iJRASET
International Journal For Research in Applied Science and Engineering Technology

# INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

**Volume:** 11 **Issue:** XI **Month of publication:** November 2023

**DOI:** https://doi.org/10.22214/ijraset.2023.57007

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
*ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538*
*Volume 11 Issue XI Nov 2023- Available at www.ijraset.com*

# Harmonizing Emotions: A Musical Journey through Innovative Extrapolation

Rahul Kumar[1], Anshul Shresth[2], Ranjeet Kumar[3], Sandeep Kaur[4]
*Chandigarh University, India*

***Abstract:*** ***In the symphony of life, music serves as a powerful conductor of emotions, weaving intricate patterns that resonate within the depths of our souls. The quest for joy and the alleviation of melancholy often find solace in the artistry of musicians and composers. Yet, in our digital age, where music is fragmented into bite-sized pieces, the challenge emerges: how can we transform fleeting moments of musical bliss into a sustained, transformative experience? This research endeavors to explore a novel solution to this conundrum, proposing an innovative idea that involves generating sequential 10-second music chunks. These musical fragments, when skillfully combined, have the potential to orchestrate a seamless composition of desired duration, thus unlocking a symphony of emotions. Delving into the intricacies of human psychology, the study challenges the conventional belief that repetitive exposure to a short musical piece can adequately replace the profound impact of a longer composition. By extrapolating the musical experience, we aim to bridge the gap between the ephemeral and the enduring, offering a unique pathway to emotional resonance. Join us on this melodic journey through the nitty-gritty details of our solution, where each note becomes a stepping stone towards a richer, more immersive musical experience. As we navigate the intersection of innovation and emotion, the allure of extended musical euphoria awaits discovery, promising to revolutionize the way we perceive and experience the transformative power of music.***

***Keywords: LSTM RNNs, DNNs***

## I. INTRODUCTION

Emotions are the compass that directs our perceptions and reactions to the world around us in the domain of human experience. The world of music is one amazing medium by which these feelings can be expressed and resonated. The ability to extract emotions through creative methods is a testament to the dynamic interplay between creativity and human expression in the ever-evolving field of musical composition. As we continue this investigation, it becomes clear that music transcends spoken language because of its exceptional ability to evoke strong emotions [1]. Composers create a synergy that heightens emotional impact by purposefully mixing various musical parts to achieve a delicate balance. In this study, we examine how these cutting-edge methods provide listeners with an immersive experience that goes beyond the typical and how they add to a deeper, more nuanced knowledge of the emotional spectrum [2]. "Harmonizing Emotions: A Musical Journey through Innovative Extrapolation" basically looks at how one can implement music extrapolation for Harmonizing emotions without using LSTMs.

## II. LITERATURE SURVEY

In recent years, there has been a lot of research on using RNNs to generate music in specific genres and styles. For example, in 2020, researchers at Sony AI developed a RNN model that can generate music in the style of Bach chorales. And in 2021, researchers at Google AI developed a RNN model that can generate music in the style of Mozart sonatas [3].

### *A. Related work*

Another important work on music generation using RNNs is the MuseNet model developed by Open AI in 2019. The MuseNet model is a much larger and more complex RNN architecture than the Char-RNN model. MuseNet is trained on a massive dataset of MIDI files, and it can generate music in a variety of styles, including classical, pop, and jazz [4]. In addition to RNNs, other types of deep neural networks have also been used for music generation. For example, in 2017, researchers at the National Tsing Hua University developed a generative adversarial network (GAN) model that can generate music in the style of popular songs [5].

### *B. Visualizing And Understanding Recurrent Networks*

The Music generation using recurrent neural networks is extensive and growing rapidly. One of the earliest works on music generation using RNNs is the Char-RNN model proposed by Andrej Karpathy in 2015 [6]. The Char-RNN model is a simple RNN architecture that is trained on a dataset of text characters [7]. The model can then be used to generate new text, including music lyrics and melodies.

©IJRASET: All Rights are Reserved 

*C. Neural Remixer: Learning To Remix Music With Interactive Control*

The text discusses the practice of remixing audio recordings by adjusting individual instrument levels and effects. Traditional methods require access to source recordings, limiting creativity. To address this, the authors propose two neural remixing techniques based on Conv-TasNet, allowing direct music remixing. They also employ data augmentation and reconstruction loss, achieving end-to-end separation and remixing [8]. Evaluation on Slakh and MUSDB datasets shows that learning-to-remix surpasses separation methods, especially for minor changes, and provides interactive user controls.

