Proceedings of the Second International Conference on Automation, Computing and Renewable Systems (ICACRS-2023)
IEEE Xplore Part Number: CFP23CB5-ART; ISBN: 979-8-3503-4023-5

2023 2nd International Conference on Automation, Computing and Renewable Systems (ICACRS) | 979-8-3503-4023-5/23/$31.00 ©2023 IEEE | DOI: 10.1109/ICACRS58579.2023.10405241

Advancements in Music Generation: Leveraging RNN and LSTM Networks for Automated Composition

Pragya Gupta
Apex Institute of Technology (CSE)
Chandigarh University
India
pragyagupta1012@gmail.com

Pulkit Dwivedi
Apex Institute of Technology (CSE)
Chandigarh University
India
pdwivedi1990@gmail.com

***Abstract*—Historically, music has been considered an analogue signal and was manually generated by humans. However, in contemporary times, technology has made it possible to automatically create instrumental compositions without mortal intervention. To achieve this, specific challenges need to be addressed, which are thoroughly discussed in this paper. The paper begins with a concise introduction to music and its various factors, citing and analyzing related work conducted by different authors. The primary objective of this paper is to propose an algorithm based on recurrent neural networks (RNN), specifically long short-term memory (LSTM) networks, for the generation of musical notes. By implementing this algorithm, we aim to explore the potential of machine-generated music and its implications in the modern era.**

***Index Terms*—Music generation, Signal Processing, Harmonic progression, Recurrent Neural Network, Long Short Term Memory, Generative Adverserial Networks.**

I. INTRODUCTION

Composing music is an art , even act of playing any music is a labor-intensive endeavour for humans. It is difficult and futile to create an algorithm that can complete both tasks simultaneously at this position of complexity and abstraction. In order to extract all the meaningful musical patterns , it is simpler to represent this as a problem of learning using composed music as training data.

To generate a composition of sounds with continuity and unity, there must be a temporal link between them . To put it another way, a musical note is any sound produced by a musical instrument or a human voice. A simple unit of music is a musical note. There are some characteristics about the quality and performance of music and its notes.Nowadays for professional music producers , the mean time taken to produce a song is at least one day, with the producer's mind put to song [9]. The sound input used to train the artificial intelligence model can be polyphonic or monodic, with each sound representing a different melodic line. The interval of pitches known as an octave is used to represent the musical notes [8]. A pitch class that describes the relative octave position is related to pitch. The model must be trained using specifics from the music, and the input's characteristics determine the model's complexity and output.

The trained model is expected to remember previous information and produce a logical piece. By playing notes with various frequencies, music is created while preserving the relationships between the notes. Genetic algorithms are one way to create music by using pre-existing music [12]. As mentioned, a genetic algorithm can emphasise each fragment's powerful rhythm and merge them into unique musical compositions. But because to the delay in each of its iterations, it is inefficient. [10] Additionally, the background is missing, making it challenging to understand the coherence and underlying rhythm [11]. As a result, we need a systemto solve the aforementioned issue, that should be able to recall the prior note sequence and anticipate the following sequence , and so forth. RNN is used and especially LSTM a special RNN is used [22] [24].

The field of automated music creation now relies heavily on deep learning. Numerous deep learning architectures have been researched for objectives including producing pop music and developing original melodies that sound like classical pieces (Pachet, and Nielsen 2017). A variety of commercial music generation systems are available, and there are tools available to support musicians in their creative process.

The two main categories of generation tools now in use are raw and symbolic models of audio. Symbolic approaches learn and work at notes of the music; they generate new melodies as MIDI or related outputs depending on current training data's melodies and notes.

RNN, especially LSTM networks [4], are a common strategy in this field (LSTMs). The LSTMs' ability to reasonably accurately capture medium-scale melodic structure in music is a benefit of this method. They also frequently produce new melodies quickly, however the tunes they produce must be interpreted by a human or synthesiser. This study develops a neural network technique based on long short term memory networks that can be used to create melodies and music automatically, without any need of human input. The main task is to build a model that can study a set of musical notes, learn, and after that produce a perfect set of notes.This job presents a significant difficulty because the proposed model must be able to recollect previous information and structure of all musical notes for future learning sequences.Model must be taught all the required originally originated sequences of music next to the previous one and transformed for the learning system.

