IJCCS (Indonesian Journal of Computing and Cybernetics Systems)
Vol.16, No.4, October 2022, pp. 379~388
ISSN (print): 1978-1520, ISSN (online): 2460-7258
DOI: https://doi.org/10.22146/ijccs.73910

Traditional Music Regional Classification using Convolutional Neural Network (CNN)

Raymond Luis*[1], Nur Rokhman[2]
[1]Bachelor Program of Computer Science, FMIPA UGM, Yogyakarta, Indonesia
[2] Department of Computer Science and Electronics, Faculty of Mathematics and Natural Sciences, Universitas Gadjah Mada, Yogyakarta, Indonesia
e-mail: ***[1]raymondluis@mail.ugm.ac.id** , [2]nurrokhman@ugm.ac.id

Abstrak

Musik tradisional Indonesia merupakan warisan budaya Indonesia yang sering dilupakan oleh masyarakat modern. Banyak masyarakat yang tidak mengetahui dari daerah mana musik tradisional tersebut berasal. Hal ini menjadi permasalahan karena banyaknya musik tradisional yang kehilangan indentitas. Teknologi Deep Learning dapat menjadi solusi permasalahan klasifikasi musik tradisional ini. Topik pengklasifikasian musik tradisional dipilih karena belum ada penelitian yang menggunakan topik ini sebelumnya.

Penelitian ini akan melakukan klasifikasi musik tradisional berdasarkan daerah asal menggunakan data dari Youtube dengan metode ekstraksi fitur Mel-Frequency Cepstral Coefficients (MFCC) dan model klasifikasi Convolutional Neural Network (CNN). Terdapat 7provinsi yang akan digunakan sebagai label klasifikasi, yaitu Riau, Papua, Daerah Khusus Ibukota Jakarta, Daerah Istimewa Yogyakarta, Sumatera Utara, Jawa Barat, dan Sulawesi Selatan.

Sistem klasifikasi yang dihasilkan dalam penelitian ini menghasilkan akurasi klasifikasi yang baik dengan nilai 74.03%.

Kata kunci*— Musik Tradisional, Audio Classification, Convolutional Neural Network,CNN, Mel-Frequency Cepstral Coefficients*

Abstract

Traditional Indonesian music is an Indonesian cultural heritage that is often forgotten by modern society. Many people do not know which area the traditional music came from. This is a problem because of the large amount of traditional music that loses its identity. Deep Learning technology can be a solution to this traditional music classification problem. The topic of traditional music classification was chosen because there has been no research using this topic before.

This research will classify traditional music based on the area of origin using data fromYoutube with the extraction method of the Mel-Frequency Cepstral Coefficients (MFCC) featureand the Convolutional Neural Network (CNN) classification model. There are 7 provinces that will be used as classification labels, namely Riau, Papua, Special Capital District of Jakarta, Special Region of Yogyakarta , North Sumatra, West Java, and South Sulawesi.

The classification system produced in this study produced good classification accuracy with a value of 74.03%.

Keywords— *Traditional Music, Audio Classification, Convolutional Neural Network, CNN, Mel-Frequency Cepstral Coefficients*

Received April 4th,2022; Revised October 31th, 2022; Accepted October 31th, 2022

1. INTRODUCTION

Music is an art form that consists of vocal sounds and instrument sounds in tandem soas to form beautiful tones. Music has become a part of people's daily lives, both using music asa medium of entertainment and as a livelihood. The majority of people listen to music every dayregardless of their background. The type of music listened to also varies in genres, there are music with pop, rock, jazz, and other genres.

The type of music consumed by Indonesian people varies greatly due to globalization. Globalization causes many outside cultures to be absorbed by the people of Indonesia. This led to the many traditional Indonesian cultures left behind by the younger generation. The abandonment of traditional Indonesian culture resulted in people losing their identity [1]. Traditional Indonesian music is one of the Indonesian cultures left behind by the younger generation.

Music distribution today uses the internet to reach listeners from anywhere and anytime.This causes music to become one of the data that is widely used for research because of its ease of search. Therefore, an automated procedure capable of processing large amounts of music datain digital format is very important, and Music Information Retrieval (MIR) has become one of the important research areas. One of mir's focuses is the issue of Music Genre Classification (AMGC). In general, a music genre is a label created by a musicologist to identify the type of music [2].

Music Information Retrieval (MIR) has a branch that mostly contains audio data analysis, audio data analysis also consists of genre classification, song identification, tone recognition, sound effects detection, atmosphere detection, and audio data feature extraction [3], [4]. These problems are not only experienced by music researchers, many people experience these problems, so there are also applications made as a solution to these problems. One of the most popular is the song search algorithm by Shazam Entertainment, an application that is able to identify songs quickly and efficiently just by listening to audio samples [5].

