

A Systematic Review of AI-Based Therapeutic Music Generation Systems



Anoushka Tandon

Abstract Music therapy is an alternative medicine used to increase the cognitive, emotional, social, or physical health of individuals over the past decades. Music therapy is most successful globally, where personalized and adequate stimuli are provided by therapists to stimulate the human brain. Traditional music therapy techniques are often limited by cost, expertise, availability, and so on. In recent years, the advancement of generative Artificial Intelligence (AI) technologies has aided in understanding general mental well-being via AI-driven music generation systems. Hence, different research on music generation systems using AI is employed to enhance music therapy. This paper presents a comprehensive review of prevailing generative AI-based music generation systems and highlights the performance of each strategy. The prevailing generative AI-based music generation approaches are categorized into Long-Short Term Memory (LSTM)-based, Variational Autoencoder (VAE)-based, Generative Adversarial Networks (GAN)-based, other deep learning-based and machine learning-based music generation systems. Here, 25 research articles are taken for categorization and the fundamental components are outlined. Moreover, the analysis proved that the deep learning-based music generation system is the most used approach, Python tools are widely utilized, and accuracy is the most used evaluation parameter.

1 Introduction

In recent years, music has become a medium and means for expressing people's emotions, and it also shows real-life experiences, feelings, and thoughts of the creator [1]. The medical health of individuals can be increased by providing physical exercise, music therapy, counseling, and other practices. Music is essential in everyday life, which also helps to resolve mood swings [2]. In general, music is a nonverbal expression and comprises four parts, namely timbre, harmony, melody, and rhythm, that help to express people's emotions as well as thoughts [3]. Music has gained

A. Tandon (✉)
Mumbai, Maharashtra, India
e-mail: anoushkatandon21@gmail.com

huge benefits that include improvement in markers of general mental well-being, like stress reduction, creating links to reduce dementia cases in elderly populations, improving cardiovascular health, and so on [4]. Music not only increases the mood of healthy individuals and contributes to positive well-being but also possesses the tendency to reduce the stress level of patients [5]. The cognitive abilities of listeners are highly influenced by the emotional contents of music [6]. Music therapy is the process of using different forms of music activities, like rhythm, playing, singing, and listening to stimulate the human brain. The stimulation is performed to relax the mind and body by singing sounds [3]. Music therapy is an alternative medicine, where music helps to increase the cognitive or emotional health, and social, and physical health of a group or individuals [5].

A comprehensive interdisciplinary subject that integrates medicine, psychology, and musicology is termed music therapy. The four major music functions, like aesthetic sharing, social integration, psychological empathy, and physiological resonance are used in music therapy intervention on individuals [1]. Music therapy is a risk-free and cost-effective alternative, which provides cultural, emotional, and physical advantages to patients [7, 8]. Although music is widely used for therapy, few smart and intelligent applications are used to assist or support medical therapists in the selection of the most suitable songs for their patients [2]. The music structures based on AI have gained more attention among researchers over the last few years due to the rapid advancement of AI technologies [6]. Over the past three decades, the music generation using AI with artistic style has broadly explored, where the music is converted into the symbolic or audio domain for the conversion of music into machine-perceivable matrices of numbers [9]. At present, musical AIs define the instrument combinations and also generate music in specific musical genres [10]. The recent advancement of deep learning and machine learning techniques in AI technology gained huge attention to the automatic generation of music. Deep learning techniques have attained huge significance in the field of automatic music generation [11], and have also demonstrated significant progress in the area of music generation [11]. A subset of machine learning, Recurrent Neural Networks (RNNs) is used for capturing sequential patterns in music. Moreover, theoretical frameworks, like LSTM networks are used to enable AI systems to comprehend intricate temporal relationships within compositions [12].

In this study, a comprehensive analysis of existing techniques on AI-based music generation systems is performed. This survey mainly delineates the aims, challenges, requirements, and important technical aspects considered for developing the music generation system. The baseline AI-based music generation systems are categorized into three distinct types, namely LSTM-based, VAE-based, GAN-based, other deep learning-based, and machine learning-based music generation systems. The categorization is performed by taking 25 research papers related to AI-based music generation systems from Google Scholar from reputed publishers, like Springer, Elsevier, Wiley, Taylor and Francis (T&F), Institute of Electrical and Electronics Engineers (IEEE), Association for Computing Machinery (ACM), and so on. Similarly, the challenges as well as issues encountered by these techniques are validated and a

brief analysis is executed regarding different evaluation metrics, published years, tools, utilized databases, etc.

The remaining part of the manuscript is arranged as follows: Sect. 2 delineates the general outline of AI-based music generative systems; the categorization of the baseline AI-based music generation system is explicated in Sect. 3. A brief description of challenges and issues encountered by research articles is illustrated in Sect. 4, and the analysis carried out is demonstrated in Sect. 5. Moreover, the article concludes with future research ideas in Sect. 6.

