Journal of Networking Technology

DLINE JOURNALS

Print ISSN: 0976-898X
Online ISSN: 0976-8998

JNT 2025: 16 (4)
https://doi.org/10.6025/jnt/2025/16/4/147-154

AI-Driven Music Composition Using Recursive Neural Networks

Meifan Ma
Department of Brewing Engineering, Qilu University of Technology, Jinan. China
Dawei8888@163.com

ABSTRACT

This paper presents an AI driven digital music creation system based on recursive neural networks (RNNs), designed to automate and enhance the music composition process. By leveraging RNNs particularly architectures like LSTM the system learns patterns from MIDI training data to generate new musical pieces that reflect specific styles, rhythms, and harmonic structures. The approach addresses the complexity and unpredictability of traditional digital music creation by reducing reliance on manual input and musical expertise. The study outlines the neural network's architecture, training process, and the use of activation functions, such as Sigmoid, alongside feature selection and data preprocessing techniques to enhance performance. Experimental results show that processed data yields an 81.2% classification accuracy, rising to 90% with paragraph based feature methods. The system demonstrates strong potential in applications such as music education and live performance, though challenges remain in expressing emotional depth and ensuring creative diversity. Future work includes integrating large language models, such as ChatGPT, to enable expressive and improvisational AI musicians.

Keywords: Recursive Neural Networks (RNN), AI-Driven Music Composition, LSTM (Long ShortTerm Memory), MIDI Data Processing, Music Style Classification, Activation Functions, Feature Selection, Digital Music Creation

Received: 10 May 2025, Revised 2 August 2025, Accepted 27 August 2025

1. Introduction

With the ongoing advancement of artificial intelligence technology, numerous sectors have started to investigate and utilize AI technologies. In the realm of digital music creation, the application of AI technology is already prevalent. Digital music creation encompasses music generated via computer, covering all elements such as melody, harmony, rhythm, and arrangement [1, 2]. This approach to music creation can significantly enhance both the efficiency and quality of the music produced while also offering increased creativity and possibilities for composers. Nonetheless, challenges and issues persist in the domain of digital music creation. The primary concern lies in the inherent complexity and unpredictability of the process. Factors such as music theory, harmonic transitions, melodic structures, and rhythmic patterns must all be considered. The interplay among these elements is intricate, necessitating that creators possess a high level of musical knowledge and skills [3]. Additionally, digital music creation requires an understanding of computer programming and algorithm design, which adds further challenges to the process. To address the issues of complexity and unpredictability in digital music creation, a recursive neural network algorithm can be employed to learn from and simulate input music data, ultimately generating new digital music compositions [4]. With this system, users can provide a piece of music data, and the system will autonomously learn and simulate based on the input, producing new digital music works. Traditional methods of digital music creation rely heavily on manual efforts, which can be inefficient and subject to the creators' personal biases. By implementing the recursive neural network artificial intelligence algorithm, our digital music creation system can autonomously accomplish music production without requiring human intervention. This can lead to a notable improvement in both the efficiency and quality of music creation. Furthermore, this system is capable of delivering more tailored digital music pieces, catering to the diverse needs and tastes of users. Additionally, the digital music creation system that utilizes the recursive neural network artificial intelligence algorithm can introduce greater creativity and opportunities within the digital music landscape. For instance, it can facilitate automatic arrangements, rhythm generation, harmony matching, and more. These applications can significantly foster development and innovation in the digital music sector, offering a range of possibilities and options for music creators. The recursive neural network (RNN) is a neural network based artificial intelligence algorithm characterized by a loop structure, enabling it to handle sequence data adaptively [5]. Within a digital music creation system, recursive neural networks can forecast and generate note sequences, enabling automated music generation. Various elements, including the quality of training data, model parameter configurations, and the stylistic attributes of the music, influence the predictive outcomes of recursive neural networks. Thus, careful tuning and optimization of parameters are essential for various application contexts.