Our objective is to produce music that is enjoyable to listen to but may not necessarily sound like music played by humans. We anticipate that the learning algorithm will locate areas where music sounds well without imposing any requirements that it follow to rules of musical theory. For the same reason, here features as notes, notation and musical chords are not used to aid the learning and generation process, instead of it we are directly dealing with the end result which are called audio waveforms.

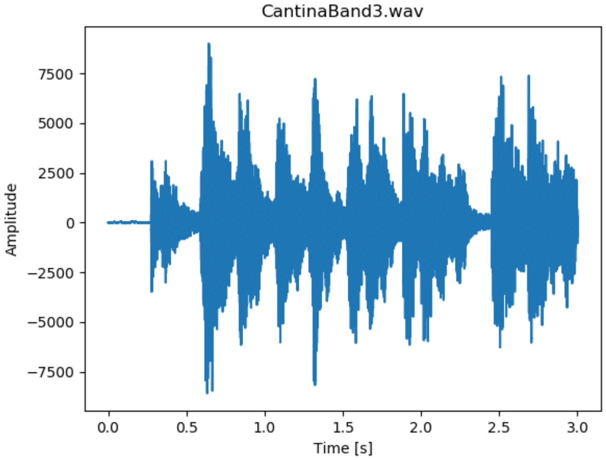

Fig. 1. visualization of sample audio waveform

Fig1 shows audio waveforms as one-dimensional signals that change with time in such a way that audio fragments from different timesteps transition smoothly from one another. A straightforward representation of the raw audio waveform is shown in Fig. 1. A recurrent neural network would be a logical choice of design to simulate a time-varying function due to its capacity to share parameters over time. To represent the signals, we will specifically use Long Short Term Memory (LSTM) networks.

There are numerous efforts to robotically compose tune. The general techniques are categorised into classes:

1) rule-primarily based totally computerized composition with musical grammar and 2) procedural composition with mathematical fashions inclusive of fractal, cell automata or L-system. However, the preceding techniques required customers to have targeted information approximately tune. Otherwise, the preceding techniques couldn't generate tune with numerous styles. Therefore, the excessive stage manage over the automatic composition of numerous tune has not been correctly achieved.

II. Related Work

Numerous studies have employed LSTMs to create music utilising musical elements including notes, chords, and notations. These studies offer encouraging findings and show that LSTMs can collect the necessary long-range data for music production. In these methods, a common architecture entails structured input of a music notation from MIDI files that is fed into an LSTM at each timestep, chen et al [1] . The next timestep's encoding is predicted by the LSTM, and so on. Negative log loss between the projected value and the actual value is used to calculate the error, Grandhe et al [2]. These methods completely rule out the possibility of noise since, during generation, the predicted notes are mapped one to one to their respective limited audio vocabulary. The LSTM can afford a larger cell size since training is done on very low dimensional vectors, extending the time-range over which they may learn. The selection of log loss also makes it simpler to train the architectures. These methods can provide reasonably enjoyable audio as a result. However, because the outputs are constrained to the low dimensionality of the input vector that the model operates on, they are limited in the type of audio they may produce. When networks operate with raw audio data, these limitations are entirely lifted, Vasanth et al [2] .

The use of machine learning (ML) or deep learning (DL) methodologies in computer music has advanced significantly over the years [20] [21] [23] [25]. Li introduced the Daubechies Wavelet Coefficient Histograms feature extraction approach for categorising the genres of music . This method concurrently extracted local and global information from music signals by constructing histograms on the Daubechies wavelet coefficients of the signals. Additionally, Li and Ogihara looked into hierarchical categorization using taxonomies and showed how we have to use linearly plotted discriminant projection to automatically create a genre taxonomies based on confusion matrix . In a different study, Kalingeri et al [2], looked at temporal feature integration, which used a multivariate auto regressive feature of this model to combine all the feature vectors in given time window into a single feature vector to capture the pertinent temporary information

.

