KEDS Knowledge Engineering and Data Science (KEDS)
Vol 6, No 1, April 2023, pp. 41–56
pISSN 2597-4602
eISSN 2597-4637

Deep Learning for Multi-Structured Javanese Gamelan Note Generator

Arik Kurniawati [a,1], Eko Mulyanto Yuniarno [a,b,2,*], Yoyon Kusnendar Suprapto [a,b,3]

[a] *Department of Electrical Engineering, Institut Teknologi Sepuluh Nopember, Jl. Teknik Kimia, Surabaya 60117, Indonesia*
[b] *Department of Computer Engineering, Institut Teknologi Sepuluh Nopember, Jl. Teknik Kimia, Surabaya 60117, Indonesia*
[1] *arikkurniawati.19071@mhs.its.ac.id;* [2] *ekomulyanto@ee.its.ac.id*;* [3] *yoyonsuprapto@ee.its.ac.id*
** corresponding author*

ARTICLE INFO

Article history:
Received 20 June 2023
Revised 07 July 2023
Accepted 15 July 2023
Published online 18 July 2023

Keywords:
Javanese *Gamelan*
Notation
CNN-LSTM
Multi-instrument
Ricikan struktural

ABSTRACT

Javanese *gamelan*, a traditional Indonesian musical style, has several song structures called *gendhing*. *Gendhing* (songs) are written in conventional notation and require *gamelan* musicians to recognize patterns in the structure of each song. Usually, previous research on *gendhing* focuses on artistic and ethnomusicological perspectives, but this study is to explore the correlation between *gendhing* as traditional music in Indonesia and deep learning technology that replaces the task of *gamelan* composers. This research proposes CNN-LSTM to generate notation of *ricikan struktural* instruments as an accompaniment to Javanese *gamelan* music compositions based on *balungan* notation, rhythm, song structure, and *gatra* information. This proposed method (CNN-LSTM) is compared with LSTM and CNN. The musical data in this study is represented using numerical notation for the main melody in *balungan* notation. The experimental results showed that the CNN-LSTM model showed better performance compared to the LSTM and CNN models, with accuracy values of 91.9%, 91.5%, and 91.2% for CNN-LSTM, LSTM, and CNN, respectively. And the value of note distance for the *Sampak* song structure is 4 for the CNN-LSTM model, 8 for the LSTM model, and 12 for the CNN model. The smaller the note distance, the closer it is to the original notation provided by the *gamelan* composer. This study provides relevance for novice *gamelan* musicians who are interested in learning *karawitan*, especially in understanding *ricikan struktural* music notation and *gamelan* art in composing musical compositions of a song.

I. Introduction

Javanese *gamelan*, one of the musical arts of Indonesia, is known for its diverse playing patterns. The technique for playing it is usually called *karawitan*. A song in Javanese *gamelan* has different patterns of presentation, as in the examples of the songs *Sampak Nem Slendro Nem* and *Srepeg Nem Slendro Nem*. What distinguishes the two songs is the type of song structure; the first is *Sampak* and the second is *Srepeg*. This song structure is like a genre in general music; this song structure is played by *ricikan struktural* instruments. This means that if the song structure pattern played is not appropriate, then the song has lost its composition. Because in Javanese *gamelan*, a song is not only based on the strength of the main melody but also on other instruments as accompaniment music, because these instruments are used to compose the composition of a song as a whole. In a song in Javanese *gamelan*, the song title reflects how the song composition is played [1][2][3][4][5][6][7][8].

One of the variations of *karawitan* patterns in Javanese *gamelan* is the Surakarta style, which has several forms of song structure [1][2]. A song is composed of various elements. These elements contribute to the overall composition. These elements include dynamics, rhythm, *laya*, *laras,* and *pathet*. Tempo plays a crucial role in controlling the rhythm of the *gendhing*, while *laya* describes the speed at which it is performed. *Pathet* expresses the specific emotion or feeling the song is trying to

https://doi.org/10.17977/um018v6i12023p41-56

convey, and *laras* refers to the scales used in that song. Dynamics, on the other hand, emphasizes the variety, balance, and dynamic nature of a song's musical components [2][3].

Song structure, a particular *karawitan* art form, uses music as a symbolic medium to represent various aspects [4][6]. The goals of song are to be complex, to entertain the audience, and to convey a range of social, moral, cultural, and spiritual values [5]. Incorrect performance of musical techniques in a song composition can lead to the loss of its aesthetic value and unique characteristics. In order to perform Javanese *gamelan* well, it is necessary to have an understanding of both the rules of *gamelan* and the emotional atmosphere conveyed by the piece of music being performed.

However, playing Javanese *gamelan* presents several challenges, especially in determining the playing pattern [3]. As a result, assistance is required to facilitate the learning process of this cultural practice for future generations [3][4]. The aim of this study is to use technology to simplify the process of playing Javanese *gamelan*.

The size of a *gendhing* (song) can be determined by calculating the number of *gatra* in each *gongan* and the total number of *gongan* in the song [1][2]. *Gendhing* is further divided into three subtypes: *ageng*(big), *sedheng (middle)*, and *alit*(small). *Gendhing alit*, consisting of *sampak, srepeg, ayak-ayakan, lancaran, bubaran, ketawang, and ladrang*, is the focus of this study [2]. This categorization is based on the design of the *ricikan struktural* instrument groupings, which include the *kenong*, *kethuk*, *kempyang*, *kempul*, and *gong*. The arrangement of the *ricikan struktural* instruments is an important factor in notation that determines the composition of the musical piece [2][6]. The musical instruments known as *kenong*, *kethuk*, and *kempul* serve as breaks in the song, while the *gong* indicates the end of the song.

There are two additional groups of Javanese *gamelan* instruments in addition to the *ricikan struktural* instruments, which are a) *ricikan balungan*, which is a group of musical instruments that play the basic melody of a song, such as *slenthem, demung, saron, and peking*; and b) *ricikan garap* as musical accompaniment like *ricikan stuktural*, which is a group of musical instruments that handle variations in song decoration, such as *rebab, gender barung, gender penerus, bonang barung, bonang penerus, gambang, siter,* and *suling* [2].

The configuration of musical pieces in Javanese *gamelan* is occasionally not only dependent on the composer's artistic expression but also matches standard notational conventions. Consequently, in order to perform a piece in Javanese *gamelan*, it is necessary to commit to memory the patterns of each composition's song structure, as complete notation for all *gamelan* instruments is not always provided. The Javanese *gamelan* notation generally consists of only the primary melody, thereby necessitating a high level of expertise among *gamelan* musicians to execute all the instruments. Nonetheless, this presents a difficulty for inexperienced musicians who require comprehensive notation for every instrument to perform *gamelan* music. Figure 1 illustrates the structure of a Javanese *gamelan* composition. *Gamelan* sheet music, as depicted in Figure 1, only displays *balungan* notation (note) and omits the notation of the other two groups of instruments, *ricikan struktural* and *ricikan garap*. This notation is typically used by *gamelan* players to perform *karawitan*, along with other information about the piece, such as the song's structure type, rhythm type, and information about the *laras* and *pathet*. *Laras* and *pathet* refer to musical scales and modes of the song.

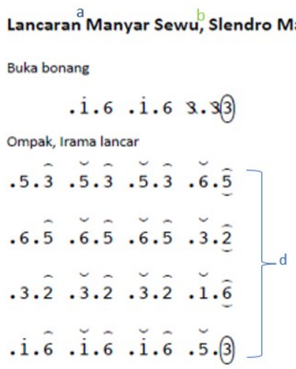


Fig. 1. Part of song in javanese *gamelan* (a) song structure, (b) title of song, (c) *laras* and *pathet*, (d) melody

Figure 2 illustrates the *ricikan struktural* instruments used in the composition of a song [1],[2],[8]. These instruments include *gong ageng, gong suwuk, kenong, kempul, kethuk,* and *kempyang*, as shown in Figure 2. The position of these instruments within a song distinguishes different types of song structure. The *gong ageng* denotes the longest cycle of a song, while the *gong suwuk* is used in all song structures except the *ketawang* and *ladrang* forms, where it is replaced by the *kempul*. The *kenong* divides the flow of the *gendhing* into musical phrases of equal length. The *kempul*, which is a smaller *gong*, often interlocks with the *kenong* in forms such as *lancaran*, *ketawang,* and *ladrang*. The *balungan* represents the melody notes of each song, which are divided into several lines, each line containing several *gatra*, each of which is made up of several notes.

