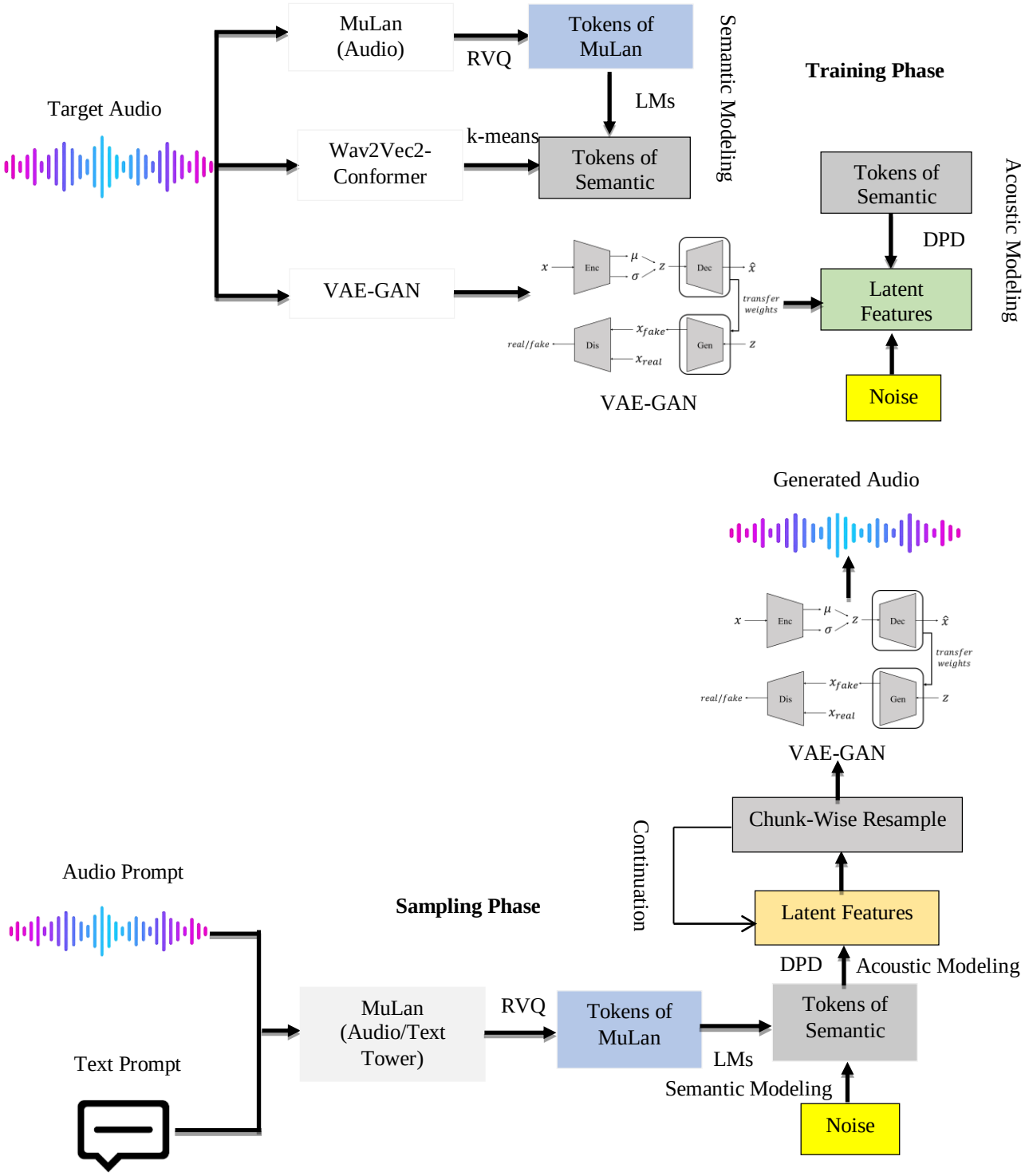


Fig. 5 The training and sampling pipelines of MeLoDy using the proposed system

The DPD model in MeLoDy is conditionally trained using the semantic tokens u1...uTST, which are produced by the LM during inference time and acquired from the SSL model during training, as shown in Figure 5. Tests show that controlling the music's semantics with token-based discrete constraints and allowing the diffusion framework to learn the embedding vector for each of the tokens on its own can greatly increase the consistency of production.

