IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received 24 December 2022, accepted 3 February 2023, date of publication 13 February 2023, date of current version 23 February 2023.

Digital Object Identifier 10.1109/ACCESS.2023.3244620

SURVEY

Music Deep Learning: Deep Learning Methods for Music Signal Processing—A Review of the State-of-the-Art

LAZAROS MOYSIS[1,2], LAZAROS ALEXIOS ILIADIS[3], (Graduate Student Member, IEEE), SOTIRIOS P. SOTIROUDIS[3], ACHILLES D. BOURSIANIS[3], (Member, IEEE), MARIA S. PAPADOPOULOU[3,6], (Member, IEEE), KONSTANTINOS-IRAKLIS D. KOKKINIDIS[4], CHRISTOS VOLOS[1], PANAGIOTIS SARIGIANNIDIS[5], (Member, IEEE), SPIRIDON NIKOLAIDIS[3], (Senior Member, IEEE), AND SOTIRIOS K. GOUDOS[3], (Senior Member, IEEE)

[1]Laboratory of Nonlinear Systems-Circuits and Complexity, School of Physics, Aristotle University of Thessaloniki, 54124 Thessaloniki, Greece
[2]Department of Mechanical Engineering, University of Western Macedonia, 50100 Kozani, Greece
[3]ELEDIA@AUTH, School of Physics, Aristotle University of Thessaloniki, 54124 Thessaloniki, Greece
[4]Department of Applied Informatics, University of Macedonia, 54636 Thessaloniki, Greece
[5]Department of Electrical and Computer Engineering, University of Western Macedonia, 50131 Kozani, Greece
[6]Department of Information and Electronic Engineering, International Hellenic University, 57400 Sindos, Greece

Corresponding authors: Lazaros Moysis (lmousis@physics.auth.gr) and Sotirios K. Goudos (sgoudo@physics.auth.gr)

This research was carried out as part of the project «Recognition and direct characterization of cultural items for the education and promotion of Byzantine Music using artificial intelligence» (Project code: KMP6-0078938) under the framework of the Action "Investment Plans of Innovation" of the Operational Program "Central Macedonia 2014 2020", that is co-funded by the European Regional Development Fund and Greece.

ABSTRACT The discipline of Deep Learning has been recognized for its strong computational tools, which have been extensively used in data and signal processing, with innumerable promising results. Among the many commercial applications of Deep Learning, Music Signal Processing has received an increasing amount of attention over the last decade. This work reviews the most recent developments of Deep Learning in Music signal processing. Two main applications that are discussed are Music Information Retrieval, which spans a plethora of applications, and Music Generation, which can fit a range of musical styles. After a review of both topics, several emerging directions are identified for future research.

INDEX TERMS Deep learning, music signal processing, music information retrieval, music generation, neural networks, machine learning.

I. INTRODUCTION

A. DEEP LEARNING IN MUSIC SIGNAL PROCESSING

Deep Learning (DL) [1], a sub-field of Machine Learning (ML), has been established as a strong computational toolbox, with applications in numerous tasks, like feature extraction, classification, and pattern recognition. Such functionalities enable the extraction of meaningful information from raw data, and thus find applications in a wide range of disciplines, including computer vision (CV) [2], natural language processing (NLP) [3], bioinformatics [4], medical diagnosis [5], speech recognition [6], image processing (IP) [7], system identification [8], recommendation systems [9], and more [10].

A research field where DL has emerged as a valuable tool over the last decade is that of audio signal processing (ASP) [11] and music signal processing (MSP) [12]. Music is a well-known art form that is a big part of the most fun and educational human activities. As a result, the music industry includes a wide range of organizations and consumers. The application of DL tools in MSP has led to a collection of successful commercial applications, the

The associate editor coordinating the review of this manuscript and approving it for publication was Pasquale De Meo.

most famous of which is Music Recommendation Systems (MRS) [13]. As shown in Fig. 1, the number of publications indexed in Scopus under the keywords "deep," "learning," and "music" demonstrate the applicability of DL in music processing. From 2014 to 2021, there are 638 publications, a sharp increase each year. This shows that scientists are becoming more interested in this field. The diversity of the field is also made apparent when looking at the subject area categorization of these works, with 567 being listed in Computer Science, 296 in Engineering, 136 in Mathematics, 74 in Physics and Astronomy, 63 in Decision Sciences, 51 in Arts and Humanities, and the rest covering disciplines such as Materials Science, Medicine, Social Sciences, Energy and more.

The broad field of DL in music-related applications could be termed Music Deep Learning (MDL) and can be divided into two categories, Music Information Retrieval (MIR) [11] and Music Generation (MG) [14]. MIR refers to the extraction of characterizing information from music data. Such information can then be exploited for a wide range of applications, such as genre classification [15], [16], music recommendation [17], [18], music source separation [19], singing voice detection [20], instrument recognition [21], music emotion recognition [22] and transcription [23]. All of the above applications aid in the digital preservation of music, by constructing and managing song databases, as well as the study of different music genres.

MG, under the framework of DL, broadly refers to the automatic generation of music content. This task is performed by first extracting valuable information from music databases using MIR techniques, and then building DL architectures to generate original music content. This has several commercial applications, like movie and game score generation. The automatic generation of music content has spun discussions on whether this new way to create art will eventually replace musicians. However, the more realistic projection for the future is that MG can serve as a valuable tool to musicians and educators alike, to explore new approaches to composition and teaching [24].

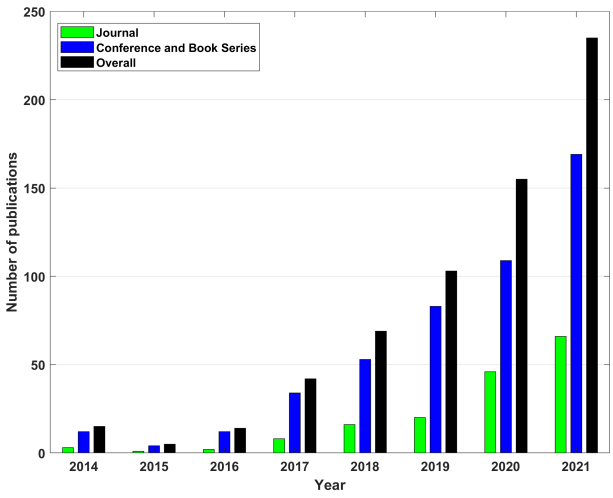

FIGURE 1. Number of publications indexed under the common keywords 'deep', 'learning', and 'music' in Scopus.

B. RELATED SURVEYS

There have been some reviews of the results so far in MDL. In [11] a review of the (at the time) current DL techniques for ASP is provided. Three types of audio are considered, speech, music, and environmental sounds, with applications like audio recognition, synthesis, and transformation. Several reviews have also considered specific applications of MDL. In [21], a tutorial on MIR is provided that is especially useful to newcomers in the field [13], [25] reviews MRS. Music Genre Classification (MGC) is reviewed in [26]. Drum transcription is reviewed in [27], focusing on non-negative matrix factorization and recurrent neural network architectures. In [28], a review of the audio signal representations for use with CNNs is given. A review of DL for speech recognition is available in [6], though the focus is not on music signals. For singing information processing, [29] reviews several aspects, like singing skill evaluation, singing voice synthesis, singing voice separation, lyric synchronization, and transcription. Specifically for singing voice detection, the review in [20] investigates the traditional and deep learning techniques available. DL for music emotion recognition is reviewed in [22].

For MG, the extensive survey in [14] offers an in-depth analysis, covering five key aspects of MG, the objective, representation, architecture, challenge, and strategy. The work [30] provides a systematic review of AI techniques in MG with valuable information regarding publications, citations, geographical distribution, and many more. A review of the composition tasks for various music generation levels is provided in [31]. Finally, [32] talks about the challenges and limitations of MG. These include, for example, the designer's creative limitations, the lack of structure, the extent of control the designer has over the generated music features, and the lack of direct user interaction. Moreover, it argues on how to address these issues.

C. MOTIVATION

From the above, it is clear that different aspects of DL in MSP have been surveyed, with many reviews being dedicated to focused topics, thus providing highly detailed insights into it. In this work, DL for both MIR and MG is discussed, which to the authors' knowledge are discussed for the first time together. The purpose of this work is to provide a more spherical overview of the current research in this field, which could serve as a guide for identifying new research trends.

For that matter, after a review of recent results on both MIR and MG, a section is dedicated to identifying future directions on MDL. Specifically, four research directions are identified, all of which can yield fruitful results in MDL. An earlier version of this study was presented in [33]. The current work

extends [33] by expanding upon the literature review, and the discussion on future topics of interest.

The main contributions of this work are summarized as follows:

1) To complement previous surveys, emphasis is given to works published in 2020 or later. In this way, the evolution of MDL into a mature field is presented.
2) To the best of the authors' knowledge, this is the first time that MIR techniques and MG processes are reviewed together, highlighting the interconnection between the two research directions.
3) Attention is given to four areas, which are identified as emerging research topics. These areas are hybrid architectures, DL in traditional music genres, MDL in medical applications, and DL for music generated from dynamic systems.

The rest of the work is outlined as follows: In Section II, the DL methods for Music Information Retrieval are presented. In Section III the field of DL-based Music Generation is discussed. Section IV identifies future research directions. Finally, Section V concludes the work. For a list of Abbreviations, see Appendix A.

II. DL METHODS FOR MIR

In this section, the application of DL for different MIR applications is reviewed. The section is divided into subsections based on the DL architecture used, and the different applications are talked about in each subsection. Table 1, summarizes all the reviewed works in MIR, organized by architecture.

First, a short description is provided of the various applications of MIR:

1) Music Recommendation Systems (MRS): MRS is the most fundamental application of MDL. Its goal is to successfully recommend new music tracks to users based on their previous listening history. For new users with no prior information, the problem is termed "cold-start MR."
2) Music classification: The goal is to identify the musical genre of a song, which is of fundamental importance in MRS. A more general goal is to identify music from other audio tracks, like speech, natural sounds, etc.
3) Emotion classification and prediction: The goal is to identify the underlying emotions that can be triggered by a song. This is again useful in MRS and music therapy.
4) Instrument/voice identification: The goal is to identify and separate the different instruments used to compose a music track. This also applies to detecting singing voices.

Several objective measures can be used to evaluate MIR architectures. These include accuracy, precision, recall, f1-score, mean absolute and square error, Area Under the Receiver Operating Characteristic Curve (ROC-AUC), and more. In the following, a note is made for each work on the accuracy achieved, or the ROC-AUC score, when provided.

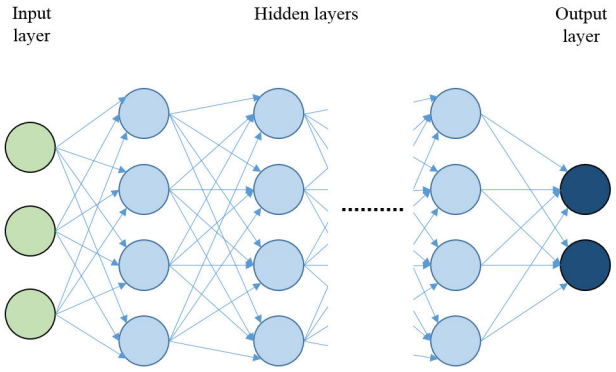

FIGURE 2. Fully connected DNN.

The reader should refer to each work for an extensive presentation of the evaluation analysis. DL-based on the dataset used for training and validation is also provided, for works that used public datasets.

A. FULLY CONNECTED DEEP NEURAL NETWORKS (FCDNN)

FCDNNs refer to the most basic type of deep neural network, where multiple hidden layers are applied and all nodes between consecutive layers are connected, as shown in Fig. 2.

For MR, in [18], an architecture termed HitMusicNet, using an FCDNN was presented, for predicting the popularity of a music recording, using inputs that incorporate text, audio, and meta-data. The authors also construct a database termed the SpotGenTrack Popularity Dataset (SPD), which unifies information from the Spotify and Genius music and lyric databases. Meta-data information that was considered was the number of an artist's followers, an artist's popularity, as well as market availability. The resulting system can reach an 83% precision score. In [34], an FCDNN was used for MR combining content-based and collaborative filtering in its input. The dataset used was the Spotify Recsys Challenge 2018 million playlist dataset [35], reaching an 88% precision score.

For emotion classification in [36], classification was performed on the Music4All dataset [37], using valence, danceability, and energy as features. The classification is binary, with happy/sad classes. The model has a mean accuracy of 98.3%.

B. RECURRENT NEURAL NETWORKS (RNN)

RNNs are a class of neural networks used for processing sequential data [10], and are thus suitable for time series input signals. In contrast to the FCDNN architectures, RNNs are composed of loops or cycles. RNNs also possess an internal memory state, that is utilized to process long sequences. There are many variants of such architectures, including Long Short Term Memory (LSTM), Gated RNN (GRU), bidirectional RNNs, Hopfield networks, etc. [10]. A simple RNN structure is shown in Fig. 3.

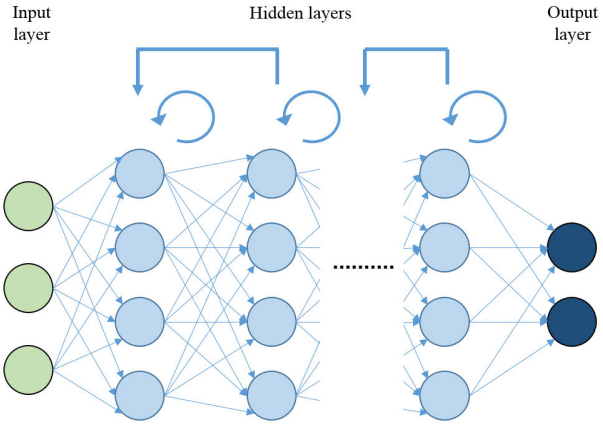

FIGURE 3. Recurrent neural network.

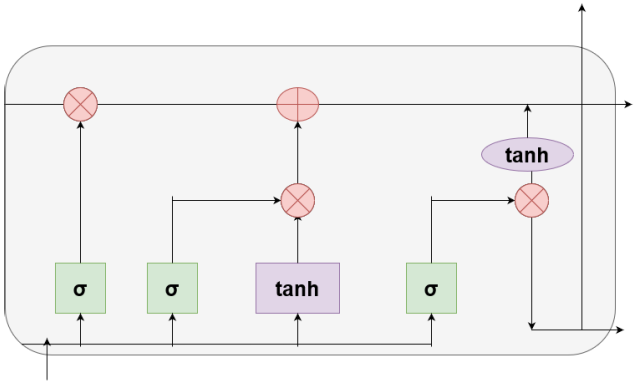

FIGURE 4. LSTM unit cell.

In [38], a tagging system is developed using RNN. A scattering transform is used to extract features from the data. The MagnaTagATune dataset [39] is used. The resulting architecture achieves an average AUC-ROC score of 0.909. In [40], a web application was developed that can take as input any YouTube video song and classify its music genre, using four different architectures. The classification is performed for individual 10-second segments of the input track. The results are visualized in a graph. The music genre samples from the Audioset database [41] are used for training. The supporting website, being highly visual, can offer great help to music composers and students, and also has the potential to be used for user feedback.

For emotion classification tasks, in [42] an RNN is proposed that uses a two-note melody trend as a music feature. Five emotion classes were considered, aggressive, bittersweet, happy, humorous, and passionate. Data files from YouTube were used, and the accuracy is up to 75.4%. In [43], emotion recognition is performed on classes of instruments. Four instrument classes are considered: string, percussion, woodwind, and brass, and four emotion classes are deemed: happy, sad, neutral, and fear. The study shows that the system recognizes more specific instrument-emotional pairings.

RNNs have also been employed for music recommendation. In [44], an RNN architecture was used, and the study showed that song order does not significantly affect the quality of playlist recommendations. The AotM-2011 [45] and 8tracks [46] playlist datasets were used.

For singing voice separation, in [47], a curriculum learning approach was considered, where the learning begins with easy examples and the difficulty is steadily increased. Three different databases were tested: MIR-1K [48], ccMixter [49], and MUSDB18 [50], with the model yielding improved performance with respect to the global normalized source to distortion ratio measure.

A piano harmony automatic arrangement architecture is proposed in [51]. The model performs three tasks, note detection, multibasic frequency estimation, and training. Apart from objective evaluation, the resulting tracks were evaluated by human listeners and were positively received.

For Music Classification, Attention Mechanism (AM) has proven to be a strong technique for improving performance and is adopted in many architectures. An RNN with an attention mechanism is used with MIDI formatted input by [52]. Five classes are considered, classical, country, dance music, folk, and metal. The accuracy achieved is 90.1%.

C. LONG SHORT-TERM MEMORY (LSTM)

Long Short-Term Memory networks (LSTM) [53] constitute a special case of RNNs, which have found applications in MIR. An LSTM unit is shown in Fig. 4.

An LSTM network can be mathematically represented as follows. For a given input vector $\mathbf{u}_k$ at time step k and N_h hidden layers, the activation vector of the forget gate is $\mathbf{f}_k \in (0, 1)^{N_h}$.

$$\mathbf{f}_k = \sigma(W_f \mathbf{u}_k^T + U_f \mathbf{q}_{k-1}^T + \mathbf{b}_f) \tag{1}$$

where W_f and U_F are weight matrices, $q_k \in (0, 1)^{N_h}$ is the vector representing the hidden state, and b_f is the bias vector.