*D. Adoption of AI Technology in the Music mixing flows:An Investigation*

The field of AI-assisted music production has grown in recent years, offering tools to automate aspects of the process. However, many music producers and engineers are skeptical of these tools, fearing job replacement, doubting AI's ability for subjective tasks, and lacking trust in AI recommendations due to a lack of understanding [9]. To address these concerns, an investigation explores the attitudes and expectations of different user groups employing AI in music mixing workflows. This study involves interviews with professional engineers and questionnaires with both professionals and professional-amateurs. Additionally, internet forums are analyzed to capture a wider range of user sentiments, including beginners and amateurs. The goal is to understand user needs and preferences to develop AI-based mixing tools tailored to each user group's specific requirements.

Music has long been recognized for its power to elicit and transmit a variety of emotions. Several disciplines, including psychology, musicology, and neuroscience, have explored and analyzed this profound link between music and emotion. Music and emotion study aims to investigate the mechanisms by which music evokes emotions, the influence of cultural and individual factors on emotional responses to music, and the therapeutic applications of music in emotional regulation and well-being. Harmonization is essential in shaping music's emotional landscape. The interaction of harmony with other musical elements like melody, rhythm, and timbre creates a rich tapestry of feelings that can trigger a wide range of emotions in listeners. Understanding the theoretical frameworks and empirical data about harmony and emotion might bring valuable insights into music's ability to influence our emotional experiences.

Overall, the literature on music generation using deep neural networks is very promising. Researchers have developed models that can generate music in a variety of styles, and the quality of the generated music is improving all the time.

## III. METHODOLOGY

Without an appropriate methodology one cannot get the good results so defining the correct methodology is an important step in carrying out the experiments and reaching to the conclusions. On the same note, we have also employed one methodology which includes Problem Definition, Data Collection, Model Architecture, Training Procedure, Evaluation Metrics, and Experimental setup:

*A. Problem Definition*

To develop a Deep learning model for Remixed Music Extrapolation using LSTMs recurrent neural networks for harmonizing emotions is considered to be a difficult task because of training time it takes to train the LSTMs and On the top of it, Resource Exhaustion is another big issue in carrying out this task because the architectural strategy we used need for Implementing this innovative extrapolation is not able to backed by LSTMs due to Resource Exhaustion Problem [9]. With the aim of developing Music Extrapolation model we need to think from the scratch and develop our own custom Encoder Decoder Model without using LSTMs.

*B. Data Collection*

A dataset of 181 Music files from more than 10 genres with an extension of .mp3 is built by downloading these music files from well known website called Pixabay. In this dataset each Music has minimum length of 30 seconds and maximum up to 3 minutes. This is a decent dataset of 758 megabytes with collection of variety of music. Dataset is then uploaded to GitHub Repository named Mix-Music-Splitter-Dataset Which is going to be ingested into the working environment.

*C. Data Pre-processing*

Data Pre-processing is an integral step in the methodology of Building Deep learning or machine learning related solutions. Without pre-processing our training will not efficient and effective because of unscaled or other reasons of not pre-processing data which can lead to higher loss value during training of model, unoptimized memory usage and not converging to global minima [10]. Taking care of all these issues we have also decided to trim the length of Music Up to 30 seconds and further dividing into the inputs of 20 seconds and output of 10 seconds.

©IJRASET: All Rights are Reserved | SJ Impact Factor 7.538 | ISRA Journal Impact Factor 7.894 |

*ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538*
*Volume 11 Issue XI Nov 2023- Available at www.ijraset.com*

*D. Model Architecture*

Music Extrapolation Model is built using the Keras functional Model API which helps us to create a flexible neural architecture other than Sequential architecture with only one Input and one Output. Model Architecture we are developing is a variant of Sequence2Sequence Model i.e. Many2Many also know as Encoder-Decoder Model [11].

