Power System Technology
ISSN:1000-3673

Received: 06-11-2024 *Revised: 15-12-2024* *Accepted: 05-01-2025*

Algorithmic Orchestration: Deep Learning Techniques in Music Generation

[1]Sampada K S, [2]Deepa S, [3]Prasanna Kumar M

[1]Associate Professor, CSE, RNS Institute of technology,

[2]Assistant Professor, CSE, RNS Institute of technology

[3]Associate Professor, CSE, RNS Institute of technology

Abstract

Music generation using Deep Learning presents a comprehensive system for generating music in ABC notation using character based recurrent neural networks (RNNs), accompanied by a conversion pipeline that transforms the output music from ABC notation to a playable audio format. The integration of character-based RNNs allows for the creation of coherent and melodic musical compositions, while the ABC- to-MIDI-to WAV conversion enhances the usability and accessibility of the generated music. The work starts by preprocessing the ABC notation dataset and representing it at the character level. A character-based RNN, such as a long short-term memory (LSTM) network, is then employed to learn the sequential dependencies within the ABC notation and generate music that follows the learned patterns and structures. The RNN model is trained on a substantial corpus of ABC- encoded music, enabling it to capture the statistical regularities and nuances of the dataset. To make the generated music readily playable, the project incorporates a conversion pipeline that translates the output from ABC notation to MIDI format. MIDI files serve as a widely supported industry-standard representation of music, making them compatible with a variety of digital audio workstations (DAWs) and synthesizers. The MIDI files are subsequently converted to WAV format, a universally recognized audio format suitable for playback on diverse platforms and devices. This project contributes to the field of AI-generated music by providing an integrated system for music generation in ABC notation, accompanied by a seamless conversion pipeline for playback in the universally supported WAV audio format. The combination of character- based RNNs, ABC notation, and the ABC-to-MIDI-to-WAV conversion offers a valuable tool for musicians, composers, and music enthusiasts, facilitating the generation, sharing, and playback of musical compositions. Future research may explore advanced synthesis techniques, incorporate additional musical features, or investigate other music notation systems to further enhance the capabilities and usability of AI generated music.

Index Terms: Music generation, AI, ABC, WAV, MIDI

Power System Technology
ISSN:1000-3673
Received: 06-11-2024 Revised: 15-12-2024 Accepted: 05-01-2025

I. INTRODUCTION

Although the boundaries of artificial intelligence(AI) re- main fugitive, computers can now perform musical tasks that were formerly associated simply with naturally intelligent mu- sicians. After a literal note, this paper sermonizes on the need for AI ways in four areas of musical exploration composition, performance, music proposition, and digital sound processing. The coming part checks recent work involving AI and music. The discussion concentrates on operations in the four areas of exploration just mentioned. The final part examines how AI ways of planning and literacy could be used to expand the knowledge base and enrich the gist of musically intelligent systems. Amper Music is a pall- grounded, AI-driven music composition platform. The system reportedly generates unique musical selections grounded on the mood, style, and duration parameters named by the stoner. Once these selections are made the stoner can make fresh edits before the composition is complete. The company claims that the platform is erected on a set of samples of music data Popgun was Launched in January 2017, Australia- grounded incipiency, reportedly uses deep literacy through a platform called ALICE to accompany or compound musical compositions. The incipiency's platoon, substantially composed of software masterminds, argues that the thing of their work is to foster collaboration and " not replace mortal capacities. " Professional quality music recording isn't cheap and certain music outfit alone can reach overhead of $ 10,000. For the average independent artist, numerous specialized factors of the music product process are cost prohibitive. Companies like LANDR are trying to level the platform by giving musical artists the capability to master or professionally complete their songs at affordable rates. By barring the need for specialized training, the music composition platforms we've bandied in this composition are situating themselves to be readily accessible to guests. still, these platforms are still in the early stages and their profit models aren't yet clear. also, it's too soon to say with certainty who the target customer will be for these companies. These factors will inform the long term viability of these music composition platforms. There are upto Four Major ways AI can be inculcated into music as per exploration into this field recommends as the major pillars of AI music generation Music Recommendation is one similar possible field. As streaming services like spotify need to make upon complex algorithms to actually personalize the data for the stoner. Major music service providers, similar as Spotify, Apple, Pandora, and Amazon, use AI to dissect the preferences of their listeners and 'prognosticate' new favorite songs. AI- driven recommendation machines are used by music streaming platforms to enhance their services. By exercising the technology, they can under- stand the tastes of their guests and offer quality recommendations for farther listening. For illustration, in order to produce accurate recommendations for the stoner, these apps analyzes the harkening history of the entire subscriber base against the songs in your heavy gyration. AI compares the playlists you frequent to the playlists of other Spotify subscribers to detect druggies with tastes that are