2 The General Outline of AI-Based Music Generation Systems

Music has proven to have an extremely high capacity and potential for automatic creation due to the rising advancement of AI. At present, the generative AI-based music generation system is considered one of the most productive sub-fields in the area of computational creativity and it helps to solve automatic music composition issues. The AI utilizes five basic steps to produce music, they are composition, arrangement, sound design, mixing, and mastering. To understand the AI-based music generation systems, the general outline of the generation techniques is depicted in Fig. 1.

The timbre, length, strength of sound, and pitch are considered the most basic elements of music, which combine to generate the tone, rhythm, harmony, and melody of music. During the music generation process, the music content given as input is encoded in digital form, and eventually the music generative systems provide different types of output, such as style transitions, audio, and chords.

(i) Representation

The input given to the generation system is in the form of music representations and this representation of music contents is categorized into symbols and audio. Here, audio is a continuous signal, and symbol is considered a discrete signal. The audio is represented in two different ways, such as spectrogram and waveform. In audio signal representation, the actual properties of music are retained by actual audio waveforms for the generation of expressive music. Moreover, in symbolic representation, the

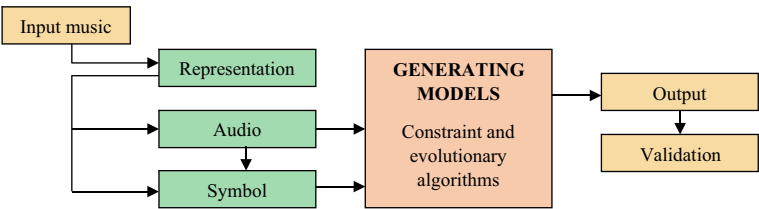


Fig. 1 General outline of AI-based music generation systems

music fragments are obtained by Musical Instrument Digital Interface (MIDI) events, piano roll, and ABC notations.

(ii) *Music Generation Systems*

The music generation system is mainly categorized into different types based on the purpose of generation, such as melody generation, style transfer, audio generation, and arrangement generation. In the music generation system, the counterpoint, lead melodies, and harmonies are generated using composition generation for producing melodies. The music is directly modeled at the audio level using generating sound segments with various styles and instruments for the generation of audio. Nowadays, automatic generation techniques, like machine learning, deep learning, rule-based systems, evolutionary computation, and so on are utilized for music generation.

(iii) *Validation*

The performance of generation techniques is validated after the generation of music. The objective and subjective evaluation is mainly performed by the existing evaluation techniques to analyze the performance of the models.

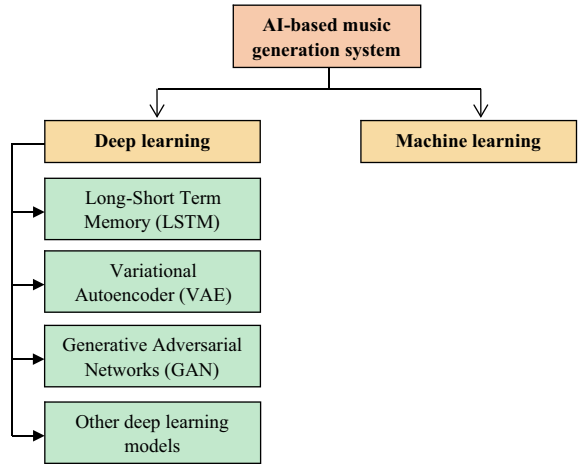
3 Categorization of the Papers

The music generation system is employed to assist therapists, which helps to personalize and customize music according to user settings. It also provides a larger volume of content than the content offered by therapists. The AI-based music generation system comprises valuable tools in recent years that provide resources to interact, innovate, and explore technology in various aspects of the music creation process. The AI-based music generation systems are generally categorized into five different types. They are LSTM-based, VAE-based, GAN-based deep learning models, and other deep learning-based and machine learning-based music generation techniques. The systematic view of categorized AI-based music generation systems is portrayed in Fig. 2.

4 Literature Review

The review of literature on baseline approaches utilized for AI-based music generation systems is briefly demonstrated in this section. The description and survey of categorized LSTM-based, VAE-based, GAN-based deep learning models, other deep learning-based, and machine learning-based music generation techniques are elaborated as follows.

Fig. 2 Systematic view of categorized AI-based music generation systems



4.1 LSTM-Based Music Generation Model

LSTM is highly used in the field of AI and deep learning for learning, processing, and classifying sequential data predominantly. The LSTM effectively learns the long-term dependency among the data time steps of data and resolves the vanishing gradient issues available in traditional networks. The review performed on existing music generation schemes based on LSTM is demonstrated as follows.

