

# Music Generation and Composition Using Machine Learning

Akanksha Dawande

Computer Science and Engineering  
UIT-RGPV, Bhopal, India

Uday Chourasia

Computer Science and Engineering  
UIT-RGPV, Bhopal, India

Priyanka Dixit

Computer Science and Engineering  
UIT-RGPV, Bhopal, India

**Abstract**— Music is derived from the greek word ‘μουσική’ (pronounced as “mousike” which means “The art of Muses”). Music is actually the arrangement of sounds in time to create a pattern which joys ears. The intention of a machine being able to create music is quite interesting. The music generation process implies the manipulation of the base-line notations to create a more complex composition. In this thesis the waveform based generation system is proposed with the help of some machine learning techniques. These raw waveforms are representing the musical bars. The audio sample’s preprocessing is performed, which involves the transformation of the waveforms (musical bars) into time-frequency representation which is very common in dealing with the music signals. The purpose of the generative model is to create music chunks analogous to those present in the dataset, which is created by 2 second long music bars. The use of the convolutional layers in the generative adversarial network known as the Deep Convolutional Generative Adversarial Network has as important significance in the model.

**Keywords**— Music Generation, Machine Learning, Deep Learning, Generative Adversarial Network, Convolutional Neural Network, Recurrent Neural Network, Long short-term memory, Reinforcement Learning, Binary Neurons.

## I. INTRODUCTION

### A. Music background information

Before jumping into the technical prospects of how the music generation system works, the basic knowledge about music is essential. Pitch is a word that used to understand the solidity and softness of any sound. There are basically seven notes present which are A, B, C, D, E, F, G. In India, people also call it as saat sur (seven sounds) with their particular meanings, sa - shadjam, re/ri - rishabham ga - gandhara, ma - madhyama, pa - panchamam, dha - dhaivatam and ni - nishadam. Each set of notes (sa to ni) is called an octave. Some of the basic terminologies of music are rhythm, melody, harmony, etc. Rhythm - The placement of sounds or notation in a particular time interval to create pattern.

Melody - The sequence (or the horizontal series) of notes of different pitches played one after another.

Harmony - It is the stacking of musical notes together at the same time to create chord and the sequence of such chords creates the chord progression which gives a pleasant feel to the listener.

### B. Neural Network

Neural Networks (NNs) or Artificial Neural Networks (ANNs) are the PC frameworks which are propelled by the natural neural design of human cerebrums. An ANN comprises of

many interconnected hubs (or fake neuron) which are gathered in various layers of the ANN, for example Info Layer, Hidden Layer and Output Layer. Every association is appointed some weight to address its relative significance. In ANN, the term called as enactment work assumes a significant part in the calculation of neurons to make the organization non-direct. The misfortune work really takes a look at the aberrations between the forecast of the calculation and the ideal yield. The inclination drop is utilized to prepare ANNs and is viewed as the most utilized iterative improvement calculation.

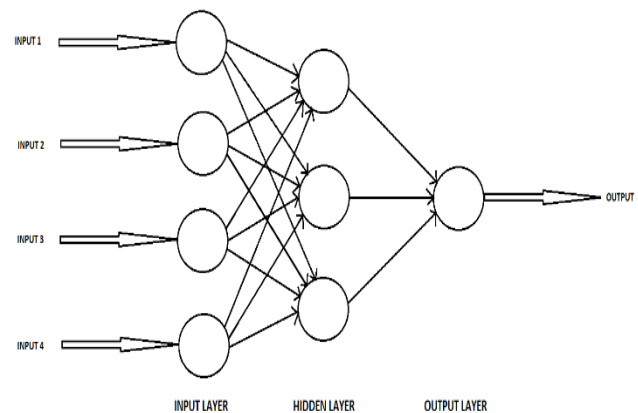


Fig .1. Structure Neural Network

### C. Backpropagation

Backpropagation algorithm is generally used in machine learning to train feedforward neural networks. The backpropagation basically computes gradient of loss function with respect to the weights associated with a single input-output in the network. It makes the use of gradient methods suitable to use gradient methods in training networks consisting multiple layers. Chain rule is used in the calculating the loss function gradient one layer at a time and also iterates in backward direction from the end/last layer. Backpropagation is one of the example of dynamic programming.