Power System Technology
ISSN:1000-3673
Received: 06-11-2024 Revised: 15-12-2024 Accepted: 05-01-2025

close to yours. After that, the AI finds song recommendations grounded on your own taste and on the music that druggies with analogous tastes hear to. Music generation is the most popular and is the extensively known operation While AI- generated music seems like a lazy and ‘ weird new ’ way of creation to some, David Bowie, one ofthe most unique and original musicians, used the AI- such like verbalizer script to induce the lyrics for his songs back in the s.s. Basically, it just provides access to sophisticated yet easy- to- use tools and measureless melodies to play around with. Whether a musician or not, you can spark your creativityand have fun with AI grounded services for music generation like Amper, AIVA, and Boomy. Music

Editing is extensively known by the music community as before songs come available for our harkening pleasure, they go through multitudinous editing way. numerous hours of attentive listening, making mixing and learning choices, with each changing the final sound of a song or a record. It’s a lengthy process that requires a lot of coffers from mortal specialists another music- related task that can be accelerated and enhanced by artificial intelligence. Music Analysis is a process of reacquiring music data and breaking songs down into their characteristics. It helps musicians, markers, directors, publishers, and playlists janitors organize and recommend music. With the help of AI, the analysis can be performed significantly briskly and more directly. It’s noted that by around 2013 a single band made the loftiest profit of 268 million bones and after which analogous music and compositions were copied contiguously till a new mega hit came by after which the process just repeated itself. This can also be criticized on a general lack of idea present in the general public but machine don’t “suppose” as humans and are seen and observed to develop new and interesting patterns in the field of AI. They may, still, increasetheir effectiveness and sound optimization capabilities with the right tools. Artificial intelligence is employed in colorful operations, including music learning, soundtrack development, and recommendation. In the following times, AI is projected to bring up a slew of new musical possibilities.

II. LITERATURE SURVEY

Although the boundaries of artificial intelligence (AI) remain elusive, computers can now perform musical tasks that were formerly associated exclusively with naturally intelligent musicians. After a historical note, this paper sermonizes on the need for AI techniques in four areas of musical research: composition, performance, music theory, and digital sound processing. The next part surveys recent work involving AI and music. The discussion concentrates on applications in the four areas of research just mentioned. The final part examines how AI techniques of planning and learning could be used to expand the knowledge base and enrich the behavior of musically intelligent systems as told by Briot, Jean-Pierre, Gaetan Hadjeres, and Francois Pachet in their survey of ”Deep learning techniques for music generation-a survey”[9].

Power System Technology
ISSN:1000-3673
Received: 06-11-2024 Revised: 15-12-2024 Accepted: 05-01-2025

Amper Music is a cloud-based, AI-driven music composition platform. The system reportedly generates unique musical selections based on the mood, style, and duration parameters selected by the user. Once these selections are made the user can make additional edits before the composition is complete. The company claims that the platform is built on a set of samples of music data Popgun was Launched in January 2017, Australia-based startup, reportedly uses deep learning through a platform called ALICE to accompany or augment musical compositions. The startup's team, mainly composed of software engineers, argues that the goal of their work is to foster collaboration and "not replace human abilities." as told by a general tech article compiled by a unanimous user.