In addition, the activation vectors for the input/update gate $\mathbf{I}_k \in (0, 1)^{N_h}$ and the output $\mathbf{O}_k \in (0, 1)^M$ are represented similarly

$$\mathbf{I}_k = \sigma(W_I \mathbf{u}_k^T + U_I \mathbf{q}_{k-1}^T + \mathbf{b}_I) \tag{2}$$

and

$$\mathbf{O}_k = \sigma(W_O \mathbf{x}_k^T + U_O \mathbf{q}_{k-1}^T + \mathbf{b}_O) \tag{3}$$

where I and O represent input and output, respectively, whereas the rest of the symbols have the same meaning as previous.

An LSTM unit also contains a cell input activation vector denoted by $\mathbf{C}_k \in (-1, 1)^{N_h}$, expressed as

$$\mathbf{C}_k = \sigma(W_C \mathbf{u}_u^T + U_C \mathbf{q}_{k-1}^T + \mathbf{b}_C) \tag{4}$$

Using the following principles, the cell state vector and the hidden state vector are updated by combining the preceding equations

$$\mathbf{S}_k = \mathbf{F}_k \circ \mathbf{S}_{k-1} + \mathbf{I}_k \circ \mathbf{C}_k \tag{5}$$

where $\circ$ is the Hadamard product and $S_0 = 0$ and $q_0 = 0$. Finally,

$$\mathbf{q}_k = O_k \circ \tanh(\mathbf{S}_k) \tag{6}$$

For music Classification, a model is proposed in [54], where the segment features are the statistics of frame features in each segment. The ISMIR database [55] is used, which includes a collection of songs from different genres. The model achieves an accuracy of 89.71%. In [56], a complex architecture is used, combining a Bidirectional Long Short-Term Memory (BLSTM) model with an attention mechanism, paired with a Graphical Convolutional Network. Three datasets are tested, GTZAN [57], ISMIR [55] and MagnaTagATune [58]. An accuracy of 93.51% is achieved.

For emotion prediction, in [59] the valence-arousal (V-A) emotion model was used to represent the dynamic emotion, using a BLSTM network. The dataset used was taken from the Emotion in Music task in MediaEval 2015 [60].

The problem of music source separation was studied using a BLSTM network for instrument detection and identification in [61]. Data augmentation was used during the training to avoid overfitting. To improve performance, the BLSTM network is combined with a feed-forward neural network, which outperforms both individual networks. The SiSEC DSD100 dataset is cited [62].

For MR, an architecture was developed in [63] that analyzes the connection between dance moves and music to recommend tracks. The database used is [64], which includes samples of synchronized dance and music. The dataset contains four classes of dance, waltz, tango, cha-cha, and rumba. The accuracy can reach up to 91.3%.

For singing voice detection in [65], a Long-Term Recurrent Convolutional Network (LRCN) was considered for electronic music. The architecture consists of a voice separation step and a feature extraction step. The CNN layer extracts the audio features, and the LSTM layer uses the CNN output to differentiate between the singing and non-singing parts. The Arcadium [66] and NCS [67] were used as sources to create "Electrobyte," a new copyright-free electronic music dataset. The model was also tested in a pop dataset Jamendo [68], yielding an accuracy score of 0.833 (Electrobyte) and 0.939 (Jamendo). In [69], an LRCN architecture was developed for the vocal separation and temporal smoothing. The CNN layer is again used for feature extraction, and the LSTM learns the time-sequence relationship. The model was tested on five datasets, RWC pop music dataset [70], Jamendo [68], MedleyDB [71], MIR-1K [48], and iKala [72], yielding accuracy as high as 0.992.

D. CONVOLUTIONAL NEURAL NETWORKS (CNN)

CNNs are models that can operate on data with a grid-like structure [10]. This is why they've had success with problems involving IP, CV, NLP, and other technologies [73]. In MIR, CNNs are often used to obtain information from music signals, which are mostly represented as two-dimensional time-frequency data.

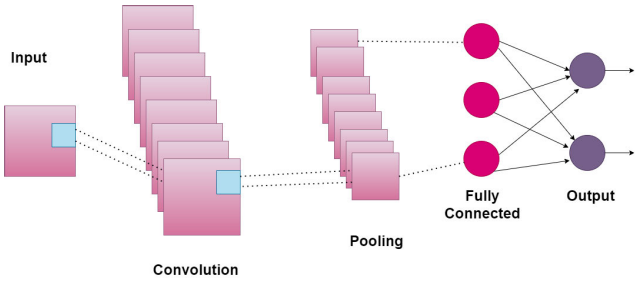

FIGURE 5. General CNN architecture.

A deep CNN model utilizes the convolution operation instead of the general matrix multiplication in at least one of its layers. In addition, the architecture consists of fully connected layers and pooling layers. The purpose of the latter is to reduce in a computationally efficient manner, the size of the incoming data. Compared to a fully connected layer, a convolutional layer is characterized by a neuron's receptive field. This receptive field indicates that every single unit receives input from only a restricted area of the previous layer. As an activation function, most CNNs in the current literature use either the rectified linear unit (ReLU) function or some kind of variant. ReLU is mathematically defined in [10] and can be expressed by

$$g(x) = \max(0, x). \tag{7}$$

A general CNN architecture is depicted in Fig. 5.

In audio Classification, an architecture was developed for spatial audio location and classification between speech and music in [74]. Two different microphone arrangements were considered. The classification can achieve an accuracy of up to 97.9%. Although audio location is not unique to music signals, it can be especially useful in MIR, such as live audio processing. In [75], different CNN architectures are used for the classification of audio videos, using a wide class of labels and a large dataset from YouTube, which is termed YouTube-100M. The ROC-AUC can reach up to 0.926. The Audioset [41] is also considered. In [76], a CNN is used for sound representation learning, using sound from an unlabeled video dataset, gathered from the Flickr website. To improve its performance, the network is trained by moving knowledge from networks that recognize images to networks that recognize sounds.

For music classification, in [15] a CNN is tested on the ISMIR dataset [55], a Latin Music Database (LMD) [77], and an African ethnic database, provided by the Royal Museum of Central-Africa (RMCA) in Belgium [78]. In all cases, the CNN performed either equally well or better than other architectures. In [16], the CNN input consists of eight music features chosen in three music dimensions: dynamics, timbre, and tonality. This outperforms the use of a spectrogram. The GTZAN dataset [57] is used for the experiments, and an accuracy of 91% is reached. In [79], sample-level CNNs were used for auto-tagging using raw waveform data. The term "sample-level" refers to learning representations from

very small waveforms, like 2-3 samples. The MagnaTagATune [39] and Million Song Dataset [80] were considered, and an AUC of over 0.905 can be achieved. In [81], a 3D convolutional denoising autoencoder architecture is built for music classification, using MIDI input format. The model gives out latent representations of the data, which are then used to classify the data with a multi-layer perceptron network. The Lakh MIDI dataset [82], [83] was used for testing, with accuracy surpassing 88% and a ROC-AUC of over 0.86.

CNNs are used for note onset detection in audio recordings in the early work [84] for sound event recognition. The use of a spectrogram as an input to the network instead of the enhanced auto-correlation yields better detection performance. The dataset used is combined from several different sources. In [85], a simple CNN was proposed for event recognition under noise, with only three layers: convolutional, pooling, and softmax. The databases used are the Real Word Computing Partnership (RWCP) Sound Scene Database in Real Acoustic Environments [86], and the NOISEX-92 database [87]. The accuracy can reach up to 99%.

For singing voice separation, in [88], a CNN architecture was successfully developed that utilized pixel-wise classification on the spectrogram image. The model is trained using the Ideal Binary Mask as the target label and cross-entropy as the objective function. The iKala database [72] was used, as well as the DSD100 dataset [62], [89].

For singing voice evaluation, in [90], a one-dimensional CNN is used, that applies fractional processing node theory for training, which reduces the training time. For the experiment, 100 music major students were selected to provide input. Accuracy can be as high as 86.3%.

For musical instrument identification, a CNN with a simple architecture is used for classification into 11 different classes in [91]. The MedleyDB database is used [71], and the accuracy surpasses 82%. In [92], three different weight-sharing strategies for CNNs are considered, temporal kernels, time-frequency kernels, and a linear combination of time-frequency kernels which are one octave apart. MedleyDB is used [71] for training and testing, with hybrid models having the best overall performance. In [93], a Temporal Convolutional Network was trained on a weakly labeled dataset. The OpenMIC-2018 [94] dataset was used for training and testing, and the MUSDB18 [50] for testing. The model slightly outperforms an LSTM model with respect to the ROC-AUC score, which indicates a strong candidate for such problems. Attention-augmented CNNs are used for instrument identification in [95]. When 25% of the filters are assigned to attention, the resulting CNN outperforms the attention-free ones. The datasets used were the London Philharmonic Orchestra Dataset [96], and the University of Iowa Musical Instrument Samples [97]. Judging from the consistently positive outcomes, it only makes sense to assume that in the future, AM-enhanced NNs will be extensively used for MIR. In [98], identification is performed for four instruments: bass, drums, piano, and guitar. The model architecture consists of four identical, independent sub-models, each catering to one instrument. The Slakh dataset is used [99], and the AUC ROC measure reached an average of 0.96, with the drums being easier to identify, and the guitar and piano being the more difficult ones.

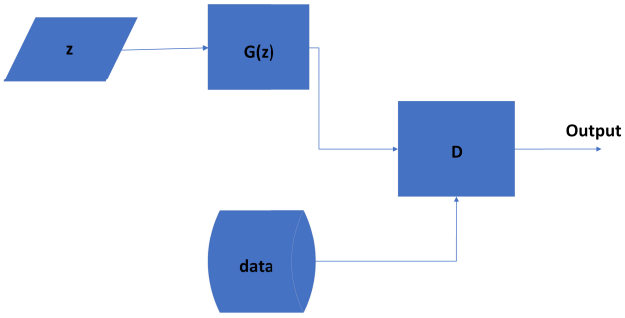

FIGURE 6. GAN architecture.

In [100], a CNN is developed for emotion classification with 18 emotion tags, using time and frequency domain information. The experiments make use of the CAL500 [288] and CAL500exp [101] datasets. In [102], classification is performed specifically for film music, with 9 emotional classes. Each class is also associated with specific colors. The Epidemic Sound Online database [103] was used. The classification is performed using 30-second excerpts of tracks.

In [104], a feature combination CNN architecture for automatic playlist continuation is proposed, with collaborative filtering integrating information from curated playlists as well as song feature vectors. The databases used are Art of the Mix [105] and 8tracks [46]. In [106], distance measuring is used for the classification system, which is then used for the recommendation system. The GTZAN database [57] is used for training, and the Emotify music dataset [107] and Music Audio Benchmark Dataset (MABD) [108] for testing. The designed system can reach a good level of accuracy on the 10-best list. In [109], a CNN architecture is tested using the MIREX database [110], along with the Baidu Music service. The model has a ROC-AUC that can exceed 0.90.

For music transcription, a toolbox termed nnAudio was developed for audio-to-spectrogram conversion using one-dimensional CNNs in [111]. The MusicNet dataset [112] is used for testing. The toolbox can significantly reduce execution time compared to the existing librosa Python library [113].

E. GENERATIVE ADVERSARIAL NETWORKS (GAN)

Despite the fact that RNNs and CNNs are the most popular MIR architectures, there have been studies that look at alternative networks for MIR. GANs (Fig. 6) were first proposed in the original version of [114]. A GAN consists of two competitive agents: a generator and a discriminator. Starting with a training set of real data, the generator is trained to generate new samples that follow the distribution of the real data, while the discriminator must identify the real from the artificial samples.

For emotion classification, a GAN is proposed in [115] that utilizes a double-channel fusion strategy to extract local and global features of an input voice or image. There are five emotion classes considered: sad, happy, quiet, lonely, and miss. The information used in the experiments comes from a number of websites, such as Kuwo Music Box, Baidu Heartlisten, and others. The recognition rates achieved are between 87.6% and 91.2% for all emotions.

In [116], an architecture combining computer vision and note recognition is proposed for music notation recognition. The experiments make use of several datasets, including the JSB Chorales [117], Maestro [118], Video Game [119], Lakh MIDI [82], [83], and another MIDI dataset. The recognition accuracy ranges from 0.88 to 0.92 for all the datasets. The proposed model's intended application is music education.

For Singing voice separation, in [120], a GAN with a time-frequency masking function is used. The databases MIR-1K [48], iKala [72], and DSD100 [62], [89] are used in the experiments, and the model outperforms a conventional DNN.

F. CONVOLUTIONAL RNNs (CRNN)

Complementary to standard models, more complex ones have been developed that utilize couplings between different architectures, often in a series interconnection, to combine their characteristics and improve performance. Convolutional RNNs (CRNNs) are one of these examples.

For music classification, a CRNN was considered in [121], which is a CNN network with the last layers replaced by an RNN. The CNN part is used for feature extraction and the RNN part as a temporal summarizer. The Million Song Dataset [80] is used for training, to predict genre, mood, instrument, and era. The model outperforms other architectures with respect to AUC-ROC.

For MR, a CRNN is used in [122] for classifying and recommending music, in the categories of classical, electronic, folk, hip-hop, instrumental, jazz, and rock music. The database used is the Free Music Archive [123]. The system was tested on a group of 30 users, and the best architecture was the one that implemented a cosine similarity, along with information on music genre.

G. CNN-LSTM

Similarly to CRNNs, some works combine the architectures of CNNs with LSTMs. For emotion classification, a model in [124], consisting of a 2d input through a CNN-LSTM and a 1d input through a DNN, combines two types of features and improves audio and lyrics classification performance. Four classes are considered, angry, happy, relaxed, and sad. The dataset used is the Last.fm tag subset of the Million Song Dataset [80], with an average accuracy of 78%. In [125], a novel database of Turkish songs is constructed for experimentation. The model uses a CNN as the feature extractor and an LSTM with a DNN as the classifier. An accuracy of over 99% is obtained. In [126], the model extracts features from the lyrics, combining a word vector and a CNN-LSTM architecture, with a word frequency weight vector along with a DNN. The outputs of the two architectures are combined on a matching attention mechanism to derive the text emotion classification. Four classes are considered, happy, sad, healing, and calm. The classification accuracy for all emotions ranges between 0.809 to 0.903.

For music score recognition, the proposed architecture takes as input an image of a music score and outputs the duration, pitch, and coordinate for each note in [127]. Data from Muse Score [128] were used for the experiments, and the model outperforms other architectures, with respect to all accuracy measures.

For sound event recognition, [129] considers polyphonic sounds, for a wide family of 61 classes, including music, taken out of a dataset of ten different daily contexts, like a sports game, a bus, a restaurant, and more [130]. The model achieves an average f1 score of around 65%.

H. ARCHITECTURE OVERVIEW

From the above review, it is clear that the "classical" DL models perform well in a variety of MIR tasks. However, the models under consideration need to be appropriately designed, so that they can achieve good results for their set problem. Thus, (and accordingly to the no free lunch theorem) there is no architecture that can be considered holistically better than the rest. On the contrary, complex architectures that incorporate layers of different types are the most promising, since they combine the best characteristics of each DL module, as discussed in Section IV.

III. DL METHODS FOR MUSIC GENERATION

In this section, the application of DL in MG is reviewed. Automatic MG utilizes the MIR techniques mentioned in the previous section to generate novel music scores of desired characteristics, like genre, rhythm, tonality, and underlying emotion. The resulting output can either be a music track in the form of audio, so it can be directly listened to, or it can be in a symbolic notation form. Along with the generation of novel tracks, some tasks can be considered adjacent to MG. One such application is Genre Transfer (GT). This refers to preserving key content characteristics of a music score and applying style characteristics that are typical of a different genre. An example would be transforming a pop song into its heavy metal cover. Another application is Music Inpainting (MI), which refers to filling a missing part of a music track, using information from the rest of its content. Again, the section is divided into subsections based on the DL architecture used. The public databases used in each work are also mentioned. Table 2, summarizes the reviewed works for MG, categorized by their architecture.

The MG architectures can be evaluated both objectively and subjectively. Objective evaluation refers to using mathematical and statistical tools, to measure the similarity of the generated music tracks to the training dataset, as well as other characteristics that can measure their similarity to real

TABLE 1. Deep learning methods for music information retrieval.

Architectures	Applications	Research Work
FCDNNs	Recommendation	[18], [34]
	Emotion Classification	[36]
RNNs	Music Classification	[38], [40], [52]
	Emotion Classification	[42], [43]
	Recommendation	[44]
	Singing voice separation	[47]
	Harmony arrangement	[51]
LSTMs	Music Classification	[54], [56]
	Emotion prediction/recognition	[59]
	Instrument detection/identification	[61]
	Recommendation	[63]
LRCN	Singing voice detection	[65], [69]
CNNs	Audio Classification	[74]–[76]
	Music classification	[15], [16], [79], [81]
	Sound event recognition	[84], [85]
	Singing voice separation	[88]
	Singing voice evaluation	[90]
	Musical instrument identification	[91]–[93], [95], [98]
	Emotion Classification	[100], [102]
	Recommendation	[17], [104], [106], [109]
	Music transcription	[111]
GAN	Emotion Classification	[115]
	Music notation recognition	[116]
	Singing voice separation	[120]
CRNN	Music Classification	[121]
	Recommendation	[122]
CNN-LSTM	Emotion Classification	[124]–[126]
	Recognition	[127]
	Sound event recognition	[129]

music. For objective evaluation, there are several measures, including the loss and accuracy of the training process, the empty bar rate, polyphonicity, note in a scale, qualified note rate, tonal distance, and note length histogram, among others. Most studies consider a subset of these measures or similar ones, so the reader can refer to each work for details.