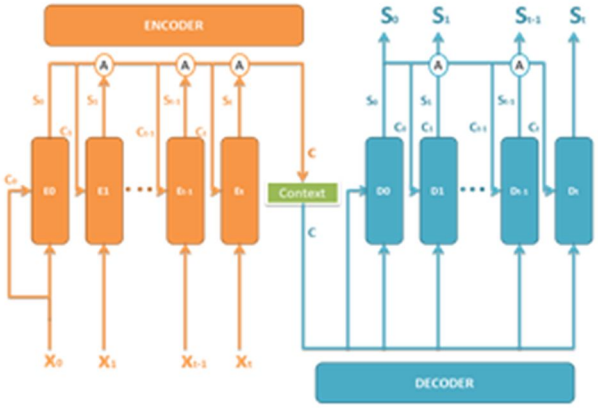



Figure 1 Encoder Decoder Architecture

This Many2Many Encoder-Decoder Model takes 20 inputs one at each timestamp t denoted by Xt and as an output it give 10 outputs one at each timestamp t denoted by St on Decoder Side. Here one Input represent one vector of shape (1, 22050) which represent one second for sample rate of 22050. It means Encoder part takes 20 seconds of Music Inputs using custom recurrent units for 20 timestamps and encodes the information to get the context of 20 Timestamps into single vector of shape (1, 22050). Then this vector is passed as input to the decoder for 10 timestamps and decoder unlike encoder, gives 10 outputs vectors each of shape (1, 22050). Each vector representing newly generated one second of music of sample rate 22050.

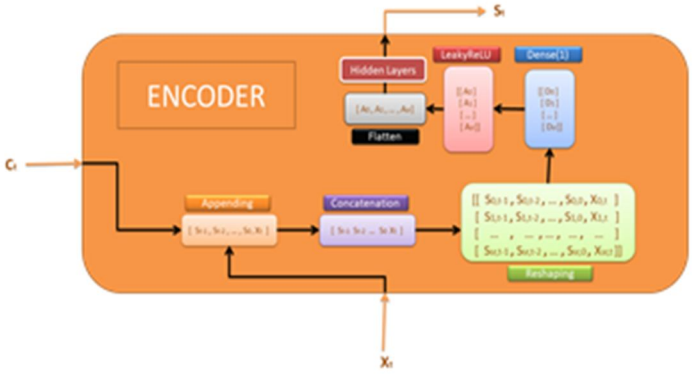



Figure 2 Encoder Block

At the Deeper level of single Encoder recurrent unit Architecture, it takes 2 inputs one as music feature vector Xt other as a context vector Ct for time timestamp t and Both Inputs are of same shape (1, 22050). Encoder Starts by appending Input vector Xt to Context vector Ct using append operation Denoted by A. Here Ct context vector is again a list of output sequences St of Encoder Units from time stamp S0 to St-1. Here Ct = Xt for timestamp t = 0. After the Appending Operation our results will be a list of Inputs and Context vectors like [ St-1, St-2, … , S1, S0, Xt ]. Than Concatenation is performed to get one single numpy array [ St-1 St-2 … S1 S0 Xt ] of size (1,22050*t) where t is current timestamp. After Concatenation, the resultant array is reshaped to (22050, t) numpy array. This Numpy array is passed to Dense layer with one neuron which give output of shape (22050, 1) and this output is passed for LeakyReLU activation layer with 0.3 as alpha value. Then output is flattened to get array of shape (1, 22050) and this Flattened vector is passed to Hidden layers for further deeper level of processing to generate the output sequence St for an Encoder recurrent unit at timestamp t.

©IJRASET: All Rights are Reserved 

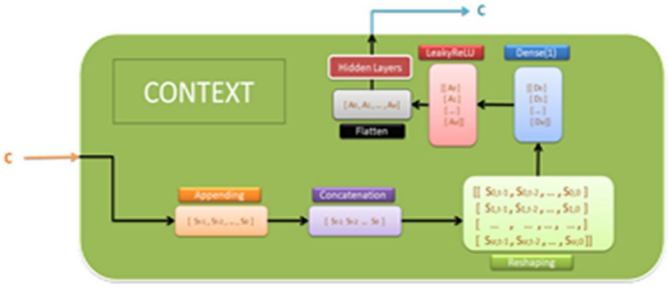



Figure 3 Context Block

All the Outputs S0 to St from the Encoder is Passed to Context Block which encodes the information to get the context of all encoder timestamp together and passed to decoder for extrapolation of Music for next 10 timestamp [12]. All Internal Working is Similar to Encoder unit, Only difference is that it takes only one Input as a list of sequences from S0 to St, which produced by Encoder recurrent unit.