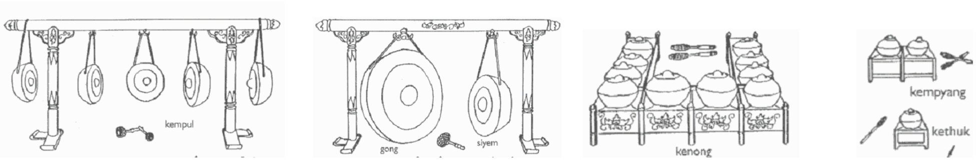


Fig. 2. *Ricikan struktural* in javanese *gamelan*

Figure 3 is an example of the detailed structure of the *gendhing lancaran* form. *Lancaran* is a form of *gendhing* that has 4 *gatra* or 16 *balungan* notations on each *gongan*. There are usually four *gongan* in a *lancaran* composition. The pattern rules for *lancaran* are as follows: (1) *kenong* occurs on the last note of each *gatra* (also known as *dhong gedhe*), and the note always matches that of the *dhong gedhe*; (2) *Kempul* occurs on the second note of each *gatra* (also known as *dhong cilik*), and there are only three *kempul* notes. The first *gatra* has no *kempul* note; (3) *Kethuk* (+) is played on the odd notes of each *gatra*; (4) *Gong suwuk* is played at the end of the fourth *gatra*.

.	5	.	3	.	5	.	3	.	5	.	3	.	6	.	5	**Balungan**
+		+		+		+		+		+		+		+		Kethuk
			3				3				5				5	Kenong
					5				5				6			Kempul
															2	Gong Suwuk

Fig. 3. Song structure of *lancaran*

Rhythm (*Irama*) refers to the tempo and rhythm in *gamelan* music. There are five types of rhythm, including *Irama Lancar*, *Irama Tanggung*, *Irama Dadi*, *Irama Wilet*, and *Irama Rangkep*. A song is typically presented in different rhythms [5], such as the *Lancaran Manyar Sewu* song, which can be presented in both the *Irama Lancar* and *Irama Tanggung* forms. In this case, the rhythm has a significant impact on the way the song is performed.

Currently, discussions of the types of *gendhing* patterns focus mainly on artistic and ethnomusicological perspectives. For example, studies have examined the *kempul* pattern in *gendhing* alit in *Klenengan* music [6], the *kenong* instrument pattern in *karawitan* style aesthetics [7], and the role of *ricikan struktural* as one of the indicators in *gendhing* formation [8]. However, the relationship between *gamelan* music and technology, especially Deep Learning (DL), has received little attention. The purpose of this study is to use DL to assist novice *gamelan* musicians in understanding the *ricikan struktural* components. This study is known as part of the music generation.

The integration of DL technology with the art of music has contributed to the development of music generators capable of creating new and unique musical compositions [9]. In recent years, the field of music composition has seen significant development due to the development of advanced deep learning techniques such as Convolutional Neural Network (CNN) and Long Short-Term Memory (LSTM).

The CNN is a special type of deep learning that has been used in the field of music composition. An example of this phenomenon is the creation of new music using audio-based music, such as MIDI [10], or symbolically represented music [11] in alternative formats. The use of CNN represents a contemporary advancement in the field of music. The use of CNN has been widely implemented in the field of image classification [12]. The previously discussed networks are purposefully constructed to detect and extract identifiable patterns and features from visual data [13]. Similar methods are used to train these networks for the purpose of recognizing patterns and features in musical sequences. In previous research related to music generation, CNN were reliable in obtaining the semantic features of music [14] and multiple feature extraction [15].

CNN is often integrated with other deep learning techniques, such as LSTM, to generate complex and sophisticated musical compositions [13]. The LSTM network is a variant of the Recurrent Neural Network (RNN), which is able to effectively capture long-term temporal dependencies in time-series data, including musical sequences. In previous studies, LSTM has been widely used for music generation because it is suitable for learning patterns from sequential music data [16][17].

The combination of CNN and LSTM networks produces both short-term and long-term musical patterns, resulting in more authentic and rationally structured music [13][18]. CNN-LSTM has several advantages, including the ability to perform temporal analysis while extracting abstract features [19], and it outperforms standard machine learning algorithms in terms of stability, accuracy, and prediction [20][21][22]. In music generation, Convolutional LSTM outperforms LSTM with more pronounced waveforms and clearer melodies [18]. It combines the advantages of CNN, which can extract effective features from data music sequences, and LSTM, which can not only discover data interdependence in time series data, but also automatically detect the ideal mode suitable for relevant data to build new sequences [23].

Many music-related studies use the combined methods of CNN and LSTM, such as music classification or music genre recognition [28][29][30][31][32], music recommendation [33], chord recognition [34][35], and music emotion recognition [36][37][38]. CNNs are used to extract audio or sheet music features, while LSTMs are used to learn temporal dependencies in music data for recognition, prediction, recommendation, and classification.

However, previous research on music generation using the CNN-LSTM combination is limited to the generation of new melodies in Turkish pop music with a certain style [13] and modern music [18] from MIDI files. In this study, the same approach is used to generate music notation for several instruments based on variations in the structure of Javanese *gamelan* songs using notation-based music datasets. However, the difference with the previous research is that this study uses a dataset with more readable notation represented as numerical notes in text format. And the focus of this study is to generate musical accompaniment for multiple instruments. In the context of *gamelan* music, CNN and LSTM have been used to create musical compositions that follow the rules and conventions of traditional *gamelan* music. The ability of the CNN network is used to extract important features from the input parameters fed into the network, such as *balungan* notation, rhythm, and *gatra* information. The LSTM network is then used to generate the notation of several *ricikan struktural* instruments as a musical accompaniment to the melodic notation of the *balungan* instrument, with the ability of the LSTM to model temporal dependencies.

According to the above statement, the issues covered in this study are:

- Writing complete notation, especially for *ricikan struktural* instruments, is very helpful for novice *gamelan* players.
- The notation patterns of the *ricikan struktural* instruments have different variations, so it will be more convenient for novice *gamelan* players to play a *gamelan* song based on the structure of the song, where the function of the notation pattern of the *ricikan struktural* instrument is used as the structure for a song.

This study aims to automatically generate notation for several instrument groups, including *kenong*, *kethuk*, *kempyang*, *kempul*, and *gong*, using CNN-LSTM. The features used in this study include the main melodic notation of the *balungan* instrument, rhythm, and *gatra* information. The main contributions of this study are presented below:

- A dataset of Javanese *gamelan* music was created based on symbol notation.
- The use of numerical notes as a simplified method of representing musical data as input.
- This study effectively generates musical accompaniment for various musical instruments, including *kenong*, *kethuk*, *kempyang*, *kempul*, and *gong*, by incorporating song characteristics such as song structure, *gatra*, and rhythm.

- To help the general public understand the various patterns of song structures and their notation for the *ricikan struktural* instrument groups.

The remaining sections of this paper are organized as follows: Section I presents the introduction and related work. Section II describes the methodology, including the details of the dataset and the proposed model. Section III presents the experiments and results. Finally, Section IV provides the conclusion of the paper.

II. Method

The objective of this study is to use CNN-LSTM to create an automatic notation generator for the *ricikan struktural* instrument. The technique in this study uses CNN for feature extraction and LSTM as the notation generator. The detailed steps for implementing the proposed method are discussed in this section.

A. Dataset

The present study employed symbol-based data, specifically numerical notes, sourced from a collection of multiple songs available at http://www.*gamelan*bvg.com, for the music dataset. The data extracted from musical compositions includes the song's notation as well as its distinctive features, such as *gatra* details, rhythmic patterns, and song structure composition. Furthermore, an annotation of certain *ricikan struktural* instruments designed by a specialist in *gamelan* from Soewidiatmaka *Gamelan* has been incorporated into the dataset.

A total of 35 songs were used in this study. These are divided into seven song structures, with five songs in each structure. The various *ricikan struktural* instruments and the notation for the *balungan* were arranged according to the *gatra* of each song. The *balungan* is often represented by four notations in one *gatra*. As a result, the dataset used in the current study contains approximately 600 *gatra* distributed across the 35 songs, as shown in Figure 4. In this dataset, 28 songs were used for training (80% of the data) and validation (20% of the data), and 7 songs were used for testing. The songs used in this study are listed in Table 1, where the table lists the song titles used as datasets with the type of song structure, type of laras (scale of the song), type of pathet (mode of the song), and the rhythm contained in the song.