For subjective evaluation, a test audience is usually given a collection of DL-generated tracks from different architectures, along with human compositions, and is asked to rate them with respect to different aspects, usually on a five-point Likert scale. Variations of this include comparing pairs of tracks and choosing which one they prefer the most or being asked to decide if a track is computer or human-made. In the following sections, we point out which works have conducted subjective evaluations, as the positive audience perception of AI music tracks is essential for the future applicability of MDL. The reader can again refer to each work for the extensive presentation of the evaluation results.

As a closing note, it is worth mentioning an issue that emerges from the field of AI-based MG, that of copyrighting [131], [132]. As AI methods use different software and sample databases, legal problems may arise when claiming authorship of the final musical product. It is thus important that legislators update the existing policies, to avoid rising such issues in the future.

A. RNNs

As with MIR, RNNs have proved popular for MG tasks. For works on classical music, the model termed SampleRNN [133] generates one audio sample at a time, with the resulting signals receiving positive evaluation from human listeners. Three different datasets were considered, one containing a female English voice actor, one containing human sounds like breathing, grunts, coughs, etc, and one containing Beethoven's piano sonatas, taken from the Internet Archive [134]. The models were evaluated by a human group, with the samples of the 3-tier model gaining the highest preference. In [117], an RNN model termed DeepBach is designed, for generating hymn-like scores mimicking the style of Bach. The dataset is taken from the music21 library [135]. The model offers some control to the user, allowing the placement of constraints like notes, rhythms, or cadences to the score. The model was evaluated by human listeners of varying expertise, who were given several samples, and had to guess between Bach or computer generated. Around 50% of the time, the computer tracks were passed as real samples, which is a very satisfying result for such complex music. The work was expanded in [136], with an architecture termed Anticipation-RNN which again offered control to the user to place defined positional constraints. The music21 library [135] was used once again.

In [137], a Graphical User Interface (GUI) system termed BachDuet was developed for promoting classical music improvisation training through user and computer interaction. The JSB chorales data from the music21 dataset [135] is used for training. The GUI was warmly received by test users, who found the improvisation interaction easy to use, enjoyable, and helpful for improving their counterpoint improvisation skills. Additionally, a second group of participants were asked to listen to music clips, rate them, and also decide whether they resulted from a human-machine improvisation using BachDuet, or human-human interaction. Both types of tracks received similar scores, and the listeners were also unable to differentiate between the duets, as they wrongly classified them around 50% of the time.

In [138] the model produces drum rhythms for a seven-piece drum kit. Natural language translation was used to express the hit sequences. An online interface was designed and evaluated by users, who gave an overall average to positive score.

In [139], the effects of different conditioning inputs on the performance of a recurrent monophonic melody generation model are studied. The model was trained on the FolkDB dataset [140] and a novel Bebop Jazz dataset. The validation Negative Log Likelihood loss (NLL) can be as low as 0.190 for the pitch and 0.045 for the duration.

In [141], the problem of inpainting was considered, which combines a VAE that takes as input past and future context sequences, with an RNN that takes as input the latent vectors from the VAE, and as output a latent vector sequence that is passed through a decoder, to create the inpainting sequence. A folk dataset from The Session [142] is used for testing. The model outperforms others with respect to the NLL measure. The architecture was also tested by users, who were given pairs of segmented sequences, and had to choose among

excerpts that fit. The model performance was on the same level as other architectures.

B. LSTMs

LSTMs have been considered for several scenarios. In [143], data preprocessing has been applied to improve the quality of the generated music, and also reduce training time.

In [144], BLSTM networks are used for chord generation. The database used was Wikifonia, which is now inactive, that included sheets for several music genres [145]. The user evaluation showed a preference for the BLSTM model over others, although the original music still received the highest score.

In [146], BLSTM is used for chord generation. The model consists of three parts: a chord generator, which uses some starting chords as input, a chord-to-note generator, which generates the melody line from the generated chords, and a music styler, which combines the chords and melody into a final music piece. Multiple music genres were used as a training database, including Nottingham [147], a collection of British and American folk tunes, Wikifonia [145], and the McGill-Billboard Chord Annotations [148]. The model was evaluated by listeners, which gave a score ranging from neutral to positive, taking into consideration harmony, rhythm, and structure.

In [149], a combination of two LSTM models, termed CLSTMS, is used to build chords that can match a given melody. One sub-model is used for the analysis of measure note information, and the other is used for chord transfer information. Wikifonia is used with data taken from [144] and [145].

In [150], a variation of Biaxial LSTM was used, and a model termed DeepJ was developed for MG. The model was tested on three types of music, baroque, classical, and romantic, with test participants being able to successfully categorize the generated samples most of the time. The Piano-MIDI dataset [151] was used. The model is also capable of mixing musical styles by tuning the values of a single input vector.

In [152], a two-stage architecture is proposed that utilizes BLSTM, where the harmony and rhythm templates are first produced, and the melody is then generated and conditioned on these templates. The Wikifonia dataset is used [145]. In the subjective evaluation, participants were given a collection of tracks and were asked to rate them according to how much they found them pleasing and coherent, and whether they believe they were human or AI-generated. The highest scores were achieved by the model where the melody generator is conditioned on an existing chord and rhythm scheme from a real song. This melody is also perceived as human-made by many participants. The authors also noted that there are high standard deviations in all answers, and slightly more so in the models rated positively, indicating that there is a much wider perception of what is considered good-sounding music, than a bad one.

In [153], an architecture combining LSTM with a Recurrent Temporal Restricted Boltzmann Machine is designed. Experiments were conducted in MuseData [154], a classical music dataset, and JSB chorales [155] dataset. The model outperforms other architectures with respect to Log-likelihood (LL) and frame-level accuracy (ACC%) measures.

In [156], variations of the LSTM are discussed, termed Tied Parallel LSTM with a neural autoregressive distribution estimator (NADE), and Biaxial LSTM. The model was tested on the datasets of JSB Chorales [155], MuseData [154], Nottingham [147], and Piano-MIDI [151], a classical piano dataset. The architectures perform well concerning the Log-likelihood measure. The architectures also have translation invariance.

In [157], an RNN-LSTM architecture is proposed, using the Meier cepstrum coefficients as features. The dataset consists of folk tunes collected by the author. The model achieves an accuracy of 99% and a loss rate of 0.03.

In [158], a model termed Chord conditioned Melody Transformer (CMT) is proposed, which generates rhythm and pitch conditioned on a chord progression. The training has two phases, first, a rhythm decoder is trained, and second, a pitch decoder is trained based on the rhythm decoder. The model was trained on a novel K-Pop dataset. In addition to various measures, like rhythm accuracy, the model was also evaluated by listeners, with respect to rhythm, harmony, creativity, and naturalness. The model outperforms the Explicitly-constrained conditional variational auto-encoder (EC^2-VAE) [159], with respect to rhythm, harmony, and naturalness. The model also has a higher score for creativity than the real dataset tracks, meaning that it can indeed generate novel melodies.

In [160], an LSTM specifically for Jazz music was designed, using a novel Jazz music dataset in MIDI format, and the Piano-MIDI [151]. The model can also generate music using only a chosen instrument. The model can achieve a very low final loss value.

In [161], a BLSTM network with attention is considered for Jazz MG. The architecture consists of a BLSTM network, an attention layer, and another LSTM layer. The Jazz ML ready MIDI dataset [162] is considered. The model outperforms simpler architectures like LSTM without attention and the attention LSTM without the BLSTM layer.

In [163], a piano composer is designed, that uses information from given composers to generate music. The datasets used were Classical Music MIDI [164] and MIDI_classic_music [165], from which tracks of Beethoven, Mozart, Bach, and Chopin were considered. The model was evaluated through a human survey, where participants had to choose the real sample among the computer-generated and composer ones. Around half the time, people mistook the model-generated music for the human-composed track, meaning that the model can generate music that is relatively indistinguishable from real samples. The generated tracks can also be perceived as fairly interesting, pleasing, and realistic.

In [166], an architecture, comprising of an LSTM paired with a Feed Forward layer, can generate drum sequences resembling a learned style, and can also match up to set

constraints. The LSTM part learns drum sequences, while the feed-forward part processes information on guitar, bass, metrical structure, tempo, and grouping. The dataset was collected from 911tabs [167], and broken into three parts, for 80s disco, 70s blues and rock, and progressive rock/metal, with the model being effective in all styles.

Finally, in [168], the MI problem was considered by combining half-toning and steganography, and various methods were compared using a dataset of various instruments, with satisfying results for the considered models.

C. CNNs

For CNN architectures, in [169], the architecture comprises an LSTM as a generator, a CNN as a discriminator, and a control network that introduces restriction rules for a particular style of music generation. The matching subset of the Lakh MIDI dataset (LMD) [82] and Piano-MIDI dataset [151] was used. The model was evaluated by music experts, with respect to melody, rhythm, chord harmony, musical texture, and emotion. The model is rated higher than other ones in all of the above aspects.

In [170], a CNN with a Bidirectional Gate Recurrent Unit (BiGRU) and attention mechanism is used for folk music generation. The ESAC dataset [171] is used for testing. The results were evaluated by listeners, who gave overall positive ratings, although lower than the real ones. There were also some exceptions of low scores, meaning that the model generation may have some inconsistencies in its performance.

In [172] a Convolution-LSTM for piano track generation is considered. The CNN layer is used for feature extraction, and the output is fed into the LSTM for music generation. Piano tracks from Midiworld [173] were used for training. The model was evaluated by listeners, who were given 10 music segments, and had to decide whether they were human-made or computer generated. In most cases, the segments were correctly identified, but the Convolution-LSTM model performed better than the simple LSTM.

D. GANs

Symbolic music is stored using a notation-based format, which makes it an easier-to-use input for training NNs. For symbolic music generation, a GAN model is proposed in [174] for piano roll generation, equipped with LSTM layers in the generator and discriminator. The generated files were evaluated by participants with respect to melody and rhythm, and the proposed model received a higher score than files generated from other architectures.

In [175], an inception model conditional GAN termed INCO-GAN is proposed that can generate variable-length music. This complex architecture consists of two phases, that of training and generation, and each phase is broken into three processes: preprocessing, CVG training, and conditional GAN training for the training stage, and CVG executing, phrase generation, and postprocessing for the generation phase. The Lakh MIDI dataset is used for the experiments [82]. The model achieves high cosine similarity with the human-composed music for the frequency vector.

In [176], the problem of symbolic music GT was studied using CycleGAN, a model consisting of two GANs that exchange data and are trained simultaneously. The model was evaluated using genre classifiers, verifying the successful style transfer.

In [177], DrumGan is proposed, an architecture for generating drum sounds (kick, snare, and cymbal). The model offers user control over the resulting score, by tuning the timbre features.

In [178], the authors generated log-magnitude spectrograms and phases directly with GAN to produce more coherent waveforms than directly generating waveforms with strided convolutions. The resulting scores are generated at a much higher speed. The NSynth dataset [179] is used, which contains single notes from many instruments, at different pitches, timbres, and volumes. The human audience rated the audio quality of the tracks, and the model was received as slightly inferior to the real tracks.

In [180], a GAN equipped with a self-attention mechanism is used to generate multi-instrument music. The self-attention mechanism is used to allow the extraction of spatial and temporal features from data. The Lakh MIDI [82] and Million Song [80] datasets were used here.

In [181], a GAN was designed for symbolic MG, along with a conditional mechanism to use available prior information, so that the model can generate melodies either starting from zero, by following a chord sequence, or by conditioning on the melody of previous bars. Pop music tabs from Theory-Tab [182] were used. The resulting system, termed MidiNet, is compared to Google's MelodyRNN and performs equally, with the test audience characterizing the results as being more interesting.

In [183], multi-track MG was considered using three different GAN models, termed the Jamming, Composer, and Hybrid. The Jamming model consists of multiple independent generators. The Composer consists of a single generator and discriminator, and a shared random input vector. In the Hybrid model, the independent generators have both an independent and a shared random input vector. The models were trained on a rock music database and used to generate piano rolls for bass, drums, guitar, piano, and strings. The database is termed Lakh Pianoroll Dataset, as it is created from the Lakh MIDI [82], by converting the MIDI files to multi-track piano rolls. A subset is also used with matched entries from the Million Song dataset [80]. Additionally to using the training database, the model can also use as an input a given music track from the user and generate four additional tracks from it. The model was evaluated by professional and casual users and received overall neutral to positive scores.

In [184], Sequence Generative Adversarial Net (SeqGAN) is proposed, which applies policy gradient update. The Nottingham folk dataset [147] is used in the experiments. The model outperforms a maximum likelihood estimation (MLE)

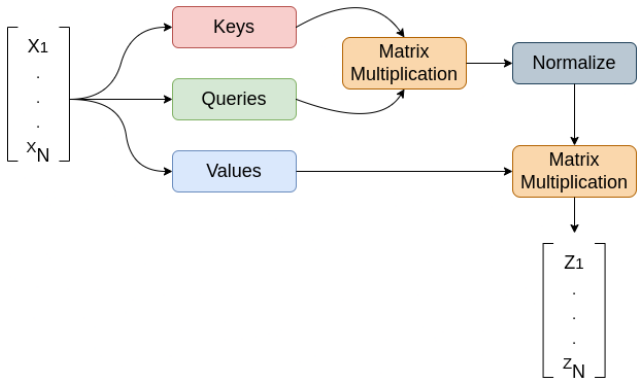

FIGURE 7. Self-attention mechanism.

trained LSTM with respect to the mean squared error and other measures.

In [185], sequence generative GANs were considered for polyphonic music generation. The method condenses the duration, octaves, and keys of melodies and chords into a one-word vector representation. The Nottingham dataset [147] was used. The results were well received by a test audience, with respect to pleasantness, realism, and interest.

In [186], a conditional GAN is proposed for long inpainting up to a few seconds. The model was trained on datasets of increasing complexity, like the Lakh MIDI [82] and Million song [80], the Maestro dataset [118], recordings of grand pianos, and free music archive dataset [123], and extensive audience experiments were performed to evaluate the model. The inpaintings were generally detectable, especially in tracks with higher complexity, but were considered slightly or non-disturbing.

E. TRANSFORMERS

Transformers constitute a relatively recent architecture [187], which has found popularity in NLP. A key aspect of transformers is self-attention, which refers to the process of weighting the relevance between different positions of a single sequence. Transformers process sequential input data, but not necessarily in order.

The transformer's architecture is basically an encoder-decoder scheme. The encoder maps the sequence of inputs $(x_1, \ldots, x_N)$ to a sequence of vector representations $(z_1, \ldots, z_N)$. The decoder then takes this vector representation and generates a sequence of outputs $(y_1, \ldots, y_M)$, one at a time.

Let $\mathbf{W}^q, \mathbf{W}^k, \mathbf{W}^v$ be the three parameter matrices that are trained. These matrices are used to define the following parameters:

- Query: $q = \mathbf{W}^q x_i$
- Key: $k = \mathbf{W}^k x_i$
- Value: $v = \mathbf{W}^v x_i$

The self-attention score is calculated as follows: For every input, our desire is to calculate how it attends to all the tokens in the sequence. To achieve this, the query vector is used and since every token becomes the query for once, we calculate

$$e_{ij} = q_i k_j, \quad \text{with } i, j \in \{1, \ldots, N\}. \tag{8}$$

To have more stable gradients, normalization is performed as

$$n_{ijk} = s_{ij} \frac{e_{ij}}{d_k^{1/2}}. \tag{9}$$

The final step is to calculate the self-attention score as

$$z_i = \sum_{j=1}^{d_k} \frac{\exp(s_{ij})}{\sum_{l=1}^{N} \exp(s_{il})} v_j. \tag{10}$$

In practice, the aforementioned procedure is performed in matrix form and is depicted in Fig. 7.

Modifications of the simple transformer are proposed in various works. In [188], a relative attention mechanism is used to generate minute-long compositions, with reduced intermediate memory requirements from quadratic to linear. The JSB chorales dataset [155] and Piano-e-Competition dataset [189] were used. The model was evaluated by listeners, who were asked to rate pairs of musical excerpts. The model outperformed other architectures and was seconded only by the real music tracks.

In [190], an adversarial transformer is proposed to generate single-track or multitrack music. The results were positively received by a test audience, who rated tracks with respect to being human-like, harmonious, rhythmic, structured, fluent, and overall quality. The model scores better compared to another architecture, and much closer to the real track scores.

In [191], sparse factorization was applied to the attention matrix, which reduced the memory and time requirements from quadratic to sub-quadratic. Five-second-long samples were generated. A piano recording dataset from [192] was used for training.

In [193] a model termed Pop Music Transformer is proposed to generate pop piano music. The model uses a beat-based music representation. The generated tracks were evaluated by experts and casual listeners and were preferred by both groups over other architectures.

In [194] a model termed Jukebox can generate music along with vocals in various musical styles. The model uses multiscale Vector Quantization - Variational Autoencoders (VQ-VAE) to compress the raw audio input to discrete codes. Then the output is generated using an auto-regressive transformer. The architecture provides lyric conditioning, to control the singing part. The Maestro dataset was used [118] for training, and the LyricWiki (now closed) to gather metadata, among others. The model can generate music in any chosen style by supplying conditioning signals during training.

In [195], a model for symbolic MG for Mandarin pop is proposed, where the transformer training considers the conditioning sequence as a thematic material. The POP909 dataset is used [196]. The model was evaluated by participants, on the aspects of theme controllability, repetition, timing, variation,

and overall structure and quality. The proposed model outperforms others in all metrics.