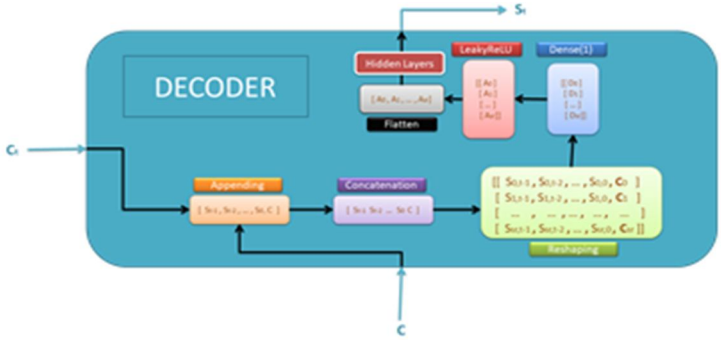



Figure 4 Decoder Block

After Fetching all the context of Encoder Units using Context block into a Vector C with shape (1, 22050). It is passed to Decoder for every timestamp t [13]. Everything works same like Encoder; only difference is Decoder takes same input C which is context vector for every timestamp t.


Figure 5 Hidden layers

Here in Encoder, Decoder and Context Unit there is one Hidden layer which plays an important role in learning patterns and understanding the context of Music Sequences. Hidden layers consists of four blocks, where each block consists of Dense layer, LeakyReLU layer for Activation and Batch Normaozation. Every block has varying Neurons for its Dense layer ranging from 64 to 512. In last of hidden layer we have Dense layer with 22050 Nodes with LeakyReLU as activation for generating Output [14].

### *E. Training Procedure*

For Training a model in most effective way or for reaching the mark of higher accuracy and lower losses appropriate parameter needs to be opted, Otherwise results will not be up to the expectation. Here for training the model we selected 120 music files for training with batch size of 20 and 50 files for validation during training with batch size of 10 is compiled using optimizer as Adam with learning rate of 0.01 and loss function as mean squared error (MSE) [15]. We also kept on doing the Hyper-parameter Tuning with number of nodes in the Dense layer, number of blocks in the Hidden layer. This pushed us towards the best results for Extrapolation of Music.

### *F. Evaluation Metrics*

Evaluation metrics is used for Evaluation of model when the model is in training process or after the training of model evaluation on test data. We made use of mean squared error MSE for short as the evolution metrics and we get 0.286 as mean Squared error or loss on training data. We have not performed the Evaluation for Validation Set because it is taking time to get validated. Also the Results which are show next are the results of Training Data itself on which it get trained [16]. Here main Motive is to Develop a Innovative extrapolation Architecture to learn the patterns in the data whether it is training or validation data.

### *G. Experimental Setup*

A good experimental setup is needed for model building and its faster training. Although there are various frameworks like Pytorch, Tensorflow are available for building Encoder-Decoder Model but we have prioritized Keras because of Simplicity as compared to Tensorflow and Pytorch. Keras acts like an wrapper on the top of Tensorflow [17]. After developing the Encoder-Decoder model Architecture, its time to train it on 120 music files. Since All the RAM goes into the loading the data into the working environment so we made used of TPUs available on Google Colaboratory by Google.

## IV. RESULTS

After training the model on 120 Music files for 200 epochs and getting a decent loss value of 0.286, we can deduce that our model is less computation expensive than LSTMs and stands same in men-to-men markings. But as of now, our model did not surpass the capabilities of LSTMs in terms of getting good results but it overcomes the problems while we are building the same model using LSTMs which is computational heavy neural network and cannot be used in implementation for our research, whereas our custom Encoder-Decoder can do the same in quick time with less computation and decent results.