4. Results and Discussion

For the study, information was used from the website investing.com. The temporal span of the data collection is 1645 days, with August 1, 2020, being the beginning and September 29, 2023, being the finish. The initial price and final price are the values of the data parameters that were employed in the analysis. The investigation employed five distinct inaccuracy indicators in all. The RMSE, MAE, MAPE, and R2 error coefficients are these ones. Each simulation has been created, trained, and assessed using a specific set of data. In Table 2 Simulation's performance throughout training is displayed. The technique outperformed the other three prediction methods (CNN, LSTM, and GRU) in terms of predicted value matching rate and proximity to the actual value. The MAE should decrease with improving forecasting. The RMSE value should decrease with increasing prediction precision. R-squared (R2) values can range from zero to one. R2 approaching signifies that the readings are quite similar to one another. However, the observational outcomes for each of the frameworks are shown in Figure 6.

The results from the hybrid technique are the most accurate and exact, with Root Mean Squared Error (RMSE) and Mean Absolute Error (MAE) levels of 0.1089 and 1.4789, respectively, and an R2 value of 0.9991, which is quite close to 1. The MAE, RMSE, and MAPE values drop whilst the R2 value rises when matching the hybrid approach to the LSTM technique.

Figure 7 shows that this hybrid approach improves accuracy while in training and validation.

Table 2. Performance measures

Model Type	Performance metrics			
	F1-Score	Accuracy	Precision	Recall
LSTM	0.32872	0.62107	0.41871	0.37398
CNN	0.26312	0.62107	021404	0.34444
Hybrid Model	0.73364	0.73515	0.73425	0.73291