By eliminating the need for technical training, the music composition platforms we have discussed in this article are positioning themselves to be readily accessible to clients. However, these platforms are still in the early stages and their revenue models are not yet clear. Additionally, it is too soon to say with certainty who the target client will be for these companies. These factors will inform the long term viability of these music composition platforms.

There are upto Four Major ways AI can be inculcated into music as per research into this field recommends as the major pillars of AI music generation Music Recommendation is one such possible field. As streaming services like spotify need to build upon complex algorithms to actually personalize the data for the user. Major music service providers, such as Spotify, Apple, Pandora, and Amazon, use AI to analyze the preferences of their listeners and 'predict' new favorite songs. AIdriven recommendation engines are used by music streaming platforms to enhance their services. By utilizing the technology, they can understand the tastes of their customers and offer quality recommendations for further listening.

Music generation is the most popular and is the widely known application While AI-generated music seems like a lazy and 'weird new' way of creation to some, David Bowie, one of the most unique and original musicians, used the AI- like verbalizer script to generate the lyrics for his songs back in the 90s. s. Essentially, it just provides access to sophisticated yet easy-to-use tools and limitless tunes to play around with. Whether a musician or not, you can spark your creativity and have fun with AI based services for music generation like Amper, AIVA, and Boomy. as told by the tech article written by Hao-Wen Dong, Wen-Yi Hsiao, Li- Chia Yang, and YiHsuan Yang. MuseGAN [8].

Following are few existing Music generation AI.

AVIA: AIVA is a music composing service that uses deep- learning algorithms to create original music and produce variations of existing songs. It was 'classically trained', having listened to large amounts of musical works written by Mozart, Bach, Beethoven, etc. to learn to capture the music theory concepts and create models and compositions of its own.

Power System Technology
ISSN:1000-3673
Received: 06-11-2024 Revised: 15-12-2024 Accepted: 05-01-2025

LALAL.AI: LALAL.AI is a source separation tool that uses machine-learning algorithms to target and divide individual audio channels from a mixture of sounds in recorded music. Breaking a mixed song into its constituent contributions is such a highly complex task, engineers and research scientists of the world have been studying it and trying to come up with effective solutions for over 50 years. Though an ideal way to separate music sources is yet to be found, LALAL.AI's technique is definitely up there, achieving state-of-the-art results.

HUMTAP: Humtap is a music-generating and sharing platform powered by artificial intelligence. It enables users to create full-fledged music compositions and music videos in real-time just by hamming a melody and tapping a beat on an iPhone screen. The application transforms creative inputs into enhanced and produced songs that can be played in different genres and in the styles of renowned artists like Depeche Mode, Drake, and Metallica. Singing and rapping recordings can be improved with a selection of vocal effects. Humtap allows creating music as a solo user and by collaboration with other community members.

Limitations of Existing AI:

- *Limited Melodic Variation*: Existing systems may struggle to generate diverse and intricate melodies, resulting in a lack of melodic variation in the generated music. This project aims to overcome this limitation by employing a character-based RNN model trained on a diverse dataset of folk tunes, enabling the generation of more nuanced and diverse melodies.

- *Lack of Realistic Audio Playback*: Existing systems may generate MIDI files but struggle to provide realistic audio playback due to limited instrument sounds and expression. This project addresses this limitation by utilizing PrettyMIDI and genre-specific Sound Font files, enabling the conversion of MIDI files to WAV with more realistic and genre- appropriate instrument sounds.

- *Complexity of Training Process*: Some existing systems require manual feature engineering or complex preprocessing steps to train the model effectively. In contrast, this project utilizes the Keras sequential model, simplifying the training process by abstracting away many of the complexities, such as handling the input sequences and managing the model architecture.

- *Limited User-Friendliness*: Existing systems may lack a user-friendly interface or require complex technical knowledge to operate effectively. In contrast, this project aims to provide a seamless workflow from music generation to audio playback, ensuring accessibility and ease of use for musicians, composers, and music enthusiasts of varying technical backgrounds.