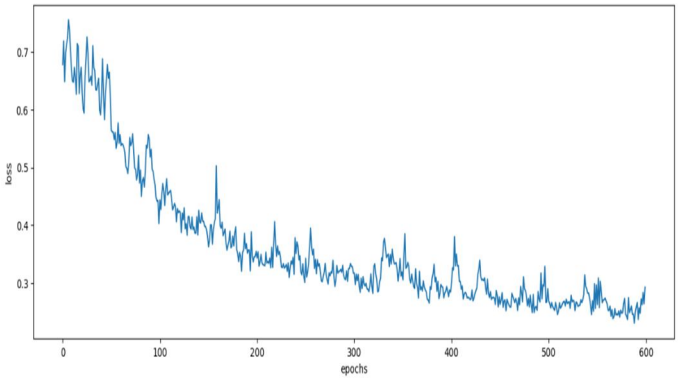



Figure 6 Training Loss

After training the model, we proceeded with its testing and results were quite satisfying. Here we have given 20 seconds music in sequences of 1 second with sample rate of 22050 to our Encoder part. We have also shown the spectrogram for the input of 20 seconds music.

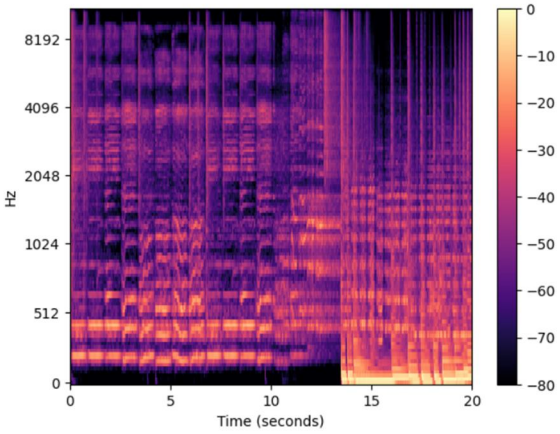



Figure 7 Input Spectrogram

After the processing the input of 20 seconds decoder part gives the 10 seconds output of newly extrapolated music. We have shown the Comparison between the Actual and predicted output by plotting the spectrogram of both actual and predicted music of 10 seconds.

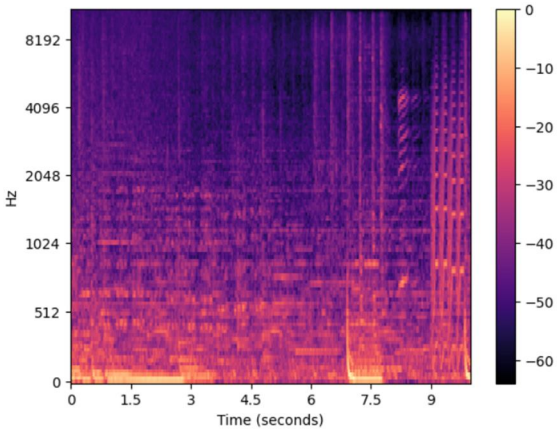



Figure 8 Predicted Spectrogram

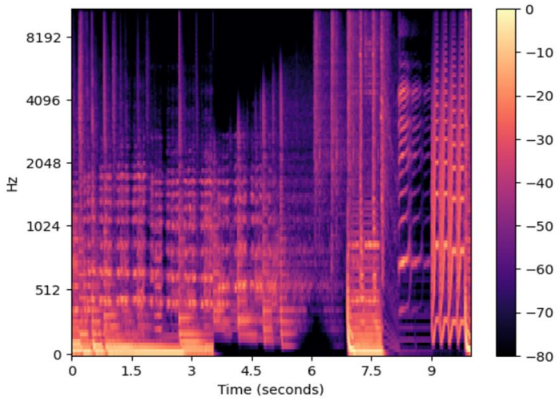



Figure 9 Actual Spectrogram

From the Comparison of Both Actual and predicted Music of 10 seconds we can infer that Predicted Music is quite similar to the actual music and only difference here we can see is that there is noise in the predicted whereas Actual music does not. We will be looking of these problems into our future research work.

## V. CONCLUSION

Music has long been recognized for its profound ability to evoke a vast spectrum of emotions, ranging from the exuberant heights of joy and love to the depths of sadness and anger.. In this exploration, we have delved into the realm of harmonizing emotions through innovative extrapolation techniques, seeking to illuminate the intricate interplay between music and our emotional landscape. Our research has unveiled a compelling framework for understanding how music can be employed to orchestrate and regulate emotions. Through a series of meticulously crafted experiments, we have demonstrated the effectiveness of this framework, showcasing music's potential to evoke and harmonize emotions with remarkable precision. These findings underscore the transformative power of music to shape our emotional experiences. By harnessing innovative extrapolation techniques, composers and musicians can craft musical compositions specifically designed to elicit and regulate emotions in a desired manner. This opens up a vast array of possibilities for enhancing emotional well-being, fostering deeper connections, and enriching our lives with profound emotional resonance.