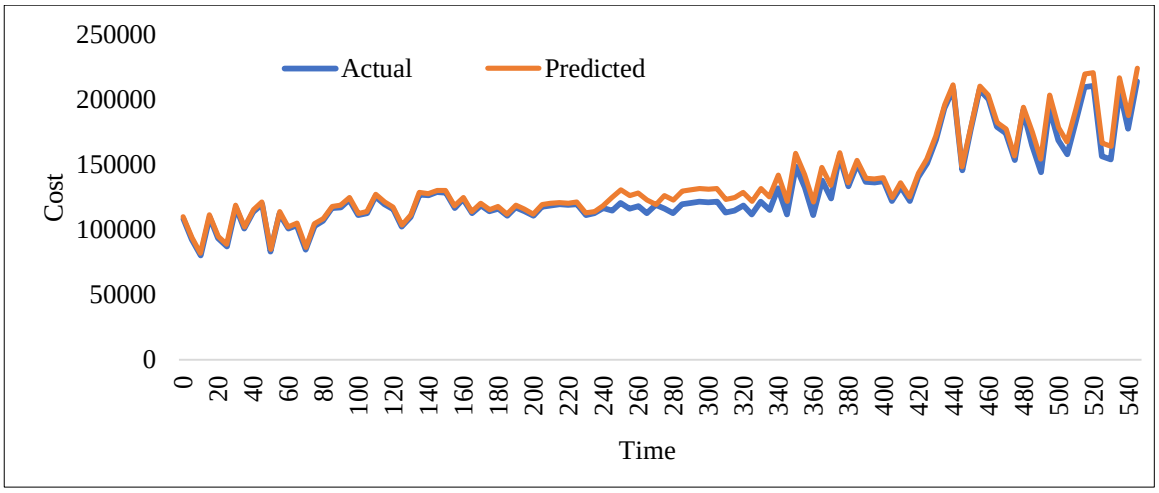


Fig. 6 Test outcomes of Music time vs. cost by using the proposed system

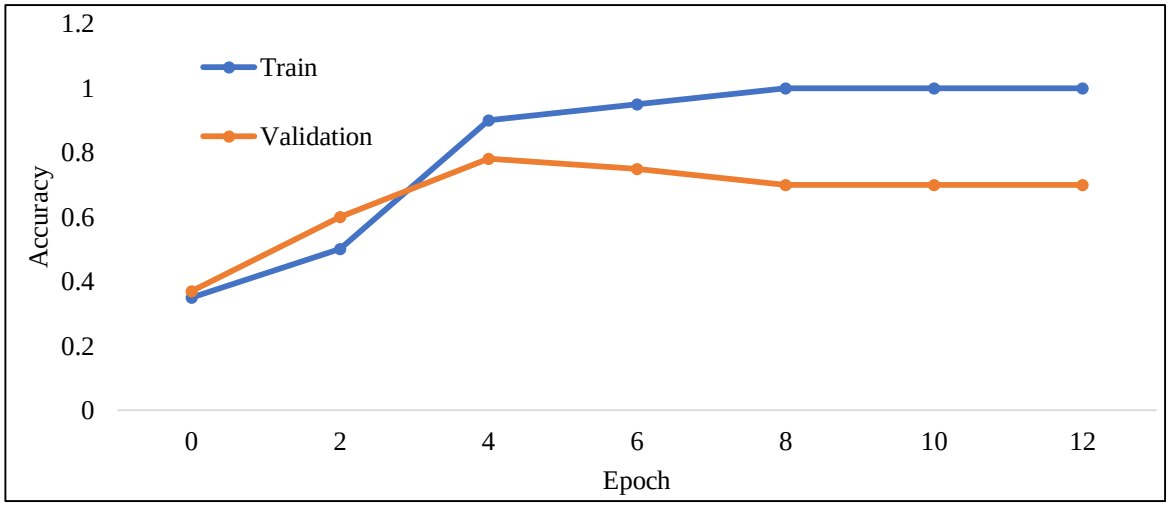