- *Limited Genre Adaptability*: Existing systems may have difficulty adapting to different musical genres, resulting in a lack of genre-specific characteristics in the generated music. This project addresses this limitation by incorporating genre- specific SoundFont files, allowing the generated compositions to align with the selected genre and produce more

Power System Technology
ISSN:1000-3673
Received: 06-11-2024 Revised: 15-12-2024 Accepted: 05-01-2025

authentic and genre-appropriate musical outcomes.

III. PROPOSED SYSTEM

The objective of this work is to develop an AI-powered music generator that can create original music in ABC notation, which is melodious and good to hear, Human-Readable Notation, Flexible Output Formats, Efficient Training and Convergence, Enhanced Creativity and Genre-Specific Sound. The model architecture as shown in Fig 1, can be divided into four main stages:

• *Data Preprocessing*: In this stage, the input dataset, consisting of 1200 folk tunes in ABC notation from the Nottingham Music Dataset, is prepared for training the character-based RNN model. The dataset is processed to extract individual tunes and split them into sequences of a fixed length. Each tune is divided into overlapping sequences, typically of size 64 characters, to capture the sequential nature of music. The sequences serve as the input and output pairs for training the model. The dataset is then divided into batches, with each batch containing 16 sequences, to facilitate efficient training.

• *Model Training*: The model training stage involves the construction and training of the character-based RNN model using the Keras sequential model framework. The model architecture consists of an embedding layer, followed by three hidden layers, each containing 256 LSTM units. The embedding layer maps the character-based input to continuous vectors, allowing the model to learn meaningful representations of the music. The LSTM layers capture long-term dependencies in the sequences, enabling the generation of coherent and musically meaningful output. The model is trained using the categorical cross-entropy loss function and optimized using the Adam optimizer. The training process is iterated over multiple epochs, typically set to 100, to ensure convergence and improved performance.

When adding to the layers present to the keras sequential model the return sequences is set to true with a hidden layer of two- fifty-six this is to ensures that for every input that is giving, will get an output. Otherwise, it'll give output at the end, only after consuming the whole sequence. The parameter stateful is also set to true to final state of the current batch is used as initial state for the next batch, instead of resetting it to zero. This helps the model learn longer sequence. As mentioned above the softmax activation layer is a multiclass classification problem since the output should be one of the 87 unique characters.

• *Music Generation*: Once the model is trained, the music generation stage involves the generation of new music in ABC notation. The process starts with providing a seed sequence, either randomly selected or user-specified, as the initial input to the model. The model then generates the next character in the sequence, which is appended to the input, forming a new input sequence. This iterative process continues, generating subsequent characters and expanding the sequence. The generation can be constrained by a maximum length or

Power System Technology
ISSN:1000-3673
Received: 06-11-2024 Revised: 15-12-2024 Accepted: 05-01-2025

determined by a predefined termination condition. The model's stochastic nature introduces randomness, resulting in the creation of diverse and unique musical compositions.

• *Audio Conversion*: In the final stage, the generated music in ABC notation is converted to both MIDI and WAV formats to enable audio playback. The ABC notation is first converted to a MIDI file using the Music21 Python library, which provides functionalities for parsing and manipulating music notation. The MIDI file contains musical events, note information, and control parameters. The PrettyMIDI library is then utilized to process the MIDI file, applying genre-specific SoundFont files that enhance the instrument sounds and expressiveness. Finally,

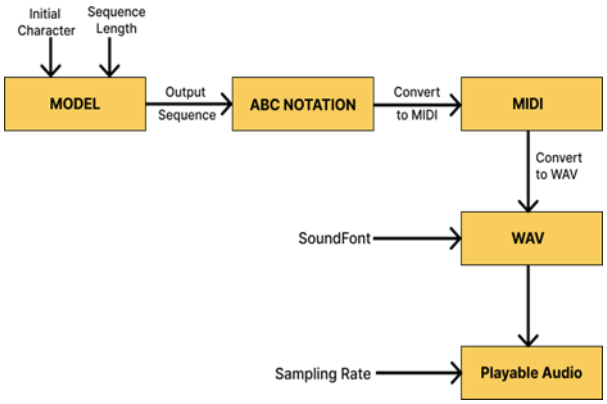

Fig 1. Model Architecture.

the MIDI file is converted to a WAV file, a universally playable audio format, allowing the generated music to be enjoyed on various platforms and devices.