4.1.1 Review of LSTM-Based Music Generation Model

Li and Li [3] designed an LSTM-based music generation and classification algorithmic approach for creating and validating the utilization of music for treatment in the medical field. In this approach, a multi-voice musing generation system was utilized for the extraction of features from music data. The LSTM-based music generation approach didn’t require many labels for training while generating and classifying melodies. However, this approach failed to utilize more precise feature extraction techniques to increase the classification accuracy during music generation.

Chen et al. [13] developed the LSTM model for tinnitus music therapy by generating particular music from music segments. Here, the original music was segmented, and the essential pitch was determined by taking into account of predominant melody pitch sequence for music generation. It effectively reduced the stress and emotional consequences without any time limitation in tinnitus patients. Meanwhile, this model only focused on pitch development and failed to consider the development of volume and rhythm to provide better results for tinnitus music therapy.

Ranjan et al. [14] established a bi-directional LSTM technique for similar types of music generation using MIDI data. The music was generated using a seq2seq attention **mechanism** that utilizes an encoder and decoder layer for generation. Here, the

bi-directional LSTM (Bi-LSTM) layer acts as the encoder layer and the regular LSTM acts as the decoder layer for decoding the output sequence. This approach effectively avoided overfitting issues that occurred during music generation and ensured the variety of generated music. However, this approach was not successful in utilizing more and large comprehensive data for music generation.

Rafraf [9] devised an LSTM scheme for automatic music generation by encoding music information based on movement other than considering absolute pitch. In this model, all the chords generated melodies at every timestamp and were encoded in a series of intervals. Also, the melodies were encoded in a series of pitches. It effectively reduced the size of training data and removed the necessity of data augmentation without any loss of information. However, this model failed to ensure the stability of music with an acceptable range of pitches to provide solutions to tonality and register problems.

4.2 VAE-Based Music Generation Model

The VAE is a general form of generative AI algorithm that uses deep learning to remove noises, determine **anomalies**, and generate new content. It is a part of the variational Bayesian method and probabilistic graphical model family. The VAE acts as a probabilistic generative model that possesses neural networks at any layer of their structure. The review of baseline techniques based on VAE for music generation is demonstrated below.

4.2.1 Review of VAE-Based Music Generation Model

Grekow [15] introduced the Conditional VAE (CVAE) model for the generation of polyphonic music contents with specific emotions. Here, the four quadrants of Russel's approach, such as relaxed, sad, angry, and happy were utilized for generating music content. Moreover, convolutional layers were employed to learn the visual representation of symbolic music for the creation of emotional music. It significantly generated music in sequences by validating the quality of music under less computational time. However, it failed to increase the generation results by utilizing large training samples related to the costly annotation process.

Grekow and Dimitrova-Grekow [16] developed the CVAE approach for the generation of monophonic music with specified emotions. In this model, the sequence processing was performed using RNN. It effectively recognized the emotion-affecting patterns available in training data and then applied it to the generated samples. This model attained less computational complexity issues while generating music sequences for provided emotions. Meanwhile, it failed to consider extra parameters, like musical structure, melody, and rhythm for the generation of music structure. Also, the length of monophonic pieces was not considered, and the emotional model was not adopted for music generation.

Guo et al. [17] designed the VAE model to effectively handle the tonal tension of the created music. Various tension feature vectors were employed in this model to add the scaled feature vectors in the latent space for providing various transformations of output music. Here, the music with high tension was generated by adding a scaled tension feature vector to the modified space variables in terms of cloud diameter and tensile strain. It attained less bass pitch loss during music generation, but it failed to utilize interactive tools for drawing tension to compose music.

4.3 GAN-Based Music Generation Model

GAN is a kind of AI model, which significantly learns any distributed data distribution, like audio, text, or images. In general, GAN utilizes an adversarial learning algorithm for the conversion of random noise samples into realistic data. The prevailing music generation models based on GAN are given below.

4.3.1 Review of GAN-Based Music Generation Model

Adhikari [18] established a GAN-powered harmonic approach to examine the validity of emotion for AI-based music generation. The GAN infused emotional awareness into generated compositions during harmonization for musical composition. It effectively maintained the integrity and emotional quality of the music in the music generation. However, it failed to consider dynamic emotion adaptation in AI-generated music for the adoption of music in a real-time context. Also, the synthesis of textural narratives or emotionally aligned visuals was not performed to obtain emotionally attained resonant multimedia experience during music generation.

Yang et al. [19] introduced the MidiNet model for music generation based on the specifications of input. The MidiNet was a CNN-GAN-based approach utilized for generating music by conditionally exploiting the prior knowledge of music. This approach was exceptionally reliable in generating melodies and was suitable for real-time applications in music generation. However, it failed to consider music information retrieval techniques, like emotion recognition and genre recognition during music generation.

Li and Sung [11] designed an inception approach, INCO-GAN to automatically generate music of complete-variable length. The quality of the generated music was increased by utilizing richer features and the whole portion to be generated was decided by using GAN. Here, the context information was considered by adding time distributed layer to the Convolutional Neural Network (CNN) for sharing the weights of each timestamp. This approach was generalized and generated diverse music without any loss. Meanwhile, the pitch of created music changes frequently resulting in characteristic changes in generated music.