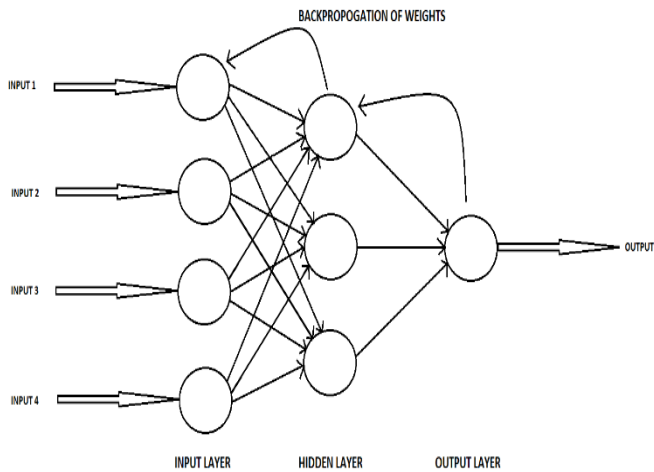


Fig.2. Backpropagation Process

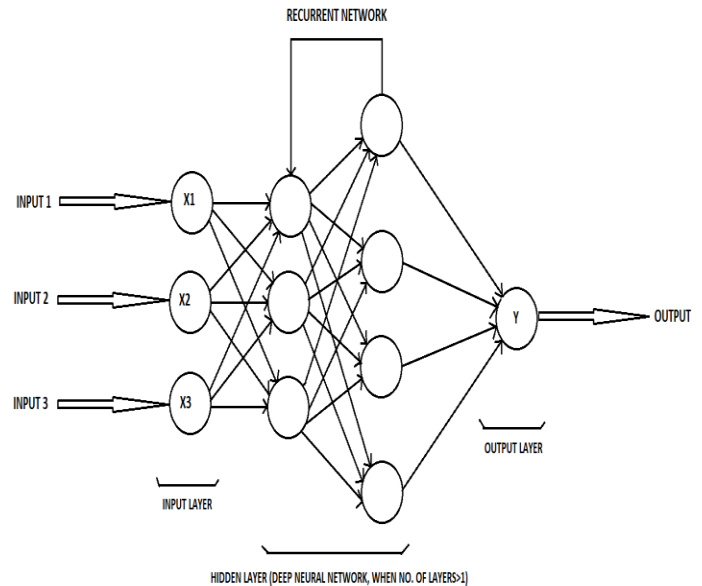


Fig.4. Structure of Recurrent Neural Network

#### D. Recurrent Neural Network

RNN represents repetitive neural organization, is one of the class of ANN (Artificial Neural Network) in which a coordinated chart is shaped and addressed utilizing the hub's associations in a specific succession. This shows worldly powerful conduct. As it is taken from feedforward neural organization, RNNs can utilize their inward state (additionally called the memory) to deal with the arrangements of sources of info. RNNs have the repetitive associations in secret layers among past and the present statuses in the neural organization. In this manner it stores the data/helpful information like a memory. This capacity of putting away data like memory make it functional in applications that incorporates penmanship acknowledgment, discourse acknowledgment, and so on The principle disadvantage of RNN is only that it stores the information of just one state before of the present status, which implies information keeping highlight expands just a single era back.

#### E. Long short-term memory

LSTM represents Long transient memory is an option of RNN design that is utilized in the field of AI. It is the arrangement of absence of long haul memory in RNN. LSTM has criticism associations. It isn't simply used to deal with single information focuses like picture, yet additionally used to handle the whole succession of information like video. LSTM is fit for catching the drawn out conditions in the information grouping. A typical LSTM unit is made out of a phone, additionally called as a memory cell. This memory cell comprises of three doors, i.e., the information entryway, the yield entryway and the neglect door. The cell keep esteems throughout erratic time frames and the three entryways expressed above directs the progression of data or helpful information into and out of the cell. In a specific cell, an information door control the measure of information being given as contribution to memory, yield entryway controls the information passed to next layer and the neglect entryway controls the misfortune in the put away memory. The neglect door can be considered as a recall vector, the yield of the neglect entryway advises the phone state which information is to keep and which is to eliminate. In the event that the yield of neglect door is 1, implies the data is kept in the cell state and gets neglected/eliminated if the yield is 0.