Fig. 7 Accuracy of the proposed model

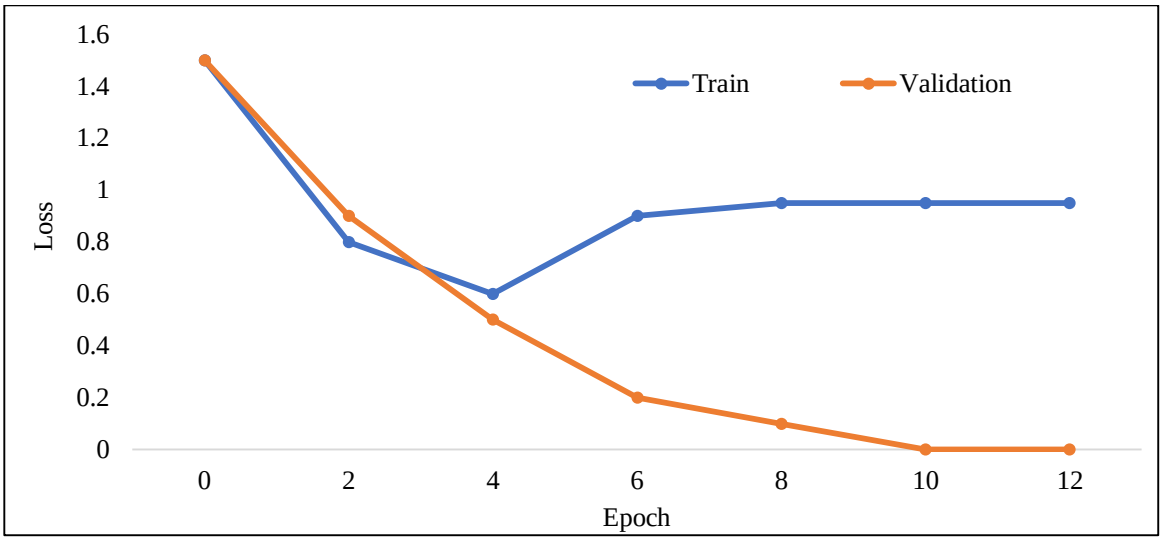
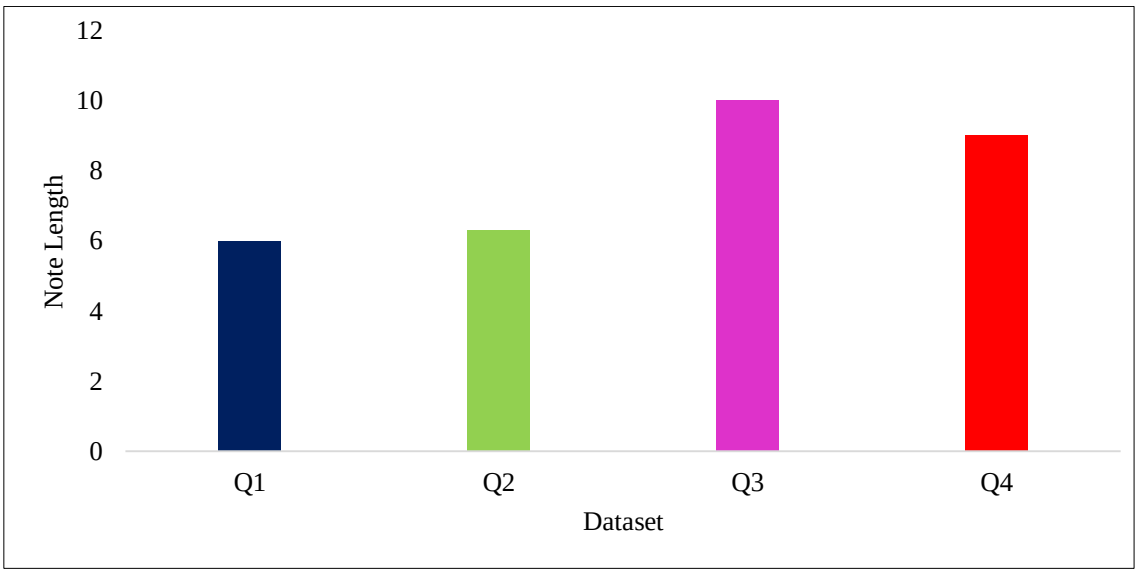


