ISSN 2442-6571

Automatic note generator for Javanese gamelan music accompaniment using deep learning

Arik Kurniawati [a,1], Eko Mulyanto Yuniarno [a,b,2,*], Yoyon Kusnendar Suprapto [a,b,3] ,
Aditya Nur Ikhsan Soewidiatmaka [c,4]

[a] Department of Electrical Engineering, Institut Teknologi Sepuluh Nopember, Surabaya, Indonesia
[b] Department of Computer Engineering, Institut Teknologi Sepuluh Nopember, Surabaya, Indonesia
[c] Soewidiatmaka Gamelan, Bandung, Indonesia
[1] arikkurniawati.19071@mhs.its.ac.id; [2] ekomulyanto@ee.its.ac.id; [3] yoyonsuprapto@ee.its.ac.id; [4] soewidiatmaka@gmail.com
* corresponding author

ARTICLE INFO

Article history
Received February 27, 2023
Revised April 24, 2023
Accepted May 21, 2023
Available online June 1, 2023

Keywords
Bidirectional LSTM
Deep Learning
Gamelan music
Javanese melody
Musical harmonic

ABSTRACT

Javanese gamelan is a traditional form of music from Indonesia with a variety of styles and patterns. One of these patterns is the harmony music of the Bonang Barung and Bonang Penerus instruments. When playing gamelan, the resulting patterns can vary based on the music's rhythm or dynamics, which can be challenging for novice players unfamiliar with the gamelan rules and notation system, which only provides melodic notes. Unlike in modern music, where harmony notes are often the same for all instruments, harmony music in Javanese gamelan is vital in establishing the character of a song. With technological advancements, musical composition can be generated automatically without human participation, which has become a trend in music generation research. This study proposes a method to generate musical accompaniment notes for harmony music using a bidirectional long-term memory (BiLSTM) network and compares it with recurrent neural network (RNN) and long-term memory (LSTM) models that use numerical notation to represent musical data, making it easier to learn the variations of harmony music in Javanese gamelan. This method replaces the gamelan composer in completing the notation for all the instruments in a song. To evaluate the generated harmonic music, note distance, dynamic time warping (DTW), and cross-correlation techniques were used to measure the distance between the system-generated results and the gamelan composer's creations. In addition, audio features were extracted and used to visualize the audio. The experimental results show that all models produced better accuracy results when using all features of the song, reaching a value of around 90%, compared to using only 2 features (rhythm and note of melody), which reached 65-70%. Furthermore, the BiLSTM model produced musical harmonies that were more similar to the original music (+93%) than those generated by the LSTM (+92%) and RNN (+90%). This study can be applied to performing Javanese gamelan music.

1. Introduction

Melody and harmony are important concepts in music. Melody refers to the tone and rhythm [1], while harmony is the use of chords or tones together [2]. They work together to create a great musical composition. Melody is usually noticed first when listening to music, followed by harmony, which completes the tune. Harmony music is used in both Western and Eastern music, including Javanese gamelan, a traditional music from Indonesia that uses harmony music as its accompaniment.

https://doi.org/10.26555/ijain.v9i2.1031

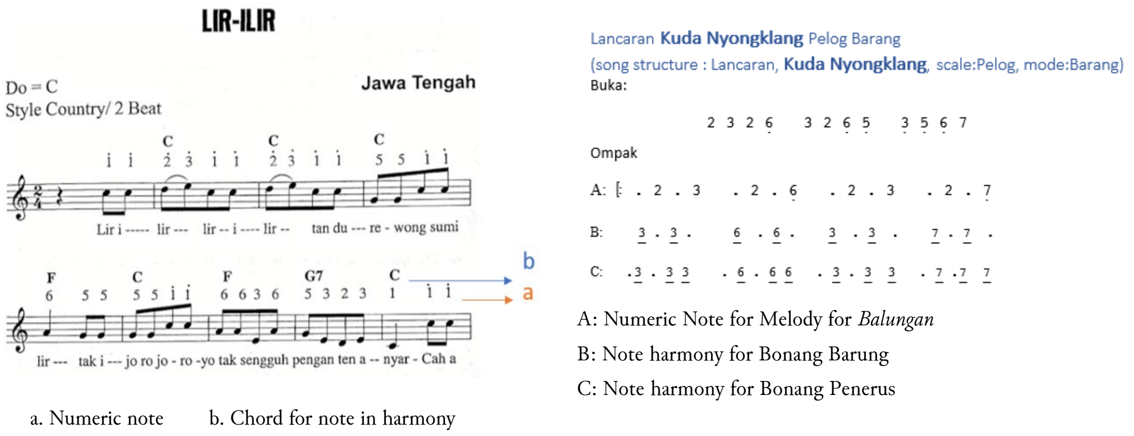
Each musical instrument has a distinct role to perform in the creation of a harmonic song. In Western music, for example, the piano, harmonica, saxophone, and violin create the melody, while the guitar, keyboard, piano, and harp give harmony, and the tambourines and drums form the rhythm. Every instrument is designed to perform a specific function based on the notes, chords, and rhythms contained within the piece. Moreover, the chord notation for all instruments is at the same pitch and is not influenced by the rhythm.

However, Javanese gamelan is unique in its harmony music, where the harmonic tones have different variations based on the rhythm, structure, and dynamics of the song [3]. Each instrument has different tones, with the Bonang instrument (consisting of Bonang Barung and Bonang Penerus) playing harmony music. In contrast, the Balungan group (Saron, Demung, and Peking) plays the melody, and the Gong and Kendhang groups are responsible for organizing the rhythmic patterns.

Song in Javanese gamelan comprises a variety of elements including rhythm, *laya*, scale (*laras*), mode (*pathet*), and dynamics. Rhythm is determined by tempo gradation, while *laya* pertains to the speed of the song. *Laras* refers to the scales used in the song, and *pathet* conveys the emotional feeling conveyed by the song. Dynamics highlight the variation, proportionality, and dynamic nature of the musical elements present in a song [4], [5]. Its purpose is not only to entertain but also to convey social, moral, cultural, and spiritual values [6]. The composition is conveyed through a pattern of playing variations, and if the resulting pattern is not appropriate, the beauty and character of the composition will be lost. In Javanese gamelan music, harmony is achieved through variations of the pattern played, which are tailored to the specific characteristics of the song, thereby showcasing the unique character of each piece. Harmony in Javanese gamelan music is unique. A single song can have several variations of harmony, all based on the specific characteristics of the piece. These variations in musical harmony are usually influenced by including rhythm, *laya*, scale (*laras*), mode (*pathet*), and dynamics of song [4], [5]. Fig. 1 shows an example of a song with the same melody but different harmony notes.

Fig. 1. Javanese *gamelan* music with the same melody but different note in harmony using (a) *Gembyang* pattern with *Irama Lancar* (b) *Imbal Sekaran* pattern with *Irama Lancar* (c) *Mipil* pattern with *Irama Tanggung*

In Indonesia, modern music uses numeric notation for the melody and letters for chords. This system makes it easier to write and read music, and is used in both classical and modern music, including Javanese gamelan. In modern music, all instruments typically use the same sheet music. However, in traditional Javanese gamelan music, each instrument has its own notation. Fig. 2 illustrates the differences in modern and traditional music notation in Indonesia, where in modern music chord notation is usually the same for all instruments, whereas in Javanese gamelan, the harmony notation can be different for each instrument.



(a) (b)

Fig. 2. An Indonesian music notation (a) modern music (b) traditional music in Javanese *gamelan*

The notation used in Javanese gamelan music is less detailed than shown in Fig. 2b, where only the main melody is written. This means that only experienced gamelan musicians can play all the instruments by reading the information in the song, such as the song structure, scale (*laras*), mode (*pathet*), and dynamics of the music, without knowing the detailed notation for all the instruments. The problem is that novice musicians need complete notation for all instruments to play gamelan music, because the harmony music may have different notation patterns. For example, the song Lancaran Angleng Pelog Barang.

- This song will use Gembyang pattern for harmony music if the dynamic of the song is for music only, and this song will use Imbal Sekaran pattern if the dynamic of the song is for singer accompaniment.
- This song will use Gembyang pattern for harmony music if the rhythm of the song uses Irama Lancar, and this song will use Mipil pattern if this song uses Irama Tanggung.

