OPEN ACCESS
VOL. 7, NO. 1, PP. 107-117, NOVEMBER 2024

ISSN : 2622-8106 (ONLINE)
DOI: 10.20895/inista.v7i1.1628

Journal of Informatics, Information System, Software Engineering and Applications (INISTA)

The Application of LSTM in the AI-Based Enhancement of Classical Compositions

Dzikri Rahadian Fudholi *1, Delfia Nur Anrianti Putri 2, Raden Bagus Muhammad AdryanPutra Adhy Wijaya3, Jonathan Edmund Kusnadi4, Jovinca Claudia Amarissa5

*1*Department of Computer Science and Electronics, Universitas Gadjah Mada*
2,3,4,5*Undergraduate Program in Computer Science, Universitas Gadjah Mada*

Gedung C, Lantai 4, Sekip Utara, Bulaksumur, Sendowo, Sinduadi, Kec. Mlati, Kabupaten Sleman, Daerah Istimewa Yogyakarta

1 dzikri.r.f@ugm.ac.id
2delfia.nur2004@mail.ugm.ac.id
3raden.bag2003@mail.ugm.ac.id
4jonathan.edmund1202@mail.ugm.ac.id
5jovinca.claudia.amarissa@mail.ugm.ac.id

Received on 15-10-2024, revised on 20-11-2024, accepted on 29-11-2024

Abstract

Music enhancement through deep learning methodologies presents an innovative approach to refining and augmenting classical compositions. This study leverages a comprehensive dataset of classical piano MIDI files to employ LSTM networks with attention mechanisms for music refinement. The model, trained on diverse compositions, demonstrates proficiency in capturing tempo nuances and integrating melodies seamlessly. However, challenges persist, particularly in accurately replicating varied pitch patterns and innovating bass lines, which slightly detracts from the generated compositions' originality and overall impact. Assessments conducted with 28 individuals revealed a positive reception, with melody integration scoring 8 out of 10, while bass line cohesion received comparatively lower ratings. These findings highlight the LSTM model's capability to generate harmonious melodies and emphasize areas for improvement. Future work could explore more advanced architectures, such as Transformers or hybrid models, to enhance pitch prediction and originality. This study contributes to advancing automated music refinement and provides a foundation for further developments in deep learning-based music generation techniques.

Keywords: Music Generation, Deep Learning, Music Enhancement, Notes, Duration

Corresponding Author:
*Dzikri Rahadian Fuhdoli
Departmenet of Computer Science and Electronics, Universitas Gadjah Mada
Gedung C, Lantai 4, Sekip Utara, Bulaksumur, Sendowo, Sinduadi, Kec. Mlati, Kabupaten Sleman, Daerah Istimewa Yogyakarta
Email: dzikri.r.f@ugm.ac.id

I. INTRODUCTION

The recent advancements in artificial intelligence (AI) and machine learning (ML) technologies have significantly impacted various domains. Those include healthcare [1], marine biology [2], the food industry [3], remote sensing [4], and music generation. These technologies have redefined how problems are approached, enabling innovative solutions to complex challenges. In music, AI has introduced transformative possibilities, shifting music generation from an exclusively human endeavour to one that leverages computational systems [5]. Automated systems for music composition utilize AI's ability to process and analyze vast datasets, offering a novel avenue for creating diverse and innovative musical pieces [6]. All of those domains are significantly affected by artificial intelligence.

AI-based music generation has made impressive advancements, but it still grapples with notable challenges. Current systems often fall short in replicating the emotional depth, creativity, and cultural significance of human-composed music. These shortcomings are especially evident in classical music—a genre celebrated for its complexity, intricate harmonies, and rich structural patterns [7]. The primary challenge lies in balancing the preservation of classical music's aesthetic essence with the introduction of innovation, a task requiring both technical precision and artistic sensitivity.

This challenge becomes even more nuanced when the goal is not to create entirely new compositions but to enhance existing ones. Enhancing classical music demands systems that can understand and retain the core elements of an original piece while enriching it with stylistic nuances and harmonic depth. Despite advancements, many AI-based music generation methods lack the refinement needed for this task. Traditional rule-based systems are too rigid to adapt to the stylistic diversity of classical music. At the same time, some deep learning models, although technically sound, often fail to deliver cohesive or emotionally resonant compositions.

Deep learning, a subset of machine learning, offers promising potential for overcoming these challenges. Its ability to learn complex patterns from large datasets makes it particularly well-suited to music generation. By analyzing and replicating intricate musical structures, styles, and sequences, deep learning models can produce music that blends technical accuracy with emotional richness [8], [9]. Among these models, Long Short-Term Memory (LSTM) networks excel due to their capability to handle sequential data, a fundamental characteristic of music. LSTMs effectively capture long-term dependencies in sequences, making them ideal for understanding and replicating the progression and flow of musical compositions [10], [11].