Min et al. [20] devised the GAN model for music generation by learning the descending direction of targeted sequences and real music by updating the parameter

over time. It increased the output and input structures of the generator and discriminator to provide solutions to model collapse and non-differentiability gradient issues that were commonly found in discriminators. It effectively increased the prediction results by reducing the dictionary size and resolved multi-classification issues with label dimension during music generation. However, this approach failed to utilize algorithmic approaches for music generation.

4.4 Other Deep Learning-Based Music Generation Model

Over the past two decades, deep learning approaches have been widely utilized for music generation due to their effectiveness in learning the audio signals and generating precise prediction results. Autonomous systems, such as AI utilize deep learning approaches to produce melodies from input audio. The deep learning approaches other than LSTM, VAE and GAN utilized for music generation are elaborated below.

4.4.1 Review of Other Deep Learning-Based Music Generation Model

Sun and Sundarasekar [6] developed the AI-Big Data (AI-BD) model for the recognition of various forms of music to generate music. Here, the brain activities were simulated by using Neural Networks (NN) and the features or styles were effectively learned by using complex algorithms, repeated processing, and big data approaches. This approach attained high prediction accuracy under less computational complexity issues and low computational cost for music generation. Meanwhile, it failed to consider time constraints for validating an AI-based music generation system.

Zhang et al. [1] designed the RNN model to provide music perception education for treating autistic children. The Electroencephalogram (EEG) was utilized by the model to gather the effect of perception and diagnosis disease of in autistic children. The RNN model was utilized to determine the interactive communication performed by human interactive robots. It effectively identified the psychological state of autistic children by providing solutions to EEG issues via music therapy. However, the model required advanced optimization algorithmic approaches to further enhance the classification effect.

Zheng et al. [21] introduced the EmotionBox technique for composing music based on given emotions under various psychological conditions. This approach did not require a labeled music database with emotions for generating various emotional music. Here, symbolism and music with specific emotions were generated by supervising the deep neural network. Moreover, the various auditory emotions and perceptions were evoked by this model by using pitch histogram and note density. It significantly generated non-repetitive music with pre-defined emotions that were utilized for music therapy applications. However, this model failed to determine unique values for some music therapy applications.

Miyamoto et al. [22] established an emotion induction system for music generation and emotion prediction using EEG. Here, the emotions were induced and predicted by real-time EEG signals using CNN, and the music was generated by a music generator utilizing predicted emotions. This technique effectively generated music by predicting emotions under less prediction time. It failed to update the music parameters by arbitrarily determining the method for music generation.

Majidi and Toroghi [23] proposed a Multi-Objective Genetic Algorithm (MO-GA) for the generation of polyphonic music pieces by taking account of listener satisfaction and grammar. A human scoring model was provided by ordinary listeners and music experts to generate music pieces based on music theory. The Bi-LSTM was used to analyze the scores of music listeners as well as experts. This approach accelerated the convergence while automatically generating the polyphonic music, but it failed to consider desirable music samples to obtain high scores for the audience.

Dai et al. [24] developed Music Frameworks that utilized deep neural networks for the generation of full-guided melodies by using long-term repetitive structure with rhythm, contour, melodic, and chord constraints. This is a multi-step generative process and a hierarchical music structure representation, which utilizes encoded useful features for music generation. It effectively resolved overfitting issues while generating melody without utilizing any musical representations. Meanwhile, it failed to utilize intelligent ways to validate the music frameworks that support polyphonic generation, harmonic generation, and richer musical vocabulary.

Hu et al. [25] designed a User-Preference Transformer (UP-Transformer) for user-preference music transfer. The user's favorite music was extracted using two steps and pattern similarity was employed to evaluate the similarity among the two music pieces. Moreover, the music was transferred by fixing and altering the input music symbols. This model attained less training time while personalizing input music for various users. However, it failed to learn event information due to the utilization of limited training samples.

Wu et al. [26] devised PopMNet, a melody structure-based model for the generation of melodies in pop music. The PopMNet was developed by incorporating RNN-based Melody Generation Net (MGN) and Convolutional Neural Network (CNN)-based Structure Generation Net (SGN). In this model, the structure of melody and their structure progressions were generated in the latter. This approach attained high-quality melodies with accurate structures by resolving complex relations among the melody segments. Meanwhile, it required advanced models for musical melody and structure generation.

4.5 Machine Learning-Based Music Generation Model

A subfield of AI, called machine learning is utilized to solve complex tasks in the same way that humans solve issues. Sequential machine learning techniques are utilized to learn music sequences for the generation of music. The prevailing machine learning models utilized for music generation are delineated as follows.

