International Journal for Multidisciplinary Research (IJFMR)

E-ISSN: 2582-2160 ● Website: www.ijfmr.com ● Email: editor@ijfmr.com

CyberTune -Dynamic Remixing and Hack your Playlist to match Beat Alchemy to transform your Sound for Human-Centric AI

Dr. M.K. Jayanthi Kannan[1], Prabhat Ranjan Srivastava[2], Anushka Sizaria[3], Elisabeth Varghese[4], Aditya Sharma[5], Divyanshu Chauhan[6]

[1]Professor, School of Computing Science Engineering and Artificial Intelligence, VIT Bhopal University, Bhopal-Indore Highway, Kothrikalan, Sehore, Madhya Pradesh - 466114

[2,3,4,5,6]Student, School of Computing Science Engineering and Artificial Intelligence, VIT Bhopal University, Bhopal-Indore Highway, Kothrikalan, Sehore, Madhya Pradesh - 466114

Abstract:

Music style transfer, a concept originating in image processing, has gained traction in the audio domain as an emerging area of research. This study explores the application of advanced machine learning models to genre transformation tasks, focusing on preserving the underlying structure of musical compositions while adapting stylistic elements like rhythm, harmony, timbre, and instrumentation. Generative models, including Generative Adversarial Networks (GANs) and Variational Autoencoders (VAEs), form the backbone of these methods, supported by novel architectures like StarGAN, CycleGAN, and Transformer-VAE hybrids. The integration of advanced feature extraction techniques, such as spectrogram-based analysis and chroma representations, enhances genre-specific adaptability. Despite their potential, challenges remain in disentangling content from style, improving training stability, and achieving computational efficiency. To address these, the study also examines techniques like spectral normalization and timbre disentanglement through supervised and self-supervised learning approaches. The outcomes of this research contribute to enhancing automated music production tools, advancing audio processing methodologies, and fostering creative applications in the entertainment industry. By analyzing existing methods and proposing innovative solutions, this study aims to further the intersection of artificial intelligence and music, paving the way for personalized and dynamic musical experiences. The research focuses on bridging the gap between music consumption and creation and the rise of the AI in tech to democratize music production.

Keywords: Dynamic Remixing, AI-driven Music, CyberTune, Redefine DJ, Personalizing Music, Intelligent Mixing, AI-enhanced remixing, AI Technique for Music Personalization, Variational Autoencoders, Spectrogram-Based Approaches, Genre Style Application using GAN and Magenta.

INTRODUCTION

In a world where music transcends barriers and serves as a universal language, imagine a groundbreaking tool that empowers listeners and creators to reimagine songs with unparalleled ease. What if you could transform a soulful ballad into an upbeat electronic anthem or morph a classical symphony into a jazz-infused masterpiece—all with a single click? This project delves into the innovative crossroads of artificial intelligence and music, leveraging advanced machine learning techniques to revolutionize how we experience and create music. By enabling the seamless transformation of a song's genre from its original form into a specified target style, this technology opens doors to boundless creative exploration and highly personalized listening experiences. By implementing a robust pipeline that preprocesses MIDI files into piano rolls and chroma features, the project bridges the gap between computational processing and musical creativity. Whether you're an artist experimenting with new sounds, a producer looking for inspiration, or a listener seeking fresh perspectives on familiar tunes, this project redefines the boundaries of musical expression. It invites everyone to become a co-creator, blending technology and artistry to craft music that resonates on a deeply personal level.

LITERATURE REVIEW AND CURRENT APPROACHES

Music style transfer has emerged as an innovative field at the intersection of artificial intelligence, machine learning, and audio processing. While the concept of style transfer originated in the image domain, its adaptation to audio presents a novel challenge, necessitating methods that disentangle and manipulate harmonic, rhythmic, and timbral features. In music, this extends beyond visual patterns to understanding the intrinsic structure of sound.

Traditional machine learning techniques initially focused on genre classification, relying on features like chromas and Mel-Frequency Cepstral Coefficients (MFCCs). However, these models struggled to encapsulate the complexities of musical structures. The advent of generative deep learning techniques, including Generative Adversarial Networks (GANs) and Variational Autoencoders (VAEs), brought significant progress in music style transfer. These approaches introduced data-driven methods capable of understanding nuanced relationships between harmonic content and stylistic elements, propelling advancements in automated music generation.