Additionally, Panagakis et al [30] used a Support Vector Machine (SVM) in a multilinear viewpoint to solve classification challenges while extracting multiscale spectro-temporal modulation data . They were motivated by a model of auditory cortex processing. In 2018, Bahuleyan contrasted the performance of four conventional machine learning classifiers that used hand-crafted features from the time and frequency domain with a deep learning technique, where a CNN model was trained end-to-end to predict just using spectrograms . In addition, it should be noted that after years of research, scientists gradually discovered a potent instrument for computer music processing, namely the spectrogram analysis.

Piano music rolls, text representations (such as ABC notation4), chord representations (such as Chord2Vec (Madjiheurem, Qu, and Walder 2016), and lead sheet representations are additional representations that have been used.In these models, it is customary to teach and generate music using the exact representation [31]; for ex- , one must train using a set of Musical Instrument Digital Interface files that encode melodies and then produce new Musical Instrument Digital Interface melodies using the learnt model,Kalingeri et al [2].

Numerous novel, models and techniques in the music generation were put out following the AI breakthrough [13]. Various AI-enabled techniques , these include probabilistic model that uses recurrent neural network ,RNN-Anticipation, and recursive artificial neural networks ,and an evolved form of artificial neural networks, for easy generation of the next note, the next note's duration, and the rhythm. Musical notes are produced by generative adversarial networks (GANs), which can contain two neural networks: a generator network that generates some random data and a breaking network that assesses the veracity of the generated random data in comparison to the original data (dataset). A generative network called MuseGAN creates musical compositions with several symbolic tracks,Mangal et al [3].

There exist no not unusual place guidelines for song in order that discords someday sound freshand nice. In the consequence, the automated composition of song subsequently fails when the musical grammar is overstressed, modak et al [3].

Early attempts to create raw audio included models primarily intended to create images, like char-rnn and LSTMs. These networks' raw audio generating outputs are typically loud and unorganised, and their capacity is constrained.To abstract more advanced audio representations, primarily a result of overfitting issues (Briot, Hadjeres, and Pachet 2018),Rachel et al [6]. It is clear from modelling polyphonic music that the chance of additional notes occurring at the same time is significantly changed when one note occurs at a specific time. In other words, because it would be exceedingly expensive to list all possible configurations of the variable to forecast, notes that appear together in linked patterns cannot be effectively characterised by a standard RNN architecture developed for multi-class classification tasks. This problem inspires energy-based models, such the restricted Boltzmann machine (RBM) [Smolensky, 1986], that enable us to represent the log-likelihood of a given configuration by any energy function, Choi et al [8].

Generative Adversarial Networks (GAN) can solve a variety of issues that are comparable [13]. GAN may be used to create sequence generation models when combined with reinforcement learning (RL) [14]. With authors seeking to replace the prior generator discriminator game with three players, GAN already excels at current face synthesis difficulties . N. Sadoughi and C. Busso claim that GAN is capable of producing results in a sequential order [16] . The results of GAN in the field of generation models have been demonstrated in numerous articles . Additionally, Dong et al [17]. introduced three models for symbolic multitrack music production within the GANS framework, trained the models on over 100,000 rock music bars, and then used them to produce piano-rolls comprising five tracks: drums,brass, guitar, piano, and strings [17]. They demonstrated that the models could produce four bars of cohesive music from scratch . We think that GAN is reasonably capable when employed in music generation, taking into account the experiments conducted by others, Rong et al [1].