4.5.1 Review of Machine Learning-Based Music Generation Model

Williams et al. [4] established a Supervised learning algorithm for the automatic generation of musical materials. The sound was synthesized in this model by a nonlinear time series of musical features by utilizing a perceptually informed musical feature similarity model. Also, the generated music undergoes subsequent psychometric validation in nonlinear soundtracking for the generation of functional music. This approach created a direct correlation among the reported negativity by reading the emotions. However, it was not successful in increasing the musical ability of the patients to express emotional states.

Siddiqui et al. [5] developed Body Area Networks (BAN) and Artificial Neural Networks (ANN) techniques to improve the safety and health of patients via music therapy. In this model, the type and duration of played music as well as its effects on the patient's body parameters during music therapy was exploited using ANN. The mood of the patient was predicted by ANN through cross-entropy and binary loss functions, Rectified Linear Unit (ReLU) activation function for the suggested type of music. It ensured the timely delivery of music therapy automatically and it didn't require any dedicated music or resources to provide music therapy services for a large population. However, it failed to include remote health management and monitoring facilities by collecting real-time patient data.

Dai et al. [27] devised a Statistical machine learning model to collect and imitate the bass style, chord, melody, and structure for generating music with high musicality. It effectively created high-quality music, which was similar to the given input song. The Statistical machine learning model created complete customized songs of different lengths and possesses hierarchical and logical music structure for the generation of music bars. This approach effectively eliminated irrelevant factors from the songs while generating music by controlling the listening environment. However, it failed to consider human suggestions for the generation of stylistic melodies automatically.

Modern [2] designed a multi-class neural network for the classification of music-based emotions into different categories. The solfeggio frequencies were considered by the model to identify the emotional and musical characteristics of persons. It effectively performed audio signal processing by considering the correlation among music features and their therapeutical effects. However, it failed to consider edge AI techniques for processing data to perform multiple options in music generation tasks.

Jin et al. [28] introduced the Music Generation Framework for Concerts in Meta-verse (MetaMGC) approach for music generation. Here, a virtual orchestra for meta-verse concerts was considered for music generation, and the virtual sound field was reconstructed using a digital audio twin part. Moreover, the natural and accurate binaural audio was synthesized using an end-to-end neural synthesis process. This approach possessed a good transient response and was highly effective in solving issues due to cross-entropy loss functions that occur during music generation. However, this model failed to consider emotional expression factors in the digital twin system for music generation.

Rahman et al. [29] designed a neural network model to provide efficient music therapy for mental healthcare. It enhanced the accuracy of classification models by identifying useful information, which also visualized the physiological signals for recording signals from video using Gingerbread Animation. It effectively determined various genres of music by resolving vanishing gradient issues and reducing computational time by utilizing a small subset of features. However, this approach failed to include the extraction of more frequency domain features for music generation.

5 Research Gaps and Issues

The limitations confronted by existing AI-based music generation systems are delineated as follows;

The challenges faced by prevailing LSTM-based music generation techniques are explicated as follows: The LSTM-based music generation and classification algorithmic approach used in [3] failed to integrate autoencoders and convolutional layers to develop a new feature extraction approach for increasing the performance. The LSTM model used in [13] failed to provide a high application of music therapy by integrating the volume, rhythm, and pitch of music for tinnitus patients. The bi-directional LSTM approach used in [14] was not successful in utilizing a large database for music generation and was not suitable for generating music with specific genres, styles, and moods. The LSTM model employed in [9] failed to generate music with excessive modulations, which also failed to modulate exceedingly to various keys for music generation.

The drawbacks encountered by baseline VAE-based music generation approaches are elaborated below: The Convolutional CVAE approach used in [15] was not successful in understanding the encoded representation of polyphonic music by considering other connection structure variants of convolutional layers. Moreover, this approach failed to create longer polyphonic music sequences during music generation for utilization in practical applications. The CVAE approach employed in [16] failed to support music composers in identifying the music sequences presented in computer games based on the background music and emotional contents in the shop. It also failed to enhance or change the emotional condition of patients by potentially applying the generated melodies to music therapy. The VAE model used in [17] failed to consider music tension factors for music generation and was not suitable for examining the coherence and long-term structure of the created music.

The different issues confronted by the GAN-based music generation model are illustrated below: The GAN-powered harmonic approach utilized in [18] was not successful in creating the immersive multimedia experience and it failed to serve as a creative collaborator, identify generative musical variations, and provide compositional suggestions during music composition. The MidiNet approach used in [19] was not successful in generating multi-track music using large and richer MIDI data by including pauses and velocity to train the model. Also, the idea of reinforcement learning was not considered for the creation of various types of music based on their

specifications. The GAN model used in [20] failed to pre-process the music database to enhance the performance of music generation.