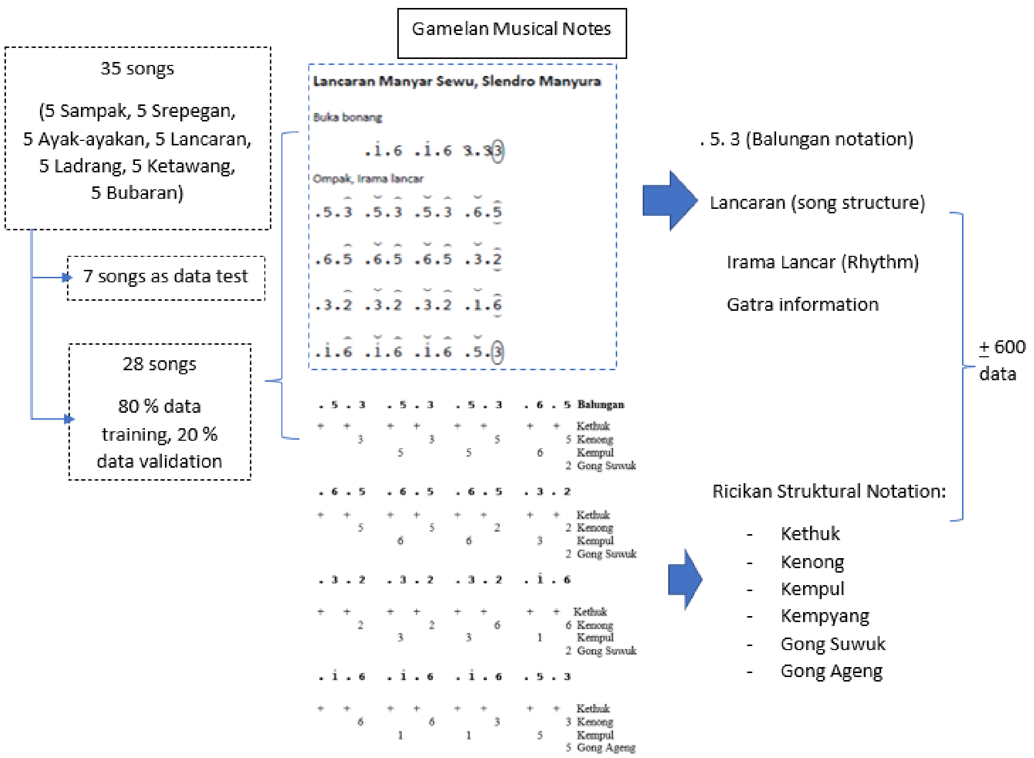


Fig. 4. Dataset Representation

Table 1. List of songs for dataset in this study

No	Song	Rhythm	Data
1	*Sampak Tlutur Slendro Manyura*	*Tanggung*	Test
2	*Sampak Manyura Slendro Manyura*	*Tanggung*	Training, Validation
3	*Sampak Nem Slendro Nem*	*Tanggung*	Training, Validation
4	*Sampak Sanga Slendro Sanga*	*Tanggung*	Training, Validation
5	*Sampak Tlutur Slendro Sanga*	*Tanggung*	Training, Validation
6	*Srepeg* Manyura Slendro Manyura	*Tanggung*	Test
7	*Srepeg* Nem Slendro Nem	*Tanggung*	Training, Validation
8	*Srepeg* Sanga Slendro Sanga	*Tanggung*	Training, Validation
9	*Srepeg* Tlutur Slendro Manyura	*Tanggung*	Training, Validation
10	*Srepeg* Tlutur Slendro Sanga	*Tanggung*	Training, Validation
11	*Ayak-Ayakan Nem Slendro Nem*	*Lancar, Tanggung, Dadi*	Test
12	*Ayak-Ayakan Manyura Slendro Manyura*	*Lancar, Tanggung, Dadi*	Training, Validation
13	*Ayak-Ayakan Pamungkas Slendro Manyura*	*Lancar, Tanggung, Dadi*	Training, Validation
14	*Ayak-Ayakan Sanga Slendro Sanga*	*Lancar, Tanggung, Dadi*	Training, Validation
15	*Ayak-Ayakan Umbul Donga Slendro Manyura*	*Lancar, Tanggung, Dadi*	Training, Validation
16	*Lancaran Manyar Sewu Slendro Manyura*	*Lancar*	Test
17	*Lancaran Kuda Nyongklang Pelog Barang*	*Lancar, Tanggung*	Training, Validation
18	*Lancaran Maesa Kurda Slendro Sanga*	*Lancar, Tanggung*	Training, Validation
19	*Lancaran Rena Rena Slendro Manyura*	*Lancar*	Training, Validation
20	*Lancaran Sarung Jagung Pelog Barang*	*Tanggung*	Training, Validation
21	*Bubaran Arum Arum Pelog Barang*	*Tanggung*	Test
22	*Bubaran Kembang Pacar Pelog Nem*	*Tanggung*	Training, Validation
23	*Bubaran Purwaka Pelog Nem*	*Tanggung*	Training, Validation
24	*Bubaran Sembunggilang Slendro Sanga*	*Tanggung*	Training, Validation
25	*Bubaran Udan Mas Pelog Barang*	*Tanggung*	Training, Validation
26	*Ketawang Ibu Pretiwi Pelog Nem*	*Tanggung, Dadi*	Test
27	*Ketawang Kinanthi Pawukir Slendro Manyura*	*Tanggung, Dadi*	Training, Validation
28	*Ketawang Kinanthi Sandhung Slendro Manyura*	*Tanggung, Dadi*	Training, Validation
29	*Ketawang Langen Gita Pelog Barang*	*Tanggung, Dadi*	Training, Validation
30	*Ketawang Subakastawa Slendro Sanga*	*Tanggung, Dadi*	Training, Validation
31	*Ladrang Kalongking Pelog Nem*	*Tanggung*	Test
32	*Ladrang Mugi Rahayu Slendro Manyura*	*Tanggung, Dadi*	Training, Validation
33	*Ladrang Pariwisata Slendro Sanga*	*Tanggung, Dadi, Wiled*	Training, Validation
34	*Ladrang Santi Mulya Pelog Lima*	*Tanggung, Dadi*	Training, Validation
35	*Ladrang Sumyar Pelog Barang*	*Tanggung, Dadi, Wiled*	Training, Validation

Laras (scale of song) : Slendro /Pelog; Pathet (mode of song): Manyura, Nem, Sanga,Barang, Lima

B. Preprocessing Data

The input data of this study consists of *balungan* notation, rhythm type, song structure type, and *gatra* information, while the output data consists of *ricikan struktural* music notation such as *kenong, kethuk, kempyang, kempul, gong ageng,* and *gong suwuk*. Preprocessing of both input and output data using one-hot encoding techniques [39], which involves converting both input and output data into binary form with careful consideration of the respective data, Figure 5 shows the preprocessing result of one-hot encoding.

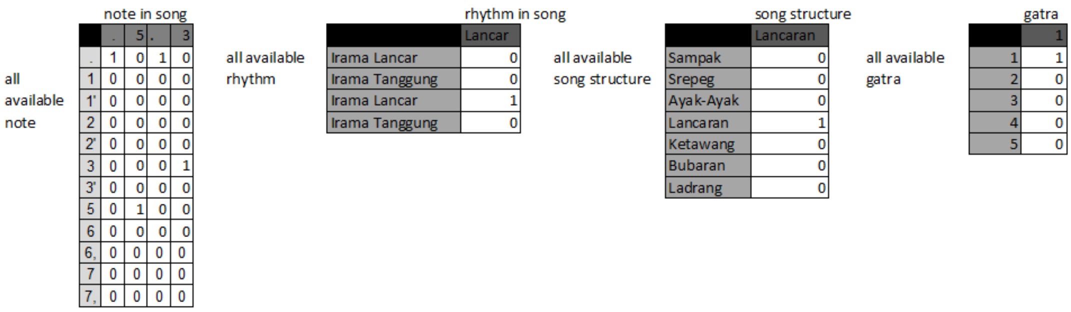


Fig. 5. One-hot encoding for note, rhythm, song structure and *gatra*

Before the input is fed into the CNN-LSTM network, a one-hot encoding process is performed on each input, which consists of *balungan* notation arranged in each *gatra*, rhythm, song structure, and *gatra* information from this note. After the encoded vector input is combined into an input sequence, it is ready to be fed into the CNN-LSTM architecture network.