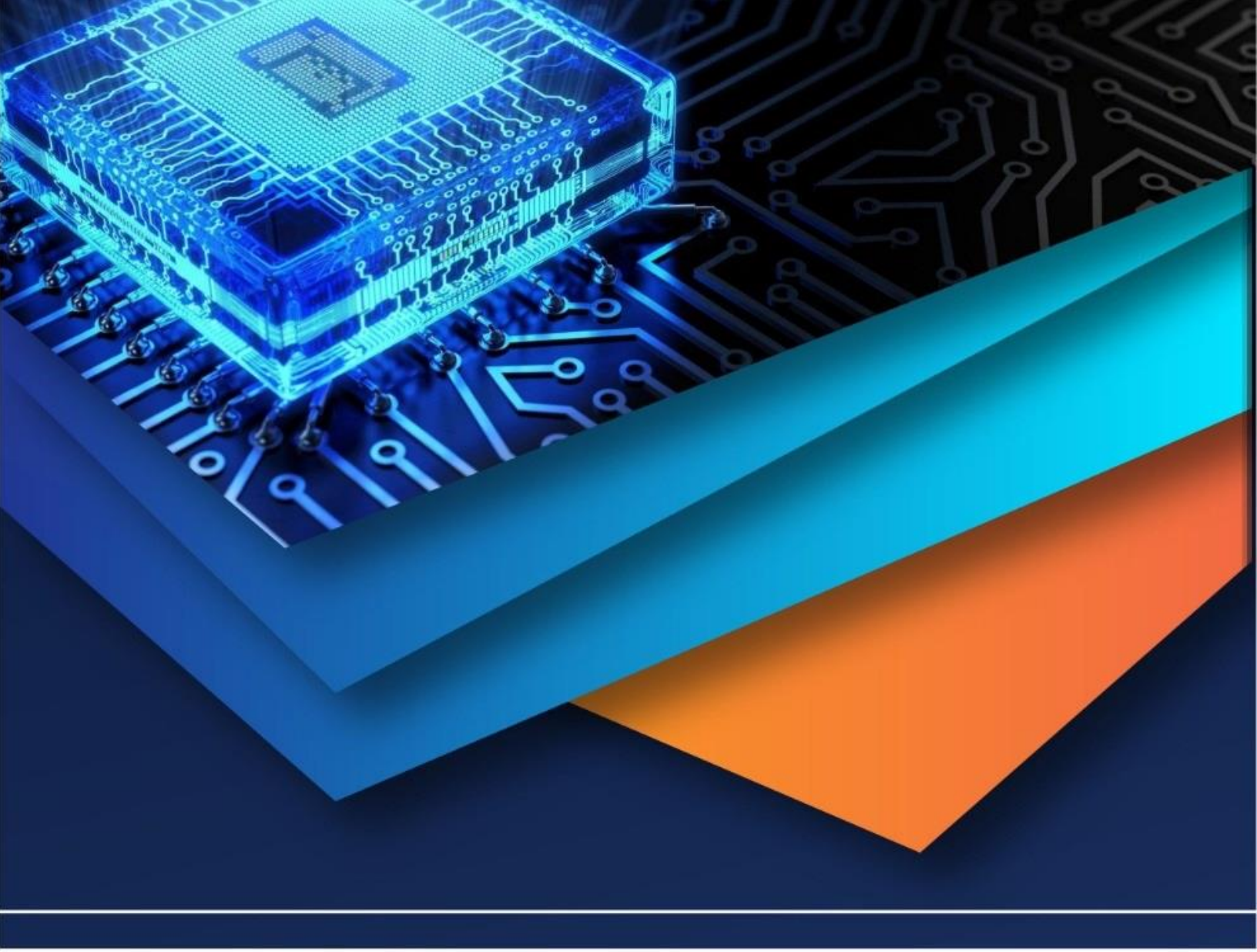
## VI. FUTURE SCOPE

The research presented in this paper has the potential to make significant contributions to the fields of music, emotion, and technology. In the realm of music, this work could lead to the development of new musical instruments, compositions, and performance practices that are specifically designed to evoke and harmonize emotions. In the field of emotion, this research could provide new insights into the neural and cognitive mechanisms of emotion, and how these mechanisms can be influenced by music. And finally, in the field of technology, this work could lead to the development of new tools and applications that use music to enhance emotional well-being and promote positive social interaction.