HMMs are one of the most widely used techniques for modelling and forecasting sequences. Given the sequence of the hidden states that determine the visible states, HMMs are predicated on the assumption that k = 1 (also known as the Markov assumption). Bach's chorales are learned using an HMM in , where 229 and 153 chorales are utilised for training and assessment, respectively. Choral harmonisation is produced after learning the chorales. In order to assist non-musicians in creating music with an HMM, chord progressions are created in to go along with a melody. 298 lead sheets, comprising pop, rock, RB, jazz, and country music, make up the HMM's training set. The system creates chords for the prediction using a 6262 chord transition probability matrix, Choi et al [8]. Given the data, computer power, and workable optimisation methodologies, HMMs were indeed the most effective for time-series modelling [7]. The 1-of-K strategy of its hidden states' inefficiency is one of HMMs' weaknesses, though. When there are N hidden states, the memory of the HMM is limited to log2(N) bits, necessitating the learning of N2 transition matrix parameters, Rachel, et al [5].

In early 2016, Deep Mind introduced WaveNet, that was generative model basically for the raw general audio, designed in very low cost for all kind of speech applications. Looking at a very high stairs , WaveNet may be a deep learning design that operates directly on a raw audio wave shape. especially, for a waveform sculptural by x = disturbance, the probability of the complete wave shape can be factorized as a product of conditional possibilities,namely

$$p(x) = \prod_{i=a}^{T} p(xt|x1, ..., xt - 1) \qquad (1)$$

Recently, Nayebi et al.have also worked with audio samples, although they focus on the frequency domain of the audio rather than trying to learn and generate from the raw audio samples. Because the network can train and predict using a collection of samples that make up the frequency domain rather than just one sample, this method is substantially faster [7]. There are no limitations on the type of music that can be produced by the frequency domain because it can still represent all audible audio signals. The samples in the Fourier domain are given as input at each timestep into a single LSTM architecture that they employ.The output of the LSTM is a Fourier representation of the signal for the following timestep.

The cost function for training the network is the mean squared difference between the anticipated output and the true frequency representation. We used this approach as the standard for all comparisons because it produced satisfactory results and required less work,Grandhe et al [2].

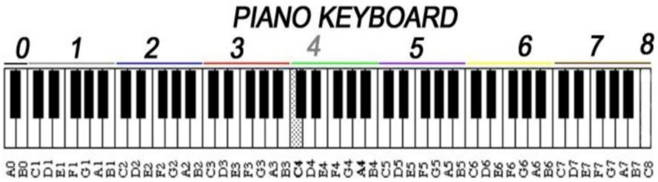
PIANO KEYBOARD
0
1
2
3
4
5
6
7
8
A0 B0 C1 D1 E1 F1 G1 A1 B1 C2 D2 E2 F2 G2 A2 B2 C3 D3 E3 F3 G3 A3 B3 C4 D4 E4 F4 G4 A4 B4 C5 D5 E5 F5 G5 A5 B5 C6 D6 E6 F6 G6 A6 B6 C7 D7 E7 F7 G7 A7 B7 C8

Fig. 2. 88 keys and 7 octaves to make music from it

III. Proposed Methodology

A. Data Source

For generating musical notes the use of a easy recurrent neural network (RNN), we are able to train a model ,using group of piano MIDI documents from the MAESTRO dataset [27]. The MAESTRO dataset is primarily based totally on recordings from the International Piano-e-Competition, a piano overall performance opposition wherein virtuoso pianists carry out on Yamaha Disklaviers with an incorporated MIDI seize system [18], [19].Given a chain of notes, your model will discover ways to expect the following word withinside the sequence. We can generate longer sequences of notes through calling the model repeatedly [26].

B. Model

Machine learning is now getting used for lots thrilling packages in a lot of fields. Music era is one of the thrilling packages of device mastering. As tune itself is sequential data, it could be modelled the use of a sequential device mastering version consisting of the recurrent neural network. This modelling can assist in mastering the tune collection and producing the collection of tune data. In this article, we're going to speak about how we are able to use neural networks, specially recurrent neural networks for computerized tune era. In a programming context, we are able to take into account song as information which can provide many insightful outcomes with the aid of using processing it with diverse procedures. Before processing it, we're required to recognize what kind of records a song report can consist of. First of all, a bit of very simple records approximately any audio or song report is that it could be made of 3 parts: 1) Pitch: It is a degree of the lowness or highness of the sound. 2) Notes: There may be seven forms of notes in a song report which may be expressed as A, B, C, D, E. F, AND G. 3) Octave: In song, we use the octave to give an explanation for the pitch variety of any word. Like what's the degree of the highness of the word A with inside the song. The photograph given in fig.2 is a illustration of the piano keyboard in which we will use 88 keys and seven octaves to make song from it.