C. CNN-LSTM

The following section provides a detailed description of the structure of the CNN-LSTM architecture model. The diagram in Figure 6 shows the different steps of this study. The proposed CNN-LSTM model consists of three main components: a Convolutional Neural Network (CNN), a Long-Short-Term Memory (LSTM) network, and a fully connected layer.

- The CNN is used to obtain a feature representation of the input music sequence, which consists of *balungan* notation divided into *gatra*, rhythm, song structure, and *gatra* information from this note. This CNN network consists of a 1D convolutional layer with 32 filters and a kernel size of 2, with padding set to the same size. This is followed by an activation layer using RELU and a 1D max-pooling layer.
- The LSTM component is responsible for modeling the temporal dependencies between the extracted features and generating musical accompaniment sequences. It consists of a single-layer LSTM with 128 hidden units and a dropout layer with a size of 0.2 to avoid overfitting.
- The fully connected layer and the output layer use a sigmoid activation function for each *ricikan struktural* instrument to predict the musical accompaniment.

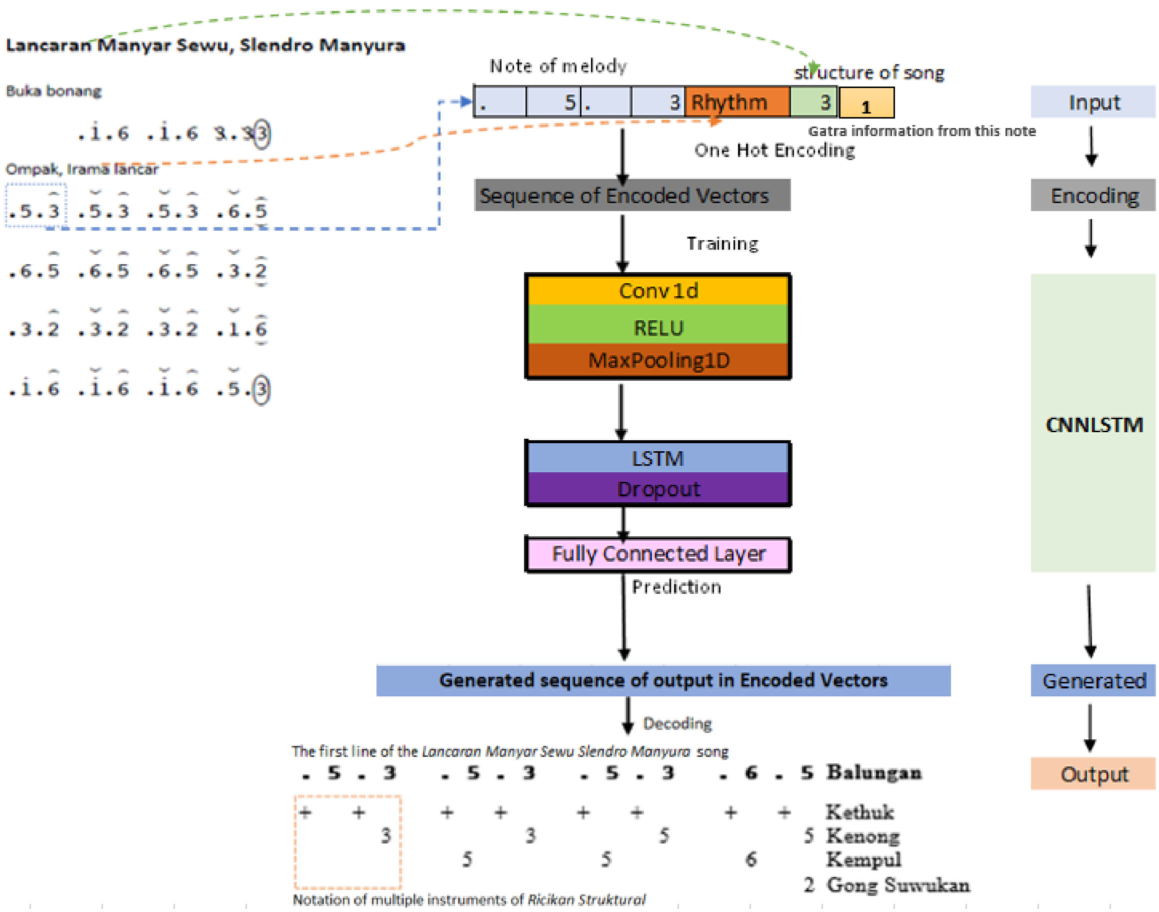
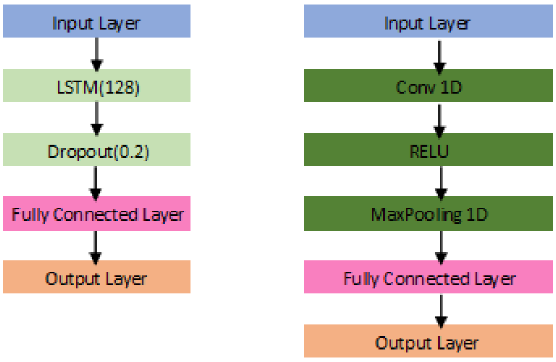


Fig. 6. Proposed method CNN-LSTM for note generator for multi-instrument

In addition, the model was trained up to 100 epochs, batch size 5, with the Adam optimizer and Binary Cross Entropy as the loss function values.

After completing the training, the CNN-LSTM network demonstrates the ability to generate a sequence of musical notes suitable for the purpose of providing accompaniment to *ricikan struktural* instruments. By first decoding the vector sequence encoded in the *ricikan struktural* instrument notation. The model uses this data to automatically predict *kenong*, *kethuk*, *kempyang*, *kempul*, *gong suwuk*, and *gong ageng* notes based on test data containing *balungan* notes, rhythm, and *gatra* information.

To provide a comparative analysis, we compared the performance of the CNN-LSTM model with that of the CNN and LSTM models. The architectural details of each model are shown in Figure 7.



(a) (b)

Fig. 7. Architecture of (a) LSTM and (b) CNN for note generator for Multi-instrument

D. Evaluation

As the first evaluation for this study, we investigated the effectiveness of our proposed CNN-LSTM model in predicting musical accompaniment notes for various *ricikan struktural* instruments. We compared its performance with that of CNN and LSTM models. To evaluate the performance of the CNN-LSTM model, we compared its predictions with the ground truth labels or desired outputs (the original notation from the *gamelan* composer). By applying the model to a specific dataset and comparing its predictions with the actual results, we were able to determine the exact values of accuracy, precision, and recall [40].

- Accuracy measures the overall prediction accuracy of a model by determining the number of correctly predicted examples. Higher accuracy indicates better performance.
- Precision is a metric that refers to the number of true positives correctly identified and the sum of true positives and false positives. An increase in precision results in a decrease in false positive accuracy. False positives indicate that the model predicts a positive outcome, but the actual outcome is negative.
- The recall metric evaluates a model's ability to reliably detect all positive cases. A lower false negative rate indicates a higher recall score. False negatives indicate that a model predicts a negative outcome when the actual outcome is positive.

The second evaluation involves applying the second scenario with different song structures by selecting a single song that is not included in the training data for each song structure. The notation generated by the song generator is then compared to the original version using music analysis methods such as note distance. In this evaluation phase, a detailed assessment of the predictive ability of the proposed model for musical accompaniment is expected.

III. Result And Discussion

This section focuses on the evaluation of the performance of the proposed CNN-LSTM model and the assessment of the generated results, with the ultimate goal of providing accompaniment music notations for different types of *ricikan struktural* instruments. The evaluation was divided into two scenarios: intensive experiments with the same song structure and experiments with different song structures.

In the first scenario, several intensive experiments were conducted to evaluate the overall performance of the model on datasets of the same type. The goal is to see how well the model performs when the song structure remains consistent throughout the test period.

In contrast, in the second scenario, the experiment was conducted by evaluating the model's performance on datasets with different types of song structure. The goal of this scenario is to evaluate the adaptability and generalizability of the model across different forms of song structure. This was intended to assess the model's ability to accurately generate musical accompaniment notes across a range of *ricikan struktural* instruments.

A. *Quantitative Analysis*

The results of the quantitative analysis of the performance of each model in the two scenarios are summarized in Table 2. The results show that the CNN-LSTM framework exhibits superior performance compared to the LSTM and CNN models in all evaluated scenarios, regardless of whether the song structures used are the same or different, as seen from the accuracy, precision, and recall values.