In [197], conditional drum generation is considered, inspired by [166]. A BLSTM encoder receives the conditioning parameter information, and a transformer-based decoder with relative global attention generates the drum sequence. A subset of rock and metal songs from the Lakh MIDI dataset is used [82]. For subjective evaluation, participants were given a set of three tracks, two being the accompanying or condition tracks, and the third being the drum track to be evaluated. They were asked to rate the drum tracks with respect to rhythm, pitch, naturalness, groove, and coherence. The tracks generated from the proposed model outperform another baseline model and are even rated higher than real compositions with respect to naturalness, groove, and coherence. The users were also asked their opinion on whether the given drum tracks each time were real compositions or computer generated. The drum tracks from the model were perceived as computer generated only 39% of the time, indicating the natural feel of the tracks.

In [198], the problem of melody harmonization was considered. The model maps lower-level melody notes into semantic higher-level chords. Three architectures are proposed using a standard transformer, variational transformer, and regularized variational transformer. The Chord Melody [199] and Hooktheory Lead Sheet [200] datasets are used. In the human evaluation conducted, participants, comprising casual music listeners and professionals, were asked to rate samples with respect to harmonicity, unexpectedness, complexity, and preference. The standard model achieved the highest scores in harmony and preference, whereas the variational model achieved the highest in unexpectedness and complexity.

F. ARCHITECTURE OVERVIEW

As with the case of MIR, it is clear that there is no single architecture that can outperform the rest in MG tasks. Multi-layered architectures though can be a path for building better models, especially when additional objectives are set, like conditioning the generated music to desired features.

TABLE 2. DL methods for MG.

DL Architectures	Applications	Research Work
RNNs	MG (Classical)	[117], [133], [136], [137]
	MG (Drums)	[138]
	MG (Folk)	[139]
	MG (Jazz)	[139]
	SG (Speech)	[133]
	SG (Human vocal sounds)	[133]
	MI	[141]
LSTMs	Multiple genres	[143], [144], [146], [149], [152]
	MG (Classical)	[150], [153], [156]
	K-Pop	[158]
	Folk	[156], [157]
	Jazz	[160], [161]
	Piano	[156], [163]
	Drums	[166]
	MI	[168]
CNN	MG Classical	[169]
	Folk	[170]
	Piano	[172]
GANs	Symbolic MG	[174]–[178], [180]
	MG (Pop)	[181]
	MG (Rock)	[183]
	MG (Folk)	[184], [185]
	MI	[186]
Transformers	MG (Longer sequences)	[188], [190], [191], [193], [194]
	MG (Mandarin Pop)	[195]
	MG (Drums)	[197]
	Melody harmonization	[198]

IV. FUTURE STUDIES IN MDL

In this section, future research directions in MDL are identified and discussed.

A. MIXED ARCHITECTURES

So far there have been multiple approaches and different architectures to address key problems in MDL. However, despite most works reporting positive results, due to the complexity of the applications under study and their peculiarities, there is no dominant method that should be followed for a given task. Thus, there is no overall superior architecture that is guaranteed to outperform all others for any given MDL problem.

On the other hand, results indicate that the best approach to constructing holistically better models, which can consistently yield improved results is to consider combined architectures, like CRNNs [121], [122] or LRCNs [65], [69]. Such approaches can harness the individual characteristics of each model to surpass their counterparts. Attention mechanism enhanced architectures is one such example [56], [95], [126], [161], [180], with more being developed [201], [202], [203], [204]. Such approaches will surely lead the advances in the MDL field.

Apart from hybrid architectures, MDL will be significantly benefited from the use fusion of diverse input modalities. This would increase performance, as the conjunction of different modalities can help build connections between different features. For example, in [76] sound signals were extracted from unlabelled video sources. In [205], the combination of singing signals along with laryngoscope images was combined for voice parts division. In [206], a system that combined heart rate measurements and facial expressions was composed to detect drowsiness in drivers, which is accompanied by a music recommendation system used as a countermeasure to avoid accidents. In [63] and [64], a synchronized music and dance dataset were used for recommendation. In [207], music emotion classification is performed for four emotional classes, combining features from lyrics and acoustics. These are indicative examples of an emerging trend of bridging the gap between different modalities.

For the above techniques, an all-present problem is the computational cost of training [208]. The increase in hardware requirements creates practical issues with energy

consumption and environmental footprint, which under the scope of the global energy and environmental crisis, are mandatory to address. Addressing the above will require the performance improvement of current architectures, or the consideration of different ones [209]. Understandably, any improvements in the computational cost will, by extension, also boost the commercialization of MDL applications.

B. TRADITIONAL MUSIC

Most of the existing works use widely available training databases, which mainly include western music genres, like classical music, pop, rock, metal, jazz, blues, etc. Using widely established music genres make sense, due to their popularity, but it is highly important to enrich and diversify the training databases by including more genres. So, while it is essential to consider new and emerging genres, especially ones that are computer-based, like electronic, synth-wave, and vaporwave [65], [210], [211], [212], another trend that is gaining popularity is the application of MDL and MG for traditional and regional music. Traditional music refers to music originating from a specific country or region and is closely tied to its culture [213]. Examples include the recitation of religious excerpts like the Holy QurBan [214], and traditional music from different regions, like Byzantine [215], Greek [216], [217], [218], Persian [219], Chinese [220], [221], Indian [222], [223], and many more.

In the development of MDL for regional and traditional music, several challenges may appear, as a result of the distinct nature of the topic. One issue is the dataset availability, which in contrast to western popular music, is in many cases hard to gather, especially in large amounts, which are required for optimal training. In most cases, the research groups take it upon themselves to build their own dataset, due to the lack of existing ones, so hopefully, in the future, more authorities will help towards building free databases [77], [78], [142], [196], [221], [224], [225], [226]. For this task, recording difficulties may arise, especially for recordings made outside a music studio, with varying acoustics, for example in religious singing. Coming along with the problem of dataset collection is that of appropriate feature tagging of the tracks. This is strenuous work that requires time, and often the collaboration of music experts, for tasks like the annotation of music features, and testing audiences, for more ambiguous characterizations, like the emotion that a track evokes.

Moreover, many musical instruments, like the guitar and piano, are present in almost all music genres, so it is easier to adopt MG architectures for a specific instrument to many different styles. This may not be the case for regional instruments, which are only used for playing a region's traditional music. So, for preserving and learning musical styles through DL, it is essential to build datasets for specific instruments [221]. Finally, many traditional music styles have a distinct musical notation, like Mensural notation, Chinese Gongche, and Organ tablature, meaning that MDL architectures for transcription, pattern recognition, and symbolic MG would have to be adjusted to fit the characteristics of each genre. This again requires the existence of appropriate databases for different musical notations.

TABLE 3. List of DL studies focused in traditional music.

Genre	Applications	Research Work
African	Classification	[15]
Arabic	Classification	[214]
Byzantine	Music recognition	[215]
Chinese	Song recognition	[220]
	Instrument recognition	[227]–[229]
	Classification	[230]–[232]
	Drum MG	[233]
	Folk MG	[170], [234]
	Pop MG	[195]
	Feature extraction	[235]
Croatian	Transcription	[236]
Ethiopian	Classification	[237]
Greek	Classification	[216]–[218]
Indian	Classification	[222], [238]–[244]
	Transcription	[223]
	Melodic framework conversion	[245]
Irish	Player recognition	[246]
	MG	[247], [248]
Korean	Classification	[249]
	MG	[158]
Persian	Classification	[219]
	Composition	[250]
	Source Separation	[251]
Scandinavian-like Folk	MG	[248], [252]
Scottish	MG	[253]
Thai	Transcription	[254]
Turkish	MG	[255]–[257]
	Classification	[258], [259]
	Emotion recognition	[125]
Vietnamese	Classification	[260]

Overall, it seems that there are still several practical challenges to fully developing DL for traditional music. These are steadily addressed by the efforts of several research groups over the world. Table 3 lists the recent works that study Traditional Music Deep Learning (TMDL), categorized by music type. These works offer great service to the preservation of history, culture, and art, as the digitization, study, and generation of traditional music will help open it up to new generations of listeners and also promote thematic (music, religious) tourism. Thus, it is expected that more research groups will contribute to regional MDL in the future, and hopefully, such research endeavors will also receive governmental support and recognition.

C. MEDICAL APPLICATIONS

The field of Music Therapy (MT) lies at the intersection of Medicine and Music. MT is an evident-based approach for treating a plethora of pathological conditions, including, among others, anxiety, depression, substance abuse, Alzheimer's, eating disorders, sleep disorders, and more [261], [262], [263]. Naturally, DL can prove a valuable tool to therapists and patients, as a complement to existing treatments. Table 4 summarizes the recent applications of DL in music therapy, categorized by architecture. The conditions that have been addressed include music remixing to

TABLE 4. DL methods for music medical applications.

DL Architectures	Applications	Research Work
DNNs	Cochlear implant	[264], [265]
RNNs	Anxiety	[266]
	Cochlear implant	[267]
LSTMs	Alzheimer's	[268]
	Tinnitus music therapy	[269]
	Depression treatment	[270]
	Multi-Voice Music Generation music therapy	[271]
Support vector machine (SVM)	Anxiety Prediction	[272]
CNN	Vocal art medicine	[205]
	Mood transformation	[273], [274]
	Depression	[275]
	Anxiety	[276]
GAN	Control of heart rate variability	[277]

TABLE 5. List of abbreviations.

Term	Abbreviation
Area Under Receiver Operating Characteristic Curve	AUC-ROC
Attention Mechanism	AM
Audio Signal Processing	ASP
Bidirectional Gate Recurrent Unit	BiGRU
Bidirectional Long Short Term Memory	BLSTM (BiLSTM)
Computer Vision	CV
Convolutional Neural Network	CNN
Convolutional Recurrent Neural Network	CRNN
Deep Learning	DL
Fully Connected Deep Neural Network	FCDNN
Generative Adversarial Networks	GAN
Genre Transfer	GT
Image Processing	IP
Long Short Term Memory	LSTM
Long-Term Recurrent Convolutional Network	LRCN
Multi-layer Perceptron	MLP
Music Deep Learning	MDL
Music Genre Classification	MGC
Machine Learning	ML
Music Generation	MG
Music Information Retrieval	MIR
Music Inpainting	MI
Music Recommendation Systems	MRS
Music Signal Processing	MSP
Music Therapy	MT
Natural Language Processing	NLP
Recurrent Neural Network	RNN
Singing Voice Detection	SVD
Speech Generation	SG
Traditional Music Deep Learning	TMDL
Vector Quantization - Variational Auto-encoders	VQ-VAE

improve cochlear implant performance, effective MRS and MG for mood transformation, including anxiety and depression, MG for stimulating the musical memory in patients with Alzheimer's, MG for relieving Tinnitus, and voice parts classification for vocal art medicine. Existing architectures of DL for tasks like music recommendation and emotion classification can be adapted to fit many of the above conditions. For example, music recommendation systems can be updated to make suggestions based on emotion and mood, using a collection of patient inputs, like facial expressions, and other physiological signals, like heart rate, temperature, respiratory rate, EEG signals, and more. By designing appropriate user interfaces [40], [117], [137], MDL architectures could also be used as an entertainment and educational tool, especially for interventions with children. Finally, it would also be interesting to see if knowledge transfer could be applied to models developed for treating conditions with overlapping symptoms, for example, anxiety and depression.

MT is a field that is constantly developing, with medical researchers turning to it as a method for effectively treating, or reducing the symptoms of many conditions. By developing proper training databases and MIR and MG architectures, DL will help in establishing open-access tools that can be used by anyone alike, without the need for increased medical expenses. Moreover, tools like MRS for mood transformation can be directly available to patients, providing daily help coverage. Overall, there are many promising future directions to be considered by researchers.

D. MUSIC GENERATED FROM DYNAMICAL SYSTEMS

Another field that would also be interesting to consider is that of chaos-based music generation [278], [279], [280], [281], [282], [283]. In this interdisciplinary field, which bridges MG with the rich area of chaos theory, the time series solution of a chaotic system is used as a high entropy source for music generation, in tuning parameters like the extraction of musical pitches, the duration of a musical note, the amplitude, and the velocity. Chaotic systems are characterized by non-periodicity, and sensitivity to parameter changes, meaning that two solutions of the same system, starting from almost identical initial configurations, will quickly diverge from each other, yielding two different, non-periodic time series. This feature can thus be exploited in MG, as it can aid in the generation of non-repeating musical patterns. So exploring DL methods in this area could give rise to applications in numerous fields, including medical treatment [284], [285], and possibly secure communications [286], and system identification [287].

V. CONCLUSION

MDL has evolved into a very active field, with an increasing number of contributions each year, addressing its vast applications. This work provided a review of the recent developments in Music Deep Learning. The review was divided into two main categories, Music Information Retrieval, and Music Generation. After reviewing each field, future research trends were identified.

The future of MDL lies in developing hybrid architectures to improve performance, while applications span a plethora of commercial, conservational, medical, and experimental applications being developed. Of these, applying DL for studying and preserving the cultural heritage of each country is of high importance. So is the exploitation of MDL for medical applications. The integration of MDL and chaos seems much more experimental, but its multi-disciplinarity will surely lead to new developments in both

fields. For all of the aforementioned applications, bringing together research groups consisting of heterogeneous and complementing researchers, like computer scientists, physicists, mathematicians, musicians, audio engineers, and medical practitioners, is the key to success. The authors hope that the present work can be of service to these researchers, by providing a clear overview of recent and emerging developments in the field.

APPENDIX A
LIST OF ABBREVIATIONS

Table 5 lists the abbreviations used throughout the text.

REFERENCES

[1] A. Shrestha and A. Mahmood, "Review of deep learning algorithms and architectures," *IEEE Access*, vol. 7, pp. 53040–53065, 2019.

[2] J. Chai, H. Zeng, A. Li, and E. W. T. Ngai, "Deep learning in computer vision: A critical review of emerging techniques and application scenarios," *Mach. Learn. Appl.*, vol. 6, Dec. 2021, Art. no. 100134.

[3] N. Fatima, A. S. Imran, Z. Kastrati, S. M. Daudpota, and A. Soomro, "A systematic literature review on text generation using deep neural network models," *IEEE Access*, vol. 10, pp. 53490–53503, 2022.

[4] M. R. Karim, O. Beyan, A. Zappa, I. G. Costa, D. Rebholz-Schuhmann, M. Cochez, and S. Decker, "Deep learning-based clustering approaches for bioinformatics," *Briefings Bioinf.*, vol. 22, no. 1, pp. 393–415, Jan. 2021.

[5] M. M. Islam, F. Karray, R. Alhajj, and J. Zeng, "A review on deep learning techniques for the diagnosis of novel coronavirus (COVID-19)," *IEEE Access*, vol. 9, pp. 30551–30572, 2021.

[6] A. B. Nassif, I. Shahin, I. Attili, M. Azzeh, and K. Shaalan, "Speech recognition using deep neural networks: A systematic review," *IEEE Access*, vol. 7, pp. 19143–19165, 2019.

[7] S. Minaee, Y. Y. Boykov, F. Porikli, A. J. Plaza, N. Kehtarnavaz, and D. Terzopoulos, "Image segmentation using deep learning: A survey," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 44, no. 7, pp. 3523–3542, Jul. 2021.

[8] L. Ljung, C. Andersson, K. Tiels, and T. B. Schön, "Deep learning and system identification," *IFAC-PapersOnLine*, vol. 53, no. 2, pp. 1175–1181, 2020.

[9] G. Gupta and R. Katarya, "Research on understanding the effect of deep learning on user preferences," *Arabian J. Sci. Eng.*, vol. 46, no. 4, pp. 3247–3286, Apr. 2021.

[10] I. Goodfellow, Y. Bengio, and A. Courville, *Deep Learning*. Cambridge, MA, USA: MIT Press, 2016.

[11] H. Purwins, B. Li, T. Virtanen, J. Schlüter, S.-Y. Chang, and T. Sainath, "Deep learning for audio signal processing," *IEEE J. Sel. Topics Signal Process.*, vol. 13, no. 2, pp. 206–219, Apr. 2019.

[12] J. P. Puig, "Deep neural networks for music and audio tagging," Ph.D. thesis, Inf. Commun. Technol., Universitat Pompeu Fabra, Barcelona, Spain, 2019.

[13] M. Schedl, "Deep learning in music recommendation systems," *Frontiers Appl. Math. Statist.*, vol. 5, p. 44, Aug. 2019.

[14] J.-P. Briot, G. Hadjeres, and F.-D. Pachet, "Deep learning techniques for music generation—A survey," 2017, *arXiv:1709.01620*.

[15] Y. M. G. Costa, L. S. Oliveira, and C. N. Silla Jr., "An evaluation of convolutional neural networks for music classification using spectrograms," *Appl. Soft Comput.*, vol. 52, pp. 28–38, Mar. 2017.

[16] C. Senac, T. Pellegrini, F. Mouret, and J. Pinquier, "Music feature maps with convolutional neural networks for music genre classification," in *Proc. 15th Int. Workshop Content-Based Multimedia Indexing*, Jun. 2017, pp. 1–5.

[17] M. Lu, D. Pengcheng, and S. Yanfeng, "Digital music recommendation technology for music teaching based on deep learning," *Wireless Commun. Mobile Comput.*, vol. 2022, pp. 1–8, May 2022.

[18] D. Martín-Gutiérrez, G. H. Penaloza, A. Belmonte-Hernandez, and F. A. Garcia, "A multimodal end-to-end deep learning architecture for music popularity prediction," *IEEE Access*, vol. 8, pp. 39361–39374, 2020.

[19] Z. Rafii, A. Liutkus, F.-R. Stöter, S. I. Mimilakis, D. FitzGerald, and B. Pardo, "An overview of lead and accompaniment separation in music," *IEEE/ACM Trans. Audio, Speech, Language Process.*, vol. 26, no. 8, pp. 1307–1335, Aug. 2018.