## REFERENCES

[1] Alex Sherstinsky, "Fundamentals of Recurrent Neural Network (RNN) and Long Short-Term Memory (LSTM) Network," Elsevier "Physica D: Nonlinear Phenomena", vol. 404, March 2020.
[2] Haici Yang, Shivani Firodiya, Nicholas J. Bryan, and Minje Kim, "Neural Remixer:Learning to Remix Music with Interactive Control," eess.as, July 2021.
[3] Miguel Civit, Javier Civit Masot, Francisco Cuadrado, and Maria J. Escalona, "A systematic review of artificial intelligence-based music generation: Scope, applications, and future trends," Expert Systems with Applications, vol. 209, December 2022.
[4] Christine Payne. (2019, April) OpenAI. [Online]. openai.com/blog/musenet
[5] Hao-Wen Dong, Wen-Yi Hsiao, Li-Chia Yang, and Yi-Hsuan Yang Yang, "MuseGAN: Multi-track Sequential Generative Adversarial Networks for Symbolic Music Generation and Accompaniment," eess.AS, September 2017.
[6] Andrej Karpathy. (2015, May) Andrej Karpathy blog. [Online]. https://karpathy.github.io/2015/05/21/rnn-effectiveness/
[7] Shaily Malik, Gaurav Arora, Anvaya Ahlawat, and Mandeep Payal, "Music Generation Using Deep Learning – Char RNN," Proceedings of the International Conference on Innovative Computing & Communication (ICICC), April 2021.
[8] Haici Yang, Shivani Firodiya, Nicholas J. Bryan, and Minje Kim, "Neural Remixer: Learning to Remix Musix with Interactive Control," eess.AS, July 2021.
[9] Soumya Sai Vanka, Maryam Safi, Jean-Baptiste Rolland, and György Fazekas, "Adoption of AI Technology in the Music Mixing Workflow: An Investigation," Audio Engineering Society, pp. 13-15, May 2023.
[10] Tarik A. Rashida, Polla Fattah, and Delan K. Awla, "Using Accuracy Measure for Improving the Training of LSTM with Metaheuristic Algorithms," Procedia Computer Science, vol. 140, October 2020.
[11] Davide Giordano. (2020, July) towards data science.
[12] Albert Zeyer, Patrick Doetsch, Paul Voigtlaender, Ralf Schlüter, and Hermann Ney, "A comprehensive study of deep bidirectional LSTM RNNS for acoustic modeling in speech recognition," IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), pp. 2462-2466, 2017.
[13] Shiliang Sun, Zehui Cao, Han Zhu, and Jing Zhao, "A Survey of Optimization Methods from a Machine Learning Perspective," cs.LG, October 2019.
[14] Hana Yousuf, Michael Lahzi, Said A. Salloum, and Khaled Shaalan, "A Systematic Review on Sequence to Sequence Neural Network and its Models," International Journal of Electrical and Computer Engineering (IJECE), vol. 11, no. 3, pp. 2315-2326, June 2021.
[15] Darrell Conklin, "Music Generation from Statistical Models," Journal of New Music Research, vol. 45(2), June 2003.
[16] Alexander Agung Santoso Gunawan, Ananda Phan Iman, and Derwin Suhartono, "Automatic Music Generator Using Recurrent Neural Network," International Journal of Computational Intelligence Systems, vol. 13(1), pp. 645-654, 2020.
[17] Ilya Sutskever, Oriol Vinyals, and Quoc V. Le, "Sequence to Sequence Learning with Neural Networks," cs.LG, vol. 3, p. 14, December 2014.
[18] Pedro Ferreira, Ricardo Limongi, and Luiz Paulo Fávero, "Generating Music with Data: Application of Deep Learning Models for Symbolic Music Composition," Applied Sciences, April 2023.

©IJRASET: All Rights are Reserved | SJ Impact Factor 7.538 | ISRA Journal Impact Factor 7.894 |



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

# INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Call : 08813907089 (24*7 Support on Whatsapp)