The selection of the musical harmony pattern is based on the characteristics of a song (such as: rhythm, *laya*, scale (*laras*), mode (*pathet*), and dynamics and the melodic notation of the Balungan instrument). Fig. 3 demonstrates how Javanese gamelan songs are written. Gamelan music uses numeric notes, with the main melody represented by Balungan notes. Fig. 3 provides an example of gamelan music notation and other information contained in songs [3]–[5], [7]–[10], including:

- The song Manyar Sewu is an example of Javanese gamelan music. Its structure is Lancaran, its scale is Slendro, and its mode is Manyura, all of which are included in the song title.
- Javanese gamelan music has various song structures, including Sampak, Srepegan, Ayak-Ayakan, Lancaran, Bubaran, Ketawang, and Ladrang. Each structure is determined by the position of the Kethuk, Kenong, Kempul, Kempyang, and Gong playing, similar to a genre in modern music.
- There are two musical scales in Javanese gamelan: Slendro, which has five notes (1, 2, 3, 5, and 6), and Pelog, which has seven notes (1, 2, 3, 4, 5, 6, and 7). The musical scale is called *laras*.

- *Pathet* is a system of musical modes, and the Slendro is composed of Manyura, Nem, and Sanga, while the Pelog comprises the notes Barang, Lima, and Nem. The *pathet* helps create the atmosphere of the song.
- Buka is the opening of the song, which is represented by the Bonang Barung note.
- Rhythm (*irama*) is usually found at the start of the section.
- Ompak is a song section represented by the Balungan note. Each part is usually organized into multiple lines called *gongan*. Depending on the song structure, there may be multiple *gatra* in a single line, and each *gatra* consists of several notes from the Balungan based on rhythm.
- In gamelan music, the Balungan notes are represented by symbols, such as a circle for Gong Ageng, smiley and frowny faces for Gong Suwuk, a smiley face for Kempul, and a frowny face for Kenong. A dot above a number and the lower octave by a dot below a number indicates the upper octave. Additionally, a dot in place of a number indicates a break or interruption.

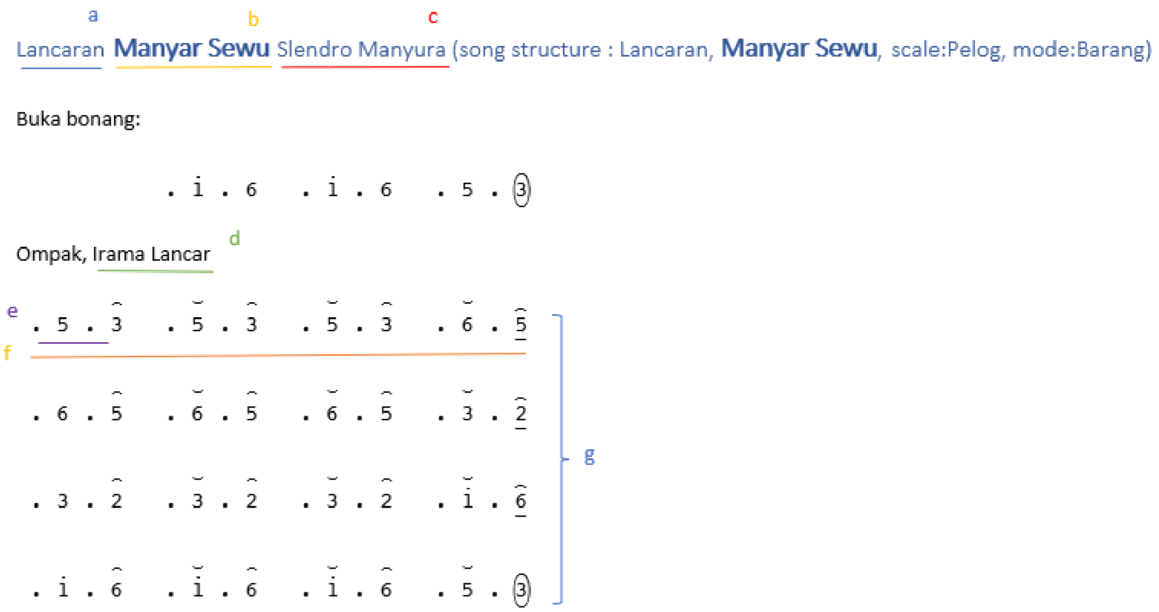


Fig. 3. A song of Javanese *gamelan* (a) song structure (b) name (c) scale and mode of the song (d) rhythm (e) *gatra* (f) *gongan* (g) melody note

All the characteristics of a Javanese gamelan song have a significant impact on the resulting variations of harmony music. Even with the same melody, differences in rhythm can result in different harmony music, as demonstrated in Fig. 1a and Fig. 1c. Likewise, variations in song dynamics can lead to different harmonies, as shown in Fig. 1b for a singer's accompaniment and Fig. 1a for music alone. Therefore, the song's characteristics that heavily influence the harmony of Javanese gamelan music are song structures, rhythm, scale (*laras*), mode (*pathet*), and song dynamics.

Based on the previous explanation, the problems discussed in this study are:

- Javanese gamelan music has various harmony note variations depending on the characteristics of the song, such as song structure, rhythm, scale (*laras*), mode (*pathet*), and dynamic of the song.
- Notation for all instruments in a Javanese gamelan song is not fully written, only the melody notation, which becomes a difficulty for novice gamelan musicians.

This study proposes a deep learning model for composing musical accompaniment as note harmony in Javanese gamelan using detailed information about a song. This study aims to assist novice gamelan musicians in learning the harmony music in Javanese gamelan. The main contributions of this paper are presented as follows:

- Proposing a simplified method for representing music data as input, utilizing numeric notes.
- Proposing a deep learning model to generate musical accompaniment as note harmony using detailed information about a song.
- Using song characteristic for features such as song structure, *gatra*, and rhythm, this study successfully generates music accompaniment for music harmony, including Bonang Barung and Bonang Penerus.

The structure of this study is as follows: Section 1 introduces the context and significance of the research, and Section 2 discusses earlier studies that have been done on Music Generation, as well as the details of BiLSTM, LSTM, and RNN. Additionally, it includes a discussion of the dataset used for this research and methods for creating an accompaniment Javanese gamelan music generator of the Bonang Barung and Bonang Penerus. Section 3 presents the outcomes of the experiments conducted and their analysis. Finally, Section 4 concludes by summarizing the findings of this study and outlining future work.

2. Method

2.1. Related Work

Deep Learning has rapidly expanded in recent years, solving many complex Artificial Intelligence issues. Deep Learning models are effective at solving various problems, including recognition [11], regression [12], semi-supervised and unsupervised problems [13] for medical diagnosis [14], natural language [15] and image processing [16], and prediction system [17]. These models learn hierarchical features from different data types, including numerical, image, text, and audio. There are different types of Deep Learning neural networks, such as Artificial Neural Networks (ANN), Convolutional Neural Networks (CNN), and Recurrent Neural Networks (RNN).

This study suggests using Deep Learning to create music with a type of RNN. Typically, only professional composers can create music, but this study aims to help beginners create their own. This is called music generation, and it uses AI to create its own music. This is done by combining AI and art, producing music without human interference.

Research on music generation is conducted to generate melodies [18], polyphonic [19], multitrack [20], and predict musical accompaniment to a given melody [21]. Most of the generators are polyphonic [19]–[23]. Data is typically separated into audio and symbols due to the different forms in which music is presented [22]–[24]. Meanwhile, audio data is used more often, and symbols are also important because melodies are read through notation [22]. Artificial neural networks (ANNs), such as RNNs and their variations, are currently used in music generation [4], [24]–[31]. ANNs can predict musical notes based on previously learned musical structures without the help of a human expert to impart musical knowledge and analysis. RNNs and their variations are particularly useful in music generation because they can remember the musical notes created in the past and use that information to predict future notes.

Recurrent neural networks (RNNs) are unique in their ability to remember past data and use it to predict future data. Therefore, many researchers have used RNNs to create automatic music. For instance, RNNs have been used to predict melodic pitch and generate new music in MIDI format [24], as well as to generate sheet music with a defined structure and style, without imposing the rules of musical composition on the model using ABC Notation [25].

LSTM networks are a type of RNN that can solve long-term dependency problems. Unlike conventional RNNs, which can struggle to link previous and new information, LSTMs store information in memory for an extended period. Although RNNs and LSTMs have similar structures, LSTMs have different memory or repeating modules. Researchers have used LSTM networks for music generation, including creating melodies and harmonies from a MIDI file using a single-layer LSTM [26]. Multiple-layer LSTMs have also been used to generate musical notes, preserving a song's characteristics, and resulting in better musical variance [27]. In another study, the dataset was preprocessed before entering the LSTM, and the results were reconstructed into sheet music in ABC format, producing pleasant-sounding music [28].