2. Related Work

In the digital age, the development and application of digital music creation systems have become a key area of focus in the music industry. The emergence of digital music creation systems offers a more convenient, efficient, and cost effective way for music creators to produce music. Also, it provides more creative possibilities and space for music lovers and professionals. Against this background, the digital music creation system based on the recursive neural network artificial intelligence algorithm has gradually attracted the attention of researchers. Gong and others discussed using deep learning technology to build a music recommendation system to recommend personalized music based on users' listening history and preferences [6]. The authors proposed a neural network based model that can learn the feature representation of music and classify and recommend music based on user preferences. Phuoc and others focused on using machine learning algorithms to identify emotions in music [7]. The authors first introduced the basic concepts and methods of emotion recognition. Then they discussed several current emotion recognition algorithms, including deep learning-based emotion recognition algorithms and sound feature based emotion recognition algorithms, examining their pros and cons. Marceau and others discussed using recursive neural networks (RNN) to generate music [8]. The authors proposed a music generation model based on LSTM (Long Short Term Memory) networks, which can learn the feature representation of music from the input music data and generate new music pieces with similar features. Yan and others discussed using collaborative filtering algorithms to build a personalized music recommendation system [9]. Collaborative filtering is a recommendation algorithm based on user behavior. It can recommend other music similar to the user's interests based on the user's historical behavior and preferences. The authors proposed a recommendation model based on collaborative filtering, which can make personalised recommendations based on the user's preferences and historical behaviour. Dao and others focused on the progress of research on digital music copyright protection technology [10]. The authors introduced the basic concepts and methods of digital music copyright protection, including watermark technology, encryption technology, and digital signatures. They discussed some current copyright protection technologies, such as block chain based copyright protection technology etc. Additionally, the study examined the application of deep learning technology in classifying music styles. A style classification model based on convolutional neural networks was proposed. This model can learn the feature representation of music from audio data and automatically classify music into various style categories, such as pop, rock, classical, and others. The design and implementation of the digital music education platform were also focused on. The basic concepts and methods of digital music education, such as the design principles of online education platforms, the characteristics of music education, etc., were introduced, and how to use modern technology to build a digital music education platform to provide efficient and personalized music education services was discussed. This work encompasses all aspects of digital music, including fundamental concepts, melody construction, harmony processing, rhythm control, and arrangement techniques. At the same time, it also focuses on research progress in the areas of recommendation systems, emotion recognition, and copyright protection, among others, in digital music.

Recent research demonstrates significant advances in AI driven music creation systems utilizing neural networks. Cao Xin [11] developed a multi characteristic music creation system using long short term memory networks based on recurrent neural networks, achieving 98.25% similarity to target music styles with 84.15% expert approval. Duan & Wang designed a semiautomatic digital creation system for electronic music using recurrent neural networks with dense connections and inception like structures, achieving 91.0% and 89.91% accuracy on GTZAN and ISMIR2004 datasets, respectively, after pre training on the Million Song Dataset. M. C investigated RNN based music generation through MIDI file analysis, demonstrating the effectiveness of these networks in creating cohesive musical compositions by learning inherent patterns and structures. Arrais & Avila explored deep learning for music composition using Tied Parallel Networks, combining recurrent and feedforward architectures, creating a real world application system with rhythm control capabilities and mobile integration for continuous music generation.

3. Materials and Methods

3.1 Neural Network Algorithms

The arrangement of a neural network progresses from the input layer to the hidden layer and subsequently to the output layer. The nodes situated between each layer remain unconnected, while every layer enjoys full connectivity. In a recursive neural network, the electrical output of the nodes in each layer is linked to both the input and the output from the previous time step. Specifically, the hidden layer's input incorporates not only the output from the input layer but also the hidden layer's output from the preceding moment. The recursive neural network framework draws inspiration from biological intelligence models, based on studies of biological neural processes. Neurons within a recursive neural network can be categorized roughly into three types: input neurons, output neurons, and hidden neurons. Input neurons are responsible for directly acquiring data or other information from the surrounding environment. Output neurons serve to communicate the constructed neural system back to the external environment. Both categories of neurons are directly tied to the outside world. In contrast, hidden neurons do not interact directly with the external environment and are confined within the framework of the neural network. Their crucial function is to gather input information from within the recursive neural network and subsequently relay the calculated output information to other neurons within the system using formula 1.