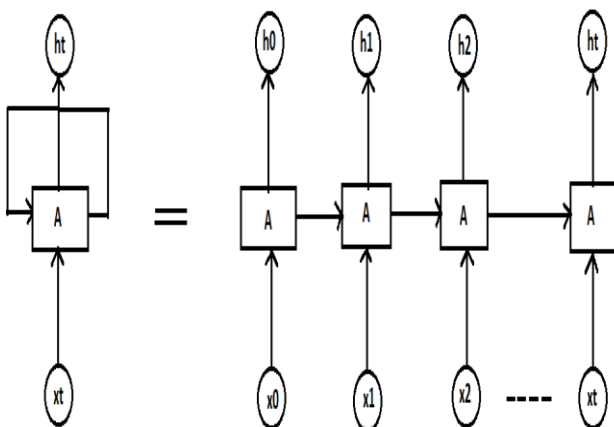


Fig.3. Recurrent Neural Network Block Diagram

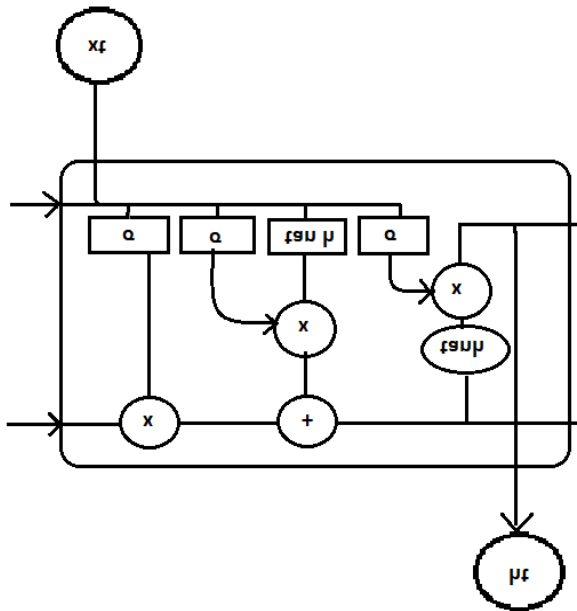


Fig.5. Structure of Long Short-Term Memory

II. METHODOLOGY

After studying all the algorithms, lstm can be consider as the best algorithm that can be used to generate music. The use of two lstms can be more preferable, which is also known as Biaxial LSTM model. One lstm can be used to predict and keep track of the time at which the node need to be played and the other lstm to predict the played node. In short lstm are used to predict and keep track of both time as well as note in a sequential order.

The main obstacle could be the representation of data. The selection of MIDI file is the major concern because it is commonly used, also it keeps the features of the songs in its metadata. Since it is commonly used so the no. of dataset needed are easily available.

As we know that the lstm came into existence because of the vanishing gradient problem faced while using the recurrent neural networks, lstm is an upgraded version of recurrent neural network.

**Vanishing Gradient Problem** - In 1980s, researchers couldn't be able to make deep neural networks because there was no RELU activation function for all neural networks, and due to the use of sigmoid activation function they were facing some problem termed as vanishing gradient problem. In The vanishing gradient problem is dicussed can be explained in the following way.