Bidirectional Long Short-Term Memory (BiLSTM) is a variant of LSTM which can process sequence information in two directions, past to future and future to past. This allows the network to preserve both past and future knowledge and generate more relevant output, making it useful in music generation tasks as well as NLP tasks. Researchers have studied the use of BiLSTM with Dynamic Time Wrapping to identify relative content and measure similarity between two temporal sequences [29], generating accurate next music with epoch variations [30], creating note sequences for classical piano [31], and generating unique music with attention [4].

The BiLSTM is a neural network that is an advanced version of the LSTM architecture and belongs to the RNN family [32]. It is capable of better contextual understanding due to its ability to process input sequences in both forward and backward directions. This allows the network to take into account contextual information from both past and future inputs, leading to more precise predictions. Furthermore, the complexity of the BiLSTM architecture makes it less prone to overfitting compared to LSTM and RNN.

This study proposes a BiLSTM model for generating music accompaniment as music harmony from numerical notes and song characteristic features, such as rhythm, melody, scale (*laras*), and mode (*pathet*) in Javanese gamelan. The proposed model aims to overcome the limitations of previous works [4], [29]–[31], which only focused on predicting the next melody to generate new notes.

2.2. Bidirectional LSTM (BiLSTM)

Composers often choose notes based on their context, both before and after, which cannot be easily handled by simple RNNs and LSTMs. To overcome this limitation, a Bidirectional LSTM combines two independent RNNs and takes data from both ends of the input flow to model sequential dependencies in both directions, as illustrated in Fig. 7. In addition, BiLSTM adds another LSTM layer that flows the input sequence backward. The outputs from the two LSTM layers can be combined in various ways, such as averaging, summing, multiplying, or concatenating. BiLSTM is particularly useful in music generation since it can preserve both future and previous knowledge, resulting in more relevant output.

The LSTM network uses activation values other than Candidate values (C). This LSTM network has two outputs, namely the current value of the memory cell ($C^{<t>}$), and the output value or hidden state ($a^{<t>}$), which is defined according to equations (1) and (2).

$$C^{<t>} = \Gamma_u * C^{N<t>} + \Gamma_f * C^{<t-1>} \tag{1}$$

$$a^{<t>} = \Gamma_o * C^{<t>} \tag{2}$$

Thus, the result of calculating the potential value of the memory cell is defined in equation (3).

$$C^{N<t>} = \tanh\,(W_c[a^{<t-1>}, x^{<t>}] + b_c) \tag{3}$$

There are separate gate equivalents (3), (4), (5) in the LSTM to control memory cells, i.e.:

$$\Gamma_u = \sigma(W_u[a^{<t-1>}, x^{<t>}] + b_u) \tag{4}$$

$$\Gamma_f = \sigma(W_f[a^{<t-1>}, x^{<t>}] + b_f) \tag{5}$$

$$\Gamma_o = \sigma(W_o[a^{<t-1>}, x^{<t>}] + b_o) \tag{6}$$

Where W and b are weight matrices and vectors, t is the current iteration of the recurrent network Γ_u, the update gate, Γ_f the forget gate, Γ_o the output gate. Therefore, if all of these components are combined, the LSTM cell is shown in Fig.7.

2.3. Dataset

The data musical notes for this study used only Lancaran song structure, which was collected from http://www.gamelanbvg.com/. The Lancaran is a short Javanese gamelan song with a variety of harmonic patterns, often used in traditional arts such as shadow puppets and traditional dances. Since the dataset used in this study only includes the Lancaran, the song structure is not a considered feature, unless the dataset is expanded to include all types of song structures. The dataset includes notes for Balungan, rhythm, melodic notes at the end of lines and melodic notes at the end of the song, scale, and mode of music were obtained from Javanese gamelan song, and the dynamic of song was determined by gamelan expert. The note of Bonang Barung and Bonang Penerus were created by gamelan expert as the output.

The data used were ten songs as training and validation data, divided between training (80% of data) and validation (20% of data) and one song as test data. Each song had three different harmonic music notation structures: Gembyang pattern with an Irama Lancar, Imbal Sekaran pattern with an Irama Lancar, and Mipil pattern with an Irama Tanggung. Therefore, a total of 33 songs, divided into approximately 400 *gatra*, were used in this study, as illustrated in Fig. 4. While Table 1 lists the songs used in this study.

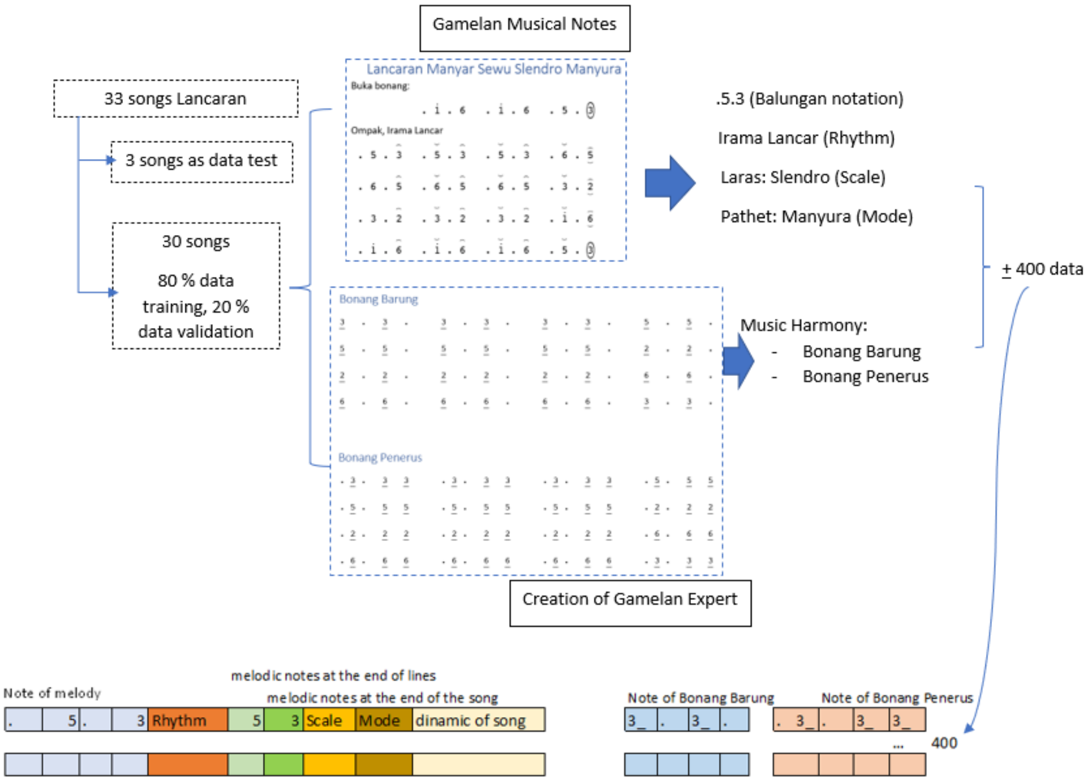
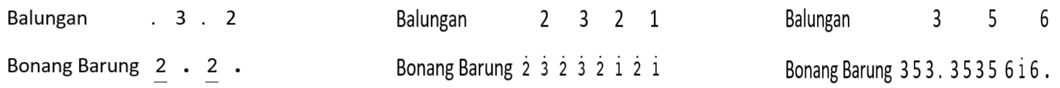


Fig. 4. Dataset Representation

Each song's melody notes, represented by the Balungan, are divided into multiple lines. Each line is composed of several *gatra*, with each *gatra* containing several notes. The number of bonang notations, including Bonang Barung and Bonang Penerus, depends on the rhythm used. For each Balungan note in each *gatra*, there is either one Bonang Barung note or 1-2 Bonang Penerus notes (Irama Lancar), two Bonang Barung notes or 2-4 Bonang Penerus notes (Irama Tanggung), or four Bonang Barung notes or 4-8 Bonang Penerus notes (Irama Dadi), as illustrated in Fig. 5.