Table 2. Performance Value of accuracy, precision, and recall for CNN-LSTM, LSTM, and CNN

Scenario		Accuracy (%)	Precision (%)	Recall (%)
CNN-LSTM (proposed)				
1	All various	91,9	92,3	91,8
2	Sampak	96,6	96,6	96,6
	Srepeg	96,6	96,6	96,6
	Ayak-Ayakan	99,1	99,1	99
	Lancaran	97,4	97,8	97
	Bubaran	98,9	99,1	98,4
	Ketawang	99	100	98,5
	Ladrang	97,6	98,3	96,1
CNN				
1	All various	91,2	91,9	91
2	Sampak	96,3	96,4	96,3
	Srepeg	96,3	96,5	96,3
	Ayak-Ayakan	99,1	99,1	98,9
	Lancaran	96,8	97,3	96,6
	Bubaran	98,7	98,8	98,2
	Ketawang	98,8	99,6	98,1
	Ladrang	97,4	97,2	95,3
LSTM				
1	All various	91,5	92	91,3
2	Sampak	96,6	96,6	96,6
	Srepeg	96,6	96,6	96,6
	Ayak-Ayakan	99,1	99,1	98,9
	Lancaran	97	97,4	96,8
	Bubaran	98,7	99,1	98,4
	Ketawang	99	99,6	98,5
	Ladrang	96,4	98,2	95,8

The CNN-LSTM model has higher accuracy, precision, and recall values compared to the CNN and LSTM models. A high accuracy score indicates better model performance. A high precision value indicates fewer false positives. And a high recall value indicates fewer false negatives.

Model performance with high values in the first scenario in Table 2 (Accuracy = 91.9; Precision = 92.3; Recall 91.8) will affect the generator results of the *ricikan struktural* instrument notation, i.e., the result of the CNN-LSTM model generator will be more similar to the original when compared to the generator results of CNN and LSTM. This will be discussed in more detail in the Music Generation Results section. While the difference in performance accuracy between the three models is comparatively small, fluctuating between a positive 0.2 and 1.2. Regarding the accuracy of the first scenario, the CNN-LSTM model achieved 91.9, while the CNN and LSTM models achieved 91.2 and 91.5, respectively. Furthermore, the second scenario tends to produce better performance results due to the homogeneity of the data used in the first scenario.

The CNN-LSTM model offers a remarkable advantage by integrating the advantageous features of both the CNN architecture, which is great at feature extraction, and the LSTM architecture, which is excellent at modeling temporal dependencies. The integration of CNN and LSTM in the model enables it to handle both micro- and macro-level musical patterns proficiently, leading to the generation of more precise and expressive musical accompaniment.

B. Music Generation Result

This section evaluates the notation generators used by analysis tools. Test data from each song structure in the second scenario, which has different song structures, will be used. The goal of this evaluation is to assess how closely the output of the generator resembles the composition provided by the *gamelan* composer. The evaluation criterion used in this evaluation phase is the measure of note distance. Note distance is a metric used to quantify the similarity between the generator's output notation ($Note_2$) and the original notation ($Note_1$) of a *gamelan* composer's creation. This distance, also referred to as the exact distance, is represented by a binary representation as written in (1).

$$N_0(Note_1, Note_2) = \begin{cases} 0 & if\ Note_1 = Note_2 \\ 1 & if\ Note_1 \neq Note_2 \end{cases} \quad (1)$$

The proposed approach, CNN-LSTM, was evaluated with a comparative analysis compared to CNN and LSTM. This evaluation was done by calculating the note distance for each instrument in each song structure. Furthermore, an in-depth analysis was conducted to investigate the relationship between input parameters such as *balungan* notation, song structure, rhythm, and *gatra* information and the output notation generated on various *ricikan struktural* instruments such as *kenong*, *kethuk*, *kempyang*, *kempul*, *gong suwuk*, and *gong ageng*.

Table 3 shows the note distance values for each instrument for the r*icikan struktural* of various song structures. The results indicate that the CNN-LSTM approach produced notations with the lowest note distance values compared to LSTM and CNN. A decrease in the note distance value indicates an increase in the degree of similarity between the notations provided by the *gamelan* composer's musical composition. The results of this study indicate that the CNN-LSTM model outperforms both the LSTM and CNN models in terms of improving overall performance, as it effectively exploits the strengths of both CNN and LSTM.

Table 3. Value of note distance from three model CNN-LSTM, CNN, LSTM

Song Structure	Method	*Kenong*	*Kethuk*	*Kempyang*	*Kempul*	*Gong Suwuk*	*Gong Ageng*	Total
Sampak	**CNN-LSTM**	**0**	**0**	**-**	**4**	**0**	**0**	**4**
	LSTM	0	0	-	8	0	0	8
	CNN	0	0	-	10	2	0	12
Srepeg	**CNN-LSTM**	**0**	**0**	**-**	**0**	**0**	**0**	**0**
	LSTM	3	0	-	3	0	0	6
	CNN	3	0	-	3	0	0	6
Ayak-Ayakan	**CNN-LSTM**	**1**	**0**	**-**	**1**	**2**	**0**	**4**
	LSTM	1	0	-	2	2	0	5
	CNN	2	0	-	1	2	0	5
Lancaran	**CNN-LSTM**	**0**	**0**	**-**	**0**	**1**	**0**	**1**
	LSTM	0	0	-	2	1	0	3
	CNN	1	0	-	2	1	0	4
Bubaran	**CNN-LSTM**	**0**	**0**	**-**	**0**	**0**	**0**	**0**
	LSTM	0	0	-	0	0	0	0
	CNN	0	0	-	0	0	0	0
Ketawang	**CNN-LSTM**	**0**	**0**	**0**	**0**	**0**	**0**	**0**
	LSTM	0	0	0	1	0	0	1
	CNN	0	0	0	2	0	0	2
Ladrang	**CNN-LSTM**	**0**	**0**	**0**	**4**	**0**	**0**	**4**
	LSTM	1	0	0	4	0	0	5
	CNN	2	0	0	3	0	0	5

The *kempyang* instrument is only present in the ketawang and ladrang song structures, and has no notation in other song structures. Table 3 shows that the *kethuk*, *kempyang*, and *gong ageng* instruments have a note distance value of 0. As a result, the generated notation from all three models across different song structures is very similar to the *gamelan* composer's original notation. The fixed notation patterns of each instrument within the song structure contribute to this similarity. Specifically, the *kethuk* instrument has a consistent notation pattern of (+), which represents a hit, while the *kempyang* instrument has a consistent notation pattern of (-), which also represents a hit. These instruments have no variations in tone. In addition, the *gong ageng* instrument serves as an indicator of the end of the song, so its notation pattern remains constant without any variations. Figure 8 shows visual representations of the notation patterns for *kethuk* and *kempyang* in each song structure.