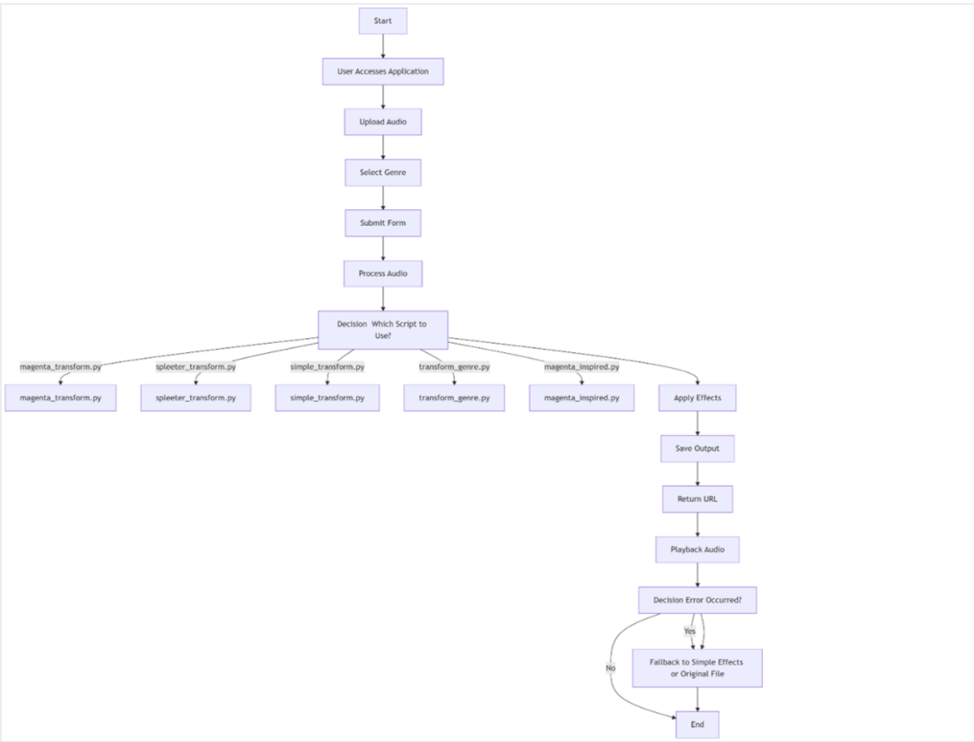
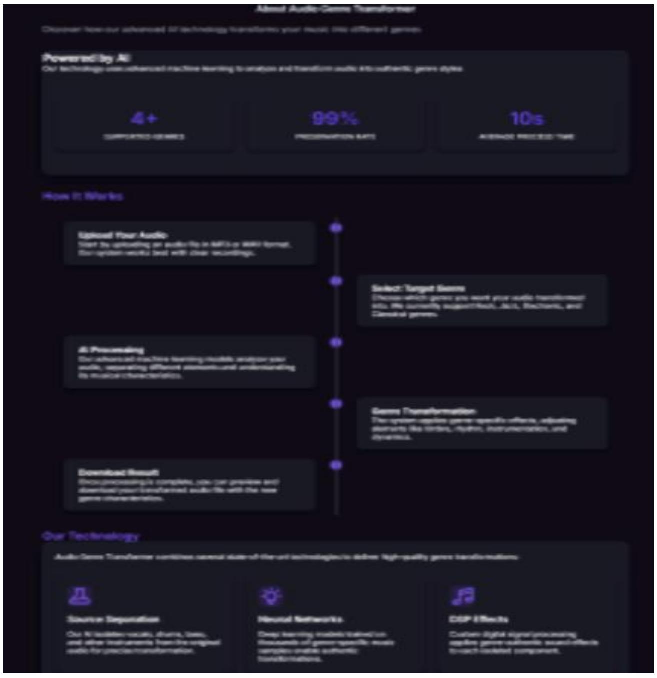
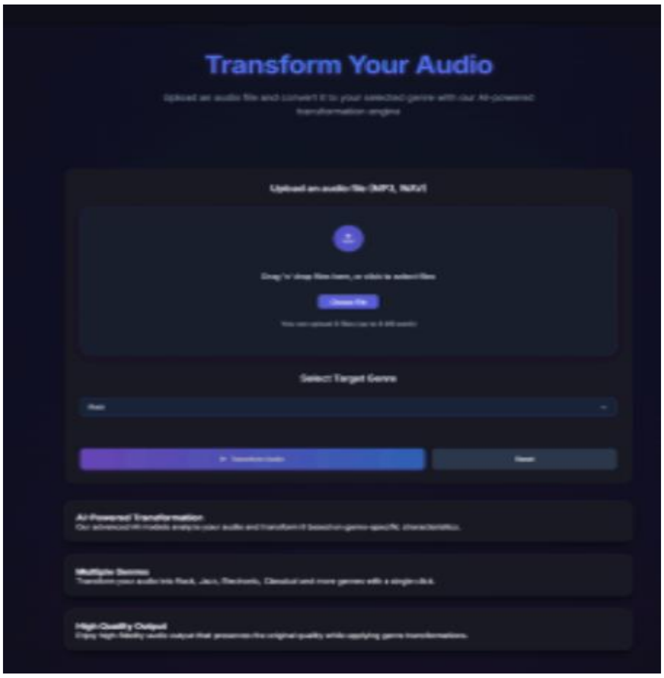
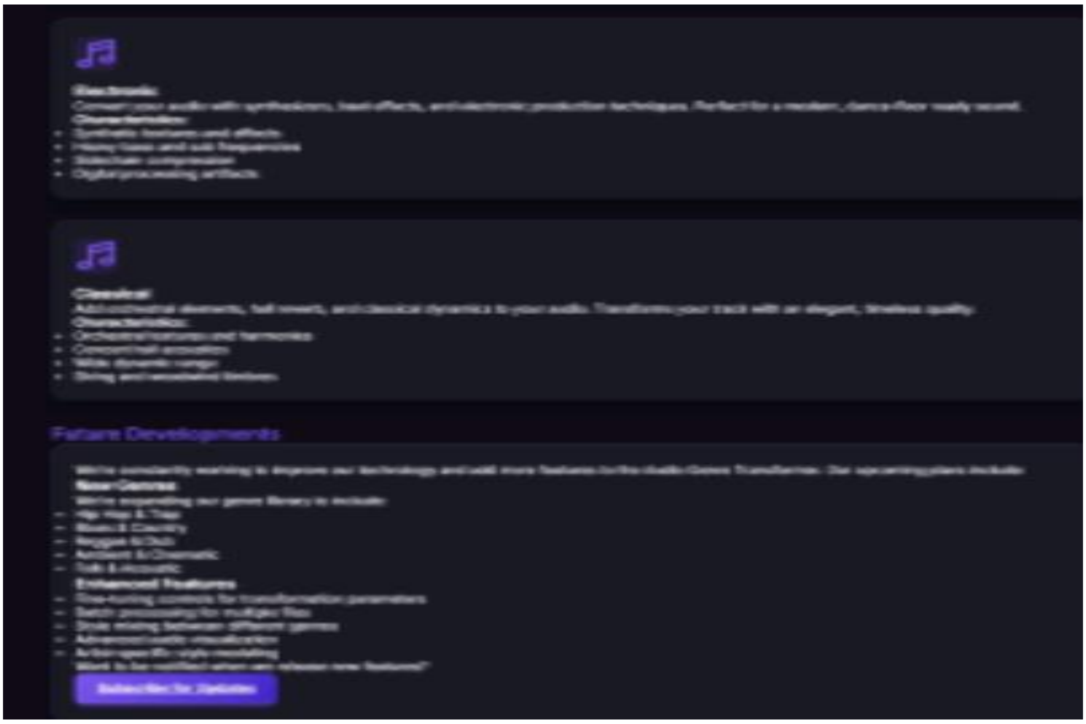