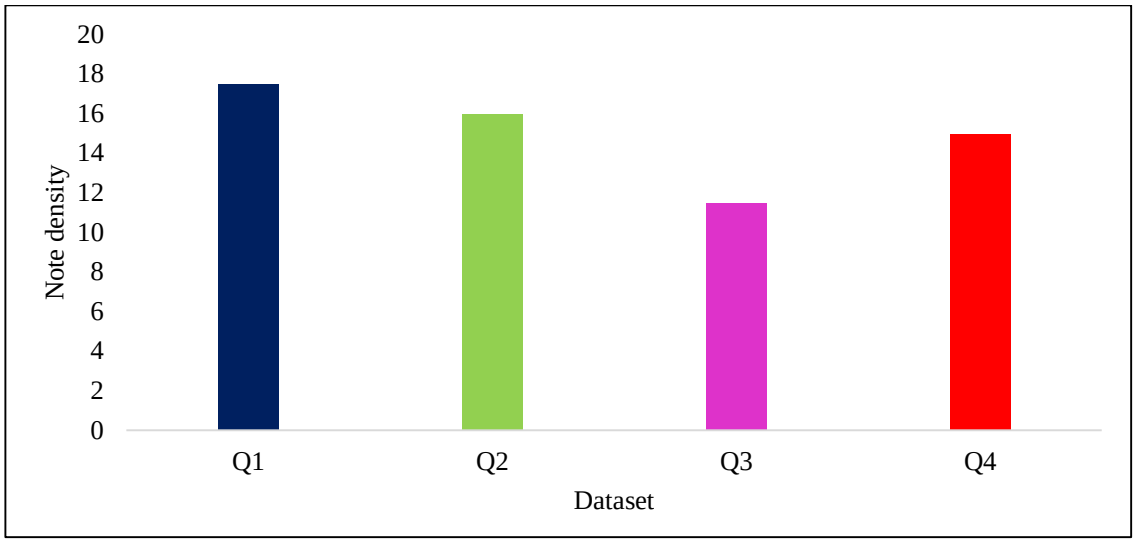
Fig. 8 Loss of proposed model

Figure 8 illustrates how overfiting occurs in deep learning. When the number of validation sets exceeds that of training sets, overfiting takes place. Overfitting in machine learning was caused by numerous variables. The absence of sufficient training data is the cause of the primary component. The second component, which may also contribute to the overfitting issue, is the selected rate of dropout and learning. The third element relates to model complexity; an overfitting issue with an easier model can arise from a complicated one. Pre-processing the data is the fourth component that may contribute to an overfitting issue. A variety of methods can overcome the overfitting graph. The second step is to apply appropriate regularization. To get around the overfitting graph, the data might be fed into the models after a suitable regularization, like L1 or L2 regularization. Last but not least, during the training phase, a dropout can be used to momentarily eliminate random neurons from the model.

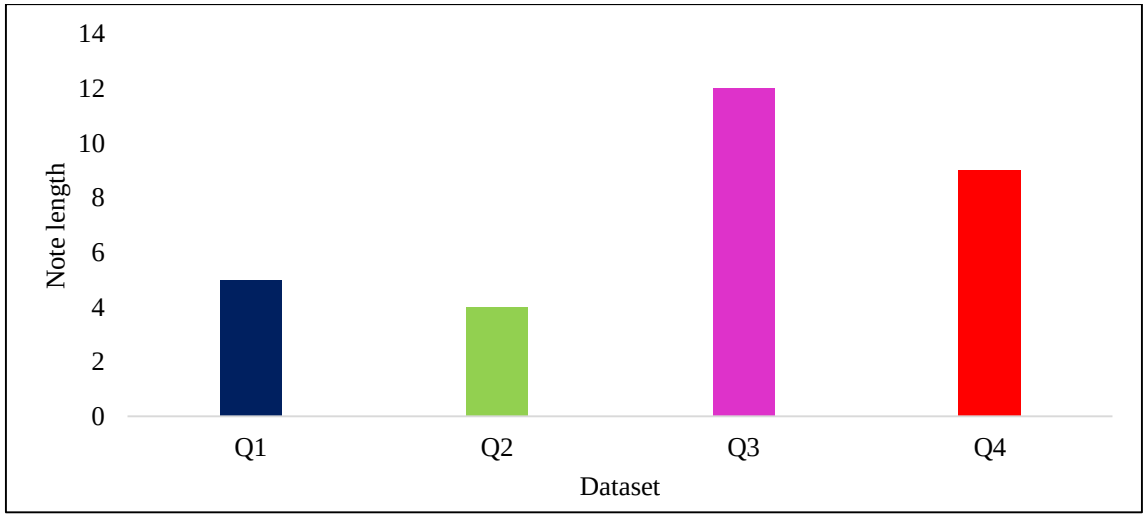
Nevertheless, a comparison of the evaluation as well as training data reveals that the combined model outperforms the other methods in both testing and training. The model's ability to effectively forecast the data is indicated by the lowest RMSE, MAE, and MAPE values in both training and testing. An elevated R2 score suggests that a significant amount of the dataset's difference can be explained by the model. Out of the three models (CNN, LSTM, and GRU), the LSTM model exhibits superior performance in both scenarios. Original and Generated dataset Note length and its density are shown in Figures 9(a) to 9(d).