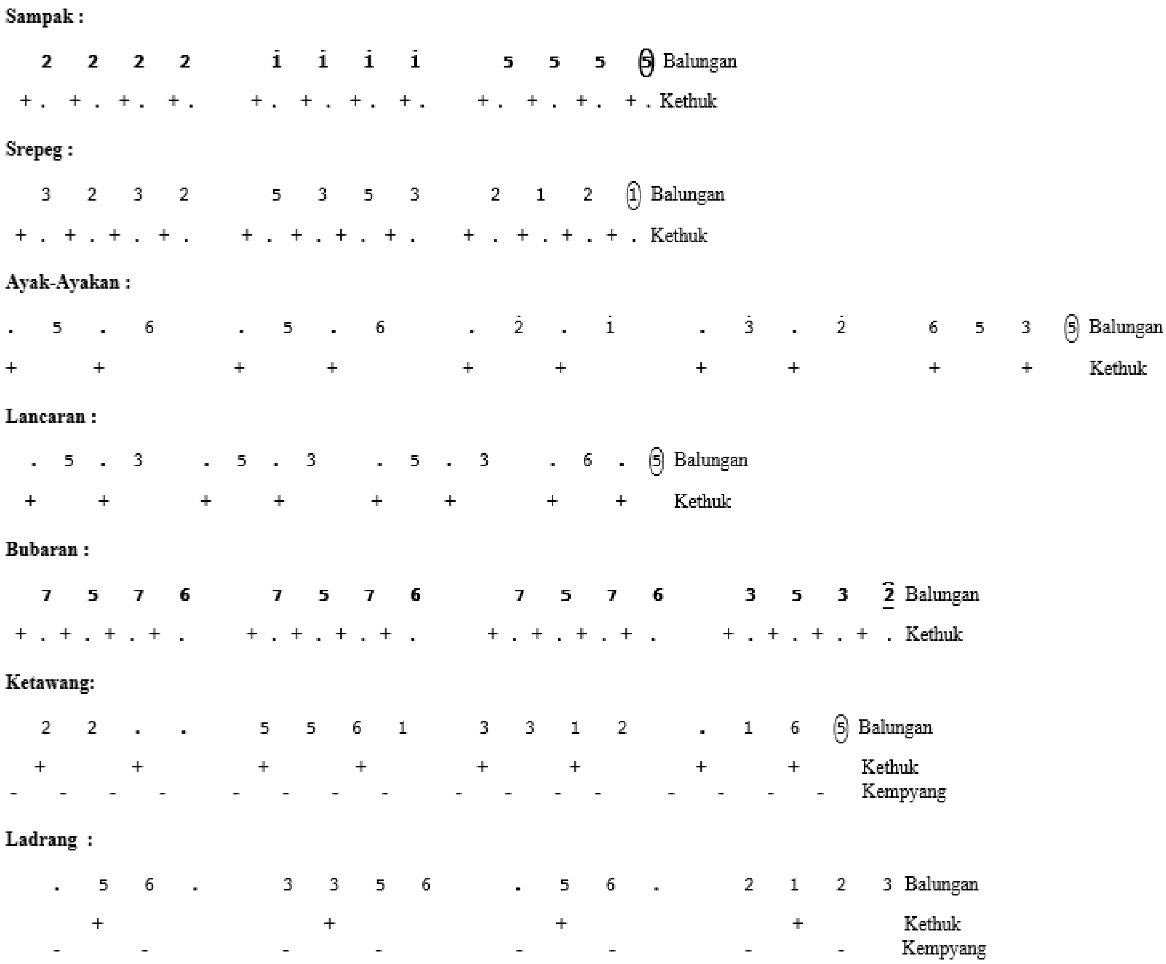


Fig. 8. Pattern of *kethuk* and *kempyang* notation for each song structure

In Table 3, both the *kenong* and *kempul* instruments show variations in note distance. The *kenong* instrument tends to have note distances close to 0 for CNN-LSTM model, indicating a close resemblance between the generated notation and the original. The notation pattern on the *kenong* instrument seems to be more consistent across different song structures compared to the *kempul* instrument. On the other hand, the note distance values for the *kempul* instrument show various variations. A value of 0 means that the generated notation is very close to the original. It should be noted, however, that in the case of *sampak,* there is a tendency for higher note distance values compared to other song structures. This is due to the notation pattern in *sampak*, where the notation for the *kempul* instrument does not always match the *balungan* notation. Such variations in the notation pattern are intentional and are often introduced by *gamelan* composers to add diversity and variation to the music.

Figure 9 and Figure 10 show the output of the notation generators using three models: the CNN-LSTM, LSTM, and CNN methods for multiple instruments in the *ricikan struktural* within the *sampak* and *bubaran* song structures. By observing these figures, we can examine the relationship between the input components, including *balungan* notation, song structure, rhythm, and *gatra* information, and the output notation of multiple instruments in the *ricikan struktural*. The following observations are possible:

- The notation for instruments such as the *kenong*, *kempul*, *gong suwuk,* and *gong ageng* is derived from the *balungan* notation within each *gatra*. However, the order in which the notes are taken is different for each instrument. For example, in *srepeg,* the notes for *kenong* are taken from the 4th tone of each *gatra*, whereas in *ketawang,* the last note of the even *gatra* is chosen.

- Song structure and rhythm determine the notation pattern for all instruments, including *kenong*, *kethuk*, *kempyang*, *kempul*, *gong suwuk*, and *gong ageng*, within each song form.
- *Gatra* information is used to determine the position of the notation for instruments such as *gong suwuk*, *gong ageng*, *kenong,* and *kempul*.

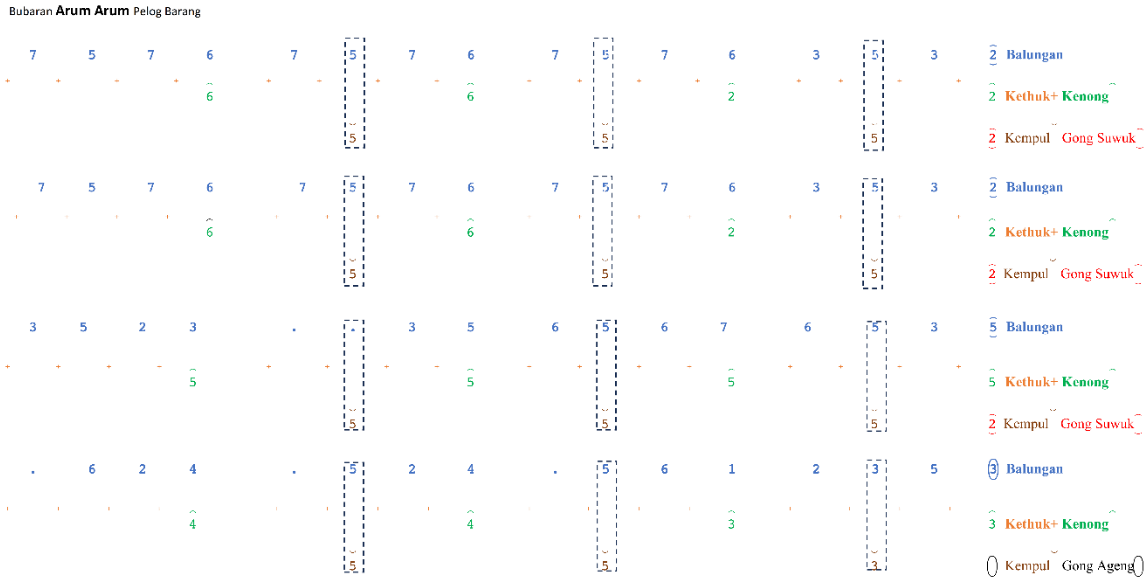


Fig. 9. Notation of *bubaran arum-arum pelog barang*

Figure 9 is the notation of the *Bubaran Arum-Arum Pelog Barang* test data, because the generator results of the three models CNN-LSTM, LSTM, and CNN for all instruments have no difference or are similar to the original notation of the *gamelan* composer, so only the original song notation is shown. If we observed, the notation pattern in Figure 9 for the *Kenong* and *Kempul* instruments has a structure pattern that is consistent with the *Balungan* notation.

However, the situation is different from what is shown in Figure 10, where the generator results of the CNN-LSTM, LSTM, and CNN models do not match the original *Kempul* notation for many *Kempul* notations. The same notation is not shown in Figure 10, while the different notation is highlighted in yellow for the CNN-LSTM model generator results, green for LSTM, and blue for CNN. In the *Sampak Tlutur Slendro Manyura* test data, there are differences in the notations generated on the k*empul* and *gong suwuk* instruments. The notation for the k*empul* and g*ong suwuk* instruments is usually derived from the *balungan* notation, but sometimes the composer substitutes variations of the notation that are different from the *balungan* notation. For example, in the 3rd *gatra* of the first line, the 5th note of the *balungan* becomes the 2nd note of the *kempul.* The generator results of CNN and LSTM are different, while the proposed method CNN-LSTM are the same notation as the original. This is consistent with the results shown in Table 3, where the note distance value for the *Kempul* instrument is smaller compared to the two models of CNN and LSTM for the *Sampak* song structure type.

Based on the results of the music notation generator shown in Table 3, Figure 9, and Figure 10, shows that the CNN-LSTM model can produce a notation generator that is more similar to the original (notation that is the creation of *gamelan* experts). With the ability of CNN in extracting important features from the input fed into the model and supported by the ability of LSTM in predicting music notation from previously learned patterns. However, in Table 3 and Figure 10, there are still some notations that are different from the original, this may still be a rule of *gamelan* notation, especially the *Kempul* instrument, which has not been used as a feature in the proposed comparison model.