[20] R. Monir, D. Kostrzewa, and D. Mrozek, "Singing voice detection: A survey," *Entropy*, vol. 24, no. 1, p. 114, Jan. 2022.

[21] K. Choi, G. Fazekas, K. Cho, and M. Sandler, "A tutorial on deep learning for music information retrieval," 2017, *arXiv:1709.04396*.

[22] D. Han, Y. Kong, J. Han, and G. Wang, "A survey of music emotion recognition," *Frontiers Comput. Sci.*, vol. 16, no. 6, pp. 1–11, Dec. 2022.

[23] B. L. Sturm, J. Felipe Santos, O. Ben-Tal, and I. Korshunova, "Music transcription modelling and composition using deep learning," 2016, *arXiv:1604.08723*.

[24] L. Casini, G. Marfia, and M. Roccetti, "Some reflections on the potential and limitations of deep learning for automated music generation," in *Proc. IEEE 29th Annu. Int. Symp. Pers., Indoor Mobile Radio Commun. (PIMRC)*, Sep. 2018, pp. 27–31.

[25] M. Kleć and A. Wieczorkowska, "Music recommendation systems: A survey," in *Recommender Systems for Medicine and Music*. Cham, Switzerland: Springer, 2021, pp. 107–118.

[26] N. Ndou, R. Ajoodha, and A. Jadhav, "Music genre classification: A review of deep-learning and traditional machine-learning approaches," in *Proc. IEEE Int. IoT, Electron. Mechatronics Conf. (IEMTRONICS)*, Apr. 2021, pp. 1–6.

[27] C.-W. Wu, C. Dittmar, C. Southall, R. Vogl, G. Widmer, J. Hockman, M. Müller, and A. Lerch, "A review of automatic drum transcription," *IEEE/ACM Trans. Audio, Speech, Language Process.*, vol. 26, no. 9, pp. 1457–1483, Sep. 2018.

[28] L. Wyse, "Audio spectrogram representations for processing with convolutional neural networks," 2017, *arXiv:1706.09559*.

[29] C. Gupta, H. Li, and M. Goto, "Deep learning approaches in topics of singing information processing," *IEEE/ACM Trans. Audio, Speech, Language Process.*, vol. 30, pp. 2422–2451, 2022.

[30] M. Civit, J. Civit-Masot, F. Cuadrado, and M. J. Escalona, "A systematic review of artificial intelligence-based music generation: Scope, applications, and future trends," *Exp. Syst. Appl.*, vol. 209, Dec. 2022, Art. no. 118190.

[31] S. Ji, J. Luo, and X. Yang, "A comprehensive survey on deep music generation: Multi-level representations, algorithms, evaluations, and future directions," 2020, *arXiv:2011.06801*.

[32] J.-P. Briot and F. Pachet, "Deep learning for music generation: Challenges and directions," *Neural Comput. Appl.*, vol. 32, no. 4, pp. 981–993, Feb. 2020.

[33] L. A. Iliadis, S. P. Sotiroudis, K. Kokkinidis, P. Sarigiannidis, S. Nikolaidis, and S. K. Goudos, "Music deep learning: A survey on deep learning methods for music processing," in *Proc. 11th Int. Conf. Modern Circuits Syst. Technol. (MOCAST)*, Jun. 2022, pp. 1–4.

[34] F. Fessahaye, L. Perez, T. Zhan, R. Zhang, C. Fossier, R. Markarian, C. Chiu, J. Zhan, L. Gewali, and P. Oh, "T-RECSYS: A novel music recommendation system using deep learning," in *Proc. IEEE Int. Conf. Consum. Electron. (ICCE)*, Jan. 2019, pp. 1–6.

[35] (2018). *Spotify RecSys Challenge*. Accessed: Sep. 30, 2022. [Online]. Available: http://www.recsyschallenge.com/2018/

[36] V. Revathy and A. S. Pillai, "Binary emotion classification of music using deep neural networks," in *Proc. Int. Conf. Soft Comput. Pattern Recognit.* Cham, Switzerland: Springer, 2021, pp. 484–492.

[37] I. A. P. Santana, F. Pinhelli, J. Donini, L. Catharin, R. B. Mangolin, V. Delisandra Feltrim, and M. A. Domingues, "Music4All: A new music database and its applications," in *Proc. Int. Conf. Syst., Signals Image Process. (IWSSIP)*, Jul. 2020, pp. 399–404.

[38] G. Song, Z. Wang, F. Han, S. Ding, and M. A. Iqbal, "Music auto-tagging using deep recurrent neural networks," *Neurocomputing*, vol. 292, pp. 104–110, May 2018.

[39] E. Law and L. von Ahn, "Input-agreement: A new mechanism for collecting data using human computation games," in *Proc. SIGCHI Conf. Hum. Factors Comput. Syst.*, Apr. 2009, pp. 1197–1206.

[40] J. R. Castillo and M. J. Flores, "Web-based music genre classification for timeline song visualization and analysis," *IEEE Access*, vol. 9, pp. 18801–18816, 2021.

[41] J. F. Gemmeke, D. P. W. Ellis, D. Freedman, A. Jansen, W. Lawrence, R. C. Moore, M. Plakal, and M. Ritter, "Audio set: An ontology and human-labeled dataset for audio events," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, Mar. 2017, pp. 776–780.

[42] W. Zhao, Y. Zhou, Y. Tie, and Y. Zhao, "Recurrent neural network for MIDI music emotion classification," in *Proc. IEEE 3rd Adv. Inf. Technol., Electron. Autom. Control Conf. (IAEAC)*, Oct. 2018, pp. 2596–2600.

[43] S. Rajesh and N. J. Nalini, "Musical instrument emotion recognition using deep recurrent neural network," *Proc. Comput. Sci.*, vol. 167, pp. 16–25, Jan. 2020.

[44] A. Vall, M. Quadrana, M. Schedl, and G. Widmer, "The importance of song context and song order in automated music playlist generation," 2018, *arXiv:1807.04690*.

[45] B. McFee and G. R. Lanckriet, "Hypergraph models of playlist dialects," in *Proc. ISMIR*. Pennsylvania, PA, USA: Citeseer, vol. 12, 2012, pp. 343–348.

[46] *8TRACKS*. Accessed: Sep. 30, 2022. [Online]. Available: https://8tracks.com/

[47] S. Kang, J.-S. Park, and G.-J. Jang, "Improving singing voice separation using curriculum learning on recurrent neural networks," *Appl. Sci.*, vol. 10, no. 7, p. 2465, Apr. 2020.

[48] *Mir-1K Dataset*. Accessed: Sep. 30, 2022. [Online]. Available: https://sites.google.com/site/unvoicedsoundseparation/mir-1k

[49] A. Liutkus, D. Fitzgerald, and Z. Rafii, "Scalable audio separation with light kernel additive modelling," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, Apr. 2015, pp. 76–80.

[50] Z. Rafii, A. Liutkus, F.-R. Stöter, S. I. Mimilakis, and R. Bittner, "The MUSDB18 corpus for music separation," Dec. 2017, doi: 10.5281/zenodo.1117372.

[51] J. Li, "Automatic piano harmony arrangement system based on deep learning," *J. Sensors*, vol. 2022, pp. 1–13, Jul. 2022.

[52] F. Zhang, "Research on music classification technology based on deep learning," *Secur. Commun. Netw.*, vol. 2021, pp. 1–8, Dec. 2021.

[53] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural Comput.*, vol. 9, no. 8, pp. 1735–1780, 1997.

[54] J. Dai, S. Liang, W. Xue, C. Ni, and W. Liu, "Long short-term memory recurrent neural network based segment features for music genre classification," in *Proc. 10th Int. Symp. Chin. Spoken Lang. Process. (ISCSLP)*, Oct. 2016, pp. 1–5.

[55] (2004). *Ismir Genre Dataset*. Accessed: Sep. 30, 2022. [Online]. Available: https://ismir2004.ismir.net

[56] S. K. Prabhakar and S.-W. Lee, "Holistic approaches to music genre classification using efficient transfer and deep learning techniques," *Exp. Syst. Appl.*, vol. 211, Jan. 2023, Art. no. 118636.

[57] G. Tzanetakis and P. Cook, "Musical genre classification of audio signals," *IEEE Trans. Speech Audio Process.*, vol. 10, no. 5, pp. 293–302, Jul. 2002.

[58] (2013). *The MagnaTagATune Dataset*. Accessed: Sep. 30, 2022. [Online]. Available: https://mirg.city.ac.uk/codeapps/the-magnatagatune-dataset

[59] X. Li, H. Xianyu, J. Tian, W. Chen, F. Meng, M. Xu, and L. Cai, "A deep bidirectional long short-term memory based multi-scale approach for music dynamic emotion prediction," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, Mar. 2016, pp. 544–548.

[60] A. Aljanaki, Y.-H. Yang, and M. Soleymani, "Emotion in music task at mediaeval 2015," in *Proc. MediaEval Workshop*, 2015, pp. 1–3.

[61] S. Uhlich, M. Porcu, F. Giron, M. Enenkl, T. Kemp, N. Takahashi, and Y. Mitsufuji, "Improving music source separation based on deep neural networks through data augmentation and network blending," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, Mar. 2017, pp. 261–265.

[62] *SiSEC DSD100*. Accessed: Sep. 30, 2022. [Online]. Available: https://sisec.inria.fr/sisec-2016/2016-professionally-produced-music-recordings/

[63] W. Gong and Q. Yu, "A deep music recommendation method based on human motion analysis," *IEEE Access*, vol. 9, pp. 26290–26300, 2021.

[64] T. Tang, J. Jia, and H. Mao, "Dance with melody: An LSTM-autoencoder approach to music-oriented dance synthesis," in *Proc. 26th ACM Int. Conf. Multimedia*, Oct. 2018, pp. 1598–1606.

[65] R. Romero-Arenas, A. Gómez-Espinosa, and B. Valdés-Aguirre, "Singing voice detection in electronic music with a long-term recurrent convolutional network," *Appl. Sci.*, vol. 12, no. 15, p. 7405, Jul. 2022.

[66] TheFatRat. (2016). *The Arcadium*. Accessed: Sep. 30, 2022. [Online]. Available: https://www.youtube.com/c/TheArcadium

[67] B. Woodford. (2011). *NCS (No Copytight Sounds)—Free Music for Content Creators*. Accessed: Sep. 30, 2022. [Online]. Available: https://www.ncs.io/

[68] M. Ramona, G. Richard, and B. David, "Vocal detection in music with support vector machines," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process.*, Mar. 2008, pp. 1885–1888.

[69] X. Zhang, Y. Yu, Y. Gao, X. Chen, and W. Li, "Research on singing voice detection based on a long-term recurrent convolutional network with vocal separation and temporal smoothing," *Electronics*, vol. 9, no. 9, p. 1458, Sep. 2020.

[70] (2012). *Rwc Pop Music Dataset*. Accessed: Sep. 30, 2022. [Online]. Available: https://staff.aist.go.jp/m.goto/RWC-MDB/

[71] R. M. Bittner, J. Salamon, M. Tierney, M. Mauch, C. Cannam, and J. P. Bello, "MedleyDB: A multitrack dataset for annotation-intensive MIR research," in *Proc. ISMIR*, vol. 14, 2014, pp. 155–160.

[72] *iKala Dataset*. Accessed: Sep. 30, 2022. [Online]. Available: https://paperswithcode.com/dataset/ikala

[73] Z. Li, F. Liu, W. Yang, S. Peng, and J. Zhou, "A survey of convolutional neural networks: Analysis, applications, and prospects," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 33, no. 12, pp. 6999–7019, Dec. 2021.

[74] T. Hirvonen, "Classification of spatial audio location and content using convolutional neural networks," in *Proc. 138th Audio Eng. Soc. Conv.*, 2015, pp. 1–10.

[75] S. Hershey, S. Chaudhuri, D. P. W. Ellis, J. F. Gemmeke, A. Jansen, R. C. Moore, M. Plakal, D. Platt, R. A. Saurous, B. Seybold, M. Slaney, R. J. Weiss, and K. Wilson, "CNN architectures for large-scale audio classification," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, Mar. 2017, pp. 131–135.

[76] Y. Aytar, C. Vondrick, and A. Torralba, "SoundNet: Learning sound representations from unlabeled video," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 29, 2016, pp. 1–9.

[77] C. N. Silla Jr., A. L. Koerich, and C. A. Kaestner, "The Latin music database," in *Proc. ISMIR*, 2008, pp. 451–456.

[78] *Royal Museum of Central-Africa (RMCA)*. Accessed: Sep. 30, 2022. [Online]. Available: https://www.africamuseum.be/en

[79] J. Lee, J. Park, K. Luke Kim, and J. Nam, "Sample-level deep convolutional neural networks for music auto-tagging using raw waveforms," 2017, *arXiv:1703.01789*.

[80] T. Bertin-Mahieux, D. P. Ellis, B. Whitman, and P. Lamere, "The million song dataset," in *Proc. 12th Int. Soc. Music Inf. Retr. Conf.*, 2011, pp. 591–596.

[81] L. Qiu, S. Li, and Y. Sung, "3D-DCDAE: Unsupervised music latent representations learning method based on a deep 3D convolutional denoising autoencoder for music genre classification," *Mathematics*, vol. 9, no. 18, p. 2274, Sep. 2021.

[82] (2004). *The Lakh Midi Dataset*. Accessed: Sep. 30, 2022. [Online]. Available: https://colinraffel.com/projects/lmd

[83] C. Raffel, "Learning-based methods for comparing sequences, with applications to audio-to-midi alignment and matching," Ph.D. thesis, Columbia Univ., New York, NY, USA, 2016, doi: 10.7916/D8N58MHV.

[84] B. Stasiak and J. Mońko, "Analysis of time-frequency representations for musical onset detection with convolutional neural network," in *Proc. Ann. Comput. Sci. Inf. Syst.*, Oct. 2016, pp. 147–152.

[85] H. Phan, L. Hertel, M. Maass, and A. Mertins, "Robust audio event recognition with 1-max pooling convolutional neural networks," 2016, *arXiv:1604.06338*.

[86] S. Nakamura, K. Hiyane, F. Asano, T. Yamada, and T. Endo, "Data collection in real acoustical environments for sound scene understanding and hands-free speech recognition," in *Proc. 6th Eur. Conf. Speech Commun. Technol. (EUROSPEECH)*, 1999, pp. 1–4.

[87] A. Varga and H. J. M. Steeneken, "Assessment for automatic speech recognition: II. NOISEX-92: A database and an experiment to study the effect of additive noise on speech recognition systems," *Speech Commun.*, vol. 12, no. 3, pp. 247–251, Jul. 1993.

[88] K. W. E. Lin, B. T. Balamurali, E. Koh, S. Lui, and D. Herremans, "Singing voice separation using a deep convolutional neural network trained by ideal binary mask and cross entropy," *Neural Comput. Appl.*, vol. 32, no. 4, pp. 1037–1050, Feb. 2020.

[89] A. Liutkus, F.-R. Stöter, Z. Rafii, D. Kitamura, B. Rivet, N. Ito, N. Ono, and J. Fontecave, "The 2016 signal separation evaluation campaign," in *Proc. Int. Conf. Latent Variable Anal. Signal Separat.* Cham, Switzerland: Springer, 2017, pp. 323–332.

[90] Y. Liusong and D. Hui, "Voice quality evaluation of singing art based on 1DCNN model," *Math. Problems Eng.*, vol. 2022, pp. 1–9, Jul. 2022.

[91] P. Li, J. Qian, and T. Wang, "Automatic instrument recognition in polyphonic music using convolutional neural networks," 2015, *arXiv:1511.05520*.

[92] V. Lostanlen and C.-E. Cella, "Deep convolutional networks on the pitch spiral for musical instrument recognition," 2016, *arXiv:1605.06644*.

[93] D. Mukhedkar, "Polyphonic music instrument detection on weakly labelled data using sequence learning models," School Elect. Eng. Comput. Sci., KTH Roy. Inst. Technol., Stockholm, Sweden, 2020.

[94] E. Humphrey, S. Durand, and B. McFee, "OpenMIC-2018: An open data-set for multiple instrument recognition," in *Proc. ISMIR*, 2018, pp. 438–444.

[95] A. Wise, A. S. Maida, and A. Kumar, "Attention augmented CNNs for musical instrument identification," in *Proc. 29th Eur. Signal Process. Conf. (EUSIPCO)*, Aug. 2021, pp. 376–380.

[96] *London Philharmonic Orchestra Dataset*. Accessed: Sep. 30, 2022. [Online]. Available: https://philharmonia.co.uk/resources/sound-samples/

[97] *University of Iowa Musical Instrument Samples*. Accessed: Sep. 30, 2022. [Online]. Available: https://theremin.music.uiowa.edu/MIS.html

[98] M. Blaszke and B. Kostek, "Musical instrument identification using deep learning approach," *Sensors*, vol. 22, no. 8, p. 3033, Apr. 2022.

[99] E. Manilow, G. Wichern, P. Seetharaman, and J. Le Roux, "Cutting music source separation some Slakh: A dataset to study the impact of training data quality and quantity," in *Proc. IEEE Workshop Appl. Signal Process. to Audio Acoust. (WASPAA)*, Oct. 2019, pp. 1–7.

[100] X. Liu, Q. Chen, X. Wu, Y. Liu, and Y. Liu, "CNN based music emotion classification," 2017, *arXiv:1704.05665*.