C. Data Processing

All the tune documents in our studies are with inside the layout of MIDI (musical tool virtual interface), that is the maximum enormous fashionable tune layout in tune arrangement. MIDI statistics tune with virtual indicators of notes, like commands for pitch and volume. Having uploaded tune data to our model, we rework them right into a single array of notes encoded as a tuple of pitch and chord. You also can generate an audio document via way of means of processing one of the different layers of tune records representations. The processing is then referred to as synthesis, because the application generates every pattern in step with the shape of every other layer, like as an example the MIDI representation, which tells what notes to play and while to play them.

D. Training and Testing

- If once a model is trained symbolic melody is attained and then we treat it as an alternate time series with our symbolic and audio model(similar we have alternate time series with asked textbook to be presented in the domain of speech). In this particular, in the Wave Net model, each sub caste features is a reopened activation unit. If x is n audio input , also at each sub caste k, it passes from following activation unit.

$$z = \tan h(Wf,\ kx)\sigma(Wg,\ kx) \quad (2)$$

Here in this equation * is convolution operator, is known as an element wise multiplication operator, (·) is called as a sigmoid function, and Wf, wg and k , k are memorable convolution filters.

- Following WaveNet's uses local conditioning, we can introduce a second time series y(in this case from the long short term memory model, to capture the long-term melody), and rather use the following activation, effectively incorporating y as an extra input.

$$z = \tan h(W * f, kx + V * f, ky)\sigma(W * g,\ kx + V * g,\ ky) \quad (3)$$

here V is considered as linear projections. By conditioning on extra time series input, we basically make the raw audios as a generation to require certain characteristics , y influences the output at all timestamps.

IV. Experimental Results

If any kind of generated piece of music can only be subjectively estimated by mortal listeners, it's grueling to quantitatively estimate the generations from this model. For case, to compare the labors of two systems, one can listen to extracts of both the music and can rate which of the two sounds is more natural(or whichever parameter is asked to be estimated).

We are using three different variables to represent a note while training the mode they are 1) pitch 2)step 3) duration. Pitch in any music is the eternal quality of this sound as a Musical Instrument Digital Interface (MIDI) note number. Basic step is the time ceased from the previous note or start of the track. The duration is time in how much the note will