$$E = \sum_{i=0}^{N_t}\sum_{j=1}^{m}\left[t_j^{(i)} - y_j^{(i)}\right]^2 \tag{1}$$

3.2 Neural Network Algorithm Process

When developing a recursive neural network model, it is essential first to gather a sufficient number of training samples, referred to as training cycles. After several of these cycles, the neural network's effectiveness will improve significantly, ultimately resulting in a stable output. Measurement indicators are required as input parameters for the network to determine the number of features. Some of these input parameters consist of quantitative values, while others are qualitative. A sublevel scoring technique is implemented to quantify qualitative values. For quantitative values, the engineering economics standardization method can be applied for processing. Initial weight settings or threshold values can be established randomly; however, if the weight value is excessively high, it may lead to rapid saturation of the network, and the initial weight will influence the network's convergence speed to some degree. The learning process of the neural network is executed using the provided sample data, and the metric representing the learning outcome is the root mean square error of the network. Typically, when the root mean square error of the recursive neural network drops below 0.1, it signifies that the training results for the provided samples are satisfactory. The flow of the recursive neural network algorithm is depicted in Figure 1.

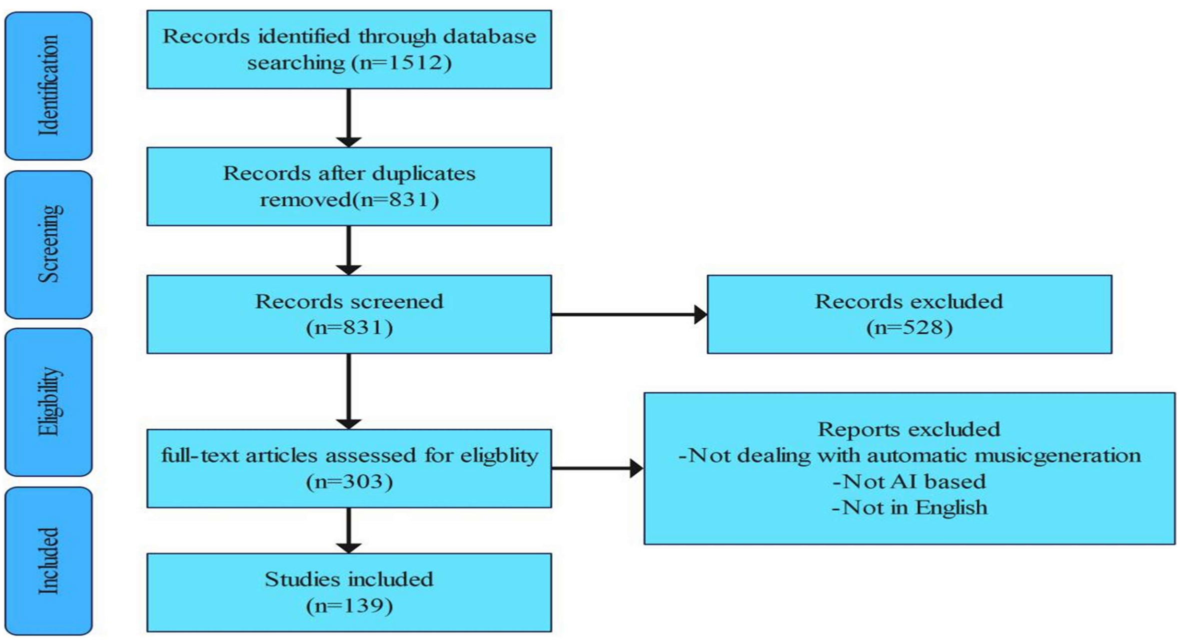

Figure 1. Recursive Neural Network Flowchart

3.3 Activation Function

The equation for weight adjustment in the recursive algorithm is derived based on the learning rate of each subset, the resultant outputs, and the necessary dataset. These subsets encompass data sourced from the original input and data refined through the network. The quantity of nodes corresponds to the number of input variables. The output layer conveys the results of information processing externally.