Figure 1: CyberTune -Dynamic Remixing Workflow Diagram







Figures 2,3, & 4: Cybertune- Different Genres and Future Feature Development Plans in our Model

Spectrogram-based approaches are spectrograms that serve as a foundation for style transfer by converting audio signals into visual time-frequency representations. Prominent methods include Convolutional Neural Networks (CNNs):- CNNs operate on spectrograms, employing transformations like Short-Time Fourier Transform (STFT) and Mel spectrograms. These highlight rhythmic and harmonic structures while encoding stylistic elements. However, performance depends on the richness of spectrogram representations. Diffusion Models, Emerging latent diffusion models reconstruct spectrograms iteratively, achieving multi-instrument timbre transformations. These models outperform GANs in audio clarity and realism, eliminating the need for preprocessing techniques such as source separation.

Variational Autoencoders, Vector-Quantized VAEs (VQ-VAE) models separate musical attributes like timbre and pitch through self-supervised learning, enabling one-shot timbre transfer. They excel in symbolic music domains where hierarchical disentanglement is essential. Transformer-VAE Hybrid, Innovations like MuseMorphose combine VAEs with Transformers to handle long-sequence modeling, offering precise control over rhythmic intensity and polyphony at bar levels. A challenge arises when genre is defined purely by the timing and pitch of the played notes, without explicit instrumentation information. Additionally, the paper highlights that only a few key notes often define the genre of a piece, suggesting that genre attribution in music may rely on specific tonal or rhythmic elements, rather than the full spectrum of the composition. A useful feature in this context is chromas, which represent the pitch class content of music and are particularly effective in genre classification. Chromas capture the harmonic structure of a piece by focusing on the twelve pitch classes of the chromatic scale, making them an excellent tool for analyzing music across genres. They are easy to compute and provide a compact, informative representation of musical content, which can be especially beneficial for tasks such as genre transfer.