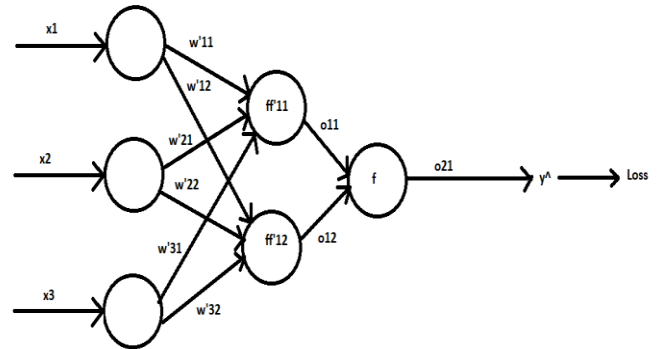


Fig.6. Neural Network Example to explain Vanishing Gradient Problem

Here,  $x_1, x_2$  and  $x_3$  are inputs,  $w'^{11}, w'^{12}, w'^{21}...$  are the weights associated,  $ff'^{11}, ff'^{12}$  and  $f$  are the functions of neurons,  $o_{11}, o_{12}, o_{21}$  are the outputs and  $y^{\wedge}$  is the final result.

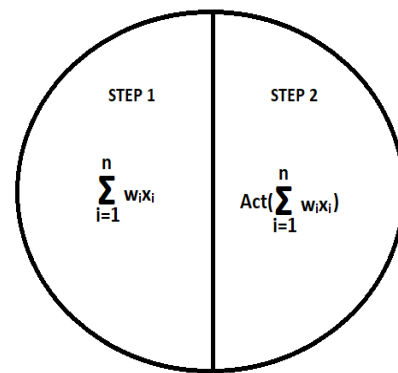


Fig.7. Structure of Neuron

We know weight updating formula is

$$w'_{11new} = w'_{11old} - \frac{\eta \cdot \partial L}{\partial w'_{11old}} \quad \dots (1)$$

$\eta$  is the learning rate.

So, for the given neural network, we need to find  $\frac{\eta \cdot \partial L}{\partial w'_{11old}}$

So by Chain rule,

$$\frac{\partial L}{\partial w'_{11old}} = \frac{\partial L}{\partial o_{21}} \cdot \frac{\partial o_{21}}{\partial o_{11}} \cdot \frac{\partial o_{11}}{\partial w'_{11}} \quad \dots (2)$$

The learning rate used is sigmoid, and the derivative of sigmoid ranges between 0 to 0.25.

The sigmoid function formula is  $\frac{1}{1+e^{-z}}$

where  $z = \Sigma(xw) + b$ .

And activation function of  $z = \sigma(z)$

So,  $\frac{\partial \sigma(z)}{\partial z} = 0$  to 0.25

or for better understanding, we can say  $0 \leq \frac{\partial \sigma(z)}{\partial z} \leq 0.25$

Now as the number of layers increases, the derivative value will decrease, which at the end will cause  $\partial w'_{11new} \approx \partial w'_{11old}$ . Because

in this way the gradient descent will never reach to the global minima, which will cause vanishing gradient problem.

### III. LITERATURE SURVEY

This piece contains the information about the researches done in the field of music and different technologies. The researches contained some models which provided interesting outcomes.