The number of nodes here equates to the number of output variables. Within the network's output layer, the error signal is defined by the discrepancy between the expected and actual outcomes, providing a direct indication of output accuracy. Conversely, in the hidden layer, error signals originate from the results of preceding layers, contrasting with those of the output layer. The Sigmoid function is the most prevalent activation function utilized in neural networks, as illustrated in formulas 2 and 3. By employing this function to "compress" the inputs from the prior layer of neurons, the output is confined to a specific range. During the development phase, the architecture of the recursive neural network is shaped according to the characteristics of the function. Utilizing the recursive neural network enables an effective integration of input sample data with output data, thus forming a model capable of accurately forecasting output results.

$$\Delta w_{ij}^{n} = \eta \delta_{j}^{n} y_{i}^{n-1} = \eta y_{i}^{n-1} \sum_{k=1}^{m_{n+1}} \delta_{k}^{n+1} w_{jk}^{n+1} f_{n-1}^{\prime}\left(\mathrm{net}_{n-1}\right) \tag{2}$$

$$y_{i} = f\left(\sum_{k=1}^{n} K_{ki} v_{k} + v_{i0}\right) \tag{3}$$

4. Results

4.1 Experimental Design

In terms of system design, this research intends to gather a substantial number of MIDI files to train the neural network model. Subsequently, a music generation model will be constructed based on the processed MIDI data. Moreover, to enhance the quality of the generated music, researchers will devise various adaptive adjustment strategies, such as optimizing the system and facilitating user interaction during the mixing process, in accordance with user defined music characteristics and mixing parameters like volume, balance, and spatial awareness, aimed at boosting the efficiency and quality of music composition. We will also perform both subjective and objective evaluations of the created music to gauge its quality and identify potential advantages and drawbacks. By delving into these research concepts, we aim to develop an AI driven digital music creation system that facilitates the automatic generation of digital music. Given the application demands in music education, high profile live performances, and other settings, a trend of rapid growth is anticipated in the future. By further merging elements like lighting and scene design, highlighting the flow of time and space, along with the color, aesthetics, and dimensions of objects in relation to the music's pitch, timbre, and intensity, we aim to express the emotional qualities of music to achieve enhanced audiovisual and aesthetic impacts. Since the chosen feature set significantly mirrors the relevant classification criteria, analyzing the interplay between the feature sets of two distinct classification standards in the standard music sample library and the user classification standard feature set will allow us to uncover the connection between user classification and the two standard classifications.

4.2 Experimental Result Analysis

The recursive neural network is used as an independent detection technology to obtain and analyze data from the neural network directly. By adopting the method of algorithm decomposition, we can preprocess most of the information in the original model and only retain a small part of the model, and then make decisions based on the output of the preprocessing results, thereby improving the model's performance. This method can greatly improve the accuracy and reliability of the model. However, training with just one set of samples often comes with a lot of noise, which makes it impossible to get satisfactory results with any parameters in each round of iteration, thus making the effect of gradient reduction techniques relatively poor. We can effectively solve linearly inseparable classification problems through the recursive neural network model. In addition, when the number of hidden layers is sufficient, this model can classify various patterns. However, since the learning method of each node is unique, their outputs often fail to achieve the expected results, as shown in Figure 2.