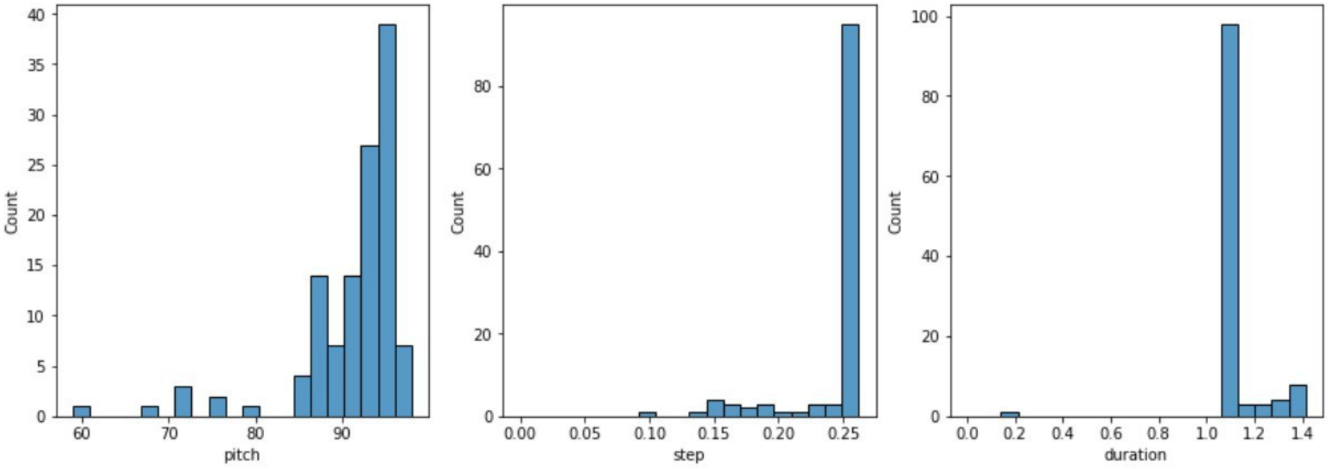

Fig. 3. distributions of pitch, step and duration

be playing in secs and is difference between the note end and starting times.

For visualizing a piece of music , here we are plotting the note pitch, starting and ending across the length of music track i.e a piano roll.

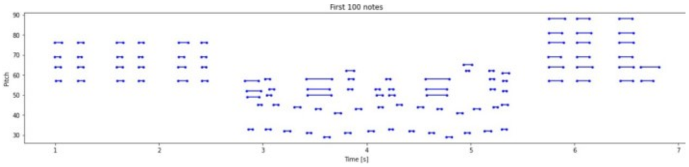

Fig. 4. First 100 notes

Figure 4 shows the starting first 100 notes of music files in graphical form .

Plotting graph for the notes of whole track of music piece. In a given pitch of note, a sample is drawn from the soft max distribution of notes which are produced by the model, and it doesn't just simply pick from the note with the loftiest probability. Always picking the note with the loftiest probability can lead to repetitious sequences of the notes being generated here.

In Figure 3 plots, you can notice the change in the way of distribution of the note variables. Since there's a feedback circle between the model's labors and inputs, the model keeps inducing analogous sequences of labors to reduce the loss. This is specially applicable for the step and duration, which uses the Mean Squared Error loss. Pitch's randomness can be increased by adding temperature in predicting next note.

V. CHALLENGES

The largest venture whilst doing any artificial intelligence layout is that the delicacy is noway one hundred and as a consequence the labors generated aren't usually authentic and correct, they may be naked prognostications which are formulated, which might also additionally or won't have an effect on into a real Affair. The affair generated on this layout won't usually have a right semantic hyperlinks among the phrases of the these days generated song lyrics. The 2d largest venture is of crisp records this is had to fed into the version this is created. Lot of records this is to be had second is uncooked and want a variety of preprocessing earlier than the usage of them. Lack of computational strength is also handled as a trouble to do any form of gadget literacy layout.

VI. CONCLUSION

Automatic music generation have always been an intriguing content for experimenters across globe. All our proposed idea in below given literature have a pivotal advantages of their own as well as limitations as well. thus no system is perfect as no mortal ever is.Markov model that's used in this design has a python perpetration.

This paper achieves the purpose of designing a version which may be used to generate tune and melodies automatically with none human intervention. The version is succesful to remember the preceding info of the dataset and generate a polyphonic tune the use of a unmarried layered LSTM version, talented sufficient to research harmonic and melodic note collection from MIDI documents of Pop tune [25]. The version layout is defined with a belief of capability and adjust ability. Induction and technique of schooling dataset for tune era is accomplished via this paintings. Furthermore, evaluation of the version is likewise imitate for higher insights and understand ability. Enrichment of version feasibility and beyond and gift opportunities also are mentioned on this paper. Future paintings will purpose to check how nicely this version scales on an awful lot large dataset, which include the Million Song Dataset. In destiny I'll preserve on experimenting with new datasets and lately created gadget gaining knowledge of fashions which could result in new tune lyrics through gaining knowledge of from the inputs fed into it. I'll certainly trial with new indigenous Languages of India. Focus can even be on integrating a speech synthesiser in order to provide the set of rules a voice.

REFERENCES

[1] J. Chen and R. Du, "Music Generation using Deep Learning with Spectrogram Analysis," 2021 2nd International Seminar on Artificial Intelligence, Networking and Information Technology (AINIT), Shanghai, China, 2021, pp. 589-595, doi: 10.1109/AINIT54228.2021.00119.