(a) (b) (c)

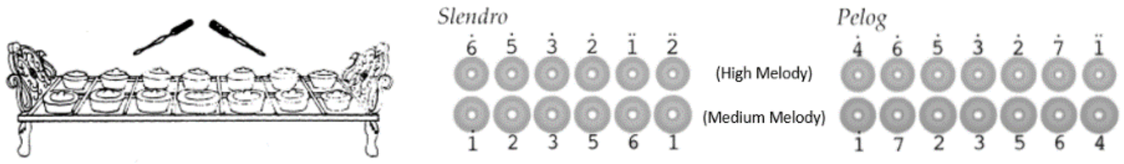
Fig. 5. Structure Notation for (a) *Irama Lancar* (b) *Irama Tanggung* (c) *Irama Dadi*

Table 1. List of songs for dataset

No	Song Titles *	Dataset		
		Training	*Validation*	*Test*
1	*Lancaran **Kuda Nyongklang** Pelog Barang*	Yes	Yes	No
2	*Lancaran **Maesa Kurda** Slendro Sanga*	Yes	Yes	No
3	*Lancaran **Manyar Sewu** Slendro Manyura*	Yes	Yes	No
4	*Lancaran **Rena Rena** Slendro Manyura*	Yes	Yes	No
5	*Lancaran **Sarung Jagung** Pelog Barang*	Yes	Yes	No
6	*Lancaran **Angleng Pelog** Barang*	Yes	Yes	No
7	*Lancaran **Bendrong** Pelog Nem*	Yes	Yes	No
8	*Lancaran **Kandhang Bubrah** Pelog Nem*	Yes	Yes	No
9	*Lancaran **Kedhu** Pelog Barang*	Yes	Yes	No
10	*Lancaran **Lasem Bubrah Pelog** Nem*	Yes	Yes	No
11	*Lancaran **Ricik Ricik** Slendro Manyura*	No	No	**Yes**

* Lancaran is song structure, Pelog/Slendro is scale, Barang/Sanga/Manyura/Nem is mode

Harmony music in Javanese gamelan is produced by the Bonang group, which includes Bonang Barung and Bonang Penerus. Both instruments have the same layout and design, differing only in their musical scale, Pelog or Slendro, as shown in Fig. 6. The Bonang Barung and Bonang Penerus have a medium melody in the bottom layout and a high melody in the top layout. They can be played using a single or double melody. Writing a note in Bonang Barung can be done with one note if played in a single melody or a double note with medium and high notes played simultaneously. For example, 3 represents playing Bonang 3 in the lower and upper octaves simultaneously.



(a) (b)

Fig. 6. (a) *Bonang* kettles (b) Layout of *Slendro* and *Pelog Bonang*

2.4. Prediction of Notes Harmony as Accompaniment Music

The proposed method in this study is the Bidirectional LSTM (BiLSTM) method for generating musical accompaniment to traditional Javanese gamelan music, as illustrated in Fig. 7. The input and output of the system are as follows.

- Input: notes for *Balungan*, rhythm, a note at the end of lines, a note at the end of the song, scale, mode, dinamic of song.
- Output: sequence of notes harmony for Bonang Barung and Bonang Penerus instruments

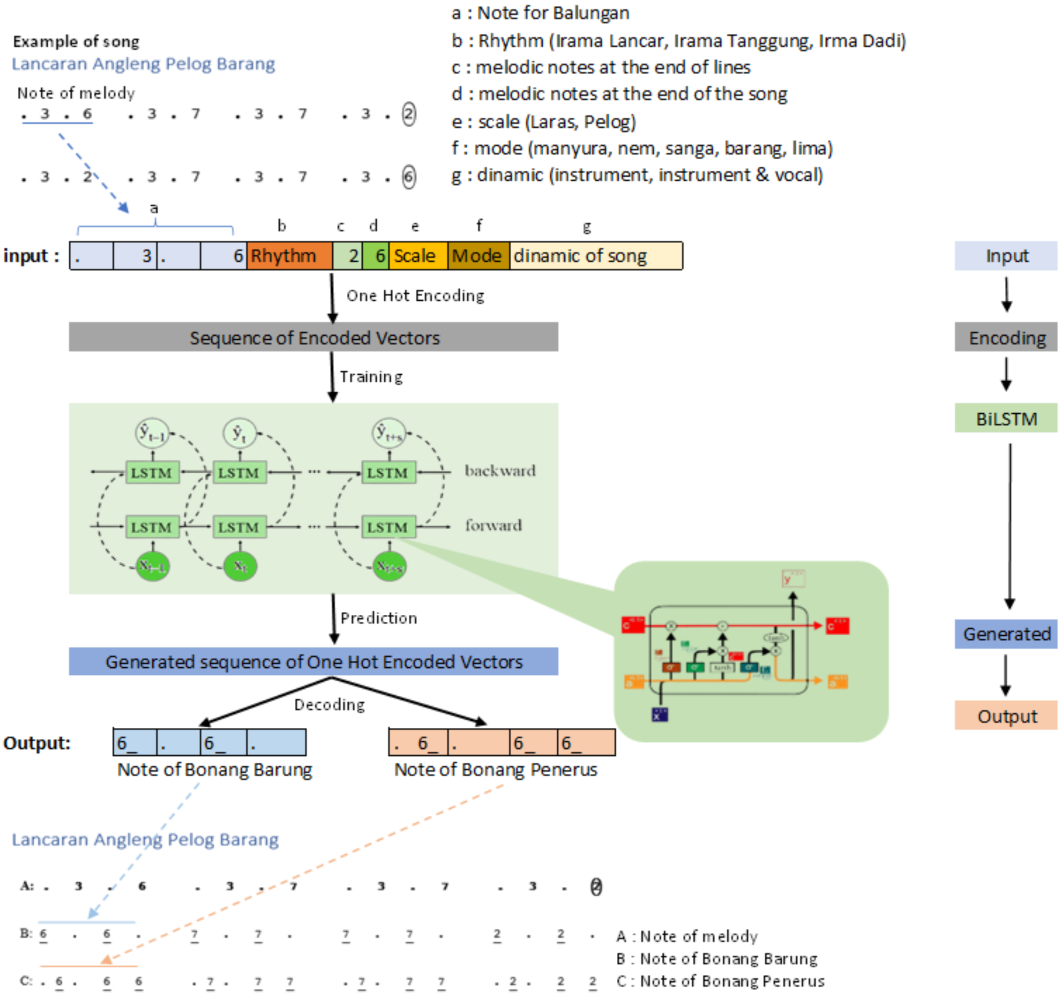
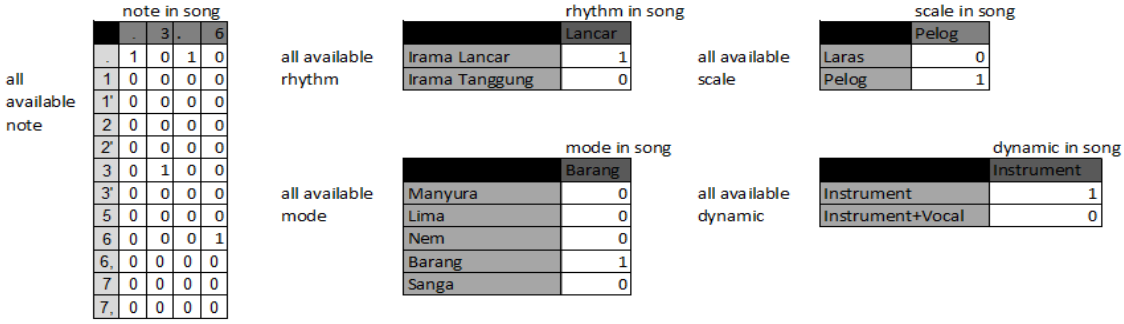


Fig. 7. Diagram of proposed method BiLSTM for prediction of notes harmony in Javanese *gamelan*

2.4.1. Data Encoding

Before using the input and output data in the BiLSTM, the input and output data need to be encoded into the correct input vector using a one-hot encoding mechanism. Each element of the input/output data is assigned a suitable true/false value. For example, if a particular note exists at a specific time step, the value of the column corresponding to that number "1" will be true, and all other settings will be false. Fig. 8 shows an illustration of a hot coding mechanism, where each column shows the vector code for a different note, rhythm, scale, mode and dynamic in the song for each data. The resulting vector is used as input to the BiLSTM, LSTM, and RNN models.



note in song (all available note)

	.	3	.	6
.	1	0	1	0
1	0	0	0	0
1'	0	0	0	0
2	0	0	0	0
2'	0	0	0	0
3	0	1	0	0
3'	0	0	0	0
5	0	0	0	0
6	0	0	0	1
6,	0	0	0	0
7	0	0	0	0
7,	0	0	0	0

rhythm in song (all available rhythm)

	Lancar
Irama Lancar	1
Irama Tanggung	0

scale in song (all available scale)

	Pelog
Laras	0
Pelog	1

mode in song (all available mode)

	Barang
Manyura	0
Lima	0
Nem	0
Barang	1
Sanga	0

dynamic in song (all available dynamic)

	Instrument
Instrument	1
Instrument+Vocal	0

Fig. 8. Encoding Technique (a)Note (b)Rhythm (c)Scale (d)Mode (e) Dinamic of song

2.5. BiLSTM Model

This section explains the structure of the BiLSTM model. Fig. 9a shows BiLSTM architecture, which includes an input layer of size 66, 256 LSTM cells, and 256 dropout cells to prevent overfitting. The output layer has 24 dense layers, with 8 for Bonang Barung and 16 for Bonang Penerus. The input includes Balungan note, rhythm, a note at the end of lines, a note at the end of the song, scale, mode, and dynamic of song. The input layer is the first layer of the model, followed by the BiLSTM (LSTM 128) and Dropout (0.2) layers. This is followed by a dense layer and a sigmoid layer as the activation layer for the output of the network. The sigmoid layer can predict accompanying notes in a piece of music. By training the model with an original dataset and testing it with new data, the model can generate accompanying music.