Based on the insights gained from these studies, our project seeks to address the limitations and leverage the strengths of both LSTM-based and attention mechanism-driven architectures for sound and music generation. While LSTMs are effective at capturing sequential dependencies, they can encounter issues such as vanishing gradients and challenges with handling long-range dependencies, especially in complex datasets like audio signals. On the other hand, attention mechanisms, as employed in Transformer-based models, excel at identifying relevant parts of input sequences but often face challenges with computational efficiency and the resource-intensive nature of training on high-dimensional audio data.

The main challenge addressed in this research is overcoming the limitations of AI-based systems in enhancing classical music compositions. While AI has made strides in music generation, existing methods often fail to balance the preservation of classical music's aesthetic and structural essence with the introduction of stylistic innovation. Classical music is particularly challenging due to its complexity, intricate harmonies, and rich patterns, requiring a system that can maintain its emotional and cultural significance while introducing creative enhancements.

The primary purpose of this research is to develop a novel approach to enhancing existing classical music compositions by leveraging a combination of Long Short-Term Memory (LSTM) networks and attention mechanisms. Instead of creating entirely new pieces, the goal is to refine and enrich existing works by integrating stylistic nuances and harmonic depth from various composers. This approach seeks to produce harmonically richer, stylistically diverse compositions that retain the original piece's essence while addressing technical and artistic limitations in current AI systems. The research also aims to contribute to the broader role of AI in preserving and augmenting cultural heritage through advanced technology.

II. LITERATURE REVIEW

Before the project execution, we reviewed papers that target sound generation, focusing on two main approaches: Generative Adversarial Networks (GANs) and Recurrent Neural Networks (RNNs). In "Comparing Representations for Audio Synthesis Using Generative Adversarial Networks" [12], the authors aimed to assess the impact of different audio signal representations on quality, training, and generation times. Utilizing a Progressive Growing GAN (P-GAN) architecture with a subset of NSynth data, they found that complex and mag-if audio representations exhibited the best-perceived quality in audio synthesis, closely matching state-of-the-art results in adversarial audio synthesis.

A comprehensive review titled "Generative Adversarial Networks for Speech Processing: A review" [13] provided a high-level overview of the latest work in speech GANs. The review highlights interesting future directions, summarizes evaluation metrics, and discusses resources required for fruitful research in speech processing.

Addressing audio-to-image synthesis is done through research titled Audio to Scene Image Synthesis Using Generative Adversarial Network" [14]. Their goal was to create an audio-to-image generator based on GAN, incorporating advanced techniques of conditional GANs. The model demonstrated improved image generation quality compared to naive conditional GAN approaches.

In the realm of music generation, [10] explored "LSTM Based Music Generation" using the Wikifonia dataset. Leveraging Google Magenta's models and introducing encoding variants, including Neighbours Distribution and Consecutive Pitch Difference, their LSTM-based architecture produced melodies comparable or superior in perceptual quality to baseline Magenta models.

A model developed for novices and musicians to create melodies without deep musical knowledge [15]. Using Recurrent Neural Networks (RNN) and a dataset of Christmas tunes in ABC notation, their model achieved a remarkable 99% accuracy in music generation and could predict and generate music in MIDI format [15].

Application of transformers and GANs for music generation [16], employing two piano datasets where the authors provided sentiment annotations. Evaluators, drawn from the researchers' professional networks, assessed the generated music based on four qualitative metrics: Human likeness, Originality, Structure, and Overall Quality. The findings indicate that the proposed GAN models perform competitively with state-of-the-art approaches across all evaluated metrics.

In music regeneration and restoration, [11] proposed an LSTM model trained on music transcriptions written in textual notation (ABC). The model, with three hidden layers, successfully learned the pattern of the track and demonstrated effective execution and performance. These studies collectively contribute to the understanding and advancement of sound and music generation techniques, showcasing the diversity of approaches and applications within the domain.

III. METHODOLOGY

A. *Proposed Solution*

The primary aim of this research is to refine and augment existing musical compositions using deep learning. A multi-step approach has been adopted to conduct this research, as shown in Figure 1. We begin with collecting a broad spectrum of classical music compositions. Subsequently, audio preprocessing techniques were applied, focusing on feature engineering to extract musical elements like notes and duration from the audio data. This process enabled the creation of a structured dataset representing the musical compositions in a machine-readable format, essential for further analysis and modelling. In this research, we train a Long Short Term Memory (LSTM) model with an attention mechanism. Additionally, transfer learning techniques are employed, leveraging the knowledge gained by the model during training to enhance particular compositions. Finally, the output of the model undergoes thorough evaluation to assess the quality of the refinements made and the effectiveness of the augmentation process, thereby contributing to the advancement of automated music refinement through deep learning methodologies.