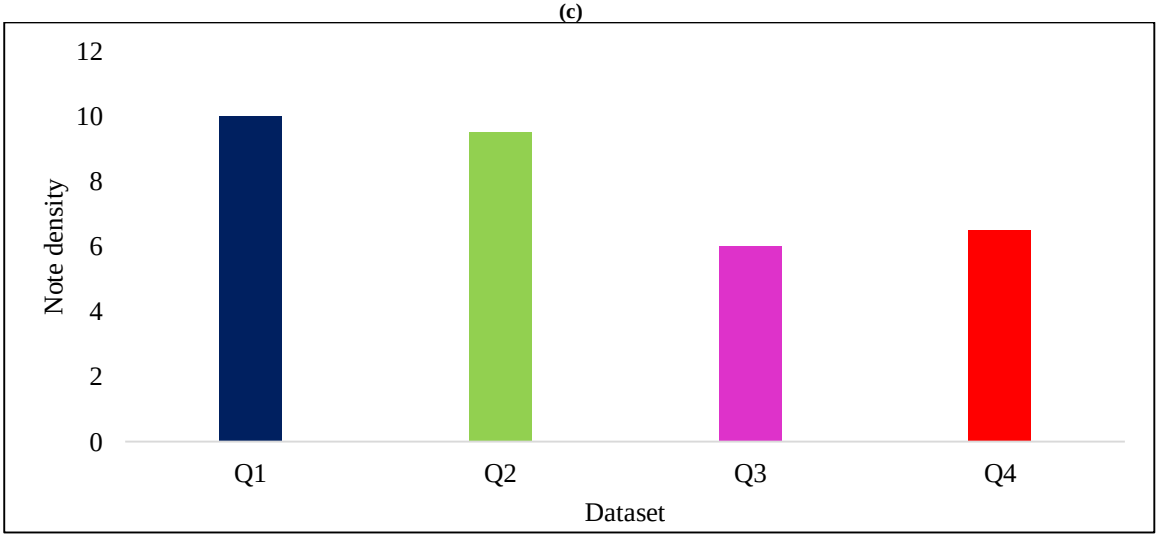
(a)



(b)



(c)



(d)

Fig. 9(a) Original data note length, (b) Original data note density, (c) Generated data note length, and (d) Generated data note density.

Conversely, the GRU model performs the worst during testing and training. These results show that because a hybrid design neither overfits nor underfits, it is typically affordable and balanced. Comparing the LSTM model's training and testing results to those of the other models reveals that it performs admirably in both domains. While the RMSE, MAE, and MAPE values are incredibly low, the R2 value is substantial.

The CNN model works brilliantly, while the GRU model performs better with the highest error rates and lowest R2 values in both instances. This suggests that the GRU model is the least suitable model for this dataset. The LSTM and CNN findings show that both models are consistent with the dataset and do not have any problems with overfitting or underfitting, even if their efficiency is lower than that of the hybrid model.

5. Conclusion

In conclusion, there is a tonne of room for creativity, innovation, and creative expression in the field of deep learning approaches for music generation. To solve current issues and optimise the potential of deep learning models in producing imaginative and high-calibre music compositions, a few crucial areas still need more research and development. First, in order to raise the calibre and inventiveness of generated music compositions, hybrid deep learning architectures that successfully blend several models such as GRU and LSTM networks must be investigated and developed. Capturing long-term dependencies, preserving context and coherence, and synthesising intricate musical structures should be the core goals of these hybrid systems.

Additionally, in order to precisely evaluate the level of creativity, musicality, and quality of generated compositions, extensive evaluation metrics and procedures are required. Expert musician qualitative evaluations should be incorporated into these evaluation processes in order to capture artistic and creative elements beyond the conventional metrics of loss and accuracy.

Additionally, in order to improve user happiness and engagement, personalised and customised music compositions should be created based on individual tastes. This will increase user engagement. In order to promote experimentation and multidisciplinary collaboration in music composition. This entails democratising the process of writing music and enabling cooperation and co-creation amongst musicians, composers, and AI systems.

All things considered, the development of innovative deep learning algorithms, architectures, and assessment methodologies calls for concentrated efforts to advance research and innovation in the field of music generation. Fully utilise deep learning models to produce imaginative and high-calibre music compositions, advancing the field of music composition and encouraging novel forms of artistic expression by filling in research gaps and expanding the parameters of artistic expression.