[2] Kalingeri, Vasanth, and Srikanth Grandhe. "Music generation with deep learning." arXiv preprint arXiv:1612.04928 (2016).

[3] Mangal, Sanidhya, Rahul Modak, and Poorva Joshi. "Lstm based music generation system." arXiv preprint arXiv:1908.01080 (2019).

[4] P. Dwivedi and A. Upadhyaya, "A Novel Deep Learning Model for Accurate Prediction of Image Captions in Fashion Industry," 2022 12th International Conference on Cloud Computing, Data Science & Engineering (Confluence), Noida, India, 2022, pp. 207-212, doi: 10.1109/Confluence52989.2022.9734171.

[5] Manzelli, Rachel, et al. "An end to end model for automatic music generation: Combining deep raw and symbolic audio networks." Proceedings of the Musical Metacreation Workshop at 9th International Conference on Computational Creativity, Salamanca, Spain. 2018.

[6] Lyu, Qi, et al. "Modelling high-dimensional sequences with lstm-rtrbm: Application to polyphonic music generation." Twenty-Fourth International Joint Conference on Artificial Intelligence. 2015.

[7] Choi, Keunwoo, George Fazekas, and Mark Sandler. "Text-based LSTM networks for automatic music composition." arXiv preprint arXiv:1604.05358 (2016).

[8] Hracs, Brian J. "A creative industry in transition: The rise of digitally driven independent music production." Growth and Change 43.3 (2012): 442-461.

[9] Robb, Sheri L. "Techniques in song writing: Restoring emotional and physical well being in adolescents who have been traumatically injured." Music Therapy Perspectives 14.1 (1996): 30-37.

[10] F. Carnovalini and A. Roda`, "A Multilayered Approach to Automatic Music Generation and Expressive Performance," 2019 International Workshop on Multilayer Music Representation and Processing (MMRP), Milan, Italy, 2019, pp. 41-48, doi: 10.1109/MMRP.2019.00016.

[11] Herremans, Dorien, and Elaine Chew. "MorpheuS: automatic music generation with recurrent pattern constraints and tension profiles." Proceedings of IEEE TENCON,-2016 IEEE Region 10 Conference. IEEE, 2016.

[12] B. Royal, K. Hua and B. Zhang, "Deep Composer: Deep Neural Hashing And Retrieval Approach To Automatic Music Generation," 2020 IEEE International Conference on Multimedia and Expo (ICME), London, UK, 2020, pp. 1-6, doi: 10.1109/ICME46284.2020.9102815.

[13] Guimaraes, Gabriel Lima, et al. "Objective-reinforced generative adversarial networks (ORGAN) for sequence generation models." arXiv preprint arXiv:1705.10843 (2017).

[14] Shen, Yujun, et al. "Faceid-gan: Learning a symmetry three-player gan for identity-preserving face synthesis." Proceedings of the IEEE conference on computer vision and pattern recognition. 2018.

[15] N. Sadoughi and C. Busso, "Speech-Driven Expressive Talking Lips with Conditional Sequential Generative Adversarial Networks," in IEEE Transactions on Affective Computing, vol. 12, no. 4, pp. 1031-1044, 1 Oct.-Dec. 2021, doi: 10.1109/TAFFC.2019.2916031.

[16] Dong, Hao-Wen, et al. "Musegan: Multi-track sequential generative adversarial networks for symbolic music generation and accompaniment." Proceedings of the AAAI Conference on Artificial Intelligence. Vol. 32. No. 1. 2018.

[17] Ycart, Adrien, and Emmanouil Benetos. "A study on LSTM networks for polyphonic music sequence modelling." ISMIR, 2017.

[18] Hawthorne, Curtis, et al. "Enabling factorized piano music modeling and generation with the MAESTRO dataset." arXiv preprint arXiv:1810.12247 (2018).

[19] Kong, Qiuqiang, et al. "Giantmidi-piano: A large-scale midi dataset for classical piano music." arXiv preprint arXiv:2010.07061 (2020).