Most research on the topic of music classification uses the music genre as a classification label. The use of music genres as classification labels is due to the fact that genre is a detail of data that is usually used by music experts to group music into groups. In addition, the music genre is also the most widely used query by the general public in search of music, as shown by several authors [6]–[8]. The popularity of the music genre as a classification label led to a research space for classifying music using other queries. One label that can be used as a classification label is the area of origin.

This research will focus on creating a traditional music classification system based on the area of origin of the music. The study used Mel-Frequency Cepstral Coefficients (MFCC) asa feature extraction method. MFCC became a popular feature extraction technique because the MFCC model was created based on variations in the critical frequency bandwidth of the human ear. The MFCC coefficient is obtained by de-correlating the output of a bank filter consisting of triangular filters, which are linear on the Mel Frequency Scale [9]. MFCC provides a collection of features often used in voice recognition problems [10]. MFCC provides an envelope curve representation of the amplitude spectrum, thus most signal energy is concentrated to the first coefficient [7].

The classification model used in this study is the Convolutional Neural Network (CNN) classification model. Convolutional Neural Network (CNN) is a deep learning algorithm that uses images as input and can find the characteristics of an image and distinguish it from other images. CNN has a wide variety of layers such as Convolutional layers, ReLU layers,

Pooling layers, and Fully-connected layers. CNN is widely used as an image classification modelbecause CNN can perform feature extraction automatically without the help of researchers [11].

The data generated by MFCC will be used as image data, therefore the CNN model that works well on image classification becomes the best choice as a classification model. In this study, a model will be created that uses CNN as a classification model and MFCC as a feature extraction method to classify traditional Indonesian music based on the area of origin of the music.

2. METHODS

The classification model used in the study is the Convolutional Neural Network (CNN). The CNN model will be combined with the Mel-Frequency Cepstral Coefficients (MFCC) feature extraction method. The classification system designed in the study will receive audio data and classify it according to existing classification labels. There are 7 provinces that will be used as classification labels, namely Riau, Papua, Special Capital District of Jakarta, Special Region of Yogyakarta , North Sumatra, West Java, and South Sulawesi. The main flow of the system is shown in Figure 1.