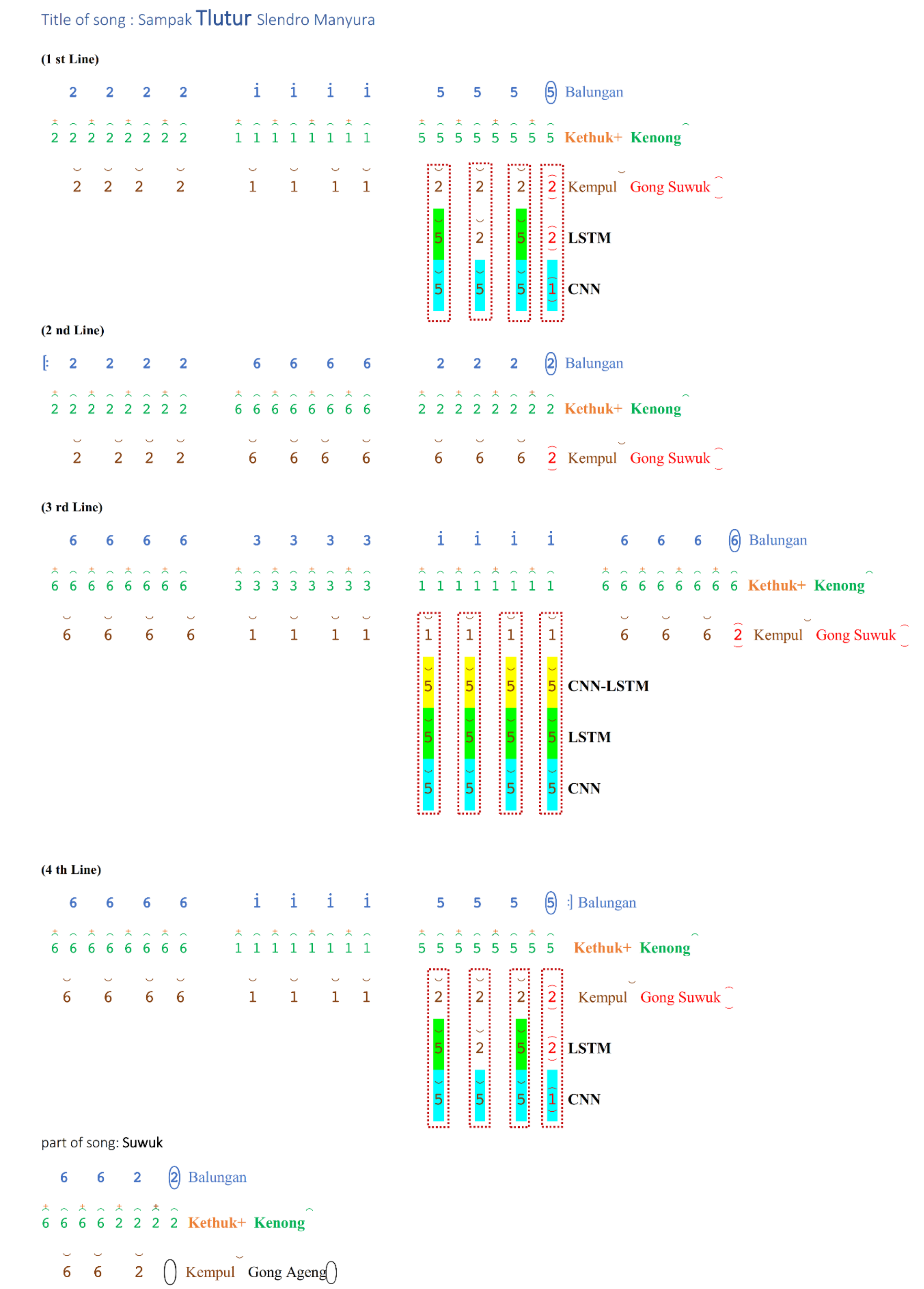


Fig. 10. Notation of *Sampak Tlutur Slendro Manyura*, the colored notation is the result of a generator notation that differs from the original notation of the composer's *gamelan* (yellow section generated by CNN-LSTM, green section generated by LSTM, and blue section generated by CNN).

The results of this study can be useful in the field of education, especially for novice *gamelan* players, in playing r*icikan struktural* instruments, because in *gamelan* songs there is only melody notation. The notation pattern of *ricikan struktural* instruments can be identified by the title of a song in Javanese *gamelan*, because in the title there is a structure of the song that affects the notation pattern of *ricikan struktural* instruments. In addition, this study is also useful in the field of *gamelan* art, with the creation of an automatic generator of *ricikan sruktural* instrument notation, it can be used to compose an automatic musical composition on a Javanese *gamelan* song as an accompaniment to melody notation.

The limitation of this research is that it only generates the notation of *ricikan struktural* instruments, it still needs to be combined with other instrument notations, such as the notation of *ricikan garap* instruments as song decorators and kendang instruments as rhythmic controllers. In order to improve the results more optimally, further investigation is needed, especially in relation to the rules on the *kempul* and *gong suwuk* instruments and its correlation with a song in Javanese *gamelan*, because the results of this study still have some notation patterns that do not match the original, especially for the *kempul* and *gong suwuk* instruments.

IV. Conclusion

This study concludes that CNN-LSTM, LSTM, and CNN models can effectively predict musical note generation for multi-instrument *ricikan struktural* Javanese *gamelan*. Experimental results show that CNN-LSTM outperforms LSTM and CNN in terms of accuracy, recall, precision, and quality of generated notations. This superiority can be attributed to the combination of the strengths of both models, resulting in improved performance.

The more homogeneous data scenario yields higher accuracy scores due to the consistent distribution of the same data, resulting in more consistent pattern generation. Note Distance, which measures the difference between the generator's notations and the composer's *gamelan* notations, shows that the third generator model (CNN-LSTM, LSTM, and CNN) produces similar notations to the original for instruments such as *kethuk*, *kempyang,* and *gong ageng*. However, instruments such as *kenong*, *kempul,* and *gong suwuk* show significant differences.

The small note distance value indicates a consistent notation pattern on the *ricikan struktural* instrument, which follows the *balungan* notation. However, the large note distance value indicates variation of pattern in the r*icikan struktural* instrument, which sometimes does not follow the *balungan* notation. This illustrates that consistency with standardized pattern rules does not always exist in Javanese *gamelan*, but sometimes *gamelan* composers change the notation of these instruments as a variation in playing *gamelan* music.

Although not all notations are exactly the same as the original, this method of music generation can still be used to supplement the notation in Javanese *gamelan* songs based on song characteristics such as the type of song structure, rhythm, melody (*balungan*) notation, and *gatra* information.

This study has benefited for novice *gamelan* players, especially in playing *ricikan struktural*, by creating an automatic *ricikan struktural* instrument notation generator. This can be used to create an automatic musical composition on Javanese *gamelan* songs, complementing the melody notation in *gamelan* songs. The study can also be applied to *gamelan* art.

This study focuses on the *ricikan struktural* generators in Javanese *gamelan*, but also explores the *ricikan garap* and kendang instruments for next study. Future studies should look at the rules of the *kenong* and *gong suwuk* instruments and how they relate to the songs, as there are notation patterns in the study that still differ from the original, especially for the *kempul* and *gong suwuk* instruments. In addition, the wide variety of Javanese *gamelan* styles provides opportunities for further study.

Acknowledgment

We are grateful to Soewidiatmaka *Gamelan* for their invaluable contributions of both knowledge and data to our study, and we express our deep appreciation to them.

Declarations

Author contribution

All authors contributed equally as the main contributor of this paper. All authors read and approved the final paper.

Funding statement

This research receive funding from the Indonesian Endowment Fund for Education (LPDP), which was provided under the BUDI DN Doctoral Scholarship Programme.

Conflict of interest

The authors declare no known conflict of financial interest or personal relationships that could have appeared to influence the work reported in this paper.

Additional information

Reprints and permission information are available at http://journal2.um.ac.id/index.php/keds.

Publisher's Note: Department of Electrical Engineering and Informatics - Universitas Negeri Malang remains neutral with regard to jurisdictional claims and institutional affiliations.