[101] S.-Y. Wang, J.-C. Wang, Y.-H. Yang, and H.-M. Wang, "Towards time-varying music auto-tagging based on CAL500 expansion," in *Proc. IEEE Int. Conf. Multimedia Expo. (ICME)*, Jul. 2014, pp. 1–6.

[102] T. Ciborowski, S. Reginis, D. Weber, A. Kurowski, and B. Kostek, "Classifying emotions in film music—A deep learning approach," *Electronics*, vol. 10, no. 23, p. 2955, Nov. 2021.

[103] *Epidemic Sound*. Accessed: Sep. 30, 2022. [Online]. Available: https://www.epidemicsound.com/

[104] A. Vall, M. Dorfer, H. Eghbal-zadeh, M. Schedl, K. Burjorjee, and G. Widmer, "Feature-combination hybrid recommender systems for automated music playlist continuation," *User Model. User-Adapted Interact.*, vol. 29, no. 2, pp. 527–572, Apr. 2019.

[105] *Art of the Mix*. Accessed: Sep. 30, 2022. [Online]. Available: http://www.artofthemix.org/

[106] M. Sheikh Fathollahi and F. Razzazi, "Music similarity measurement and recommendation system using convolutional neural networks," *Int. J. Multimedia Inf. Retr.*, vol. 10, no. 1, pp. 43–53, Mar. 2021.

[107] M. Zentner, D. Grandjean, and K. R. Scherer, "Emotions evoked by the sound of music: Characterization, classification, and measurement," *Emotion*, vol. 8, no. 4, pp. 494–521, 2008.

[108] H. Homburg, I. Mierswa, B. Möller, K. Morik, and M. Wurst, "A benchmark dataset for audio classification and clustering," in *Proc. ISMIR*, 2005, pp. 31–528.

[109] H. Gao, "Automatic recommendation of online music tracks based on deep learning," *Math. Problems Eng.*, vol. 2022, pp. 1–8, Jun. 2022.

[110] J. S. Downie, K. West, A. Ehmann, and E. Vincent, "The 2005 music information retrieval evaluation exchange (MIREX 2005): Preliminary overview," in *Proc. 6th Int. Conf. Music Inf. Retr. (ISMIR)*, 2005, pp. 320–323.

[111] K. W. Cheuk, H. Anderson, K. Agres, and D. Herremans, "NnAudio: An on-the-fly GPU audio to spectrogram conversion toolbox using 1D convolutional neural networks," *IEEE Access*, vol. 8, pp. 161981–162003, 2020.

[112] J. Thickstun, Z. Harchaoui, and S. Kakade, "Learning features of music from scratch," 2016, *arXiv:1611.09827*.

[113] B. McFee, C. Raffel, D. Liang, D. Ellis, M. McVicar, E. Battenberg, and O. Nieto, "Librosa: Audio and music signal analysis in Python," in *Proc. 14th Python Sci. Conf.* Pennsylvania, PA, USA: Citeseer, 2015, pp. 18–25.

[114] G. Ian, J. Pouget-Abadie, M. Mirza, B. Xu, and D. Warde-Farley, "Generative adversarial nets," in *Proc. Adv. Neural Inf. Process. Syst.*, 2014, pp. 1–9.

[115] I.-S. Huang, Y.-H. Lu, M. Shafiq, A. Ali Laghari, and R. Yadav, "A generative adversarial network model based on intelligent data analytics for music emotion recognition under IoT," *Mobile Inf. Syst.*, vol. 2021, pp. 1–8, Nov. 2021.

[116] N. Li, "Generative adversarial network for musical notation recognition during music teaching," *Comput. Intell. Neurosci.*, vol. 2022, pp. 1–9, Jun. 2022.

[117] G. Hadjeres, F. Pachet, and F. Nielsen, "DeepBach: A steerable model for Bach chorales generation," in *Proc. Int. Conf. Mach. Learn.*, 2017, pp. 1362–1371.

[118] C. Hawthorne, A. Stasyuk, A. Roberts, I. Simon, C.-Z. A. Huang, S. Dieleman, E. Elsen, J. Engel, and D. Eck, "Enabling factorized piano music modeling and generation with the MAESTRO dataset," 2018, *arXiv:1810.12247*.

[119] C.-Z. Anna Huang, C. Hawthorne, A. Roberts, M. Dinculescu, J. Wexler, L. Hong, and J. Howcroft, "The bach doodle: Approachable music composition with machine learning at scale," 2019, *arXiv:1907.06637*.

[120] Z.-C. Fan, Y.-L. Lai, and J.-S.-R. Jang, "SVSGAN: Singing voice separation via generative adversarial network," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, Apr. 2018, pp. 726–730.

[121] K. Choi, G. Fazekas, M. Sandler, and K. Cho, "Convolutional recurrent neural networks for music classification," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, Mar. 2017, pp. 2392–2396.

[122] A. A. S. Gunawan and D. Suhartono, "Music recommender system based on genre using convolutional recurrent neural networks," *Proc. Comput. Sci.*, vol. 157, pp. 99–109, Jan. 2019.

[123] M. Defferrard, K. Benzi, P. Vandergheynst, and X. Bresson, "FMA: A dataset for music analysis," 2016, *arXiv:1612.01840*.

[124] C. Chen and Q. Li, "A multimodal music emotion classification method based on multifeature combined network classifier," *Math. Problems Eng.*, vol. 2020, pp. 1–11, Aug. 2020.

[125] S. Hizlisoy, S. Yildirim, and Z. Tufekci, "Music emotion recognition using convolutional long short term memory deep neural networks," *Eng. Sci. Technol., Int. J.*, vol. 24, no. 3, pp. 760–767, Jun. 2021.

[126] X. Jia, "Music emotion classification method based on deep learning and improved attention mechanism," *Comput. Intell. Neurosci.*, vol. 2022, pp. 1–8, Jun. 2022.

[127] M. Liang, "Music score recognition and composition application based on deep learning," *Math. Problems Eng.*, vol. 2022, pp. 1–9, Jun. 2022.

[128] (2012). *Musescore*. Accessed: Sep. 30, 2022. [Online]. Available: https://musescore.org/en

[129] G. Parascandolo, H. Huttunen, and T. Virtanen, "Recurrent neural networks for polyphonic sound event detection in real life recordings," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, Mar. 2016, pp. 6440–6444.

[130] T. Heittola, A. Mesaros, A. Eronen, and T. Virtanen, "Audio context recognition using audio event histograms," in *Proc. 18th Eur. Signal Process. Conf.*, 2010, pp. 1272–1276.

[131] O. Bulayenko, J. Quintais, D. J. Gervais, and J. Poort. (2022). *AI Music Outputs: Challenges to the Copyright Legal Framework*. [Online]. Available: https://ssrn.com/abstract=4072806

[132] R. B. Abbott and E. Rothman, "Disrupting creativity: Copyright law in the age of generative artificial intelligence," Aug. 2022. [Online]. Available: https://ssrn.com/abstract=4185327, doi: 10.2139/ssrn.4185327.

[133] S. Mehri, K. Kumar, I. Gulrajani, R. Kumar, S. Jain, J. Sotelo, A. Courville, and Y. Bengio, "SampleRNN: An unconditional end-to-end neural audio generation model," 2016, *arXiv:1612.07837*.

[134] *The Internet Archive*. Accessed: Sep. 30, 2022. [Online]. Available: https://archive.org/

[135] M. S. Cuthbert and C. Ariza, "music21: A toolkit for computer-aided musicology and symbolic music data," in *Proc. 11th Int. Soc. Music Inf. Retr. Conf. (ISMIR)*, 2010, pp. 637–642.

[136] G. Hadjeres and F. Nielsen, "Interactive music generation with positional constraints using anticipation-RNNs," 2017, *arXiv:1709.06404*.

[137] C. Benetatos, J. VanderStel, and Z. Duan, "BachDuet: A deep learning system for human-machine counterpoint improvisation," in *Proc. Int. Conf. New Interfaces Musical Expression*, 2020, pp. 1–6.

[138] P. Hutchings, "Talking drums: Generating drum grooves with neural networks," 2017, *arXiv:1706.09558*.

[139] B. Genchel, A. Pati, and A. Lerch, "Explicitly conditioned melody generation: A case study with interdependent RNNs," 2019, *arXiv:1907.05208*.

[140] *Folkdb*. Accessed: Sep. 30, 2022. [Online]. Available: https://github.com/IraKorshunova/folk-rnn/tree/master/data

[141] A. Pati, A. Lerch, and G. Hadjeres, "Learning to traverse latent spaces for musical score inpainting," 2019, *arXiv:1907.01164*.

[142] *The Session*. Accessed: Sep. 30, 2022. [Online]. Available: https://thesession.org/
[143] S. Agarwal, V. Saxena, V. Singal, and S. Aggarwal, "LSTM based music generation with dataset preprocessing and reconstruction techniques," in *Proc. IEEE Symp. Ser. Comput. Intell. (SSCI)*, Nov. 2018, pp. 455–462.
[144] H. Lim, S. Rhyu, and K. Lee, "Chord generation from symbolic melody using BLSTM networks," 2017, *arXiv:1712.01011*.
[145] *Wikifonia Subset Dataset*. Accessed: Sep. 30, 2022. [Online]. Available: http://marg.snu.ac.kr/chord_generation/
[146] H. H. Tan, "ChordAL: A chord-based approach for music generation using Bi-LSTMs," in *Proc. ICCC*, 2019, pp. 364–365.
[147] *Nottingham Dataset*. Accessed: Sep. 30, 2022. [Online]. Available: https://ifdo.ca/~seymour/nottingham/nottingham.html
[148] *Mcgill-Billboard Chord Annotations*. Accessed: Sep. 30, 2022. [Online]. Available: https://ddmal.music.mcgill.ca/research/SALAMI/
[149] W. Yang, P. Sun, Y. Zhang, and Y. Zhang, "CLSTMS: A combination of two LSTM models to generate chords accompaniment for symbolic melody," in *Proc. Int. Conf. High Perform. Big Data Intell. Syst. (HPBDIS)*, May 2019, pp. 176–180.
[150] H. H. Mao, T. Shin, and G. Cottrell, "DeepJ: Style-specific music generation," in *Proc. IEEE 12th Int. Conf. Semantic Comput. (ICSC)*, Jan. 2018, pp. 377–382.
[151] *Classical Piano-Midi Dataset*. Accessed: Sep. 30, 2022. [Online]. Available: http://piano-midi.de/
[152] C. D. Boom, S. V. Laere, T. Verbelen, and B. Dhoedt, "Rhythm, chord and melody generation for lead sheets using recurrent neural networks," in *Proc. Joint Eur. Conf. Mach. Learn. Knowl. Discovery Databases*. Cham, Switzerland: Springer, 2019, pp. 454–461.
[153] Q. Lyu, Z. Wu, and J. Zhu, "Polyphonic music modelling with LSTM-RTRBM," in *Proc. 23rd ACM Int. Conf. Multimedia*, Oct. 2015, pp. 991–994.
[154] *Musedata*. Accessed: Sep. 30, 2022. [Online]. Available: https://musedata.org/
[155] *Johann Sebastian Bach Chorales Dataset*. Accessed: Sep. 30, 2022. [Online]. Available: https://github.com/czhuang/JSB-Chorales-dataset
[156] D. D. Johnson, "Generating polyphonic music using tied parallel networks," in *Proc. Int. Conf. Evol. Biologically Inspired Music Art*. Cham, Switzerland: Springer, pp. 128–143, 2017.
[157] M. Liang, "An improved music composing technique based on neural network model," *Mobile Inf. Syst.*, vol. 2022, pp. 1–10, Jul. 2022.
[158] K. Choi, J. Park, W. Heo, S. Jeon, and J. Park, "Chord conditioned melody generation with transformer based decoders," *IEEE Access*, vol. 9, pp. 42071–42080, 2021.
[159] R. Yang, D. Wang, Z. Wang, T. Chen, J. Jiang, and G. Xia, "Deep music analogy via latent representation disentanglement," 2019, *arXiv:1906.03626*.
[160] P. S. Yadav, S. Khan, Y. V. Singh, P. Garg, and R. S. Singh, "A lightweight deep learning-based approach for jazz music generation in MIDI format," *Comput. Intell. Neurosci.*, vol. 2022, pp. 1–7, Aug. 2022.
[161] G. Keerti, A. Vaishnavi, P. Mukherjee, A. S. Vidya, G. S. Sreenithya, and D. Nayab, "Attentional networks for music generation," *Multimedia Tools Appl.*, vol. 81, no. 4, pp. 5179–5189, 2022.
[162] *Jazz ML Ready Midi*. Accessed: Sep. 30, 2022. [Online]. Available: https://www.kaggle.com/datasets/saikayala/jazz-ml-ready-midi
[163] O. Yadav, D. Fernandes, V. Dube, and M. D'Souza, "Apollo: A classical piano composer using long short-term memory," *IETE J. Educ.*, vol. 62, no. 2, pp. 60–70, Jul. 2021.
[164] *Classical Music Midi—Kaggle*. Accessed: Sep. 30, 2022. [Online]. Available: https://www.kaggle.com/datasets/soumikrakshit/classical-music-midi
[165] *Midi Classic Music—Kaggle*. Accessed: Sep. 30, 2022. [Online]. Available: https://www.kaggle.com/datasets/blanderbuss/midi-classic-music
[166] D. Makris, M. Kaliakatsos-Papakostas, I. Karydis, and K. L. Kermanidis, "Conditional neural sequence learners for generating drums' rhythms," *Neural Comput. Appl.*, vol. 31, no. 6, pp. 1793–1804, Jun. 2019.
[167] *911TABS*. [Online]. Accessed: Sep. 30, 2022. Available: https://www.911tabs.com/
[168] Z. Cheddad and A. Cheddad, "ARMAS: Active reconstruction of missing audio segments," 2021, *arXiv:2111.10891*.
[169] C. Jin, Y. Tie, Y. Bai, X. Lv, and S. Liu, "A style-specific music composition neural network," *Neural Process. Lett.*, vol. 52, no. 3, pp. 1893–1912, Dec. 2020.
[170] Y. Su, R. Han, X. Wu, Y. Zhang, and Y. Li, "Folk melody generation based on CNN-BiGRU and self-attention," in *Proc. 4th Int. Conf. Commun., Inf. Syst. Comput. Eng. (CISCE)*, May 2022, pp. 363–368.
[171] *ESAC*. Accessed: Sep. 30, 2022. [Online]. Available: http://www.esac-data.org/
[172] Y. Huang, X. Huang, and Q. Cai, "Music generation based on convolution-LSTM," *Comput. Inf. Sci.*, vol. 11, no. 3, pp. 50–56, 2018.
[173] *Midiworld*. Accessed: Sep. 30, 2022. [Online]. Available: https://www.midiworld.com/
[174] S. M. Tony and S. Sasikumar, "Generative adversarial network for music generation," in *High Performance Computing and Networking*. Cham, Switzerland: Springer, pp. 109–119, 2022.
[175] S. Li and Y. Sung, "INCO-GAN: Variable-length music generation method based on inception model-based conditional GAN," *Mathematics*, vol. 9, no. 4, p. 387, Feb. 2021.
[176] G. Brunner, Y. Wang, R. Wattenhofer, and S. Zhao, "Symbolic music genre transfer with CycleGAN," in *Proc. IEEE 30th Int. Conf. Tools Artif. Intell. (ICTAI)*, Nov. 2018, pp. 786–793.
[177] J. Nistal, S. Lattner, and G. Richard, "DrumGAN: Synthesis of drum sounds with timbral feature conditioning using generative adversarial networks," 2020, *arXiv:2008.12073*.
[178] J. Engel, K. Krishna Agrawal, S. Chen, I. Gulrajani, C. Donahue, and A. Roberts, "GANSynth: Adversarial neural audio synthesis," 2019, *arXiv:1902.08710*.
[179] J. Engel, C. Resnick, A. Roberts, S. Dieleman, M. Norouzi, D. Eck, and K. Simonyan, "Neural audio synthesis of musical notes with wavenet autoencoders," in *Proc. Int. Conf. Mach. Learn.*, 2017, pp. 1068–1077.
[180] F. Guan, C. Yu, and S. Yang, "A GAN model with self-attention mechanism to generate multi-instruments symbolic music," in *Proc. Int. Joint Conf. Neural Netw. (IJCNN)*, Jul. 2019, pp. 1–6.
[181] L.-C. Yang, S.-Y. Chou, and Y.-H. Yang, "MidiNet: A convolutional generative adversarial network for symbolic-domain music generation," 2017, *arXiv:1703.10847*.
[182] *Theorytab*. Accessed: Sep. 30, 2022. [Online]. Available: https://www.hooktheory.com/theorytab
[183] H.-W. Dong, W.-Y. Hsiao, L.-C. Yang, and Y.-H. Yang, "MuseGAN: Multi-track sequential generative adversarial networks for symbolic music generation and accompaniment," in *Proc. AAAI Conf. Artif. Intell.*, vol. 32, 2018, pp. 1–8.
[184] L. Yu, W. Zhang, J. Wang, and Y. Yu, "SeqGan: Sequence generative adversarial nets with policy gradient," in *Proc. AAAI Conf. Artif. Intell.*, vol. 31, 2017, pp. 1–7.
[185] S.-G. Lee, U. Hwang, S. Min, and S. Yoon, "Polyphonic music generation with sequence generative adversarial networks," 2017, *arXiv:1710.11418*.
[186] A. Marafioti, P. Majdak, N. Holighaus, and N. Perraudin, "GACELA: A generative adversarial context encoder for long audio inpainting of music," *IEEE J. Sel. Topics Signal Process.*, vol. 15, no. 1, pp. 120–131, Jan. 2020.
[187] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, Ł. Kaiser, and I. Polosukhin, "Attention is all you need," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 30, 2017, pp. 1–11.
[188] C.-Z. A. Huang, A. Vaswani, J. Uszkoreit, N. Shazeer, I. Simon, C. Hawthorne, A. M. Dai, M. D. Hoffman, M. Dinculescu, and D. Eck, "Music transformer," 2018, *arXiv:1809.04281*.
[189] *Piano-E-Competition Dataset*. Accessed: Sep. 30, 2022. [Online]. Available: https://www.piano-e-competition.com/
[190] N. Zhang, "Learning adversarial transformer for symbolic music generation," *IEEE Trans. Neural Netw. Learn. Syst.*, early access, Jul. 2, 2020, doi: 10.1109/TNNLS.2020.2990746.
[191] R. Child, S. Gray, A. Radford, and I. Sutskever, "Generating long sequences with sparse transformers," 2019, *arXiv:1904.10509*.
[192] S. Dieleman, A. V. D. Oord, and K. Simonyan, "The challenge of realistic music generation: Modelling raw audio at scale," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 31, 2018, pp. 1–11.
[193] Y.-S. Huang and Y.-H. Yang, "Pop music transformer: Beat-based modeling and generation of expressive pop piano compositions," in *Proc. 28th ACM Int. Conf. Multimedia*, Oct. 2020, pp. 1180–1188.
[194] P. Dhariwal, H. Jun, C. Payne, J. Wook Kim, A. Radford, and I. Sutskever, "Jukebox: A generative model for music," 2020, *arXiv:2005.00341*.
[195] Y.-J. Shih, S.-L. Wu, F. Zalkow, M. Müller, and Y.-H. Yang, "Theme transformer: Symbolic music generation with theme-conditioned transformer," *IEEE Trans. Multimedia*, early access, Mar. 23, 2022, doi: 10.1109/TMM.2022.3161851.