METHODOLOGY

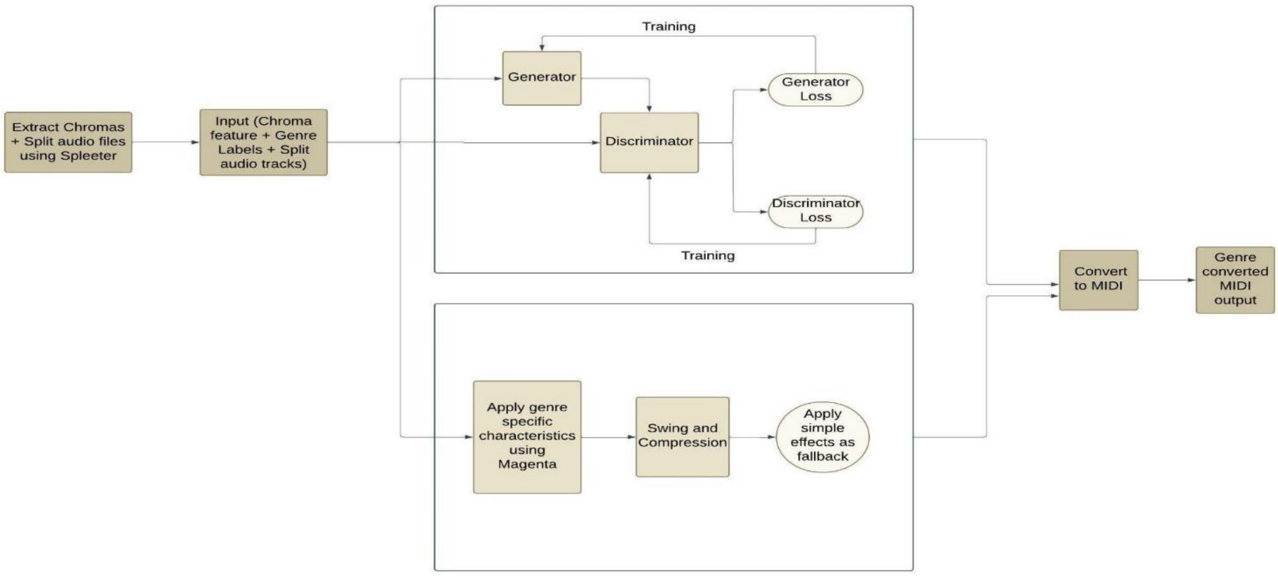


Figure 5: CyberTune -Dynamic Remixing Architecture of Morph My Tune Model

The architecture for the project utilizes a combination of machine learning techniques and audio processing methods to convert songs from one genre to another. The process begins with audio tracks, which are separated into individual components (such as vocals, drums, and bass) using Spleeter's 4-stem separation model. From these stems, we extract chroma features using LibROSA, capturing the harmonic content of the music while minimizing timbral differences. To enhance our model's understanding of genre characteristics, we add genre labels (eg. indicating whether the music is Pop or Classical) and style embeddings derived from reference tracks using Magenta's MusicVAE. The core Generative Adversarial Network (GAN) model is designed to learn and transform musical styles. After the GAN processes the input, we convert the transformed features back into MIDI format using a chroma-to-MIDI mapping technique. This is followed by dynamic range compression to enhance audio quality.

Figure 6: CyberTune -Dynamic Remixing and Hack your Playlist Prototype design of user interface