References

[1] Martongprawit, "Catatan Pengetahuan Karawitan," *Surakarta: Akademi Seni Karawitan Indonesia (ASKI)*, 1975.
[2] R Supanggah, "Bothèkan Karawitan I," . *Jakarta: Masyarakat Seni Pertunjukan Indonesia*, 2002.
[3] A. Setyoko and Z. W. Pratama, "Faktor-Faktor Kesulitan Pembelajaran Praktik Karawitan Jawa Program studi Etnomusikologi Fakultas Ilmu Budaya Universitas Mulawarman," *Jurnal Mebang: Kajian Budaya Musik dan Pendidikan Musik*, vol. 1, no. 2, pp. 81–92, 2021.
[4] S. Ananda and N. Scorviana Herminasari, "Minat Generasi Muda Kepada Pelestarian *Gamelan* Jawa Di Komunitas *Gamelan* Muda Samurti Andaru Laras," *Jurnal Studi Budaya Nusantara*, 2022.
[5] D. P. Prasetyo, "Ragam Garap Kendhang Kalih Ladrang Dalam Karawitan Gaya Surakarta," *Skripsi Institut Seni Indonesia Surakarta*, 2016.
[6] V. Melinda, "Garap Tabuhan *Kempul* Pada *Gendhing* Alit Dalam Klenèngan," *Skripsi Fakultas Seni Pertunjukan ISI Yogyakarta*, 2019.
[7] D. Purwanto, "Permainan Ricikan *Kenong* Dalam Karawitan Jawa Gaya Surakarta," *Gelar: Jurnal Seni Budaya*, 11(2), 2013.
[8] Supardi, "*Ricikan struktural* Salah Satu Indikator Pada Pembentukan Gending Dalam Karawitan Jawa," *Keteg*, vol. 13, no. 1, 2013.
[9] J. P. Briot, G. Hadjeres, and F. D. Pachet, "Deep Learning Techniques for Music Generation*," Heidelberg: Springer*, 2020.
[10] R. Madhok, S. Goel, and S. Garg, "SentiMozart: Music Generation based on Emotions," *International Conference on Agents and Artificial Intelligence,* Vol 2, pp. 501-506, 2018.
[11] Yang L. C., Chou S. Y., and Yang Y. H., "MidiNet: A Convolutional Generative Adversarial Network for Symbolic-Domain Music Generation," *arXiv preprint arXiv:1703.10847*, 2017.
[12] Q. Li, W. Cai, X. Wang, Y. Zhou, D. D. Feng, and M. Chen, "Medical Image Classification with Convolutional Neural Network," *International Conference on Control Automation Robotics & Vision (ICARCV)*, pp. 844-848, IEEE, 2014.
[13] S. Tanberk and D. B. Tükel, "Style-specific Turkish Pop Music Composition with CNN and LSTM Network," *World Symposium on Applied Machine Intelligence and Informatics (SAMI),* pp.181-185, IEEE, 2021.
[14] J. Chen, "Construction of Music Intelligent Creation Model Based on Convolutional Neural Network," *Computational Intelligence and Neuroscience*, 2022.
[15] F. Minglei, "Application of Music Industry Based on The Deep Neural Network," *Scientific Programming,* pp.1-6, 2022.
[16] F. Shah, T. Naik and N. Vyas, "LSTM Based Music Generation," *International Conference on Machine Learning and Data Engineering (iCMLDE)*, pp. 48-53, 2019.
[17] S. Mangal, R. Modak, & P. Joshi, "LSTM Based Music Generation System," *International Advanced Research Journal in Science, Engineering and Technology*, Vol. 6, Issue 5, 2019.
[18] Y. Huang, X. Huang, and Q. Cai, "Music Generation Based on Convolution-LSTM," *Computer and Information Science*, 11(3), 50-56, 2018.
[19] S. Liang, B. Zhu, Y. Zhang, S. Cheng, & J. Jin, "A Double Channel CNN-LSTM Model for Text Classification," *IEEE 22nd International Conference on High Performance Computing and Communications; IEEE 18th International Conference on Smart City; IEEE 6th International Conference on Data Science and Systems*, pp. 1316-1321, 2020.
[20] A. Agga, A. Abbou, M. Labbadi, Y. El Houm, & I. H. O. Ali, "CNN-LSTM: An Efficient Hybrid Deep Learning Architecture for Predicting Short-Term Photovoltaic Power Production," *Electric Power Systems Research*, 208, 2022.
[21] T. Liu, J. Bao, J. Wang, Y. Zhang, "A Hybrid CNN-LSTM Algorithm for Online Defect Recognition of CO_2 Welding," *Sensors (Basel),* 18(12):4369, 2018.
[22] W. Tan, J. Zhang, J. Wu, H. Lan, X. Liu, K. Xiao, & P. Guo, "Application of CNN and Long Short-Term Memory Network in Water Quality Predicting," *Intelligent Automation & Soft Computing*, *34*(3), 1943-1958, 2022.
[23] W. Lu, J. Li, Y. Li, A. Sun, & J. Wang, "A CNN-LSTM-Based Model to Forecast Stock Prices," *Complexity*, pp 1-10, 2020.
[24] A. M. Syarif, A. Azhari, S. Suprapto, & K. Hastuti, "Human and Computation-Based Music Representation for *Gamelan* Music," *Malaysian Journal of Music*, 9, 82-100, 2020.
[25] K. Hastuti, & K. Mustafa, "A Method for Automatic *Gamelan* Music Composition," *International Journal of Advances in Intelligent Informatics*, 2(1), 26-37, 2016.

[26] K. Hastuti, A. Azhari, A. Musdholifah, & R. Supanggah, "Rule-Based and Genetic Algorithm for Automatic *Gamelan* Music Composition," *International Review on Modelling and Simulations*, 10(3), pp 202-212, 2017.
[27] A. Kurniawati, E. M. Yuniarno, Y. K. Suprapto, & A. N. I. Soewidiatmaka, "Automatic Note Generator for Javanese *Gamelan* Music Accompaniment using Deep Learning," *International Journal of Advances in Intelligent Informatics*, 9(2), pp 231-248, 2023.
[28] M. Ashraf, F. Abid, M. Atif, and S. Bashir, "The Role of CNN and RNN in the Classification of Audio Music Genres," *VFAST Transactions on Software Engineering*, 2022.
[29] M. Ashraf, F. Abid, I. U. Din, J. Rasheed, M. Yesiltepe, S. F. Yeo, & M. T. Ersoy, "A Hybrid CNN and RNN Variant Model for Music Classification," *Applied Sciences*, 13(3), 1476, 2023.
[30] X. Luo, "Automatic Music Genre Classification based on CNN and LSTM", *Highlights in Science, Engineering and Technology*, pp39, 61-66, 2023.
[31] R. Gupta, S. Ashish, H. Shekhar, and M. D. S. Dominic, "Music Genre Classification Using CNN and RNN-LSTM," *Micro-Electronics and Telecommunication Engineering: Proceedings of 5th ICMETE 2021* (pp. 729-745). Singapore: Springer Nature Singapore, 2021.
[32] D. Kostrzewa, P. Kaminski, & R. Brzeski, "Music Genre Classification: Looking for The Perfect Network," *International Conference on Computational Science*, pp.55-67, Cham: Springer International Publishing, 2021.
[33] R. T. Irene, C. Borrelli, M. Zanoni, M. Buccoli, & A. Sarti, "Automatic Playlist Generation using Convolutional Neural Networks and Recurrent Neural Networks," *27th European Signal Processing Conference (EUSIPCO),* pp. 1-5, 2019.
[34] T. Ito and S. Arai, "Harmonic Representation for CNN-LSTM Automatic Chord Recognition," *3rd International Conference on Cybernetics and Intelligent System (ICORIS)*, pp. 1-5, 2021.
[35] S. B. Puri , S. P. Mahajan, "Automatic Note and Chord Recognition for Harmonium Music: A Deep Learning Approach," *Journal of Critical Reviews,* 7(15), 2020.
[36] S. Hizlisoy, S.Yildirim, & Z. Tufekci, "Music Emotion Recognition using Convolutional Long Short Term Memory Deep Neural Networks," *Engineering Science and Technology*, 24(3), pp760-767, 2021.
[37] S. Sheykhivand, Z. Mousavi, T. Y. Rezaii, & A. Farzamnia, "Recognizing Emotions Evoked by Music using CNN-LSTM Networks on EEG Signals,'' *IEEE Access*, 8, 139332-139345, 2020.
[38] S. Ayadi and Z. Lachiri, "A combined CNN-LSTM Network for Audio Emotion Recognition using Speech and Song attributs," *6th International Conference on Advanced Technologies for Signal and Image Processing (ATSIP)*, pp. 1-6, 2022.
[39] A. Ranjan, V. N. J. Behera, and M. Reza, "Using a Bi-directional LSTM Model with Attention Mechanism trained on MIDI Data for Generating Unique Music," *Artificial Intelligence for Data Science in Theory and Practice*, Nov. 2022.
[40] J. Gareth, W. Daniela, H. Trevor, and T. Robert, J. Gareth, W. Daniela, H. Trevor, & T. Robert, "An Introduction to Statistical Learning: with Applications in R," *Spinger*, 2013.