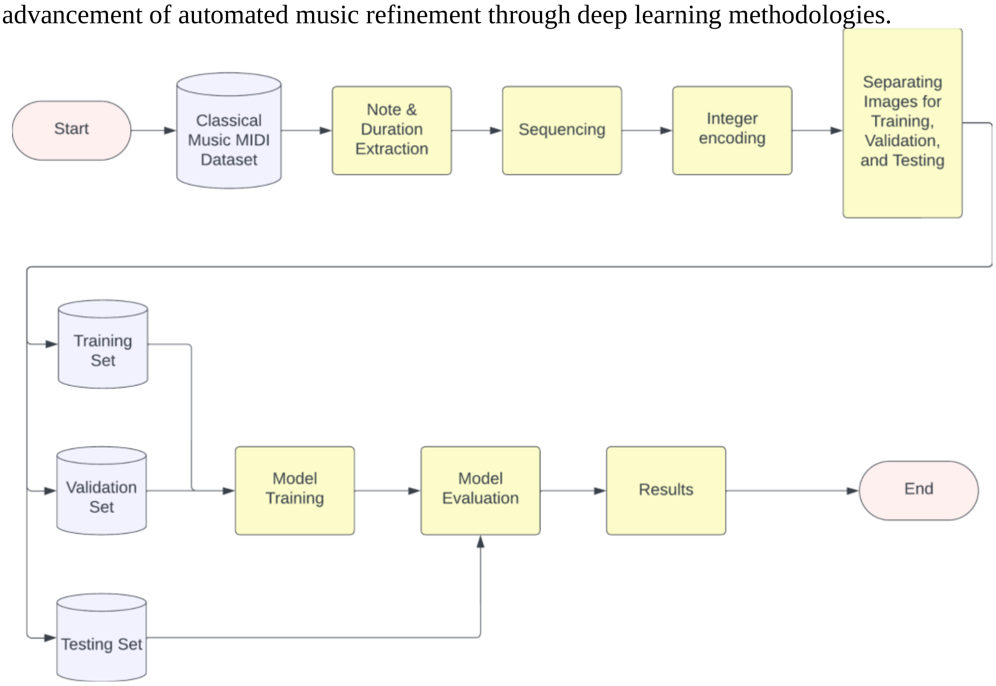

Fig. 1. Flowchart of the Research Method Conducted.

B. Dataset

The dataset utilized in this research is the "Classical Music MIDI" collection. The data collection was sourced from http://www.piano-midi.de, as shown in Figure 2. This dataset comprises classical piano MIDI files featuring compositions from 19 well-known composers. The MIDI files were scraped from the website, providing a diverse and comprehensive selection of classical piano pieces. The website's HTML introduction section, authored by Bernd Krueger, outlines technical information about the MIDI sequences, emphasizing the step-sequencing process used for musical interpretation. The author's background and equipment are briefly discussed, including a PC with a Soundblaster Live Card and a Roland HP 330e digital piano. The dataset is valuable for its richness in classical piano compositions and detailed attributes embedded in the MIDI format, making it suitable for various analytical purposes. Bernd Krueger's active contribution to the MIDI community and the technical information provided add depth to the dataset's context and reliability for further analysis.

In this research, our dataset comprises 100 classical piano compositions sampled from the original dataset. These compositions span the works of five renowned composers: Schubert (25 songs), Beethoven (3 songs), Chopin (48 songs), Mozart (21 songs), and Bach (3 songs). The MIDI format of these pieces offers a rich and varied collection, enabling comprehensive analysis and modelling for our study.