[20] T. P. Kancharlapalli and P. Dwivedi, "A Novel Approach for Age and Gender Detection using Deep Convolution Neural Network," 2023 10th International Conference on Computing for Sustainable Global Development (INDIACom), New Delhi, India, 2023, pp. 873-878.

[21] P. Dwivedi and B. Sharan, "Deep Inception Based Convolutional Neural Network Model for Facial Key-Points Detection," 2022 International Conference on Computing, Communication, and Intelligent Systems (ICCCIS), Greater Noida, India, 2022, pp. 792-799, doi: 10.1109/ICCCIS56430.2022.10037639.

[22] S. Mehta, U. Rastogi and P. Dwivedi, "Deep CNN and LSTM Architecture-Based Approach for COVID-19 Detection," 2023 10th International Conference on Signal Processing and Integrated Networks (SPIN), Noida, India, 2023, pp. 421-426, doi: 10.1109/SPIN57001.2023.10117454.

[23] P. Dwivedi and B. Islam, "An Item-based Collaborative Filtering Approach for Movie Recommendation System," 2023 10th International Conference on Computing for Sustainable Global Development (INDIACom), New Delhi, India, 2023, pp. 153-158.

[24] C. Lala and P. Dwivedi, "Hate Speech Detection Network Using LSTM," 2023 International Conference for Advancement in Technology (ICONAT), Goa, India, 2023, pp. 1-6, doi: 10.1109/ICONAT57137.2023.10080786.

[25] M. Panchal, B. Sharan and P. Dwivedi, "Comprehensive Analysis of Machine Learning Approaches for Breast Cancer Detection and Classification," 2023 10th International Conference on Computing for Sustainable Global Development (INDIACom), New Delhi, India, 2023, pp. 867-872.

[26] Choi, Kristy, et al. "Encoding musical style with transformer autoencoders." International Conference on Machine Learning. PMLR, 2020.

[27] Q. Kong, B. Li, X. Song, Y. Wan and Y. Wang, "High-Resolution Piano Transcription With Pedals by Regressing Onset and Offset Times," in IEEE/ACM Transactions on Audio, Speech, and Language Processing, vol. 29, pp. 3707-3717, 2021, doi: 10.1109/TASLP.2021.3121991.

[28] A. Meng, P. Ahrendt, J. Larsen and L. K. Hansen, "Temporal Feature Integration for Music Genre Classification," in IEEE Transactions on Audio, Speech, and Language Processing, vol. 15, no. 5, pp. 1654-1664, July 2007, doi: 10.1109/TASL.2007.899293.

[29] Panagakis, Ioannis, Emmanouil Benetos, and Constantine Kotropoulos. "Music genre classification: A multilinear approach." ISMIR. 2008.

[30] P. Khunarsal, C. Lursinsap and T. Raicharoen, "Singing voice recognition based on matching of spectrogram pattern," 2009 International Joint Conference on Neural Networks, Atlanta, GA, USA, 2009, pp. 1595-1599, doi: 10.1109/IJCNN.2009.5179014.

[31] George, Joe, and Lior Shamir. "Computer analysis of similarities between albums in popular music." Pattern Recognition Letters 45 (2014): 78-84.

[32] P. Neammalai, S. Phimoltares and C. Lursinsap, "Speech and music classification using hybrid Form of spectrogram and fourier transformation," Signal and Information Processing Association Annual Summit and Conference (APSIPA), 2014 Asia-Pacific, Siem Reap, Cambodia, 2014, pp. 1-6, doi: 10.1109/APSIPA.2014.7041658.

[33] Tao Li and M. Ogihara, "Music genre classification with taxonomy," Proceedings. (ICASSP '05). IEEE International Conference on Acoustics, Speech, and Signal Processing, 2005., Philadelphia, PA, 2005, pp. v/197-v/200 Vol. 5, doi: 10.1109/ICASSP.2005.1416274.

[34] Li, Tao, Mitsunori Ogihara, and Qi Li. "A comparative study on content-based music genre classification." Proceedings of the 26th annual international ACM SIGIR conference on Research and development in informaion retrieval. 2003.