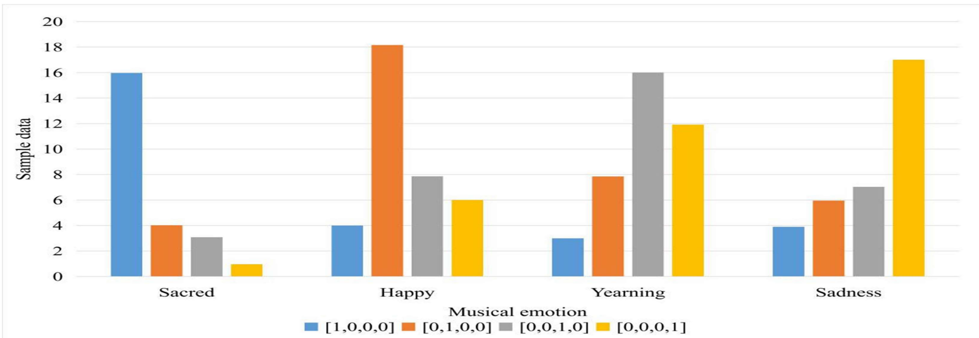

Figure 2. Recognition Result of the Neural Network

From the results, it can be concluded that the classification result of the neural network after data processing (with an accuracy rate of 81.2%) is significantly better than the result before data processing (73%). However, although the data has been processed, the actual error is still inevitable, so the computational accuracy of 81.2% is an excellent recognition result. By comparing two different music style classification methods, we found that the method based on paragraph features is more effective. We established two other models and evaluated their classification results. We found that the accuracy rate of this method in the test set is as high as 90%, proving its effectiveness. As the dimensionality of the features continues to increase, the training time for the two modes will correspondingly become longer, especially when the dimensionality exceeds 4000. However, the research found that even during the training process, the improved feature selection mode can obtain higher accuracy at a lower cost; therefore, it can still provide high efficiency. To achieve better accuracy, we should try to reduce the input for training, which is also a reasonable choice. Removing some redundant features within a proper range can improve computational efficiency without much impact on accuracy. When the dimensionality is below 4000, the time changes are not too significant. Therefore, sacrificing some training time to improve accuracy is acceptable. When the training set data changes between 1% and 10%, the accuracy of the proposed method remains high, fluctuating between 85% and 99%, with minimal variation. The adoption of text serialization techniques and the attention mechanism has increased the model's accuracy from 40% to 90%, significantly improving the model's performance. When the training data is small, the text vector feature representation exhibits high dimensional sparsity, and the model's learning ability is relatively weak. Despite using two different modes to handle large amounts of text information, due to the lack of sufficient redundancy, their classification effects still have uncertain factors, which cause the final results to be biased. To address this issue, we employed the method of embedding perturbation elements in to the input layer, thereby enhancing the model's robustness, as illustrated in Figure 3.

The probability estimate of the logistic regression model is used as the final style authenticity index. The logistic regression model achieves comparable performance in style classification compared to SVM, and its predictions have a probability interpretation. The index ranges from 0 to 1, with higher values representing the L1 writing style and vice versa. We map the style index to a bipolar color axis to make it easier for users to understand the trend of writing styles. Cross validation results show that, in the worst case scenario, where

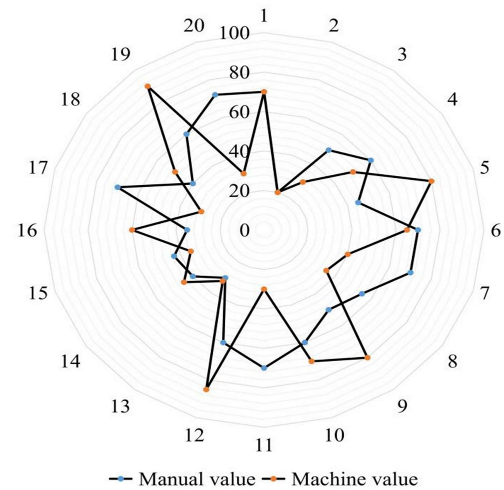

Figure 3. Changes in Music Evaluation Values

the sample length is 70 sentences, the logistic regression model's accuracy and recall rates exceed 10%. We obtain a multinomial logistic regression model for each of the five selected features and train them simultaneously, using the model's prediction output as a measure of the user text's writing style in the five categories of features. The overall range is from 0 to 1, with higher values indicating closer to the L1 writing style and vice versa.