[196] Z. Wang, K. Chen, J. Jiang, Y. Zhang, M. Xu, S. Dai, X. Gu, and G. Xia, "POP909: A pop-song dataset for music arrangement generation," 2020, *arXiv:2008.07142*.

[197] D. Makris, G. Zixun, M. Kaliakatsos-Papakostas, and D. Herremans, "Conditional drums generation using compound word representations," in *Proc. Int. Conf. Comput. Intell. Music, Sound, Art Design (EvoStar)*. Cham, Switzerland: Springer, 2022, pp. 179–194.

[198] S. Rhyu, H. Choi, S. Kim, and K. Lee, "Translating melody to chord: Structured and flexible harmonization of melody with transformer," *IEEE Access*, vol. 10, pp. 28261–28273, 2022.

[199] *Chord Melody Dataset*. Accessed: Sep. 30, 2022. [Online]. Available: https://github.com/shiehn/chord-melody-dataset

[200] *Hooktheory Lead Sheet Dataset*. Accessed: Sep. 30, 2022. [Online]. Available: https://www.hooktheory.com/

[201] M. Ashraf, G. Geng, X. Wang, F. Ahmad, and F. Abid, "A globally regularized joint neural architecture for music classification," *IEEE Access*, vol. 8, pp. 220980–220989, 2020.

[202] Y. V. Koteswararao and C. B. Rama Rao, "An efficient optimal reconstruction based speech separation based on hybrid deep learning technique," *Defence Sci. J.*, vol. 72, no. 3, pp. 417–428, Jul. 2022.

[203] H. Zhang, S. Kandadai, H. Rao, M. Kim, T. Pruthi, and T. Kristjansson, "Deep adaptive AEC: Hybrid of deep learning and adaptive acoustic echo cancellation," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, May 2022, pp. 756–760.

[204] Y. Ghatas, M. Fayek, and M. Hadhoud, "A hybrid deep learning approach for musical difficulty estimation of piano symbolic music," *Alexandria Eng. J.*, vol. 61, no. 12, pp. 10183–10196, Dec. 2022.

[205] L. Chen, C. Zhao, Y. Liu, and P. Zhuang, "A multi-modal joint voice parts division method based on deep learning," in *Proc. 15th Int. Symp. Med. Inf. Commun. Technol. (ISMICT)*, Apr. 2021, pp. 35–40.

[206] J. Lin, "Integrated intelligent drowsiness detection system based on deep learning," in *Proc. IEEE Int. Conf. Power, Intell. Comput. Syst. (ICPICS)*, Jul. 2020, pp. 420–424.

[207] S. Bisht, H. T. Kanakia, and P. Thakur, "Music emotion prediction based on hybrid approach combining lyrical and acoustic approaches," in *Proc. 6th Int. Conf. Intell. Comput. Control Syst. (ICICCS)*, May 2022, pp. 1656–1660.

[208] N. C. Thompson, K. Greenewald, K. Lee, and G. F. Manso, "The computational limits of deep learning," 2020, *arXiv:2007.05558*.

[209] H. Chen, Y. Wang, C. Xu, B. Shi, C. Xu, Q. Tian, and C. Xu, "AdderNet: Do we really need multiplications in deep learning?" in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2020, pp. 1468–1477.

[210] L. Glitsos, "Vaporwave, or music optimised for abandoned malls," *Popular Music*, vol. 37, no. 1, pp. 100–118, Jan. 2018.

[211] P. Ballam-Cross, "Reconstructed nostalgia: Aesthetic commonalities and self-soothing in chillwave, synthwave, and vaporwave," *J. Popular Music Stud.*, vol. 33, no. 1, pp. 70–93, 2021.

[212] N. Chauhan, "Is it possible to programmatically generate Vaporwave?" IndiaRxiv, Mar. 2020. [Online]. Available: http://indiarxiv.org/9um2r, doi: 10.35543/osf.io/9um2r.

[213] Wikipedia. *List of Cultural and Regional Genres of Music*. Accessed: Sep. 30, 2022. [Online]. Available: https://en.wikipedia.org/wiki/List_of_cultural_and_regional_genres_of_music

[214] S. Shahriar and U. Tariq, "Classifying maqams of Qur'anic recitations using deep learning," *IEEE Access*, vol. 9, pp. 117271–117281, 2021.

[215] P. Kritopoulou, A. Stergiaki, and K. Kokkinidis, "Optimizing human computer interaction for byzantine music learning: Comparing HMMs with RDFs," in *Proc. 9th Int. Conf. Modern Circuits Syst. Technol. (MOCAST)*, Sep. 2020, pp. 1–4.

[216] N. Bassiou, C. Kotropoulos, and A. Papazoglou-Chalikias, "Greek folk music classification into two genres using lyrics and audio via canonical correlation analysis," in *Proc. 9th Int. Symp. Image Signal Process. Anal. (ISPA)*, Sep. 2015, pp. 238–243.

[217] E. Fotiadou, N. Bassiou, and C. Kotropoulos, "Greek folk music classification using auditory cortical representations," in *Proc. 24th Eur. Signal Process. Conf. (EUSIPCO)*, Aug. 2016, pp. 1133–1137.

[218] K. Tsoulou, "Feature-based machine learning techniques towards Greek folk music classification," M.S. thesis, School Sci. Technol., Int. Hellenic Univ., Thermi, Greece, 2020.

[219] N. Farajzadeh, N. Sadeghzadeh, and M. Hashemzadeh, "PMG-Net: Persian music genre classification using deep neural networks," *Entertainment Comput.*, vol. 44, Jan. 2023, Art. no. 100518.

[220] Y. Yang and X. Huang, "Research based on the application and exploration of artificial intelligence in the field of traditional music," *J. Sensors*, vol. 2022, pp. 1–9, Jul. 2022.

[221] X. Liang, Z. Li, J. Liu, W. Li, J. Zhu, and B. Han, "Constructing a multimedia Chinese musical instrument database," in *Proc. 6th Conf. Sound Music Technol. (CSMT)*. Singapore: Springer, 2019, pp. 53–60.

[222] A. K. Sharma, G. Aggarwal, S. Bhardwaj, P. Chakrabarti, T. Chakrabarti, J. H. Abawajy, S. Bhattacharyya, R. Mishra, A. Das, and H. Mahdin, "Classification of Indian classical music with time-series matching deep learning approach," *IEEE Access*, vol. 9, pp. 102041–102052, 2021.

[223] B. S. Gowrishankar and N. U. Bhajantri, "Deep learning long short-term memory based automatic music transcription system for carnatic music," in *Proc. IEEE Int. Conf. Distrib. Comput. Electr. Circuits Electron. (ICDCECE)*, Apr. 2022, pp. 1–6.

[224] D. Makris, I. Karydis, and S. Sioutas, "The Greek music dataset," in *Proc. 16th Int. Conf. Eng. Appl. Neural Netw. (INNS)*, Sep. 2015, pp. 1–7.

[225] *Thrace and Macedonia*. Accessed: Sep. 30, 2022. [Online]. Available: http://epth.sfm.gr/

[226] M. K. Karaosmanoğlu, "A Turkish makam music symbolic database for music information retrieval: SymbTr," in *Proc. 13th Int. Soc. Music Inf. Retr. Conf.* Porto, Portugal: International Society for Music Information Retrieval (ISMIR), Oct. 2012, pp. 223–228.

[227] X. Gong, Y. Zhu, H. Zhu, and H. Wei, "ChMusic: A traditional Chinese music dataset for evaluation of instrument recognition," in *Proc. 4th Int. Conf. Big Data Technol.*, Sep. 2021, pp. 184–189.

[228] P. Cao, "Identification and classification of Chinese traditional musical instruments based on deep learning algorithm," in *Proc. 2nd Int. Conf. Comput. Data Sci.*, Jan. 2021, pp. 1–5.

[229] R. Li and Q. Zhang, "Audio recognition of Chinese traditional instruments based on machine learning," *Cognit. Comput. Syst.*, vol. 4, no. 2, pp. 108–115, Jun. 2022.

[230] K. Xu, "Recognition and classification model of music genres and Chinese traditional musical instruments based on deep neural networks," *Sci. Program.*, vol. 2021, pp. 1–8, Jun. 2021.

[231] J. Li, J. Luo, J. Ding, X. Zhao, and X. Yang, "Regional classification of Chinese folk songs based on CRF model," *Multimedia Tools Appl.*, vol. 78, no. 9, pp. 11563–11584, May 2019.

[232] Q. Chen, W. Zhao, Q. Wang, and Y. Zhao, "The sustainable development of intangible cultural heritage with AI: Cantonese opera singing genre classification based on CoGCNet model in China," *Sustainability*, vol. 14, no. 5, p. 2923, Mar. 2022.

[233] H. Wang, J. Li, Y. Lin, W. Ru, and J. Wu, "Generate Xi'an drum music based on compressed coding," in *Proc. 40th Chin. Control Conf. (CCC)*, Jul. 2021, pp. 8679–8683.

[234] J. Luo, X. Yang, S. Ji, and J. Li, "MG-VAE: Deep Chinese folk songs generation with specific regional styles," in *Proc. 7th Conf. Sound Music Technol. (CSMT)*. Singapore: Springer, 2020, pp. 93–106.

[235] Z. Xu, "Construction of intelligent recognition and learning education platform of national music genre under deep learning," *Frontiers Psychol.*, vol. 13, May 2022, Art. no. 843427.

[236] A. Skoki, S. Ljubic, J. Lerga, and I. Štajduhar, "Automatic music transcription for traditional woodwind instruments sopele," *Pattern Recognit. Lett.*, vol. 128, pp. 340–347, Dec. 2019.

[237] E. A. Retta, R. Sutcliffe, E. Almekhlafi, Y. K. Enku, E. Alemu, T. D. Gemechu, M. A. Berwo, M. Mhamed, and J. Feng, "Kinit classification in Ethiopian chants, azmaris and modern music: A new dataset and CNN benchmark," 2022, *arXiv:2201.08448*.

[238] V. S. Pendyala, N. Yadav, C. Kulkarni, and L. Vadlamudi, "Towards building a deep learning based automated Indian classical music tutor for the masses," *Syst. Soft Comput.*, vol. 4, Dec. 2022, Art. no. 200042.

[239] S. John, M. S. Sinith, R. S. Sudheesh, and P. P. Lalu, "Classification of Indian classical carnatic music based on raga using deep learning," in *Proc. IEEE Recent Adv. Intell. Comput. Syst. (RAICS)*, Dec. 2020, pp. 110–113.

[240] S. T. Madhusudhan and G. Chowdhary, "DeepSRGM—Sequence classification and ranking in Indian classical music with deep learning," in *Proc. 20th Int. Soc. Music Inf. Retr. Conf.*, 2019, pp. 533–540.

[241] S. Nag, M. Basu, S. Sanyal, A. Banerjee, and D. Ghosh, "On the application of deep learning and multifractal techniques to classify emotions and instruments using Indian classical music," *Phys. A, Stat. Mech. Appl.*, vol. 597, Jul. 2022, Art. no. 127261.

[242] A. Krishnan, A. Vincent, G. Jos, and R. Rajan, "Multimodal fusion for segment classification in folk music," in *Proc. IEEE 18th India Council Int. Conf. (INDICON)*, Dec. 2021, pp. 1–7.

[243] S. Chowdhuri, "PhonoNet: Multi-stage deep neural networks for raga identification in Hindustani classical music," in *Proc. Int. Conf. Multimedia Retr.*, Jun. 2019, pp. 197–201.