For comparison, this BiLSTM model is also compared with LSTM and RNN models. The structure of the LSTM and RNN architectures is shown in Fig. 9b and Fig. 9c, which are similar to the BiLSTM architecture. The results section discusses the performance of these models.

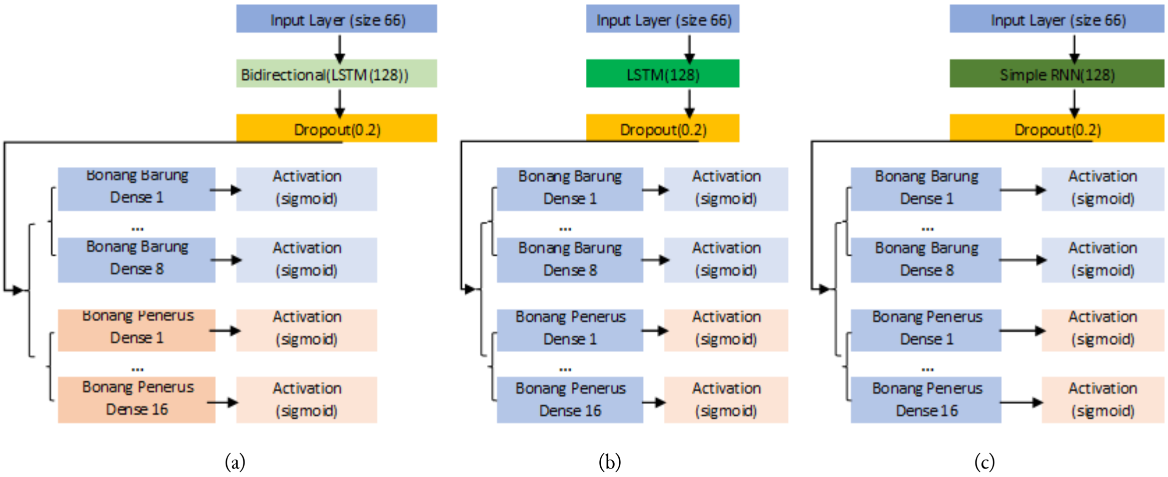


Fig. 9. Diagram architecture (a) BiLSTM (b) LSTM (c) RNN for prediction of notes harmony Javanese *gamelan*

2.6. Generated of sequences of note in harmony

After the BiLSTM network is trained, it can generate a sequence of musical notes for harmonic accompaniment. A dataset with different variations in the composition of musical notes was used to improve prediction and generate diverse sequences with three harmonic musical patterns. The model automatically predicts Bonang Barung and Bonang Penerus notes based on test data containing Balungan note, rhythm, a note at the end of lines, a note at the end of the song, scale, mode, and dynamic of the song.

3. Results and Discussion

In this study, we compared our proposed BiLSTM model with RNN and LSTM models to predict harmonious notes in Javanese gamelan music. To analyze its performance and the resulting harmonic notes, we evaluated the model's effectiveness using prediction accuracy and loss values. Then, we selected a sample song (Lancaran Ricik Ricik Slendro Manyura, not included in the training data) to compare with the generated and then evaluated the similarity using musical analysis.

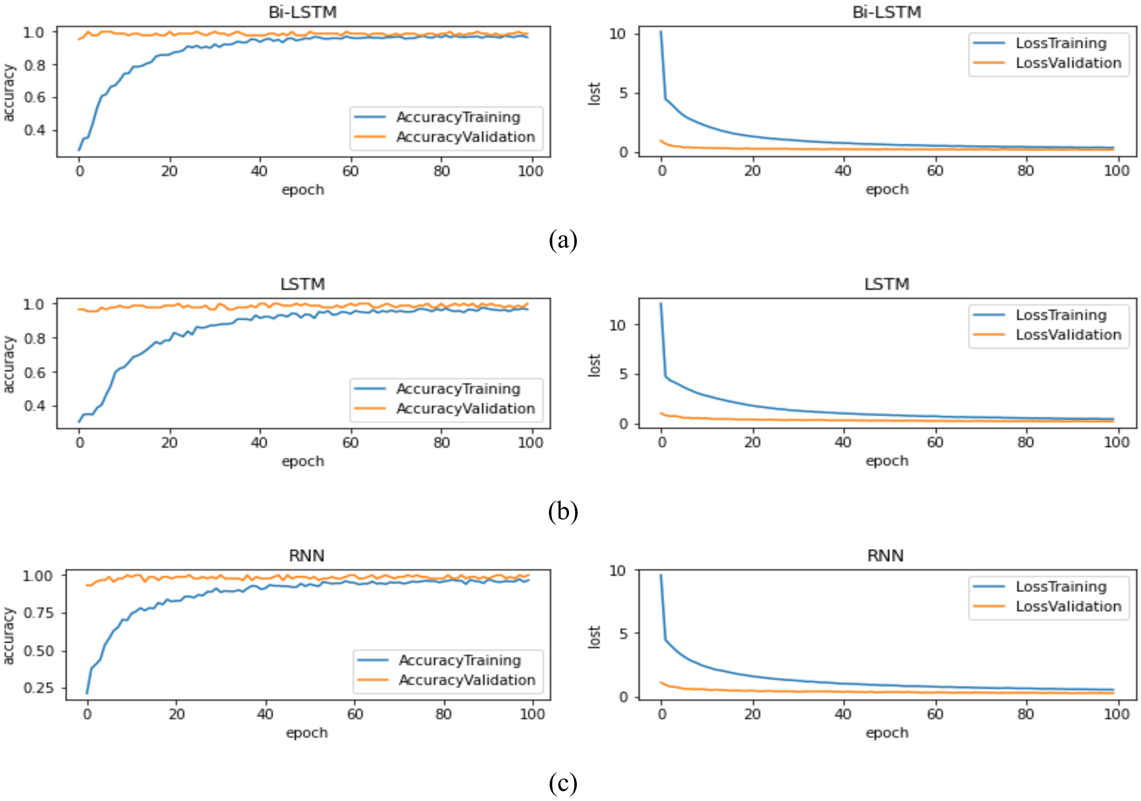
This study presents two experiment scenarios to evaluate the performance of the proposed model. In the first scenario, all features (including notes for Balungan, rhythm, a note at the end of lines, a note at the end of the song, scale, mode, and dynamic of the song) are used as input for the model. In the second scenario, only two features—the note of melody and rhythm—are used. In Javanese gamelan, melody and rhythm are typically the most important factors in determining the harmony of a song. However, sometimes other factors must be considered because different songs with the same melody and rhythm may produce different harmonic notes. This is due to various song characteristics, such as the types of modes and scales, the dynamic of songs (accompaniment of singers or just instruments), and the final note of each line and the song.

3.1. Quantitative Analysis

This section compares the accuracy and loss values of three models: RNN, LSTM, and BiLSTM in two scenarios (with all features and with only two features). Table 2 shows their loss values and prediction accuracies during the training phase. The dataset consisted of 400 data points and was trained with all three models. From Table 2 we can see that BiLSTM has the highest accuracy values compared to LSTM and RNN for all scenarios and all instruments, although the difference is small, around 0.01-0.05 for accuracy values, and the lowest loss value for all scenarios and all instruments, with a difference of around 0.01-0.1. These results were obtained after training the models for 100 epochs with a batch size of 4.

Table 2 also shows that using all song features improved the accuracy of all models. This shows that using all the features of a song can improve the success rate in generating musical accompaniment compared to using only 2 features (Balungan notes as melody notes and rhythm).

Fig. 10 shows the training and validation accuracy and loss values of the 400 data points trained using BiLSTM, LSTM and RNN. The three graphs show that BiLSTM, LSTM, and RNN have similar results, reaching stability at around epoch 40 in the training process.