IV. THE WORKING

The Dataset used for this project is the Notingham Dataset which is a collection of 1200 American and British folk songs.

A. *Formatting Input File*

The basic crux of the music generation is by using the keras sequential model with a batch size of 15 and epochs of 100 with the famous Adam optimizer with a softmax optimizer this allows the trained a neural network to solve multiclass classification problem that the raw outputs of the neural networks often struggle to interpret from hence the softmax activation function simplifies this by making the neural network's outputs easier to interpret. The softmax activation function transforms the raw outputs of the neural network into a vector of probabilities, essentially a probability distribution over the input classes. The dataset is converted into dictionary by converting every character in the Dataset to a index where ch is the key and index in key1: index1, key2: index2, ... format we create a dictionary which

Power System Technology
ISSN:1000-3673
Received: 06-11-2024 Revised: 15-12-2024 Accepted: 05-01-2025

encompasses all the eighty- six alphabets in ABC-notations sheets with numbers ranging from zero to eighty-five this is necessary as the machine cannot understand text and tokenization of text cannot be done for the ABC notations present in the dataset.

B. Creating Batches

We will feed the RNN with the sequence of characters one by one and it will output the next character in the sequence. This allows us to generate a new tune one character at a time. First we will encode each character in the input into a vector using one-hot encoding and feed them into RNN one at a time.

C. Model Architecture

When adding to the layers present to the keras sequential model the return sequences is set to true with a hidden layer of two-fifty-six this is to ensures that for every input that is giving, will get an output. Otherwise, it'll give output at the end, only after consuming the whole sequence.

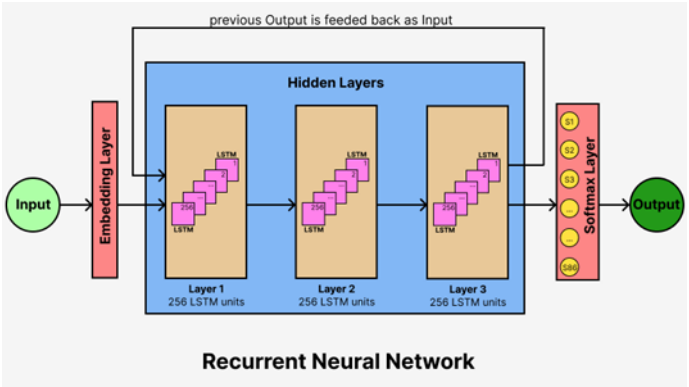

Fig.2. RNN Model with 3 hidden LSTM layers

The parameter stateful is also set to true to final state ofthe current batch is used as initial state for the next batch, instead of resetting it to zero. This helps the model learn longersequence.

As mentioned above the softmax activation layer is a multi-class classification problem since the output should be one of the 87 unique characters. So the last layer is a SoftMax layer with 87 activations.

D. Limitations and future works

It lack the human creativity behind music generation. The originality of the music can controversial since the is trained on a dataset and it will eventually mimic the patterns from that dataset. It may lack the unique sound created by a human.

AI has the potential to replace musicians and composers and cause job loss and reduce

Power System Technology
ISSN:1000-3673
Received: 06-11-2024 Revised: 15-12-2024 Accepted: 05-01-2025

music quality. AI can causesaturation in musical style since algorithms generate musicsimilar to existing ones rather than completely original music. It can be difficult to control and fit the music into a specific style or genre. Since the algorithm used to generate the music may not be own by the human, copyrights and monetization of AI generated music can be an ethical issue.