1. A thesis named “**Music Generation using Generative Adversarial Network**” under the supervision of Prof. Rodrigo Martins de Matos Ventura contained three different models to generate music. Using the class of machine learning called Generative Adversarial Network, when included convolutional layers in the GAN, shown amazing results by creating beautiful music pieces [1].
2. **Nabil Hewahi, Salman AlSaigal and Sulaiman AlJanahi** investigated the utilization of long transient memory neural organization in producing music sections and proposed a model. The proposed model takes the midi records, changes over them into melodies documents and the encode them to be proper to take care of contribution to the neural organization. An increase cycle is done prior to giving contribution to the neural organization which incorporates expansion of the document into various keys. Then, at that point, the record is taken care of into the neural organization for preparing. Furthermore, the last advance is music age. The fundamental goal was to furnish the neural organization with an arbitrary note and afterward the neural organization begins changing it steadily until delivering a decent piece of music. Different tests have been led to investigate the best upsides of boundaries that can be chosen to acquire great music ages. The outcomes were astounding for certain documents as created music pieces were fitting as far as mood and amicability [2].
3. **Natasha Jaques, Shixiang Gu, RichardE. Turner, Douglas Eck** have proposed a model for successive preparing where the grouping indicator is separated and refined by enhancing some forced award capacities, simultaneously keeping up with the great prescient properties gained from the information. They investigated the helpfulness of their methodology with regards to music age. A LSTM is prepared on a huge corpus of 30,000 MIDI tunes to foresee the following note in a melodic succession. This Note-RNN is then refined utilizing support realizing, where the prize capacity is a blend of remunerations dependent on rules of music hypothesis, just as the yield of one more prepared Note-RNN. The outcomes shown that this mix of AI and support learning can create additional satisfying songs, yet that it can fundamentally decrease undesirable practices and disappointment methods of the RNN [3].
4. **Manan Oza, Himanshu Vaghela, Kriti Srivastava** have accumulated data about the enhancements in GANs which have shown energizing outcomes, adding layers after the past ones have united has demonstrated to help in better by and large union and soundness of model just as decreasing the preparation time by an adequate sum. Subsequently they utilized this preparation method to prepare the model dynamically in the time and pitch area. Likewise they utilized a layer of deterministic twofold neurons toward the finish of the generator to get paired esteemed yields rather than fragmentary qualities existing somewhere in the range of 0 and 1, as it is demonstrated in some recently proposed models that deterministic parallel neurons help in further developing outcomes [4].
5. **Mohit Dua, Rohit Yadav, Divya Mamgai, Sonali Brodiya** had presented an improved an improved version of the sheet music system that already exists. The use of Recurrent neural network and lstm has played an essential role in the work. Particularly two modules were used in order to achieve the goal, as the final results with these modules were better than the ones used before [5].
6. **Wiktor Kania, Ewa Klapcińska, Mateusz Groblewski** developed a report which aims to describe the application of song generation and all the ways in which it is created. The goal of the application is to allow everyone to generate songs. The effective causes of computer generated music for not only the musicians but other people like game developers, youtubers, etc. inspired the team to build an AI-trained music generator. With the use of lstm they could built an application that has two layers, Front-end - Interact with users and Back-end – Loads a model with appropriate dictionary for selected musical genre and feeds the model sequences of the notes. The model tries to predict the next note in the given sequence which then appended to the sequence and the oldest note is forgotten, that make a bit different new sequence. This process continues till the formation of desired length of the song and then converted to the MIDI and JSON [6].
7. **Sanidhya Mangal, Rahul Modak, Poorva Joshi** used a fully trained model to produce a music suite. Experiments and model training were exercised on google Colab and code implementation on Keras. Their paper serves the purpose of creating a model that can be used to create music and melodies without any human interference [7].
8. **Hongyu Chen, Xueyuan Yin, Qinyin Xiao** proposed a production model which generates note sequences using Generative Adversarial Network framework. The use of convolutional neural network played an important role as by optimizing it according to the characteristics of the musical notes. The optimization algorithm helped the CNN to focus on learning all the music attributes and fasten the experimentation process [8].
9. **Tianyu Jiang, Xueyuan Yin, Qinyin Xiao** used the recurrent neural network in their proposed work. They used piano roll which widely used to represent polyphonic music generation. They used a model to produce a unique melody track, simultaneously allowing multiple notes to be played. They introduced bidirectional LSTM network with the aim of producing symmetric music. By learning the context data of notes from vertical and horizontal level

bidirectionally, the quality of model gradually improved. The loss function was got redesigned to accelerate the optimization process by avoiding generation of meaningless results [9].