(a)

(b)

(c)

Fig. 10. Graph accuracy and loss value of architecture (a) Bi LSTM (b) LSTM (c) RNN

3.2. Musical Analysis

Musical analysis is used in this study, including note distance, spectrogram [33], dynamic time warping (DTW) [34] and cross-correlation [35]. The aim is to provide additional insight, analyze the effects of musical harmony, and evaluate the similarity between the original music generated by the models, BiLSTM, LSTM, and RNN respectively. This multi-faceted approach ensures a more comprehensive analysis and a more accurate assessment of the similarity of the generated music to the original music.

Based on Table 2, the accompaniment music generator analyzed using note distance, spectrograms, dynamic time warping (DTW), and cross-correlation are scenarios using all features because the accuracy results are better when compared to those using 2 features.

3.2.1. Music Generation Result

After training the three BiLSTM, LSTM, and RNN models, the system can produce harmonic music in numerical notation for both Bonang Barung and Bonang Penerus notations, when all features are used as input. In this study, the Lancaran Ricik Ricik Slendro Manyura song was used as test data for three different harmonic music models: Gembyang pattern for Irama Lancar, Imbal Sekaran pattern for Irama Lancar, and Mipil pattern for Irama Tanggung. Fig. 11, Fig. 12, Fig.13 show the results of testing the Bonang Barung and Bonang Penerus notes for this song with three patterns.

- In the Gembyang pattern for the Lancaran Ricik Ricik Slendro Manyura song, the system generated identical harmonic music for the Bonang Barung and Bonang Penerus instruments as created by the gamelan composer (Pangrawit), as shown in Fig. 11.
- The Mipil pattern for the Lancaran Ricik Ricik Slendro Manyura song also produces harmony music of the Bonang Barung and Bonang Penerus instruments that are similar to those created by Pangrawit, as shown in Fig. 12.
- The Imbal Sekaran for the Lancaran Ricik Ricik Slendro Manyura song, the system produced harmony music for the Bonang Barung and Bonang Penerus instruments, shown in Fig. 13.

The Imbal Sekaran pattern is unique because it produces different notes despite having the same Balungan notes and rhythm as the Gembyang pattern. This is because the musical harmony notes are influenced by various song features such as the song's use, final note of each line, last note of the song, and its scale and mode. It can be seen in Fig. 13a and Fig.13b that there is a slight difference between the notation created by the gamelan composer and the result of the music generator produced by BiLSTM, LSTM, and RNN. However, according to the analysis of gamelan experts, this difference does not significantly affect the sound of the gamelan produced later, because the notation that is the Dhing Dhong gamelan (marked with a red box) is still the same as the gamelan composer's creation.

To further analyze the results of the system, the harmonic tones generated by the system are combined with the melodic tones of the Balungan instrument using the Gamelan Synthesis application. This evaluation is done because there is no difference between the instruments used in the Gembyangan and Mipil patterns created by the gamelan composer and the results of the generating system. Therefore, the Imbal Sekaran pattern for the Lancaran Ricik Ricik Slendro Manyura song is evaluated to assess the performance of the system's generator. This study visualizes audio features using Spectrogram and Amplitude vs. Frequency and calculates the correlation between two audio files using Note Distance, DTW, and Cross Correlation to evaluate similarities between the original music and the resulting results from BiLSTM, LSTM, and RNN.

Table 2. Accuracy and lost value of prediction for different models

Model	Loss and Accuracy	Scenario			
		All Features		*2 Features*	
		Bonang Barung	Bonang Penerus	Bonang Barung	Bonang Penerus
BiLSTM	Loss	0.042	0.034	0.120	0.110
	Accuracy	0.912	0.949	0.656	0.727
LSTM	Loss	0.041	0.033	0.112	0.106
	Accuracy	0.911	0.947	0.663	0.728
RNN	Loss	0.046	0.038	0.110	0.108
	Accuracy	0.906	0.943	0.665	0.732

3.2.2. Note Distances

Note distances are used for calculations to analyze the similarity of musical harmony between the original music and the music produced. This distance is called the exact distance, and its value contains a binary indicator.

$$N_o(Note_1, Note_2) = \begin{cases} 0 \; if \; Note_1 = Note_2 \\ 1 \; if \; Note_1 \neq Note_2 \end{cases} \quad (7)$$

As shown in Fig. 13, these values were obtained by calculating the difference between the original notation of the gamelan composer and the generator results of the BiLSTM, LSTM, and RNN models of both instruments, Bonang Barung and Bonang Penerus. Based on equation (7), the note distance value for the test data of the Gembyang and Mipil patterns is 0, as shown in the generator results in Fig. 11 and Fig. 12, where for all models BiLSTM, LSTM, and RNN are similar to the creations of gamelan composers.

In contrast, for the Imbal Sekaran pattern, the value of the note distance varies depending on the generator model. The Bi-LSTM model has a note distance value of 7, the LSTM model has a value of 9 and the RNN model has a value of 10. This means that the result generated by BiLSTM is closer to the original song than LSTM and RNN.

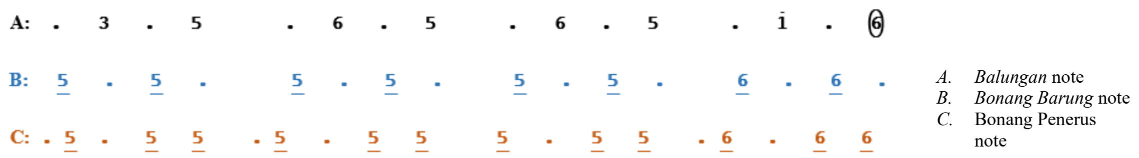


Fig. 11. Part of *Lancaran Ricik Ricik Slendro Manyura* song using *Irama Lancar* and *Gembyang* pattern

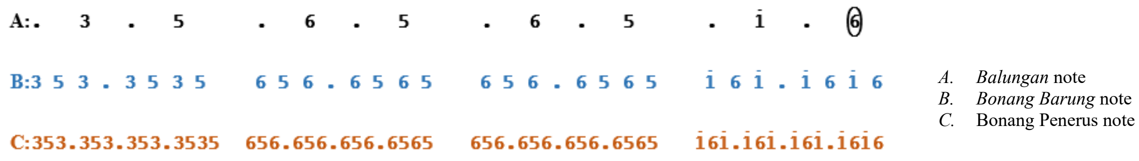


Fig. 12. Part of *Lancaran Ricik Ricik Slendro Manyura* song using *Irama Tanggung* and *Mipil* pattern

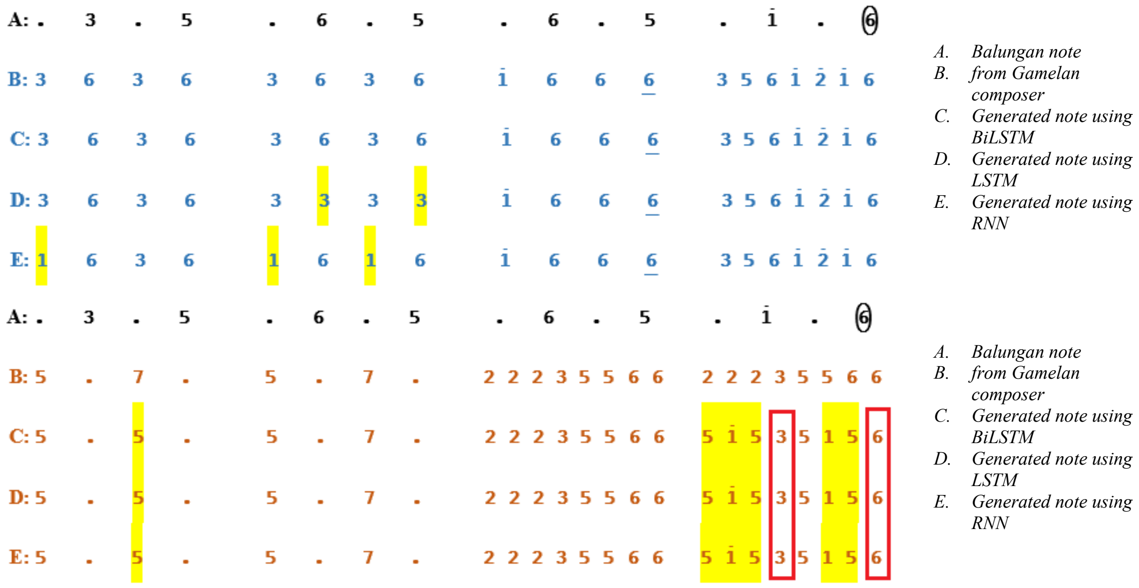
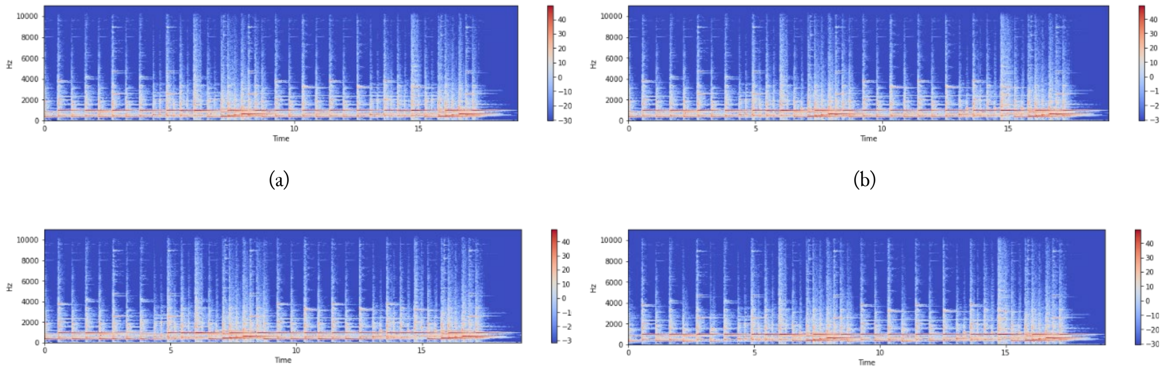