V. RESULT

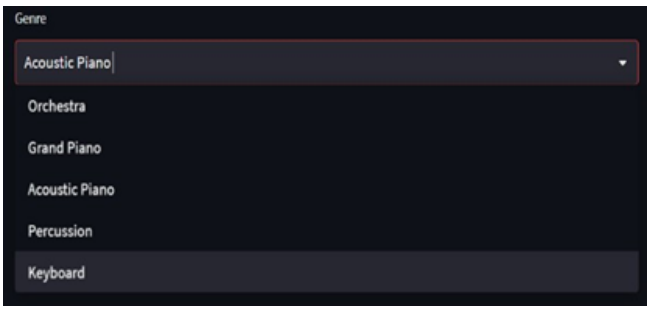

Fig 3-7 shows the various output

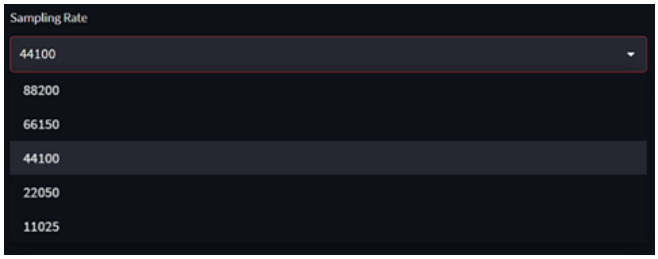

Fig 3: Different Genres available to the User

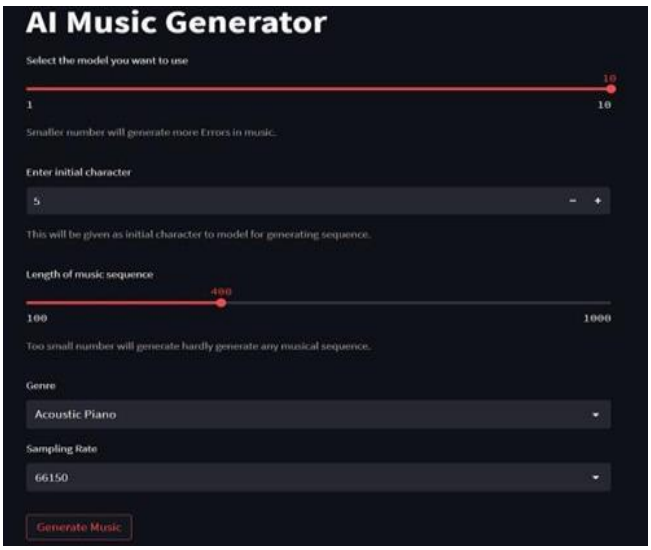

Fig 4: Different sampling rates available to the user

Power System Technology
ISSN:1000-3673
Received: 06-11-2024 Revised: 15-12-2024 Accepted: 05-01-2025

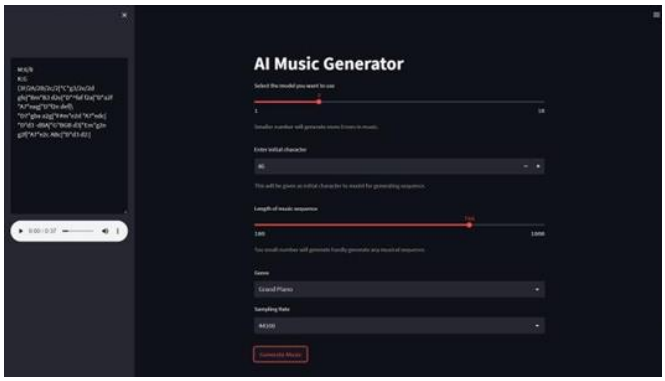

Fig.5. Music Generator Input given by the User

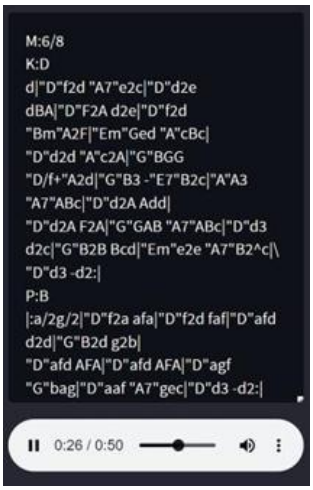

Fig 6: Playable Audio generated by Music Generator along with ABC Notations.