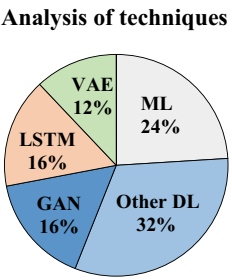
The challenges of other deep learning-based music generation schemes are briefly demonstrated as follows: The AI-BD model used in [6] failed to consider reorchestrated and rearrangement audio rather than static recordings for music generation. The RNN model used in [1] was not successful in reflecting the real-time applications for applying music perception. It failed to validate the effectiveness and applicability of the model under music perception education using optimization algorithmic techniques and human–computer interaction mode. The EmotionBox used in [21] was not successful in identifying appropriate features to transfer the valence of music for music generation based on psychological conditions. The emotion induction system employed in [22] failed to identify the emotional changes while listening to the generated music by controlling the inputs to the music generator. The Music Frameworks model used in [24] was not successful in generating music by imitating and capturing the musical styles.

The issues encountered by machine learning-based music generation techniques taken for the analysis are given as follows: The Supervised learning algorithm used in [4] failed to consider advanced methodologies for validating the success and failure state for music generation. The ANN techniques used in [5] were not successful in utilizing high-level prototypes to transmit real-time data for providing music therapy services to patients requiring physical sessions. The Statistical machine learning model used in [27] failed to consider deep learning techniques for the generation of stylistic melodies. The multi-class neural network employed in [2] was not effective in selecting suitable moods to increase the performance of present applications. The MetaMGC technique used in [28] failed to perform realistic emotion-rich live concert scenarios by considering a meta-universe concert intelligent music generation framework for music generation. In [29], a neural networks model was used for music generation, which failed to increase and redefine the overlapped colors and increase the edges to more blurred and more natural color mixtures.

6 Results and Discussion

The analysis performed on different AI-based music generation systems, namely LSTM-based, VAE-based, GAN-based deep learning models, and other deep learning-based and machine learning-based music generation techniques is elaborated in this section. Moreover, the validation is performed by taking into account of year of publication, categorized technique, tools utilized, and performance metrics.

Fig. 3 Validation using categorized techniques



6.1 Analysis Based on Categorized Techniques

Figure 3 portrays the validation using categorized approaches, where the generative AI-based music generation approaches are categorized into LSTM-based, VAE-based, GAN-based deep learning models, deep learning-based, and machine learning-based music generation techniques. Among all these techniques, the other deep learning-based approaches are highly utilized for automatic music generation in most of the research papers. The other deep learning-based music generation systems are utilized by 32% of research papers and about 24% of research papers utilize machine learning-based music generation systems. Moreover, nearly 16% of research papers utilize GAN and LSTM-based music generation approaches and VAE-based music generation approaches are utilized by 12% of research papers.

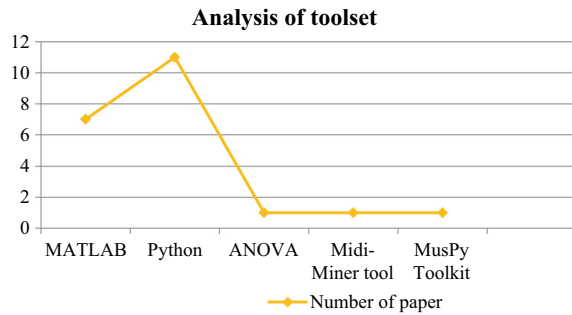
6.2 Analysis Based on Published Year

The validation of prevailing research papers based on AI-based music generation systems regarding published years is delineated in Table 1. It is estimated that about nine research papers based on AI-based music generation systems were published in the year 2022. Moreover, in 2020 about six research papers related to AI-based music generation systems were published in reputed journals. Also, about five research articles were published in the year 2023, and four articles based on music generation systems were published in 2021.

Table 1 Validation using published year

Published year	Published papers
2017	[19]
2020	[4, 5, 13, 17, 26, 29]
2021	[11, 16, 24, 27]
2022	[1, 3, 9, 14, 20, 21, 22, 25, 28]
2023	[2, 6, 15, 23]

Fig. 4 Validation using employed tools



6.3 Analysis Based on Tools Utilized

The validation of baseline techniques utilized for AI-based music generation systems based on employed tools is shown in Fig. 4. MATLAB, Python, Analysis of Variance (ANOVA), Midi-Miner tool, and MusPy Toolkit are the different toolsets used by existing music generation systems. The analysis proved that about 11 research articles utilized Python tools for the implementation of music generation systems and about seven articles utilized MATLAB tools for the implementation of the research work. Furthermore, the ANOVA, Midi-Miner tool, and MusPy Toolkit have been used in one article each.

6.4 Analysis Based on Journal Publisher

The analysis of prevailing music generative systems using journal publishers is portrayed in Fig. 5. The publishers, such as Frontiers, Elsevier, IEEE, MDPI, Springer, ACM, and Taylor and Francis (T&F) are considered to collect the research papers for the analysis. Out of 25 research papers, about six research papers are taken from Frontiers, and five research papers are taken from Elsevier and IEEE each. Moreover, four papers based on AI-based music generation systems are taken from MDPI publishers and about two articles are considered from Springer publishers. Similarly, about one paper based on music generation systems based on generative AI is taken from ACM and T&F publishers.