5. Conclusion

The rapid development of artificial intelligence technology provides more possibilities for innovation in digital music. In the digital music creation system, artificial intelligence can assist in directly creating music. This paper introduces the main work and innovation points of the digital music creation system, focusing on its application of artificial intelligence, and highlights existing problems and future research directions. However, compared with traditional music creation, digital music based on artificial intelligence still faces challenges and difficulties. For example, existing artificial intelligence algorithms struggle to express and convey emotion when generating music, lacking the nuanced connotations and meaning inherent in human music creation. Moreover, due to the limitations of large amounts of training data, digital music creation systems may lack sufficient diversity and creativity. This direction also needs to be improved and optimized in the digital music creation system. Therefore, future research aims to enhance music style classification and emotion analysis, improve the performance of music generation algorithms, increase the diversity and creativity of generated music, and make the generated music more closely resemble human music creation, thereby conveying more emotions and meanings. It can be combined with ChatGPT to empower virtual people and robots, transforming them into singers who can create and perform music improvisationally according to human needs. We need to optimize the digital music creation system based on artificial intelligence while also researching how the system can be integrated and empowered with other products.

References

[1] Ku, H. H., Liu, C. H., Wang, W. C. (2022). Design of an artificial intelligence of things based indoor planting model for *Mentha spicata*. *Processes, 10*(1), 116.

[2] Ichimasa, K., Nakahara, K., Kudo, S. E., et al. (2022). Novel "resect and analysis" approach for T2 colorectal cancer with use of artificial intelligence. *Gastrointestinal Endoscopy, 96*(4), 665-672.

[3] Crowell, B., Cug, J., Michalikova, K. F. (2022). Smart wearable internet of medical things technologies, artificial intelligence-based diagnostic algorithms, and real-time healthcare monitoring systems in COVID-19 detection and treatment. *American Journal of Medical Research, 9*(1), 17-32.

[4] Amann, J., Vetter, D., Blomberg, S. N., et al. (2022). To explain or not to explain?—Artificial intelligence explainability in clinical decision support systems. *PLOS Digital Health, 1*(2), e0000016.

[5] Junejo, A. R., Li, X., Madiha, H., et al. (2021). Molecular communication networks: Drug target scalability based on artificial intelligence prediction techniques. *Journal of Nanoparticle Research, 23*, 1-23.

[6] Gong, X., Li, X. (2023). Human–robot interactive communication and cognitive psychology intelligent decision system based on artificial intelligence—Case study. *International Journal of Humanoid Robotics, 20*(02n03), 2240005.

[7] Phuoc, N. V. (2022). The critical factors impacting artificial intelligence applications adoption in Vietnam: A structural equation modeling analysis. *Economies, 10*(6), 129.

[8] Marceau, G., Daniele, F. (2022). Artificial intelligence and international economic law by Shin-yi Peng, Ching-Fu Lin, and Thomas Streinz. *World Trade Review, 21*(4), 516-520.

[9] Yan, L., Song, H., Guo, Y., et al. (2022). HLDnet: Novel deep learning-based artificial intelligence tool fuses acetic acid and Lugol's iodine cervicograms for accurate pre-cancer screening. *Biomedical Signal Processing and Control, 71*, 103163.

[10] Van Dao, D., Bui, Q. A. T., Nguyen, D. D., et al. (2022). Prediction of interlayer shear strength of double-layer asphalt using novel hybrid artificial intelligence models of ANFIS and metaheuristic optimizations. *Construction and Building Materials, 323*, 126595.

[11] Cao, X. (2024). Research on music creation system based on computer artificial intelligence technology. In *2024 IEEE 3rd International Conference on Electrical Engineering, Big Data and Algorithms*.

[12] Duan, Y., Wang, J. (2022). Design of semiautomatic digital creation system for electronic music based on recurrent neural network. *Computational Intelligence and Neuroscience*.

[13] Devadharshini, C., Nishanth, S., Hemavathi, R. (2024). Music generation using neural network. *International Journal of Scientific Research in Engineering and Management*.

[14] Arrais de Souza, V. A., Fontes de Avila, S. E. (2018). Deep neural networks for generating music. *Technical Report — IC-PFG-18-16, University of Campinas Institute of Computing*.