[244] D. P. Shah, N. M. Jagtap, P. T. Talekar, and K. Gawande, "Raga recognition in Indian classical music using deep learning," in *Proc. Int. Conf. Comput. Intell. Music, Sound, Art Design (EvoStar)*. Cham, Switzerland: Springer, 2021, pp. 248–263.
[245] R. Surana, A. Varshney, and V. Pendyala, "Deep learning for conversions between melodic frameworks of Indian classical music," in *Proc. 2nd Int. Conf. Adv. Comput. Eng. Commun. Syst.* Cham, Switzerland: Springer, 2022, pp. 1–12.
[246] I. Ali-MacLachlan, C. Southall, M. Tomczak, and J. Hockman, "Player recognition for traditional Irish flute recordings," in *Proc. 8th Int. Workshop Folk Music Anal.*, 2018, pp. 3–8.
[247] A. Kolokolova, M. Billard, R. Bishop, M. Elsisy, Z. Northcott, L. Graves, V. Nagisetty, and H. Patey, "GANs & reels: Creating Irish music using a generative adversarial network," 2020, *arXiv:2010.15772*.
[248] B. L. Sturm and O. Ben-Tal, "Folk the algorithms: (Mis) applying artificial intelligence to folk music," in *Handbook of Artificial Intelligence for Music*. Cham, Switzerland: Springer, pp. 423–454, 2021.
[249] J. Lee, M. Lee, D. Jang, and K. Yoon, "Korean traditional music genre classification using sample and midi phrases," *KSII Trans. Internet Inf. Syst.*, vol. 12, no. 4, pp. 1869–1886, 2018.
[250] M. Ebrahimi, B. Majidi, and M. Eshghi, "Procedural composition of traditional Persian music using deep neural networks," in *Proc. 5th Conf. Knowl. Based Eng. Innov. (KBEI)*, Feb. 2019, pp. 521–525.
[251] S. S. Hashemi, M. Aghabozorgi, and M. T. Sadeghi, "Persian music source separation in audio-visual data using deep learning," in *Proc. 6th Iranian Conf. Signal Process. Intell. Syst. (ICSPIS)*, Dec. 2020, pp. 1–5.
[252] E. Hallström, S. Mossmyr, B. Sturm, V. Vegeborn, and J. Wedin, "From jigs and reels to schottisar OCH polskor: Generating Scandinavian-like folk music with deep recurrent networks," in *Proc. 16th Sound Music Comput. Conf.*, Malaga, Spain, May 2019, pp. 1–8.
[253] F. Marchetti, C. Wilson, C. Powell, E. Minisci, and A. Riccardi, "Convolutional generative adversarial network, via transfer learning, for traditional Scottish music generation," in *Proc. Int. Conf. Comput. Intell. Music, Sound, Art Design (EvoStar)*. Cham, Switzerland: Springer, pp. 187–202, 2021.
[254] A. Huaysrijan and S. Pongpinigpinyo, "Automatic music transcription for the Thai xylophone played with soft mallets," in *Proc. 19th Int. Joint Conf. Comput. Sci. Softw. Eng. (JCSSE)*, Jun. 2022, pp. 1–6.
[255] A. Aydingun, D. Bagdatlioglu, B. Canbaz, A. Kokbiyik, M. F. Yavuz, N. Bolucu, and B. Can, "Turkish music generation using deep learning," in *Proc. 28th Signal Process. Commun. Appl. Conf. (SIU)*, Oct. 2020, pp. 1–4.
[256] I. H. Parlak, Y. Çebi, C. Işikhan, and D. Birant, "Deep learning for Turkish makam music composition," *TURKISH J. Electr. Eng. Comput. Sci.*, vol. 29, no. 7, pp. 3107–3118, Nov. 2021.
[257] S. Tanberk and D. B. Tukel, "Style-specific Turkish pop music composition with CNN and LSTM network," in *Proc. IEEE 19th World Symp. Appl. Mach. Intell. Informat. (SAMI)*, Jan. 2021, pp. 181–185.
[258] M. A. Kızrak and B. Bolat, "A musical information retrieval system for classical Turkish music makams," *Simulation*, vol. 93, no. 9, pp. 749–757, Sep. 2017.
[259] M. A. Kizrak and B. Bolat, "Classification of classic Turkish music makams by using deep belief networks," in *Proc. 23rd Signal Process. Commun. Appl. Conf. (SIU)*, May 2015, pp. 1–6.
[260] T. P. Van, N. T. N. Quang, and T. M. Thanh, "Deep learning approach for singer voice classification of Vietnamese popular music," in *Proc. 10th Int. Symp. Inf. Commun. Technol. (SoICT)*, 2019, pp. 255–260.
[261] T. Stegemann, M. Geretsegger, E. Phan Quoc, H. Riedl, and M. Smetana, "Music therapy and other music-based interventions in pediatric health care: An overview," *Medicines*, vol. 6, no. 1, p. 25, Feb. 2019.
[262] M. de Witte, G.-J. Stams, X. Moonen, A. E. R. Bos, and S. van Hooren, "Music therapy for stress reduction: A systematic review and meta-analysis," *Health Psychol. Rev.*, vol. 16, no. 1, pp. 134–159, Nov. 2020.
[263] H. L. Lam, W. T. V. Li, I. Laher, and R. Y. Wong, "Effects of music therapy on patients with dementia—A systematic review," *Geriatrics*, vol. 5, no. 4, p. 62, 2020.
[264] S. Tahmasebi, T. Gajecki, and W. Nogueira, "Design and evaluation of a real-time audio source separation algorithm to remix music for cochlear implant users," *Frontiers Neurosci.*, vol. 14, p. 434, May 2020.
[265] J. Gauer, A. Nagathil, K. Eckel, D. Belomestny, and R. Martin, "A versatile deep-neural-network-based music preprocessing and remixing scheme for cochlear implant listeners," *J. Acoust. Soc. Amer.*, vol. 151, no. 5, pp. 2975–2986, May 2022.
[266] Y.-J. Hong, J. Han, and H. Ryu, "The effects of synthesizing music using AI for preoperative management of Patients' anxiety," *Appl. Sci.*, vol. 12, no. 16, p. 8089, Aug. 2022.
[267] T. Gajęcki and W. Nogueira, "Deep learning models to remix music for cochlear implant users," *J. Acoust. Soc. Amer.*, vol. 143, no. 6, pp. 3602–3615, Jun. 2018.
[268] J. Singh and A. Ratnawat, "Algorithmic music generation for the stimulation of musical memory in Alzheimer's," in *Proc. 4th Int. Conf. Comput. Commun. Autom. (ICCCA)*, Dec. 2018, pp. 1–4.
[269] J. Chen, F. Pan, P. Zhong, T. He, L. Qi, J. Lu, P. He, and Y. Zheng, "An automatic method to develop music with music segment and long short term memory for tinnitus music therapy," *IEEE Access*, vol. 8, pp. 141860–141871, 2020.
[270] Y. Heping and W. Bin, "Online music-assisted rehabilitation system for depressed people based on deep learning," *Prog. Neuro-Psychopharmacology Biol. Psychiatry*, vol. 119, Dec. 2022, Art. no. 110607.
[271] Y. Li, X. Li, Z. Lou, and C. Chen, "Long short-term memory-based music analysis system for music therapy," *Frontiers Psychol.*, vol. 13, Jun. 2022, Art. no. 928048.
[272] G. Kruthika, P. Kuruba, and N. Dushyantha, "A system for anxiety prediction and treatment using Indian classical music therapy with the application of machine learning," in *Intelligent Data Communication Technologies and Internet of Things*. Cham, Switzerland: Springer, pp. 345–359, 2021.
[273] S. Shaila, V. Gurudas, R. Rakshita, and A. Shangloo, "Music therapy for mood transformation based on deep learning framework," in *Proc. Comput. Vis. Robot.* Singapore: Springer, pp. 35–47, 2022.
[274] S. Shaila, T. Rajesh, S. Lavanya, K. Abhishek, and V. Suma, "Music therapy for transforming human negative emotions: Deep learning approach," in *Proc. Int. Conf. Recent Trends Comput.* Singapore: Springer, pp. 99–109, 2022.
[275] Q. Ding, "Evaluation of the efficacy of artificial neural network-based music therapy for depression," *Comput. Intell. Neurosci.*, vol. 2022, pp. 1–6, Aug. 2022.
[276] Z. Hu, Y. Liu, G. Chen, S.-H. Zhong, and A. Zhang, "Make your favorite music curative: Music style transfer for anxiety reduction," in *Proc. 28th ACM Int. Conf. Multimedia*, Oct. 2020, pp. 1189–1197.
[277] E. Idrobo-Ávila, H. Loaiza-Correa, F. Muñoz-Bolaños, L. van Noorden, and R. Vargas-Cañas, "Development of a biofeedback system using harmonic musical intervals to control heart rate variability with a generative adversarial network," *Biomed. Signal Process. Control*, vol. 71, Jan. 2022, Art. no. 103095.
[278] A. E. Coca, G. O. Tost, and L. Zhao, "Characterizing chaotic melodies in automatic music composition," *Chaos, Interdiscipl. J. Nonlinear Sci.*, vol. 20, no. 3, Sep. 2010, Art. no. 033125.
[279] A. E. Coca, D. C. Corrêa, and L. Zhao, "Computer-aided music composition with lstm neural network and chaotic inspiration," in *Proc. Int. Joint Conf. Neural Netw. (IJCNN)*, 2013, pp. 1–7.
[280] M. A. Kaliakatsos-Papakostas, M. G. Epitropakis, A. Floros, and M. N. Vrahatis, "Chaos and music: From time series analysis to evolutionary composition," *Int. J. Bifurcation Chaos*, vol. 23, no. 11, Nov. 2013, Art. no. 1350181.
[281] B. Sobota, F. Majcher, M. Sivy, and M. Hudak, "Chaos simulation and audio output," in *Proc. IEEE 15th Int. Sci. Conf. Informat.*, Nov. 2019, pp. 000137–000142.
[282] E. Berdahl, E. Sheffield, A. Pfalz, and A. T. Marasco, "Widening the razor-thin edge of chaos into a musical highway: Connecting chaotic maps to digital waveguides," in *Proc. Int. Conf. New Interfaces Musical Expression (NIME)*, 2018, pp. 390–393.
[283] M. Skarha, V. Cusson, C. Frisson, and M. M. Wanderley, "Le bâton: A digital musical instrument based on the chaotic triple pendulum," in *Proc. NIME*, 2021, pp. 1–17.
[284] S.-T. Lin and R.-F. Hsu, "Chaotic signal synthesizer applied on portable devices for tinnitus therapy," in *Proc. Int. Symp. Intell. Signal Process. Commun. Syst. (ISPACS)*, Nov. 2021, pp. 1–2.
[285] J.-M. Chen, P.-Y. He, and F. Pan, "Research on synthesizing music for tinnitus treatment based on chaos," in *Proc. 12th Int. Conf. Signal Process. (ICSP)*, Oct. 2014, pp. 2286–2291.
[286] T.-L. Liao, H.-C. Chen, C.-Y. Peng, and Y.-Y. Hou, "Chaos-based secure communications in biomedical information application," *Electronics*, vol. 10, no. 3, p. 359, Feb. 2021.

[287] E. Bollt, "On explaining the surprising success of reservoir computing forecaster of chaos? The universal machine learning dynamical system with contrast to VAR and DMD," *Chaos, Interdiscipl. J. Nonlinear Sci.*, vol. 31, no. 1, Jan. 2021, Art. no. 013108.

[288] D. Turnbull, L. Barrington, D. Torres, and G. Lanckriet, "Towards musical query-by-semantic-description using the CAL500 data set," in *Proc. 30th Annu. Int. ACM SIGIR Conf. Res. Develop. Inf. Retr.*, 2007, pp. 439–446.

LAZAROS MOYSIS received the B.Sc., M.Sc., and Ph.D. degrees from the Department of Mathematics, Aristotle University of Thessaloniki, Greece, in 2011, 2013, and 2017, respectively. He is currently a Researcher with the Physics Department, Aristotle University of Thessaloniki, and the Laboratory of Nonlinear Systems, Circuits and Complexity. His research interests include the theory of control systems, descriptor systems, chaotic systems, and their applications (notable examples include observer design, synchronization, chaotification, chaos encryption, and chaotic path planning).

LAZAROS ALEXIOS ILIADIS (Graduate Student Member, IEEE) received the B.Sc. degree in physics and the M.Sc. degree in electronic physics (radioelectrology) from the Aristotle University of Thessaloniki, in 2017 and 2021, respectively, where he is currently pursuing the Ph.D. degree. His research interests include development of the sixth-generation communications systems (6G), antenna design and electromagnetics, artificial intelligence techniques (evolutionary algorithms, machine learning, and deep learning methods), and computer vision.

SOTIRIOS P. SOTIROUDIS received the B.Sc. degree in physics and the M.Sc. degree in electronics from the Aristotle University of Thessaloniki, in 1999 and 2002, respectively, the B.Sc. degree in informatics from Hellenic Open University, in 2011, and the Ph.D. degree in physics from the Aristotle University of Thessaloniki, in 2018. From 2004 to 2010, he worked with the Telecommunications Center, Aristotle University of Thessaloniki. From 2010 to 2022, he worked as a Teacher of physics and informatics with the Greek Ministry of Education. He joined the Department of Physics, Aristotle University of Thessaloniki, in 2022, where he has been involved in several research projects. His research interests include wireless communications, radio propagation, optimization algorithms, computer vision, and machine learning.

ACHILLES D. BOURSIANIS (Member, IEEE) received the B.Sc. degree in physics, the M.Sc. degree in electronic physics (radioelectrology) in the area of electronics telecommunications technology, and the Ph.D. degree in telecommunications from the School of Physics, Aristotle University of Thessaloniki, in 2001, 2005, and 2017, respectively.

Since 2019, he has been a Postdoctoral Researcher and an Academic Fellow with the School of Physics, Aristotle University of Thessaloniki. He is currently a member of the ELEDIA@AUTH Research Group. He is the author or coauthor of more than 70 articles in international peer-reviewed journals and conferences. His research interests include wireless sensor networks, the Internet of Things (IoT), antenna design and optimization, 5G and beyond communication networks, radio frequency energy harvesting, and artificial intelligence.

Dr. Boursianis is a member of the Hellenic Physical Society and the Scientific Committee of the National Association of Federation des Ingenieurs des Telecommunications de la Communaute Europeenne (FITCE). He is a member of the Editorial Board of the *Telecom* journal. He serves as a reviewer for several international journals and conferences and as a member of the technical program committees for various international conferences, which are technically sponsored by IEEE.

MARIA S. PAPADOPOULOU (Member, IEEE) received the B.Sc. degree in physics, the M.Sc. degree in electronics, and the Ph.D. degree in nonlinear circuits from the School of Physics, Aristotle University of Thessaloniki (AUTh). She is currently an Assistant Professor with the Department of Information and Electronic Engineering, International Hellenic University. She is also a member of the ELEDIA@AUTH Research Group, ELEDIA Research Center Network. She has authored or coauthored several peer-reviewed journals and conferences. Her research interests include RF energy harvesting, wireless sensor networks, the Internet of Things, nonlinear dynamics, and electronic design and optimization. She is a member of the Hellenic Physical Society. She serves as a reviewer for several international journals and conferences and as a member of the technical program committees for various international conferences.

KONSTANTINOS-IRAKLIS D. KOKKINIDIS received the B.A. degree from Hellenic Open University and the M.B.A. and Ph.D. degrees from the University of Macedonia, Thessaloniki, Greece. He is currently a Special Teaching/Technical Personnel with the Department of Applied Informatics, University of Macedonia. He has published numerous articles in academic conferences and journals, such as the International Conference on Modern Circuits and Systems Technologies (MOCAST) on Electronics and Communications, the International Conference on Movement and Computing (MOCO), and the *International Journal of Mechanical and Mechatronics Engineering* (IJMME-IJENS). His research interests include human-centered computing with special interests in human–computer interaction, machine learning, the Internet of Things (IoT), gesture and audio signal processing and identification, and sensorimotor learning, with a focus on sound and image processing.

CHRISTOS VOLOS received the Diploma degree in physics, the M.Sc. degree in electronics, and the Ph.D. degree in chaotic electronics from the Physics Department, Aristotle University of Thessaloniki, Greece, in 1999, 2002, and 2008, respectively. He is currently an Associate Professor with the Physics Department, Aristotle University of Thessaloniki. He is also a member of the Laboratory of Nonlinear Systems, Circuits and Complexity, Physics Department, Aristotle University of Thessaloniki. His current research interests include the design and study of analog and mixed signal electronic circuits, chaotic electronics and their applications (secure communications, cryptography, and robotics), experimental chaotic synchronization, chaotic UWB communications, and measurement and instrumentation systems.

PANAGIOTIS SARIGIANNIDIS (Member, IEEE) received the B.Sc. and Ph.D. degrees in computer science from the Aristotle University of Thessaloniki, Thessaloniki, Greece, in 2001 and 2007, respectively. He has been an Associate Professor with the Department of Electrical and Computer Engineering, University of Western Macedonia, Kozani, Greece, since 2016. He has been involved in several national, European, and international projects. He is currently the Project Coordinator of three H2020 projects, namely, a) H2020-DS-SC7-2017 (DS07-2017), "SPEAR: Secure and PrivatE smArt gRid," and H2020-LC-SC3-EE2020-1 (LC-SC3-EC-4-2020); b) "EVIDENT: bEhaVioral Insgihts anD Effective eNergy policy acTions" and H2020-ICT-2020-1 (ICT-56-2020); and c) "TERMINET: nexT gEneRation sMart INterconnectEd ioT," while he coordinates the Operational Program "MARS: sMart fArming With dRoneS" (Competitiveness, Entrepreneurship, and Innovation). He serves as a Principal Investigator for the H2020-SU-DS-2018 (SU-DS04-2018-2020), "SDN-microSENSE: SDN-microgrid reSilient Electrical eNergy SystEm," and the Erasmus+ KA2 "ARRANGE-ICT: pArtneRship foR AddressiNG mEgatrends in ICT" (Cooperation for Innovation and the Exchange of Good Practices). He has published over 180 papers in international journals, conferences, and book chapters, including the IEEE Communications Surveys and Tutorials, the IEEE Internet of Things Journal, the IEEE Transactions on Broadcasting, the IEEE Systems Journal, the *IEEE Wireless Communications Magazine*, the IEEE/OSA Journal of Lightwave Technology, IEEE Access, and *Computer Networks*. His research interests include telecommunication networks, the Internet of Things, and network security. He participates in the editorial boards of various journals, including the *International Journal of Communication Systems* and the *EURASIP Journal on Wireless Communications and Networking*.

SPIRIDON NIKOLAIDIS (Senior Member, IEEE) received the Diploma and Ph.D. degrees in electrical engineering from Patras University, Greece, in 1988 and 1994, respectively. Since September 1996, he has been with the Department of Physics, Aristotle University of Thessaloniki, Greece, where he is currently a Full Professor. From 2003 to 2017, he was also a contract teaching staff of Hellenic Open University. He has worked in the areas of digital circuits and system design. He is the author or coauthor of more than 200 scientific articles in international journals and conference proceedings, while his work has more than 2300 references (Google Scholar, H-index=23). Two articles presented at international conferences achieved honorary awards. His current research interests include the design of high-speed and low-power digital circuits and embedded systems, modeling the operations of basic CMOS structures, modeling the power consumption of embedded processors, and development of algorithms for leak detection and localization in pipelines. He was a member of the organization committees of three international conferences. He is the founder and organizer of the Annual International Conference on Modern Circuit and System Technologies (MOCAST) since 2012. He also organized the 27th International Symposium on Power and Timing Modeling, Optimization and Simulation (PATMOS), in 2017. He contributes or has contributed to a number of research projects funded by the European Union and the Greek Government, for many of which he has scientific responsibility.

SOTIRIOS K. GOUDOS (Senior Member, IEEE) received the B.Sc. degree in physics, the M.Sc. degree in electronics, the Ph.D. degree in physics, and the Diploma degree in electrical and computer engineering from the Aristotle University of Thessaloniki, Thessaloniki, Greece, in 1991, 1994, 2001, and 2011, respectively, and the M.Sc. degree in information systems from the University of Macedonia, Greece, in 2005.

He is currently an Associate Professor with the Department of Physics, Aristotle University of Thessaloniki. He is also the Director of the ELEDIA@AUTH and a Laboratory Member of the ELEDIA Research Center Network. He has participated in more than 16 national and European-funded projects and has been a principal investigator of five national funded research projects. He is the author of the book titled *Emerging Evolutionary Algorithms for Antennas and Wireless Communications* (The Institution of Engineering and Technology, 2021). His research interests include antenna and microwave structures design, evolutionary algorithms, wireless communications, machine learning, and semantic web technologies.

Prof. Goudos is a member of the IEICE, the Greek Physics Society, the Technical Chamber of Greece, and the Greek Computer Society. He is also a member of the editorial boards of the *International Journal of Antennas and Propagation* (IJAP), the *EURASIP Journal on Wireless Communications and Networking*, and the *International Journal on Advances on Intelligent Systems*. He is also a member of the Topic Board of the open access journal *Electronics*. He has also served as a member of the technical program committees for several IEEE and non-IEEE conferences. He is the founding Editor-in-Chief of the open access journal *Telecom* (MDPI publishing). He is serving as an Associate Editor for the IEEE Transactions on Antennas and Propagation, IEEE Access, and the IEEE Open Journal of the Communication Society. He was honored as an IEEE Access Outstanding Associate Editor, in 2019, 2020, and 2021. He has participated as a guest editor or a lead guest editor of more than 20 special issues of international journals. He has co-organized four special sessions in international conferences. He is also serving as the Chapter/AG Coordinator for the IEEE Greece Section. He has been elected as the IEEE Greece Section Secretary, in 2022.