6.5 Analysis Based on Performance Metrics

Figure 6 displays the review of prevailing techniques deployed for AI-based music generation systems using performance metrics. The performance metrics, such as accuracy, F1-score, similarity score, loss, Time, Root Mean Square Error (RMSE), R2, Pitch range, user preferences, and emotional quality are utilized to analyze

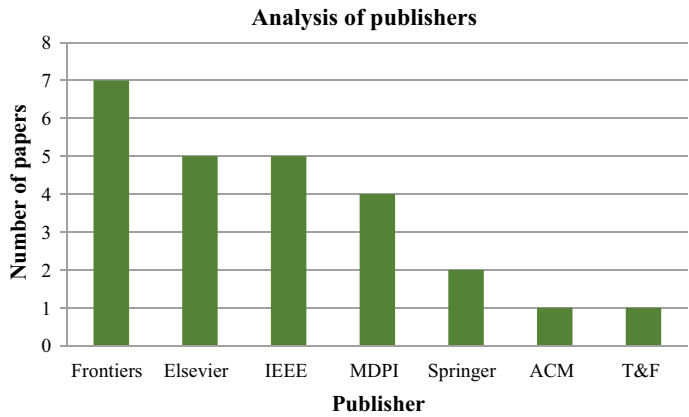
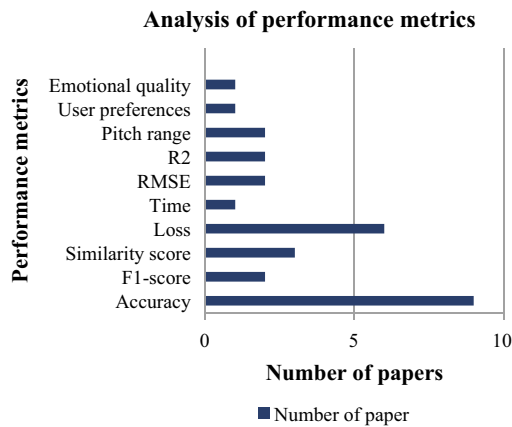


Fig. 5 Validation using journal publisher

Fig. 6 Validation using performance metrics



the performance of prevailing approaches. Here, about nine research papers utilize accuracy as an evaluation metric, and the loss function is used by six research articles.

7 Conclusion

This study briefly illustrates the analysis of baseline generative AI-based music generation systems utilized in music therapy. The analysis is performed by considering 25 research papers based on generative AI-based music generation systems. All the articles are selected from various web sources, like IEEE, Elsevier, Springer, MDPI, T&F, ACM, and so on. During analysis, the manuscripts based on generative AI are categorized into LSTM-based, VAE-based, GAN-based deep learning models,

other deep learning, and machine learning-based techniques. Then, the various limitations and issues encountered by the categorized manuscripts are enumerated. The advantages as well as disadvantages of prevailing generative AI-based music generation systems are also analyzed effectively. Moreover, the validation is carried out based on the utilized database, employed tools, published year, evaluation parameters, and utilized techniques. Moreover, deep learning is the most commonly used AI approach. The survey gives a clear picture of how generative AI can be incorporated into music therapy and the problems encountered by the various techniques. Further, identifying the challenges can assist in devising techniques that can effectively assist in AI-based therapeutic music generation with minimal computational effort and high accuracy.