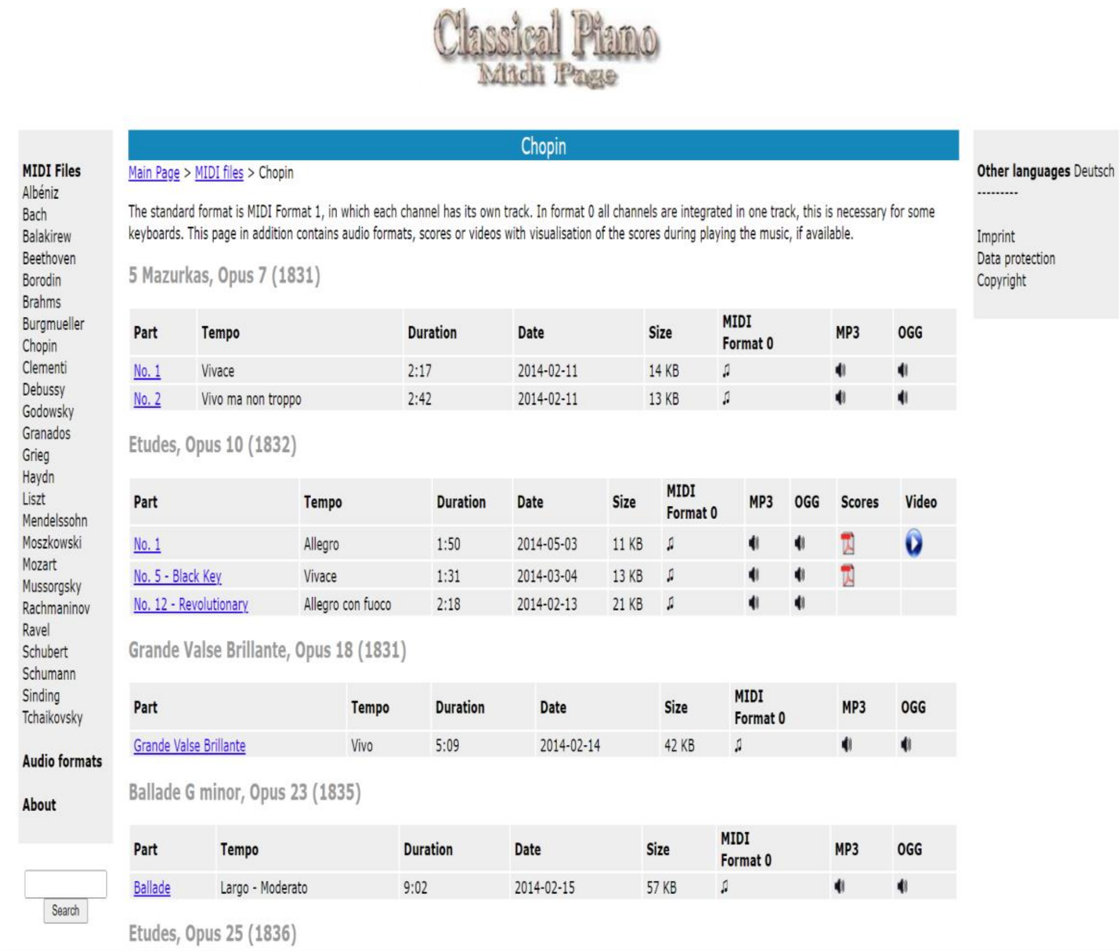

Fig. 2. Dataset Sources from Piano Midi

C. Data Preprocessing

In the first step, preprocessing is done on the dataset by extracting the pitch and duration contained in the processed song. Each pitch and duration of the song is represented in the form of integer embedding, a vector that the RNN-LSTM model can accept. Before processing the model, a sequencing process is performed on pitch and duration and slicing to determine features and labels. In this process, the length of the feature sequence is 32 notes/duration, and the label is 1 note/duration after it. The extracted notes and durations are then encoded into integer representations. After applying data preprocessing, the total data used in model training is 23948 sequences of notes with 32 elements per sequence length.

D. Modelling

The proposed neural network for music generation utilizes a sequence-to-sequence approach, incorporating Long Short-Term Memory (LSTM) units and attention mechanisms. The model receives two inputs: musical notes and durations, as shown in Figure 3, where the two inputs are in sequence. Each input sequence is embedded into continuous representations using embedding layers. As shown in Figure 4, the embedded sequences are concatenated from 100 layers each to become 200 layers. After that, the concatenated layer is fed into the first LSTM layer, which processes the combined information and generates sequential outputs.

The attention mechanism enhances the model's ability to selectively focus on specific elements within the input sequences (musical notes and durations). A second LSTM layer is employed with a focus on the attention mechanism, even though both LSTM has the same configuration, as shown in Figure 3. The attention mechanism calculates attention scores using a Dense layer. Reshape and normalization process done afterwards ensures that the scores are transformed into attention weights (alpha) that range between 0 and 1 and sum up to 1 across the sequence. To align the attention weights (alpha) with the sequence length, they are repeated and permuted to match the LSTM output sequence.

The first LSTM output is multiplied element-wise with the repeated and permuted attention weights to accentuate elements according to their importance, generating a weighted sequence representation. Finally, a context vector is derived by summing the weighted elements along the sequence axis with a lambda layer. This vector encapsulates crucial information from the input sequences, emphasizing elements deemed significant by the attention mechanism. The final output is generated through two different Dense layers, producing two outputs: predicted musical notes and durations with softmax activation functions, as shown in the bottom part of both Figure 3 and Figure 4.

This model is trained using a categorical cross-entropy loss function for pitch and duration outputs. RMSprop optimizer with a learning rate of 0.001 is utilized. Callbacks for model checkpointing and early stopping are employed to monitor the model's performance and save the best weights during training. The training involves iterating through the dataset for a specified number of epochs, with a batch size of 32, aiming to minimize the loss function. A portion of the dataset is reserved for validation to assess the model's generalization.

During the inference phase, the model receives a preprocessed piece of music with parameters such as temperature, number of extra notes, and sequence length setting. The process begins by initializing note and duration sequences with placeholders, ensuring a consistent sequence length for modeling. Subsequently, the algorithm traverses a loop, predicting subsequent notes and durations iteratively based on the learned patterns from the trained model. Each iteration generates new notes and durations using a softmax function to select the most probable next note and duration. Incorporating a temperature parameter introduces controlled randomness, influencing the diversity of predictions made by the model.