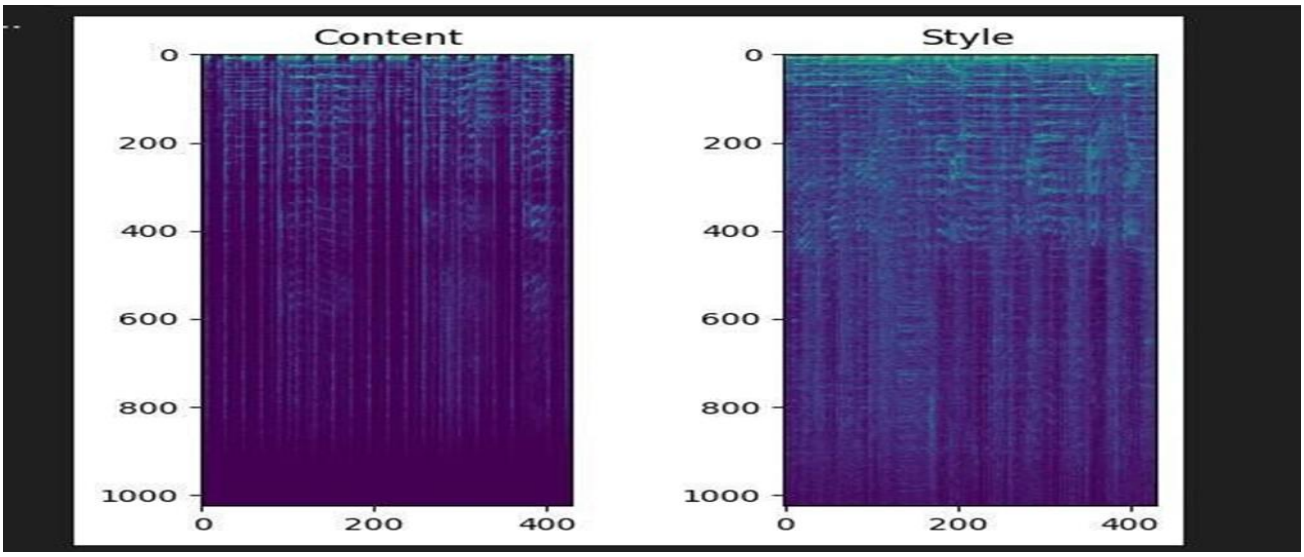


Figure 7: Visualization of Style Transferred Audio CyberTune -Dynamic Remixing and Hack your Playlist

PROJECT MODULES IMPLEMENTATION

This section elaborates on the structured approach adopted for implementing the core modules of the project. Data Preprocessing. This module involves handling the initial processing of audio files. MIDI/Audio Loading: Loads MIDI files or audio waveforms, using LibROSA to handle audio formats and extract relevant features if necessary. HPSS (Harmonic/Percussive Source Separation): Decomposes the audio into harmonic and percussive components using LibROSA's hpss function. This allows separate processing of tonal and rhythmic elements. Feature Extraction: Extracts features crucial for genre transformation, including chroma features (using librosa.feature.chroma_cqt), onset detection (using librosa.onset.onset_strength), and tempo estimation (librosa.beat.beat_track). Spleeter Integration: Audio tracks are split into stems (vocals, drums, bass, etc.).

4.1 Genre Style Application using GAN and Magenta

The core module of the project uses a Generative Adversarial Network (GAN) to map chroma features from one genre to another. The generator learns to modify the harmonic structure while maintaining musical coherence, and the discriminator ensures the output aligns with the target genre. For other features of the audio, Magenta is integrated for high-level stylistic transformations. It includes conditional logic to apply simplified harmonic enhancements and style-specific coloration. This module functions that apply audio effects commonly associated with specific genres (such as distortion for rock, swing for jazz, and synthesized sounds for electronic music. Audio Effect Processing, this module includes functions for common audio processing tasks used in style transformation. A simple audio compressor reduces the dynamic range of the audio signal, enhancing loudness and punchiness. Training and Learning, This step involves optimizing the GAN by fine-tuning hyperparameters, adjusting the loss functions, and ensuring stability in training. Proper training helps the model produce realistic genre-transformed outputs while preventing mode collapse or overfitting. Output generation, Once the models generate the music, this module converts the generated content back into MIDI format. The final MIDI files retain the structure and rhythm of the original composition while reflecting the characteristics of the target genre, making them playable on standard MIDI software and instruments. User Interface, This module involves developing the interface to display the results of the model and to try it with other music from different genres

THE IMPLEMENTATION SCREENSHOTS

```
[33]: def generator(z):
          # First hidden layer
          G_h1 = tf.nn.leaky_relu(tf.matmul(z, G_W1) + G_b1, alpha=0.2)
          G_h1 = tf.nn.dropout(G_h1, rate=0.3)

          # Second hidden layer
          G_h2 = tf.nn.leaky_relu(tf.matmul(G_h1, G_W2) + G_b2, alpha=0.2)
          G_h2 = tf.nn.dropout(G_h2, rate=0.3)

          # Output layer
          G_log_prob = tf.matmul(G_h2, G_W3) + G_b3
          G_prob = tf.nn.sigmoid(G_log_prob)

          return G_prob
```

```
[35]: def discriminator(x, c):
          D_h1 = tf.nn.leaky_relu(tf.matmul(tf.concat([x, c], 1), D_W1) + D_b1, alpha=0.2)
          D_h1 = tf.nn.dropout(D_h1, rate=0.3)  # Dropout for regularization

          # Second hidden layer
          D_h2 = tf.nn.leaky_relu(tf.matmul(D_h1, D_W2) + D_b2, alpha=0.2)
          D_h2 = tf.nn.dropout(D_h2, rate=0.3)

          # Output layer
          D_logit = tf.matmul(D_h2, D_W3) + D_b3
          D_prob = tf.nn.sigmoid(D_logit)

          return D_prob, D_logit
```