Fig.7. ABC Music and playable Audio generated by our model for musicians and composers to explore new musical ideas and foster creative expression. By simplifying the music generation process and offering accessible output formats, This system or tool democratizes music composition and invites users of all technical backgrounds to engage in creative exploration. Future work could focus on expanding the dataset, refining the audio conversion process, and further improving the model's music generation capabilities.

Future work

- Expanding the Dataset: Increase the size and diversity of the training dataset by incorporating music from different genres and styles, enabling the model to learn from a broader range of musical patterns.

- Integrating User Preferences: Allow users to provide input or preferences during the music generation process, such as specifying chord progressions, melodies, or desired moods, to enable more personalized and tailored compositions.

Power System Technology
ISSN:1000-3673
Received: 06-11-2024 Revised: 15-12-2024 Accepted: 05-01-2025

- Integration with Music Production Software: Integrate the system with popular music production software or digital audio workstations (DAWs), allowing seamless integration of the generated music with other elements of a composition or production workflow.
- Exploring Lyrics Generation: Extend the system's capabilities to include lyrics generation alongside music generation, enabling the creation of complete songs with both melodies and lyrics.
- Collaborative Music Generation: Implement collaborative features that enable multiple users to contribute to a music composition in real-time, fostering collaborative creativity and enabling remote collaboration among musicians and composers.

The output generated is tested in all possible edge cases and the Music Generator works as intended for all possible test and use cases provided to it. The music created is tested with SHAZAM an existing app to identify the music.

VI. CONCLUSION

In conclusion, The Music Generation Tool created using Deep Learning works successfully in developing a system for generating music in ABC notation using a character based RNN. The integration of MIDI and WAV conversion capabilities further enhances the system's functionality and usability. By leveraging the sequential nature of music and training on a diverse dataset, This system produces coherent and melodic compositions.

This system offers several advantages compared to existing solutions, including human-readable notation, flexible output formats, genre-specific sound, and a rich training dataset. The use of ABC notation enables easy collaboration and sharing of compositions, while the conversion to MIDI and WAV formats allows for audio playback on various platforms and devices.

The model's architecture, implemented with Keras, ensures efficient training and captures intricate musical patterns. The inclusion of genre-specific SoundFont files enhances the audio output, creating a more immersive and realistic experience.

Overall, The Music generation tool provides a valuable tool

REFERENCES

[1] Jose D. Fern´andez and Francisco Vico. AI methods in algorithmic composition: A comprehensive survey. Journal of Artificial Intelligence Research, 2013.

[2] Briot, Jean-Pierre, Gae¨tan Hadjeres, and Franc¸ois Pachet. "Deep learning techniques for music generation-a survey." arXiv preprint arXiv:1709.01620 (2017).

[3] Hao-Wen Dong, Wen-Yi Hsiao, Li-Chia Yang, and YiHsuan Yang. MuseGAN: Multi-track sequential generative adversarial networks for symbolic music generation and

Power System Technology
ISSN:1000-3673
Received: 06-11-2024 Revised: 15-12-2024 Accepted: 05-01-2025

accompaniment. In Proc. AAAI, 2018.

[4] Eck D, Schmidhuber J. A first look at music composition using lstm recurrent neural networks[J]. Istituto Dalle Molle Di Studi Sull Intelli- genza Artificiale, 2002, 103.

[5] J. D.D., "Generating polyphonic music using tied parallel networks," Correia J., Ciesielski V., Liapis A. (eds) Computational Intelligence in Music, Sound, Art and Design. EvoMUSART 2017., 2017. [Online]. Available: https://link.springer.com/chapter/10.1007/ 978-3- 319- 55750-2 9.