The model generates a new note for each prediction iteration by assessing the predicted probabilities across various MIDI pitches. This newly generated note is appended to the overall predictions. Simultaneously, the model determines the predicted note and duration based on their probability distributions. Ultimately, this phase yields the generated sequence, encompassing notes and their corresponding durations. This outcome showcases the model's predictive capacity based on the input piece and the specified parameters, offering insights into its musical generation capabilities.

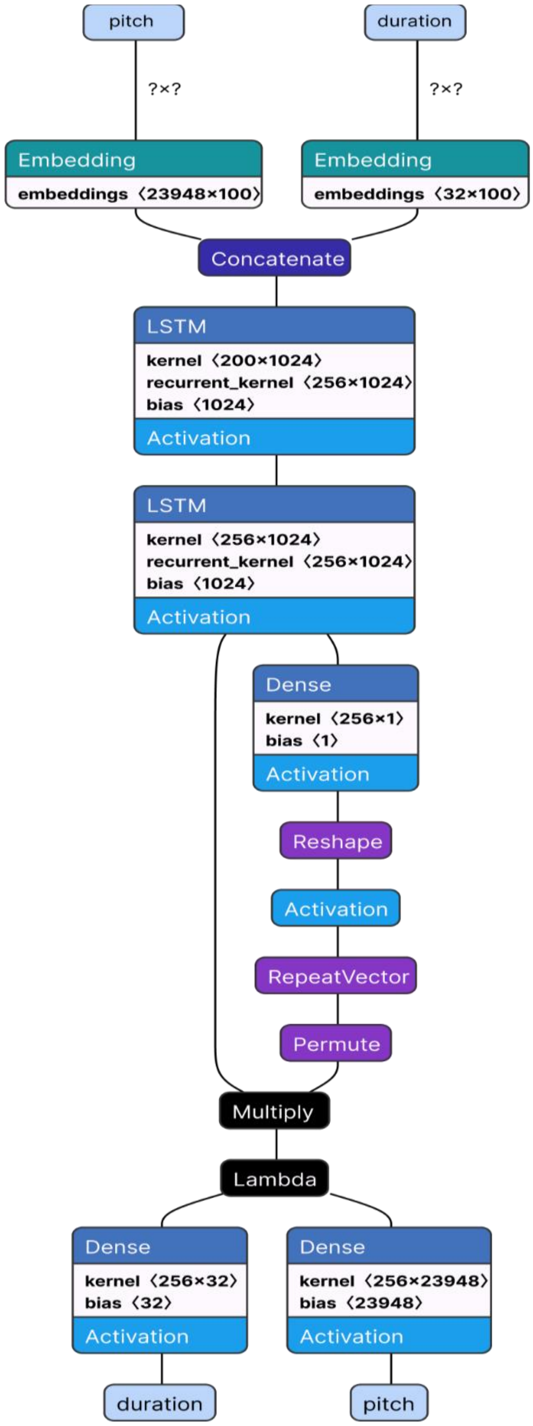

Fig. 3. LSTM Architecture for Music Enhancement.

```
Model: "model_4"
__________________________________________________________________________________________________
 Layer (type)                   Output Shape         Param #     Connected to
==================================================================================================
 input_5 (InputLayer)           [(None, None)]       0           []

 input_6 (InputLayer)           [(None, None)]       0           []

 embedding_4 (Embedding)        (None, None, 100)    2394800     ['input_5[0][0]']

 embedding_5 (Embedding)        (None, None, 100)    3200        ['input_6[0][0]']

 concatenate_2 (Concatenate)    (None, None, 200)    0           ['embedding_4[0][0]',
                                                                  'embedding_5[0][0]']

 lstm_4 (LSTM)                  (None, None, 256)    467968      ['concatenate_2[0][0]']

 lstm_5 (LSTM)                  (None, None, 256)    525312      ['lstm_4[0][0]']

 dense_2 (Dense)                (None, None, 1)      257         ['lstm_5[0][0]']

 reshape_2 (Reshape)            (None, None)         0           ['dense_2[0][0]']

 activation_2 (Activation)      (None, None)         0           ['reshape_2[0][0]']

 repeat_vector_2 (RepeatVector)  (None, 256, None)   0           ['activation_2[0][0]']

 permute_2 (Permute)            (None, None, 256)    0           ['repeat_vector_2[0][0]']

 multiply_2 (Multiply)          (None, None, 256)    0           ['lstm_5[0][0]',
                                                                  'permute_2[0][0]']

 lambda_2 (Lambda)              (None, 256)          0           ['multiply_2[0][0]']

 pitch (Dense)                  (None, 23948)        6154636     ['lambda_2[0][0]']

 duration (Dense)               (None, 32)           8224        ['lambda_2[0][0]']

==================================================================================================
Total params: 9,554,397
Trainable params: 9,554,397
Non-trainable params: 0
__________________________________________________________________________________________________
```

Fig. 4. The parameter of the LSTM Model.