Figure 8: The Generator and Discriminator Logic of Morph My Tune using GANs

```python
def apply_simple_effects(audio, sr, genre):
    """Apply simple audio effects based on genre when Magenta fails"""
    print(f"[PYTHON] Applying simple effects for {genre} genre")

    if genre.lower() == "rock":
        # Apply distortion effect for rock
        audio = np.tanh(audio * 2.0) * 0.7

        # Apply basic EQ
        b, a = librosa.filters.butter(4, 120/(sr/2), btype='highpass')
        audio = librosa.filtfilt(b, a, audio)

    elif genre.lower() == "jazz":
        # Apply warmth
        b, a = librosa.filters.butter(4, 300/(sr/2), btype='lowshelf')
        audio = librosa.filtfilt(b, a, audio)

    elif genre.lower() == "electronic":
        # Apply basic beat emphasis
        y_harmonic, y_percussive = librosa.effects.hpss(audio)
        audio = y_harmonic * 0.6 + y_percussive * 1.4

        # Add sub bass
        b, a = librosa.filters.butter(4, 80/(sr/2), btype='lowshelf')
        audio = librosa.filtfilt(b, a, audio)

    elif genre.lower() == "classical":
        # Apply reverb
        ir_length = int(sr * 1.5)
        ir = np.exp(-np.linspace(0, 8, ir_length))
        ir = ir / np.sum(ir)
        audio = np.convolve(audio, ir, mode='same')

    # Apply normalization
    audio = audio / np.max(np.abs(audio)) * 0.9

    return audio
```

Figure 9: Applying simple effects to audio track as fallback

```
Iter 195000: D_loss: 0.8013, G_loss: 12.1789
Iter 195000: D_loss: 0.8850, G_loss: 9.1929
Iter 195000: D_loss: 0.8686, G_loss: 12.1402
Iter 195000: D_loss: 0.6217, G_loss: 4.5343
Iter 196000: D_loss: 1.1054, G_loss: 2.5215
Iter 196000: D_loss: 0.7763, G_loss: 4.1651
Iter 196000: D_loss: 0.8827, G_loss: 2.1574
Iter 196000: D_loss: 1.1836, G_loss: 1.0449
Iter 197000: D_loss: 0.2328, G_loss: 8.1033
Iter 197000: D_loss: 0.0416, G_loss: 6.8467
Iter 198000: D_loss: 0.0562, G_loss: 16.1392
Iter 198000: D_loss: 0.2334, G_loss: 5.5489
Iter 198000: D_loss: 0.2417, G_loss: 11.9180
Iter 198000: D_loss: 1.7105, G_loss: 2.7716
Iter 199000: D_loss: 0.5540, G_loss: 6.1118
Iter 199000: D_loss: 0.8504, G_loss: 6.3358
Iter 199000: D_loss: 0.6992, G_loss: 8.4520
Iter 199000: D_loss: 0.9432, G_loss: 5.4383
Iter 199000: D_loss: 0.8274, G_loss: 6.7732
Iter 199000: D_loss: 0.6989, G_loss: 6.5673
Iter 199000: D_loss: 1.1027, G_loss: 7.3757
Iter 200000: D_loss: 1.0188, G_loss: 2.1252
Iter 200000: D_loss: 1.0018, G_loss: 4.2642
Iter 200000: D_loss: 1.1709, G_loss: 2.0874
Iter 200000: D_loss: 0.3161, G_loss: 11.4296
```

Figure 10 The loss values for every iteration in 200000 epochs

RESULT FINDINGS

The results of the genre transformation process show that the generated audio retains similarity to the original input, with variations in relation to the intended target genres. While the system successfully captures some genre-specific characteristic features, it has a lot to offer in terms of optimization, which can potentially improve transformation fidelity. Key Observations are, Audio Similarity: The transformed audio bears some resemblance to the original piece, which shows that the model can successfully transform harmonic constructs and stylistic properties to fit other genres. However, variation is often so subtle that further work in that direction is warranted to achieve more prominent genre traits. Noise and Artifacts:

Output quality is great; however, some generated audio includes random notes and noisy data. This would imply that further filtering or refining techniques should improve clarity and coherency for wearing the final output. Processing Time: The processing time for transforming audio files is quite efficient, usually taking only a few minutes per file. However, it should be noted that the longer the audio, the longer the processing time, thus affecting users. User Interface Performance: The user interface performs smoothly and effectively to provide user interaction such as uploading files and selecting genres. The clear interface also makes it easy for the user to understand the process of transformation. Genre Limitations: Currently, the system supports transformation in only four genres: jazz, rock, electronic, and classical. More genres would require additional training data, as well as some changes to the model architecture.