References

1. Y. Zhang, C. Zhang, M. Qi, The use of deep learning-based gesture interactive robot in the treatment of autistic children under music perception education. *Front. Psychol.* **13**, 762701 (2022)
2. H.A. Modran, T. Chamunorwa, D. Ursuțiu, C. Samoilă, H. Hedeșiu, Using deep learning to recognize therapeutic effects of music based on emotions. *Sensors* **23**(2), 986 (2023)
3. Y. Li, X. Li, Long short-term memory-based music analysis system for music therapy. *Front. Psychol.* **13**, 928048 (2022)
4. D. Williams, V.J. Hodge, C.Y. Wu, On the use of AI for generation of functional music to improve mental health. *Front. Artif. Intell.* **3**, 497864 (2020)
5. S. Siddiqui, R. Nesbitt, M.Z. Shakir, A.A. Khan, A.A. Khan, K.K. Khan, N. Ramzan, Artificial neural network (ANN) enabled internet of things (IoT) architecture for music therapy. *Electronics* **9**(12), 2019 (2020)
6. W. Sun, R. Sundarasekar, Research on pattern recognition of different music types in the context of AI with the help of multimedia information processing. *ACM transactions on Asian and low-resource language information processing* (2022)
7. Y.J. Hong, J. Han, H. Ryu, The effects of synthesizing music using ai for preoperative management of patients' anxiety. *Appl. Sci.* **12**(16), 8089 (2022)
8. D. Zhang Study on the effect of decompression music on anxiety patients based on emotion classification algorithm. *Informatica* **47**(10), (2024)
9. H. Rafrat, Differential music: automated music generation using LSTM networks with representation based on melodic and harmonic intervals. University of Florida (2022)
10. J. Sun, J. Yang, G. Zhou, Y. Jin, J. Gong, Understanding human-AI collaboration in music therapy through co-design with therapists. *arXiv preprint [arXiv:2402.14503](https://arxiv.org/abs/2402.14503)* (2024)
11. S. Li, Y. Sung, INCO-GAN: variable-length music generation method based on inception model-based conditional GAN. *Mathematics* **9**(4), 387 (2021)
12. L. Moysis, L. A. Iliadis, S. P. Sotiroidis, A. D. Boursianis, M. S. Papadopoulou, K. I. D. Kokkinidis, C. Volos, P. Sarigiannidis, S. Nikolaidis, S. K. Goudos, Music deep learning: deep learning methods for music signal processing-a review of the state-of-the-art. *IEEE Access* (2023)
13. J. Chen, F. Pan, P. Zhong, T. He, L. Qi, J. Lu, P. He, Y. Zheng, An automatic method to develop music with music segment and long short term memory for tinnitus music therapy. *IEEE Access.* **8**, 141860–141871 (2020)
14. A. Ranjan, V. N. J. Behera, M. Reza, Using a bi-directional LSTM model with attention mechanism trained on midi data for generating unique music. *arXiv preprint [arXiv:2011.00773](https://arxiv.org/abs/2011.00773)* (2020)

15. J. Grekow, Generating polyphonic symbolic emotional music in the style of bach using convolutional conditional variational autoencoder. *IEEE Access* (2023)
16. J. Grekow, T. Dimitrova-Grekow, Monophonic music generation with a given emotion using conditional variational autoencoder. *IEEE Access*. **9**, 129088–129101 (2021)
17. R. Guo, I. Simpson, T. Magnusson, C. Kiefer, D. Herremans, A variational autoencoder for music generation controlled by tonal tension. *arXiv preprint [arXiv:2010.06230](https://arxiv.org/abs/2010.06230)* (2020)
18. T. Adhikari, MeloHarmony: exploring emotion in crafting AI-generated music with generative adversarial network powered harmony. Available at SSRN 4575257 (2023)
19. L. C. Yang, S. Y. Chou, Y. H. Yang, MidiNet: a convolutional generative adversarial network for symbolic-domain music generation. *arXiv preprint [arXiv:1703.10847](https://arxiv.org/abs/1703.10847)* (2017)
20. J. Min, Z. Liu, L. Wang, D. Li, M. Zhang, Y. Huang, Music generation system for adversarial training based on deep learning. *Processes* **10**(12), 2515 (2022)
21. K. Zheng, C. Zheng, J. Sang, J. Wang, EmotionBox: a music-element-driven emotional music generation system based on music psychology. *Front. Psychol.* **13**, 841926 (2022)
22. K. Miyamoto, H. Tanaka, S. Nakamura, Online EEG-based emotion prediction and music generation for inducing affective states. *IEICE Trans. Inf. Syst.* **105**(5), 1050–1063 (2022)
23. M. Majidi, R.M. Toroghi, A combination of multi-objective genetic algorithm and deep learning for music harmony generation. *Multimed. Tools Appl.* **82**(2), 2419–2435 (2023)
24. S. Dai, Z. Jin, C. Gomes, R. B. Dannenberg, Controllable deep melody generation via hierarchical music structure representation. *arXiv preprint [arXiv:2109.00663](https://arxiv.org/abs/2109.00663)* (2021)
25. Z. Hu, Y. Liu, G. Chen, Y. Liu, Can machines generate personalized music a hybrid favorite-aware method for user preference music transfer. *IEEE Trans. Multimed.* (2022)
26. J. Wu, X. Liu, X. Hu, J. Zhu, PopMNet: generating structured pop music melodies using neural networks. *Artif. Intell.* **286**, 103303 (2020)
27. S. Dai, X. Ma, Y. Wang, R.B. Dannenberg, Personalised popular music generation using imitation and structure. *J. New Music. Res.* **51**(1), 69–85 (2022)
28. C. Jin, F. Wu, J. Wang, Y. Liu, Z. Guan, Z. Han, MetaMGC: a music generation framework for concerts in metaverse. *EURASIP J. Audio, Speech, Music. Process.* **1**(31) (2022)
29. J.S. Rahman, T. Gedeon, S. Caldwell, R. Jones, Z. Jin, Towards effective music therapy for mental health care using machine learning tools: human affective reasoning and music genres. *J. Artif. Intell. Soft Comput. Res.* **11**(1), 5–20 (2021)