E. *Evaluation*

The model that has been developed will be evaluated using loss metric calculations, accuracy metrics, precision, recall, F1 score metrics, and subjective evaluation.

1. *Loss Metrics:*

Loss metrics are crucial indicators when evaluating the performance of machine learning models for music generation. Pitch and duration losses provide insights into the model's learning dynamics. These losses can be calculated using two formulas. The first is the softmax Equation (1) where σ is softmax, $\dot{Z}$ is the input vector, e^{zi} is the standard exponential function for the input vector, e^{zj} is the standard exponential function for the output vector, and K is the number of classes in the multi-class classifier.

$$\sigma(\vec{z})_i = \frac{e^{z_i}}{\sum_{j=1}^{K} e^{z_j}} \tag{1}$$

The second is categorical cross entropy in Equation (2) where CE is Cross Entropy

$$CE = -log(Softmax\ Function) \tag{2}$$

2. *Accuracy Metrics*

   Accuracy metrics are central in providing a general measure of correctness in the model's predictions. Formulas for accuracy metrics, such as overall accuracy, pitch accuracy, and duration accuracy, involve dividing the number of correct predictions by the total number of instances. The accuracy formula used in this model is defined in Equation (3), where T is True, F is False, P is Positive, and N is Negative.

$$Accuracy = \frac{TP + TN}{TP + FP + TN + FN} \tag{3}$$

3. *Subjective Evaluation*

   Subjective evaluation, incorporating user surveys or ratings, introduces a human-centric perspective. While not expressed in a specific formula, this qualitative feedback complements the quantitative metrics, capturing aesthetic qualities and stylistic fidelity.

IV. RESULT AND DISCUSSION

The LSTM with attention mechanism model demonstrated marked improvements in pitch and duration performance metrics, as evidenced by the training results. Although the pitch accuracy remained relatively low at 5%, the addition of an attention mechanism yielded a substantial enhancement in duration prediction, achieving a final duration loss of 0.29 (down from 0.77) and an accuracy of 91.42%, up from an initial 74.97%. It can be said that the model can better learn the tempo played in the song rather than the diverse pitch patterns of each composer. This issue may stem from the limitations of the attention mechanism, which prioritizes only specific pitch variations, potentially leading to overfitting. Consequently, the model becomes less capable of accurately predicting other pitch variations, as the attention mechanism's focus on certain features inhibits a balanced representation across the broader range of pitch variations.

To address these limitations in future iterations, several key modifications are planned. First, we can enhance the attention mechanism by implementing a multi-head attention approach, which can improve the model's ability to capture a broader range of pitch variations simultaneously. By distributing attention across multiple heads, the model can prioritize diverse pitch patterns without overfitting to a subset, thus offering a more balanced prediction. Second, regularization techniques such as dropout and weight decay will be introduced to mitigate overfitting. Dropout layers within the attention module can reduce the risk of the model becoming overly dependent on specific patterns, while weight decay can constrain model complexity, encouraging it to generalize better across various pitch variations.

When comparing this model to a baseline LSTM without an attention mechanism, the LSTM with attention showed superior performance in capturing temporal dependencies within the duration data. Results indicated that the standard LSTM reached only 83.45% duration accuracy and 3% pitch accuracy. Standard LSTMs tend to struggle with balancing focus on specific time steps, which can hinder their ability to predict complex pitch and duration variations accurately. In contrast, the attention-enhanced model overcomes this limitation by dynamically adjusting its focus to essential segments, which reduces loss in duration predictions more effectively.