PREDICTIONS

As the field of audio processing and machine learning develops further, a number of predictions can be made about future features and capabilities of the audio transformation system. With the advancements in machine learning methods, especially the use of diffusion models and transformer architectures, we expect a considerable boost in the quality of audio generated. These models will be able to generate more complex and subtle transformations, resulting in outputs that closely match the target genre features while reducing artifacts and noise. A further developed version of the system can provide users with more control over particular audio aspects like rhythm, timbre, and dynamics. This can be done by incorporating sophisticated algorithms that provide context-sensitive controls in accordance with user choices or particular musical contexts. Additionally, including contextual data—lyrics, mood sensing, and visual information (e.g., album covers)—to enable more specific and emotionally connected genre transformations may result in a more customized user experience where transformations closely match the user's creative vision. The integration of stronger evaluation criterias that encompass quantitative measures (such as spectral similarity and signal fidelity) in combination with qualitative assessments through user-focused testing will raise the system's dependability. This will give stronger indications of performance as well as satisfaction on the user side. Future development can also involve a feature that learns from users' listening habits, enabling the system to adjust its transformations according to personal preferences over time. This can be achieved through machine learning methods that observe user interactions and feedback to make continuous improvements in genre adaptations. Furthermore, with the model being trained on more varied datasets, the support for more genres is expected to spread beyond the existing four (rock, jazz, electronic, classical). This will make the system more versatile and appealing to a larger audience. Ultimately, computational power and algorithmic enhancements are likely to decrease processing time dramatically, making real-time audio transformation possible even for longer files without any loss of quality.

CONCLUSION

The project *Morph My Tune* provides a newly developed contribution at the intersection of music creation and machine learning. This genre conversion tool illustrates how approaches— such as traditional feature extraction, novel feature extraction, neural style transfer, and generative algorithms (including generative adversarial networks (GANs), can produce a different version of a particular piece of music in a different style. Additionally, the tool incorporates generative architecture and can demonstrate the merging of spatial and temporal processing with the retention of musical aspects in transformations. Research problems remain in the area, including separating content from style in a complex audio source, and a danger of losing musical identity. Issues in practical use include limited data sets and significant computational costs to train these models, including considerations for real-time use. Evaluating musician results of genre conversion is a challenge, as objective metrics can represent limited aspects of the subjective nature of musical beauty. Some possible future suggestions for the improvements may involve newer architectures (e.g., diffusion models, newer transformers, or some new combination of recurrent neural networks (RNNs) with something else), gradual improvement in the stability of training by optimizing GANs, and possibly some user interactive feedback in the process.

REFERENCES

1. Gatys, Leon A. "A neural algorithm of artistic style." arXiv preprint arXiv:1508.06576 (2015).
2. Siddavatam, Irfan & Dalvi, Ashwini & Gupta, Dipen & Farooqui, Zaid & Chouhan, Mihir. (2020). Multi Genre Music Classification and Conversion System. International Journal of Information Engineering and Electronic Business. 12. 30-36. 10.5815/ijieeb.2020.01.04.
3. Kavitha, E., Tamilarasan, R., Poonguzhali, N., Kannan, M.K.J. (2022). Clustering gene expression data through modified agglomerative M-CURE hierarchical algorithm. Computer Systems Science and Engineering, 41(3), 1027-141. https://doi.org/10.32604/csse.2022.020634
4. Lei Zhang and R. Mo. 2022. Analysis of Guzheng Music Style Transformation Based on Generative Adversarial Networks. Mob. Inf. Syst. 2022 (2022).
5. P. Jain, I. Rajvaidya, K. K. Sah and J. Kannan, "Machine Learning Techniques for Malware Detection- a Research Review," 2022 IEEE International Students' Conference on Electrical, Electronics and Computer Science (SCEECS), BHOPAL, India, 2022, pp. 1-6, doi: 10.1109/SCEECS54111.2022.9740918.
6. Recurrent GAN Models for Music Generation. Azaan Rehman, Jeffrey Tsaw (2020).
7. M. K. Jayanthi, "Strategic Planning for Information Security -DID Mechanism to befriend the Cyber Criminals to assure Cyber Freedom," 2017 2nd International Conference on Anti-Cyber Crimes (ICACC), Abha, Saudi Arabia, 2017, pp. 142-147, doi: 10.1109/Anti-Cybercrime.2017.7905280.
8. Hamanaka, M., Hirata, K., & Tojo, S. (2022). Implementation of melodic morphing based on the generative theory of tonal music. Journal of New Music Research, 51(1), 86–102.
9. Kumar, K.L.S., Kannan, M.K.J. (2024). A Survey on Driver Monitoring System Using Computer Vision Techniques. In: Hassanien, A.E., Anand, S., Jaiswal, A., Kumar, P. (eds) Innovative Computing and Communications. ICICC 2024. Lecture Notes in Networks and Systems, vol 1021. Springer, Singapore. https://doi.org/10.1007/978-981-97-3591-4_21
10. Leif Sulaiman & Sebastian Larsson: Genre style transfer, Symbolic genre style transfer utilising GAN with additional genre-enforcing discriminators, © July 2, 2022.
11. Suresh Kallam , M K Jayanthi Kannan , B. R. M. , (2024). A Novel Authentication Mechanism with Efficient Math Based Approach. International Journal of Intelligent Systems and Applications in Engineering, 12(3), 2500–2510. Retrieved from https://ijisae.org/index.php/IJISAE/article/view/5722
12. O. Cífka, A. Ozerov, U. Şimşekli and G. Richard, "Self-Supervised VQ-VAE for One-Shot Music Style Transfer," ICASSP 2021 - 2021 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Toronto, ON, Canada, 2021, pp. 96-100, doi: 10.1109/ICASSP39728.2021.9414235.
13. M. K. J. Kannan, "A bird's eye view of Cyber Crimes and Free and Open Source Software's to Detoxify Cyber Crime Attacks - an End User Perspective," 2017 2nd International Conference on Anti-Cyber