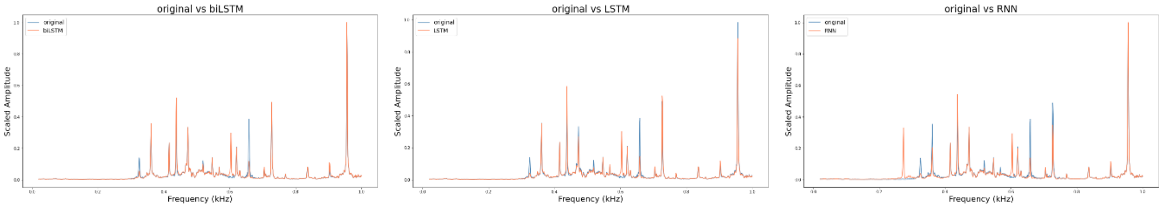
Fig. 13. *Part of Lancaran Ricik Ricik Slendro Manyura* using *Imbal Sekaran* pattern, the yellow sections on parts C, D, and E are the note generator system results that differ from the *gamelan* composer, the red box is the *dhing dhong gamelan* (a) note harmony for Bonang Barung (b) note harmony for Bonang Penerus



(a) (b)

(c) (d)

Fig. 14. Spectrogram of the audio data test (a) original (b) generated result using (c) BiLSTM (d) LSTM (e) RNN



(a) (b) (c)

Fig. 15. Visualization of original vs system audio produced by (a) biLSTM (b) LSTM (c) RNN

3.2.3. Audio Visualization

3.2.3.1. Spectrogram

This study uses a spectrogram to extract musical features from audio and measure how similar or different two sounds are. The spectrogram shows the frequency of sound over time, using color to represent signal intensity. Fig. 14 shows the time of the clip on the horizontal axis and the intensity of the original instrument and the generated system on the vertical axis using different colors for the original music from the gamelan composer and the results generated by BiLSTM, LSTM and RNN. The four images show the similarity between the original music and the generator results.

3.2.3.2. Amplitude vs Frequency

The second technique for visualizing audio is to compare the relative amplitudes of the two audios (the original instrument and the generated system) at different frequencies. Fig. 15 shows a plot where the y-axis ranges from 0 to 1. This makes it possible to compare the amplitudes of the original instrument with the generated system, BiLSTM, LSTM, and RNN on the same scale.

From the three plots in Fig. 15, it can be seen that between the original audio (blue) and the generator (red), the results of the BiLSTM generator are more similar to the original shown in Fig. 15a, where the blue plot is less visible compared to the plots of the LSTM and RNN generator results.

3.2.4. Dynamic Time Wrapping (DTW)

DTW is a method for comparing two time series of audio or signals that may have different lengths or distortions [34]. It finds the best alignment or distortion path between the two series, which reduces the distance between corresponding points along the path.

Typically, the DTW is used to calculate the distance between two identical data, with lower values indicating more similarity. In this study, the BiLSTM model has the lowest DTW value of 58198, while the LSTM and RNN models have DTW values of 77964 and 117480, respectively. These results indicate that the BiLSTM produces music that is closer to the original music than the LSTM and RNN models.

3.2.5. Cross-Correlation

Cross-correlation is used to evaluate the similarity between two audio files by comparing their spectra in the spatial and frequency domains, where a high similarity value indicates similarity with the comparison data [35].

In this study, the analysis of the results shows that the biLSTM model produces audio that is 93.00% similar to the original. Meanwhile, the LSTM model produces audio that is 92.26% similar, and the RNN model produces audio that is 90.09% similar.

4. Conclusion

This study aims to generate harmony notes as accompaniment for Javanese gamelan music using three different recurrent neural networks—BiLSTM, LSTM, and RNN. The models were trained using numerical notation to represent the melody for Balungan and Bonang notation. The use of numerical notation is a simplified way of representing musical data. Because the data used has balanced variations, the three different recurrent neural networks—BiLSTM, LSTM and RNN models—can learn the patterns of musical notes to generate harmonic music. Since the dataset used is only Lancaran type, the song structure is not a feature in this study. Based on the experimental results, it was found that all the models performed better in generating harmonies when using all the features of the song (Balungan notation, rhythm, end of lines and songs, scale, mode and dynamic of the song) than when using only two features (rhythm and Balungan notation). Among the three models, BiLSTM is able to produce harmonies that are more similar to the original music, as shown by the results of note distance analysis, audio visualization, and similarity measurements using DTW and cross-correlation techniques, compared to the LSTM and RNN models. This is because BiLSTM has 2 directions, forward and backward, in learning the training data. However, this study still has limitations in the dataset used, namely having unbalanced variations. Thus, it is necessary to add BiLSTM architecture to increase the output produced, one of which is to use BiLSTM with attention.

Future research can use audio data to produce harmonic music and explore more varied song structures in gamelan music, such as Ketawang, Ladrang, Bubaran, and Ayak-Ayakan, which also have variations in notation for their harmonic patterns. The characteristics of a song greatly affect the success of producing musical accompaniment. Therefore, in future research it is still necessary to study the characteristics of each song according to its song structure, because the diversity of Javanese gamelan styles is not only limited to playing music, but to a more philosophical art. From the results of this study, it appears that the results of the generator system can replace gamelan creators in completing song notation which usually only consists of melodic notation for several instruments. One of them is the harmony notation of the Bonang Barung and Bonang Penerus instruments. Therefore, this study aims to help novice gamelan players learn to play the gamelan, one of which is the harmony music of the Bonang Barung and Bonang Penerus instruments which have various patterns, without the need to understand the rules in Javanese gamelan.

Acknowledgment

This research was funded by the Indonesian Endowment Fund for Education (LPDP) through the BUDI DN Doctoral Scholarship Programme and the Soewidiatmaka Gamelan as gamelan experts in providing data.

Declarations

Author contribution. Arik Kurniawati: conception, methodology, analysis, writing. Eko Mulyanto Yuniarno: Validated, reviewed. Yoyon Kusnendar Suprapto: supervised, reviewed. Aditya Nur Ikhsan Soewidiatmaka: data provision, gamelan theory. All authors read and approved the final version of the manuscript.

Funding statement. This research was supported by LPDP.
Conflict of interest. The authors declare no conflict of interest.
Additional information. No additional information is available for this paper.

References

[1] M. Cousineau, S. Carcagno, L. Demany, and D. Pressnitzer, "What is a melody? On the relationship between pitch and brightness of timbre," *Front. Syst. Neurosci.*, vol. 7, no. Jan, p. 127, Jan. 2014, doi: 10.3389/fnsys.2013.00127.

[2] J. D. Lomas and H. Xue, "Harmony in Design: A Synthesis of Literature from Classical Philosophy, the Sciences, Economics, and Design," *She Ji J. Des. Econ. Innov.*, vol. 8, no. 1, pp. 5–64, 2022, doi: 10.1016/j.sheji.2022.01.001.

[3] J. Becker, "Traditional Music in Modern Java," *University of Hawaii Press*, p. 253, 2019, doi: 10.2307/j.ctv9zcjt8.

[4] A. Ranjan, V. N. J. Behera, and M. Reza, "Using a Bi-Directional Long Short-Term Memory Model with Attention Mechanism Trained on MIDI Data for Generating Unique Music," in *Studies in Computational Intelligence*, vol. 1006, Springer Science and Business Media Deutschland GmbH, 2022, pp. 219–239, doi: 10.1007/978-3-030-92245-0_10.