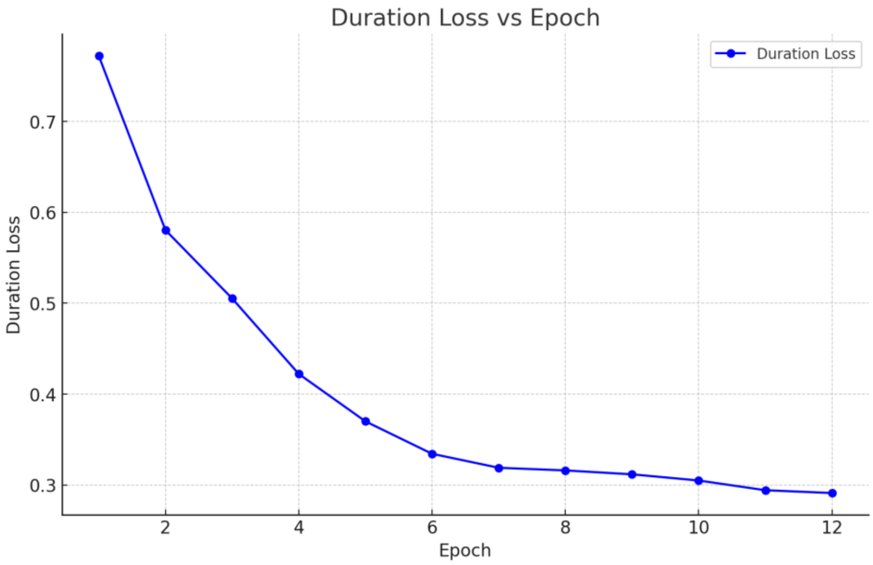

Fig. 5. The line plot duration loss per epoch

After the training, the model went through fine-tuning for 4 different pieces of music. The music pieces assessed include Beethoven's "Moonlight Sonata" (both Adagio and Presto Agitato), J.S. Bach's "Prelude," and Chopin's "Waltz in E flat major.". The assessment of music pieces generated by the LSTM model, conducted by 28 individuals with diverse backgrounds, offers insightful results. The assessors were asked to rate three aspects: the integration of piano melody with the overall piece, the originality and freshness of the piano bass notes in classical music, and the overall impression of the piano bass notes. The overall score is shown in Figure 6 and Figure 7.

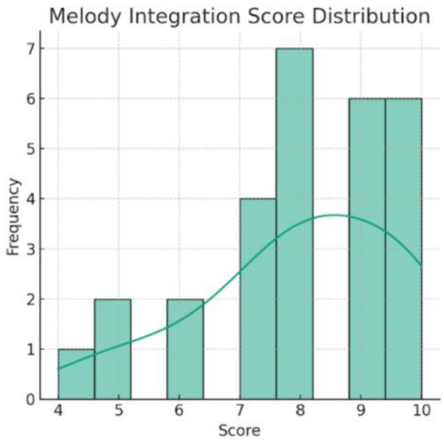

Fig. 6. The distribution of the Melody Integration Score

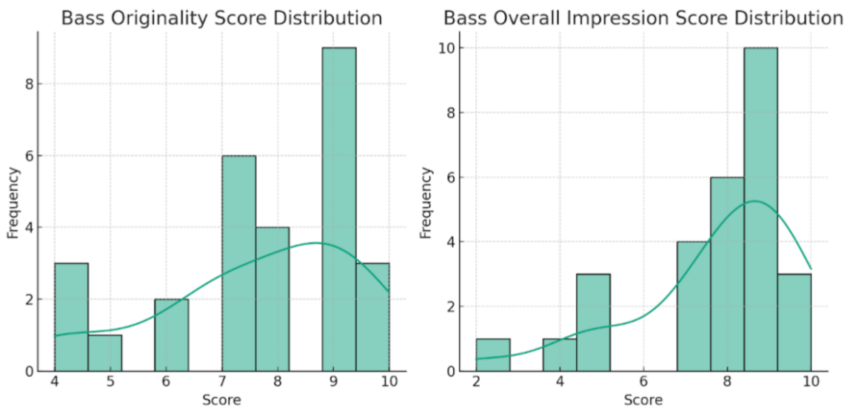

Fig. 7. The distribution of the Bass Originality Score

The average scores for melody integration, bass originality, and overall bass impression were 8.00, 7.64, and 7.75 out of 10, respectively. This indicates a generally positive reception of the LSTM-generated music. The melody integration aspect scored particularly well, with a mean score of 8, suggesting that the LSTM model effectively created harmonious and cohesive piano melodies that blend well with the overall composition. The slightly lower scores for bass originality and overall bass impression, with means of 7.64 and 7.75, may imply less innovation or impact in the bass lines than the melody.

In conclusion, the assessment reveals a favourable reception of LSTM-generated music, particularly in melody integration. The slightly lower scores for bass originality and overall impression suggest areas for improvement or refinement in the LSTM model's approach to generating bass lines in classical compositions. These findings could guide further development and refinement of LSTM models in music generation, particularly in enhancing the originality and impact of bass lines while maintaining the successful integration of melodies. The accompanying visualizations, displaying the distribution of scores across the three assessed areas, further illustrate these insights.

V. CONCLUSION

This research introduces innovation in automatic music generation by modifying existing songs through the LSTM-attention model. The model training results reveal a pitch loss of 9.29 in the last epoch, highlighting the model's enhanced adaptation to learning tempo rather than diverse pitch patterns.

Furthermore, qualitative assessments conducted with 28 individuals indicate a positive reception to the LSTM-generated music, particularly in melody integration, with an average score of 8. This underscores the model's effectiveness in creating harmonious and cohesive piano melodies that seamlessly blend with the overall composition. However, slightly lower scores for bass originality (7.64) and overall bass impression (7.75) suggest areas that require refinement. In the training process, adjustments are also necessary to address pitch variations among composers, such as balancing the dataset or focusing the model on a specific composer for targeted learning.

In summary, while this research has successfully achieved its objectives, there are notable areas for improvement. For instance, enhancing the model's adaptability to a broader range of musical genres could contribute to its versatility. Additionally, incorporating a more nuanced understanding of subtle variations in pitch patterns specific to different composers may further refine the model's generative capabilities. These considerations highlight the potential for future research and development, offering opportunities to elevate the performance of the LSTM model in music generation.