References

[1] Shuyu Li, and Yunsick Sung, "MRBERT: Pre-Training of Melody and Rhythm for Automatic Music Generation," *Mathematics*, vol. 11, no. 4, pp. 1-14, 2023. [CrossRef] [Google Scholar] [Publisher Link]

[2] Miguel Civit et al., "A Systematic Review of Artificial Intelligence-Based Music Generation: Scope, Applications, and Future Trends," *Expert Systems with Applications*, vol. 209, pp. 1-16, 2022. [CrossRef] [Google Scholar] [Publisher Link]

[3] Zongyu Yin et al., "Deep Learning's Shallow Gains: A Comparative Evaluation of Algorithms for Automatic Music Generation," *Machine Learning*, vol. 112, no. 5, pp. 1785-1822, 2023. [CrossRef] [Google Scholar] [Publisher Link]

[4] Luntian Mou et al., "Memomusic Version 2.0: Extending Personalized Music Recommendation with Automatic Music Generation," *2022 IEEE International Conference on Multimedia and Expo Workshops (ICMEW)*, Taipei City, Taiwan, pp. 1-6, 2022. [CrossRef] [Google Scholar] [Publisher Link]

[5] Shulei Ji, Xinyu Yang, and Jing Luo, "A Survey on Deep Learning for Symbolic Music Generation: Representations, Algorithms, Evaluations, and Challenges," *ACM Computing Surveys*, vol. 56, no. 1, pp. 1-39, 2023. [CrossRef] [Google Scholar] [Publisher Link]

[6] Yi-Jen Shih et al., "Theme Transformer: Symbolic Music Generation with Theme-Conditioned Transformer," *IEEE Transactions on Multimedia*, vol. 25, pp. 3495-3508, 2022. [CrossRef] [Google Scholar] [Publisher Link]

[7] Dyasha Dash, and Kathleen Agres, "AI-Based Affective Music Generation Systems: A Review of Methods, and Challenges," *arXiv*, pp. 1-26, 2023. [CrossRef] [Google Scholar] [Publisher Link]

[8] Prasant Singh Yadav et al., "A Lightweight Deep Learning-Based Approach for Jazz Music Generation in MIDI Format," *Computational Intelligence and Neuroscience*, vol. 2022, no. 1, pp. 1-7, 2022. [CrossRef] [Google Scholar] [Publisher Link]

[9] Jacopo de Berardinis et al., "Measuring the Structural Complexity of Music: From Structural Segmentations to the Automatic Evaluation of Models for Music Generation," *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 30, pp. 1963-1976, 2022. [CrossRef] [Google Scholar] [Publisher Link]

[10] Shobhan Banerjee et al., "Music Generation Using Time Distributed Dense Stateful Char-RNNs," *IEEE 7th International Conference for Convergence in Technology*, pp. 1-5, 2022. [CrossRef] [Google Scholar] [Publisher Link]

[11] Jingwei Zhao, Gus Xia, and Ye Wang, "Domain Adversarial Training On Conditional Variational Auto-Encoder For Controllable Music Generation," *arXiv*, pp. 1-8, 2022. [CrossRef] [Google Scholar] [Publisher Link]

[12] Chenfei Kang et al., "EmoGen: Eliminating Subjective Bias in Emotional Music Generation," *arXiv*, pp. 1-12, 2023. [CrossRef] [Google Scholar] [Publisher Link]

[13] Pedro Sarmento et al., "GTR-CTRL: Instrument and Genre Conditioning For Guitar-Focused Music Generation with Transformers," *International Conference on Computational Intelligence in Music, Sound, Art, and Design*, pp. 260-275, 2023. [CrossRef] [Google Scholar] [Publisher Link]

[14] Zongyu Yin et al., "Measuring When a Music Generation Algorithm Copies Too Much: The Originality Report, Cardinality Score, and Symbolic Fingerprinting by Geometric Hashing," *SN Computer Science*, vol. 3, no. 5, pp. 1-18, 2022. [CrossRef] [Google Scholar] [Publisher Link]