[5] S. Hastanto, "Konsep *Pathet* Dalam Karawitan Jawa," 2009. [Online]. Available: https://repositori.kemdikbud.go.id/13325/1/Konsep *pathet* dalam karawitan jawa.pdf.

[6] J. Becker, "Earth, Fire, Sakti, and the Javanese Gamelan," *Ethnomusicology*, vol. 32, no. 3, p. 385, 1988, doi: 10.2307/851938.

[7] A. N. (Andrew N. Weintraub, "Unplayed Melodies: Javanese Gamelan and the Genesis of Music Theory (review)," *Notes*, vol. 63, no. 1, pp. 87–90, 2006, doi: 10.1353/not.2006.0123.

[8] B. Drummond, "Instruments of the Gamelan," *Boston Village Gamelan*, pp. 1–14, 2014. [Online]. Available: https://www.gamelanbvg.com/gendhing/gamelanGlossary.pdf.

[9] J. Hilder, "Central Javanese Gamelon Handbook." p. 51, 1992, [Online]. Available: https://gamelan.org.nz/wp-content/uploads/2015/02/Jo-Hilder-Central-Javanese-Gamelan.pdf.

[10] W. Widodo, B. Susetyo, S. Walton, and W. Appleton, "Implementation of Kupingan Method in Javanese Karawitan Music Training for Foreigners," *Harmon. J. Arts Res. Educ.*, vol. 21, no. 1, pp. 105–114, Jun. 2021, doi: 10.15294/harmonia.v21i1.29993.

[11] O. S. Olufunso, A. E. Evwiekpaefe, and M. E. Irhebhude, "Gender recognition based fingerprints using dynamic horizontal voting ensemble deep learning," *Int. J. Adv. Intell. Informatics*, vol. 8, no. 3, p. 324, Nov. 2022, doi: 10.26555/ijain.v8i3.927.

[12] S. Lathuiliere, P. Mesejo, X. Alameda-Pineda, and R. Horaud, "A Comprehensive Analysis of Deep Regression," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 42, no. 9, pp. 2065–2081, Sep. 2020, doi: 10.1109/TPAMI.2019.2910523.

[13] B. Kiran, D. Thomas, and R. Parakkal, "An Overview of Deep Learning Based Methods for Unsupervised and Semi-Supervised Anomaly Detection in Videos," *J. Imaging*, vol. 4, no. 2, p. 36, Feb. 2018, doi: 10.3390/jimaging4020036.

[14] M. Bakator and D. Radosav, "Deep Learning and Medical Diagnosis: A Review of Literature," *Multimodal Technol. Interact.*, vol. 2, no. 3, p. 47, Aug. 2018, doi: 10.3390/mti2030047.

[15] H. Li, "Deep learning for natural language processing: advantages and challenges," *Natl. Sci. Rev.*, vol. 5, no. 1, pp. 24–26, Jan. 2018, doi: 10.1093/nsr/nwx110.

[16] T. Yampaka, S. Vonganansup, and P. Labcharoenwongs, "Feature selection using regression mutual information deep convolution neuron networks for COVID-19 X-ray image classification," *Int. J. Adv. Intell. Informatics*, vol. 8, no. 2, p. 199, Jul. 2022, doi: 10.26555/ijain.v8i2.809.

[17] S. H. Ing, A. A. Abdullah, M. Y. Mashor, Z.-A. Mohamed-Hussein, Z. Mohamed, and W. C. Ang, "Exploration of hybrid deep learning algorithms for covid-19 mrna vaccine degradation prediction system," *Int. J. Adv. Intell. Informatics*, vol. 8, no. 3, p. 404, Nov. 2022, doi: 10.26555/ijain.v8i3.950.

[18] S. Bhardwaj, S. M. Salim, D. Talha Ali Khan, and S. JavadiMasoudian, "Automated Music Generation using Deep Learning," in *2022 International Conference Automatics and Informatics (ICAI)*, Oct. 2022, pp. 193–198, doi: 10.1109/ICAI55857.2022.9960063.

[19] O. Peracha, "Improving Polyphonic Music Models with Feature-Rich Encoding," in *Proceedings of the 21st International Society for Music Information Retrieval Conference, ISMIR 2020*, Nov. 2019, p. 7. [Online]. Available: https://arxiv.org/abs/1911.11775.

[20] J. Ens and P. Pasquier, "MMM : Exploring Conditional Multi-Track Music Generation with the Transformer," *arxiv Comput. Sci.*, p. 10, Aug. 2020, Accessed: May 27, 2023. [Online]. Available: https://arxiv.org/abs/2008.06048v2.

[21] A. Kurniawati, Y. K. Suprapto, and E. M. Yuniarno, "Multilayer Perceptron for Symbolic Indonesian Music Generation," in *2020 International Seminar on Intelligent Technology and Its Applications (ISITIA)*, Jul. 2020, pp. 228–233, doi: 10.1109/ISITIA49792.2020.9163723.

[22] J.-P. Briot, G. Hadjeres, and F.-D. Pachet, "Deep Learning Techniques for Music Generation," *Cham: Springer International Publishing*, p. 284 , 2020, doi: 10.1007/978-3-319-70163-9.

[23] M. Civit, J. Civit-Masot, F. Cuadrado, and M. J. Escalona, "A systematic review of artificial intelligence-based music generation: Scope, applications, and future trends," *Expert Syst. Appl.*, vol. 209, p. 118190, Dec. 2022, doi: 10.1016/j.eswa.2022.118190.

[24] J. Xie, "A Novel Method of Music Generation Based on Three Different Recurrent Neural Networks," *J. Phys. Conf. Ser.*, vol. 1549, no. 4, p. 042034, Jun. 2020, doi: 10.1088/1742-6596/1549/4/042034.

[25] N. Agarwala, Y. Inoue, and A. Sly, "Music composition using neural network," p. 10, 1992. [Online]. Available: https://web.stanford.edu/class/archive/cs/cs224n/cs224n.1174/reports/2762076.pdf.

[26] S. Mangal, R. Modak, and P. Joshi, "LSTM Based Music Generation System," *IARJSET*, vol. 6, no. 5, pp. 47–54, May 2019, doi: 10.17148/IARJSET.2019.6508.

[27] F. Shah, T. Naik, 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.

[28] S. Agarwal, V. Saxena, V. Singal, and S. Aggarwal, "LSTM based Music Generation with Dataset Preprocessing and Reconstruction Techniques," in *2018 IEEE Symposium Series on Computational Intelligence (SSCI)*, Nov. 2018, pp. 455–462, doi: 10.1109/SSCI.2018.8628712.

[29] S. Chikkamath and N. S R, "Melody generation using LSTM and BI-LSTM Network," in *2021 International Conference on Computational Intelligence and Computing Applications (ICCICA)*, Nov. 2021, pp. 1–6, doi: 10.1109/ICCICA52458.2021.9697286.

[30] S. S. Azmi, C. S. Shreekara, and S. Baliga, "Music generation using Bidirectional Recurrent Neural Nets," *Int. Res. J. Eng. Technol.*, vol. 7, no. May, pp. 6863–6866, 2020, [Online]. Available: https://www.academia.edu/download/64615427/IRJET-V7I51292.pdf.

[31] T. Jiang, Q. Xiao, and X. Yin, "Music Generation Using Bidirectional Recurrent Network," in *2019 IEEE 2nd International Conference on Electronics Technology (ICET)*, May 2019, pp. 564–569, doi: 10.1109/ELTECH.2019.8839399.

[32] H. Elfaik and E. H. Nfaoui, "Deep Bidirectional LSTM Network Learning-Based Sentiment Analysis for Arabic Text," *J. Intell. Syst.*, vol. 30, no. 1, pp. 395–412, Dec. 2020, doi: 10.1515/jisys-2020-0021.

[33] N. M R and S. Mohan B S, "Music Genre Classification using Spectrograms," in *2020 International Conference on Power, Instrumentation, Control and Computing (PICC)*, Dec. 2020, pp. 1–5, doi: 10.1109/PICC51425.2020.9362364.

[34] L. Lerato and T. Niesler, "Feature trajectory dynamic time warping for clustering of speech segments," *EURASIP J. Audio, Speech, Music Process.*, vol. 2019, no. 1, p. 6, Dec. 2019, doi: 10.1186/s13636-019-0149-9.

[35] C. Jiang, D. Yang, and X. Chen, “Similarity Learning For Cover Song Identification Using Cross-Similarity Matrices of Multi-Level Deep Sequences,” in *ICASSP 2020 - 2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, May 2020, vol. 2020-May, pp. 26–30, doi: 10.1109/ICASSP40776.2020.9053257.