Comparison table of various papers discussed above –

#### IV. COMPARISON TABLE

References	Author Name	Name of paper with year	Algorithm Used	Learning Model	Type of Input	Dataset	Performance
[1]	Prof. Rodrigo Martins de Matos Ventura	Music Generation using Generative Adversarial Network (2018)	Generative Adversarial Network	Unsupervised	MIDI	10000 songs	Intermediate
[2]	Nabil Hewahi	Generation of Music pieces using machine learning: long short-term memory neural network approach (2019)	Long Short-Term Memory	Supervised	MIDI	Bach's 'Well-Tempered Clavier Book II'	High
[3]	Natasha Jaques	Generating Music by Fine-Tuning Recurrent Neural Networks with Reinforcement Learning (2016)	Recurrent Neural Network	Supervised	MIDI	Monophonic melodies from a corpus of 30,000 MIDI songs	Intermediate
[4]	Manan Oza	Progressive Generative Adversarial Binary Networks for Music Generation (2019)	Generative Adversarial Network	Unsupervised	Multi track piano-roll	Lakh Piano roll Dataset (LPD - cleansed)	High
[5]	Mohit Dua	An Improved RNN-LSTM based Novel Approach for Sheet Music Generation (2020)	Long Short-term Memory, Recurrent Neural Network	Supervised	Instrumental	DSD100 dataset	High
[6]	Wiktor Kania	FRIML - Music Generation using Machine Learning (2021)	Long Short-term Memory	Supervised	MIDI	The Lakh MIDI Dataset, NES-MDB	Intermediate
[7]	Sanidhya Manga	LSTM Based Music Generation System (2019)	Long Short-term Memory, Recurrent Neural Network	Supervised	MIDI	Million Song Dataset	High
[8]	Hongyu Chen	Generating Music Algorithm with Deep Convolutional Generative Adversarial Networks (2019)	Generative Adversarial Networks	Unsupervised	MIDI	The Lakh MIDI Dataset	High
[9]	Tianyu Jiang	Music Generation using Bidirectional Recurrent Network (2019)	Bidirectional Recurrent Neural Network	Supervised	MIDI	Classical Piano Dataset with 295 MIDI files	Intermediate

#### V. CONCLUSION

After having a deep investigation of the papers bring up in this paper, it is observed that the music generation and composition take a very vast and much more explorative topic when is combined with machine learning and its various algorithms. Generative adversarial neural network helps to produce the new data on the basis of training data and the data that is already present. Recurrent neural network is vast topic used in the applications like speech recognition and hand writing recognition as it keeps the output of one before neuron to feed as input to the next neuron, hence has the ability to keep the information as memory. But the limitation of Recurrent neural network is that it can only keeps output of only one neuron, and this limitation is overcome by Long Short-Term Memory which keeps more information as memory. Hence, LSTM is more preferred over Recurrent Neural Network. Also, the discussed

the Vanishing Gradient Problem which can be ignored by using RELU activation function.

#### VI. REFERENCES

- <https://fenix.tecnico.ulisboa.pt/downloadFile/844820067125714/FINAL%20TESE.pdf>
- [https://www.researchgate.net/publication/334999618\\_Generation\\_of\\_music\\_pieces\\_using\\_machine\\_learning\\_long\\_short-term\\_memory\\_neural\\_networks\\_approach](https://www.researchgate.net/publication/334999618_Generation_of_music_pieces_using_machine_learning_long_short-term_memory_neural_networks_approach)
- <https://research.google/pubs/pub45871/>
- [https://www.researchgate.net/publication/331700709\\_Progressive\\_Generative\\_Adversarial\\_Binary\\_Networks\\_for\\_Music\\_Generation](https://www.researchgate.net/publication/331700709_Progressive_Generative_Adversarial_Binary_Networks_for_Music_Generation)
- [https://www.researchgate.net/publication/341907780\\_An\\_Improved\\_RNN-LSTM\\_based\\_Novel\\_Approach\\_for\\_Sheet\\_Music\\_Generation](https://www.researchgate.net/publication/341907780_An_Improved_RNN-LSTM_based_Novel_Approach_for_Sheet_Music_Generation)
- <https://card-file.onaft.edu.ua/handle/123456789/17586>
- [https://www.researchgate.net/publication/333604736\\_LSTM\\_Based\\_Music\\_Generation\\_System](https://www.researchgate.net/publication/333604736_LSTM_Based_Music_Generation_System)
- <https://ieeexplore.ieee.org/document/8839521>
- <https://ieeexplore.ieee.org/abstract/document/8839399>