Crimes (ICACC), Abha, Saudi Arabia, 2017, pp. 232-237, doi: 10.1109/Anti-Cybercrime.2017.7905297.
14. Dr. Naila Aaijaz, Dr. K. Grace Mani, Dr. M. K. Jayanthi Kannan and Dr. Veena Tewari (Feb 2025), The Future of Innovation and Technology in Education: Trends and Opportunities, ASIN : B0DW334PR9, S&M Publications, Mangalore, Haridwar, India-247667, ISBN-13 :978-8198488824, https://www.amazon.in/gp/product/B0DW334PR9/ref=ox_sc_act_title_1?smid=A2DVPTOROMUBNE&psc=1#detailBullets_feature_div
15. G., D. K., Singh, M. K., & Jayanthi, M. (Eds.). (2016). Network Security Attacks and Countermeasures. IGI Global. https://doi.org/10.4018/978-1-4666-8761-5
16. B. R M, S. Kallam and M. K. Jayanthi Kannan, "Network Intrusion Classifier with Optimized Clustering Algorithm for the Efficient Classification," 2024 5th International Conference on Intelligent Communication Technologies and Virtual Mobile Networks (ICICV), Tirunelveli, India, 2024, pp. 439-446, doi: 10.1109/ICICV62344.2024.00075.
17. Wu, Shih-Lun, and Yi-Hsuan Yang. "MuseMorphose: Full-song and fine-grained piano music style transfer with one transformer VAE." IEEE/ACM Transactions on Audio, Speech, and Language Processing 31 (2023): 1953-1967.
18. R M, B.; M K, J.K. Intrusion Detection on AWS Cloud through Hybrid Deep Learning Algorithm. Electronics 2023, 12, 1423. https://doi.org/10.3390/electronics12061423
19. Dr. M.K. Jayanthi Kannan, Aditya Sharma, Anushka Sizaria, Elisabeth Varghese, Divyanshu Chauhan, Prabhat Ranjan Srivastava, (Dec 2024), Morph My Tune: Mix. Master. Unleash Your Inner DJ to Redefine Your Rhythm, International Journal of Emerging Technologies and Innovative Research (2349-5162), (www.jetir.org | UGC and issn Approved), ISSN:2349-5162, Vol.11, Issue 12, page no. ppc358-c365, December-2024, Available at : http://www.jetir.org/papers/JETIR2412240.pdf
20. Python for Data Analytics: Practical Techniques and Applications, Dr. Surendra Kumar Shukla, Dr. Upendra Dwivedi, Dr. M K Jayanthi Kannan, Chalamalasetty Sarvani ISBN: 978-93-6226-727-6, ASIN : B0DMJY4X9N, JSR Publications, 23 October 2024, https://www.amazon.in/gp/product/B0DMJY4X9N/ref=ox_sc_act_title_1?smid=A29XE7SVTY6MCQ&psc=1
21. Balajee RM, Jayanthi Kannan MK, Murali Mohan V. Image-Based Authentication Security Improvement by Randomized Selection Approach. InInventive Computation and Information Technologies 2022 (pp. 61-71).
22. Brunner, Gino, Mazda Moayeri, Oliver Richter and Roger Wattenhofer. “Neural Symbolic Music Genre Transfer Insights.” PKDD/ECML Workshops (2019).