REFERENCES

[1] D. A. Tyas and T. Ratnaningsih, “Analisis Segmentasi Sel Darah Merah berbasis Mask-RCNN,” *Jurnal Informasi, Sistem Informasi, dan Rekayasa Perangkat Lunak (INISTA)*, vol. 5, no. 1, pp. 1–7, Nov. 2022, doi: 10.20895/inista.v5i1.766.

[2] D. A. Wibisono and A. Buhanuddin, "Classification of Marine Fish Based on Image Using Convolutional Neural Network Algorithm and VGG16," *Jurnal Informasi, Sistem Informasi, dan Rekayasa Perangkat Lunak (INISTA)*, vol. 6, no. 2, pp. 116–123, Jun. 2024, doi: 10.20895/inista.v6i2.1466.

[3] P. D. N. A. Wijaya R. B. M. A. A. and D. R. Fudholi, "Smart GreenGrocer: Automatic Vegetable Type Classification Using the CNN Algorithm," *Indonesian Journal of Computing and Cybernetics Systems (IJCCS)*, vol. 17, no. 3, Jul. 2023, doi: 10.22146/ijccs.82377.

[4] W. R. B. M. A. A. N. Putri D. N. A. and Wahyono, "U-Net Based Vegetation Segmentation for Urban Green Space (UGS) Regulation in Yogyakarta City," in *2023 Eighth International Conference on Informatics and Computing (ICIC)*, Dec. 2023, pp. 1–6. doi: 10.1109/ICIC60109.2023.10381938.

[5] S. Malloch and C. Trevarthen, "The Human Nature of Music," *Front Psychol*, vol. 9, Oct. 2018, doi: 10.3389/fpsyg.2018.01680.

[6] V. Velardo and M. Vallati, "Automated Planning and Music Composition: an Efficient Approach for Generating Musical Structures," in *Proceedings of the 11th International Symposium on Computer Music Multidisciplinary Research (CMMR 2015)*, Plymouth, UK: Springer, 2015, pp. 778–790.

[7] K. J. Rolla V. and L. Velho, "The complexity of classical music networks," *Europhys Lett*, vol. 121, no. 3, p. 38005, Apr. 2018, doi: 10.1209/0295-5075/121/38005.

[8] K. S. S. Y. V. G. P. Yadav P. S. and R. S. Singh, "A Lightweight Deep Learning-Based Approach for Jazz Music Generation in MIDI Format," *Comput Intell Neurosci*, vol. 2022, p. 2140895, Aug. 2022, doi: 10.1155/2022/2140895.

[9] Y. C. G. S. Z. G. Possik J. and A. D'Ambrogio, "A Model Based Systems Engineering Approach to Automated Music Arrangement," in *2021 Annual Modeling and Simulation Conference (ANNSIM)*, Jul. 2021, pp. 1–12. doi: 10.23919/ANNSIM52504.2021.9552105.

[10] N. T. Shah F. and N. Vyas, "LSTM Based Music Generation," in *2019 International Conference on Machine Learning and Data Engineering (iCMLDE)*, Dec. 2019, pp. 48–53. doi: 10.1109/iCMLDE49015.2019.00020.

[11] T. V Mohanty M. N. and R. K. Pattanaik, "Music Regeneration with RNN Architecture Using LSTM," in *2023 International Conference in Advances in Power, Signal, and Information Technology (APSIT)*, Jun. 2023, pp. 538–542. doi: 10.1109/APSIT58554.2023.10201708.

[12] L. S. Nistal J. and G. Richard, "Comparing Representations for Audio Synthesis Using Generative Adversarial Networks," in *2020 28th European Signal Processing Conference (EUSIPCO)*, Jan. 2021, pp. 161–165. doi: 10.23919/Eusipco47968.2020.9287799.

[13] A. et al. Wali, "Generative adversarial networks for speech processing: A review," *Comput Speech Lang*, vol. 72, p. 101308, Mar. 2022, doi: 10.1016/j.csl.2021.101308.

[14] C. S.-P. Wan C.-H. and H.-Y. Lee, "Towards Audio to Scene Image Synthesis Using Generative Adversarial Network," in *ICASSP 2019 - 2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, May 2019, pp. 496–500. doi: 10.1109/ICASSP.2019.8682383.

[15] D. S. Sajad S. and M. Meleet, "Music Generation for Novices Using Recurrent Neural Network (RNN)," in *2021 International Conference on Innovative Computing, Intelligent Communication and Smart Electrical Systems (ICSES)*, Sep. 2021, pp. 1–6. doi: 10.1109/ICSES52305.2021.9633906.

[16] F. J. Neves P. L. T. and J. B. Florindo, "Generating Music with Sentiment Using Transformer-GANs," in *Proceedings of the 23rd ISMIR Conference*, Dec. 2022, pp. 717–725. doi: 10.5281/zenodo.7342704.