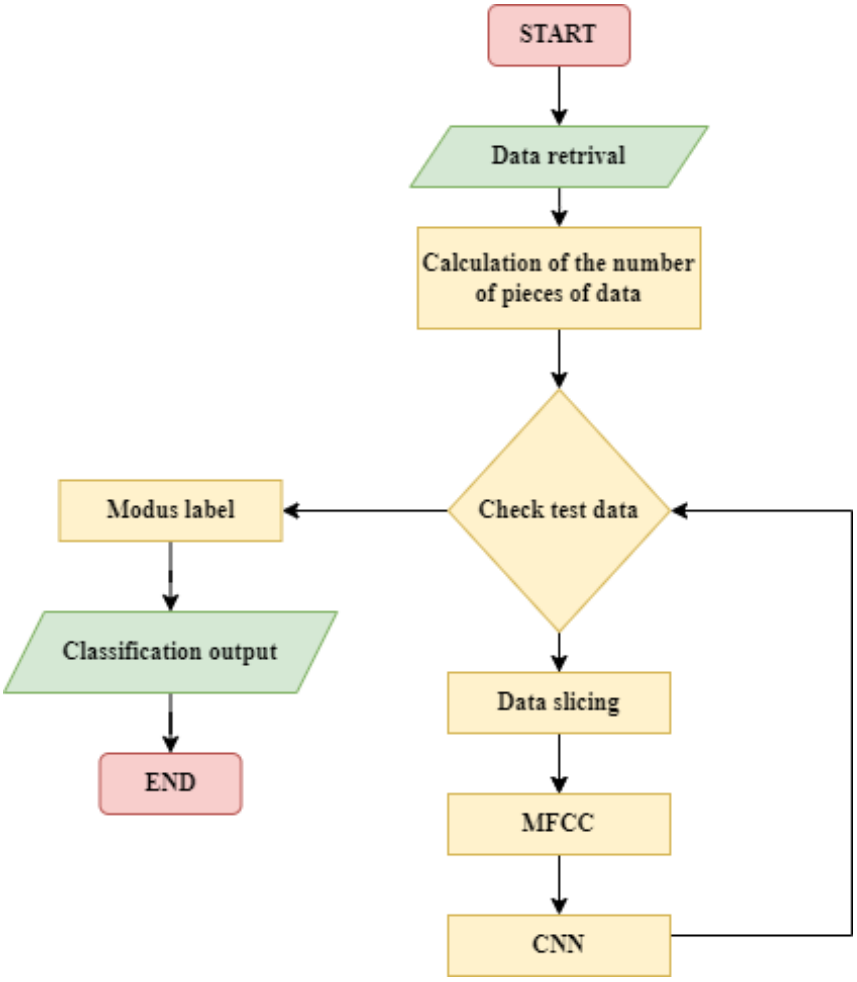

Figure 1. Flow of the system

2.1 Data Collection

The data used in this study is traditional Indonesian music data obtained from youtube. The traditional music data used is traditional music originating from Riau, Papua, Special Capital District of Jakarta, Special Region of Yogyakarta, North Sumatra, West Java, and South Sulawesi. The data that has been collected will be divided into training data, validation data, and test data.

2.2 Data Slicing

The data that has been collected will be broken down into 15-second pieces of music audio data for each file. The number of data pieces for training data is 795 data, the number of data pieces for validation data is 242 data, and the number of data pieces for test data is 362 data.

Table 1. Amount of data

Data set	Amount of data
Training data	795
Test data	362
Validation data	242

2.3 Feature Extraction

In this study, pieces of data that have been obtained will be included in the feature extraction process. The feature extraction method used in this study is Mel-Frequency Cepstral Coefficients (MFCC). The extraction method of the MFCC feature will receive audio data and convert it into a vector containing information about the audio.

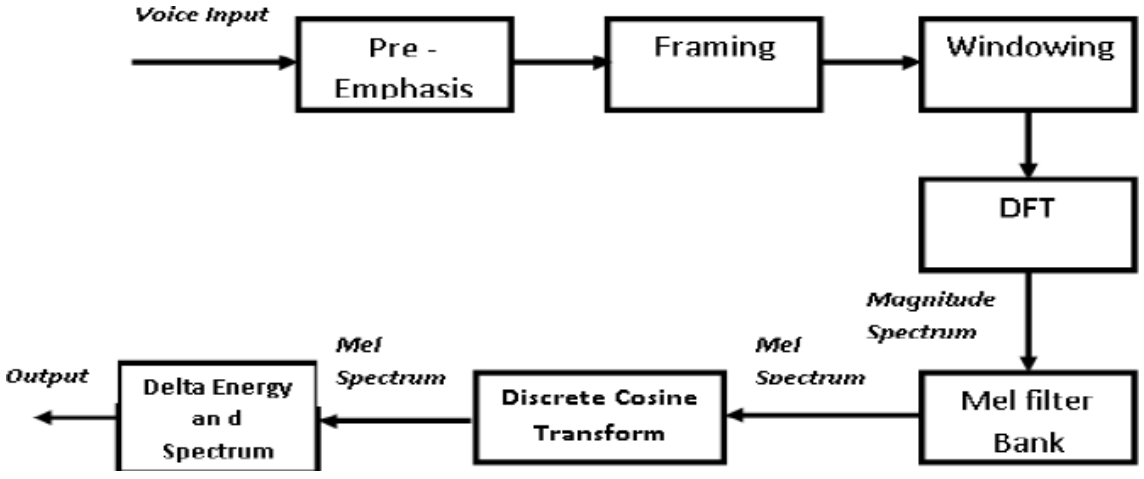

Figure 2. MFCC

2.4 Convolutional Neural Network

The extraction of MFCC features obtained from training data and validation data will be used to train the designed CNN classification model. The CNN model architecture used in this study is 3 layers of convolution, 3 layers of pooling, and 1 layer of fully-connected network. The hyperparameter used by the classification model is obtained by performing the hyperparameter tuning process. The hyperparameter tuning process is done by trying a combination of hyperparameters in order to obtain the combination of hyperparameters with the best results. The combination of hyperparameters tested can be seen in Table 2.

Table 2. Hyperparameter tuning

Hyperparameter	**Value**
Convolution layer filter size 1	Min = 32, Max =128
Kernel size of convolution layer 1	3,5,7
Kernel size of pooling layer 1	2,3,4
Dropout rate 1	Min = 0.1, Max =0.3
Convolution layer filter size 2	Min = 32, Max =128
Kernel size of pooling layer 2	3,5,7
Ukuran kernel lapisan pooling tiga	2,3,4
Dropout rate 2	Min = 0.1, Max =0.3
Convolution layer filter size 3	Min = 32, Max =128
Kernel size of convolution layer 3	3,5,7
Kernel size of pooling layer 3	2,3,4
Dropout rate 3	Min = 0.1, Max =0.3
Fully-connected layer size	Min = 32, Max =128
Learning Rate	Min = 0.001, Max =0.01

The CNN model architecture can be seen in Figure 3 and the hyperparameter configuration used can be seen in Table 3.

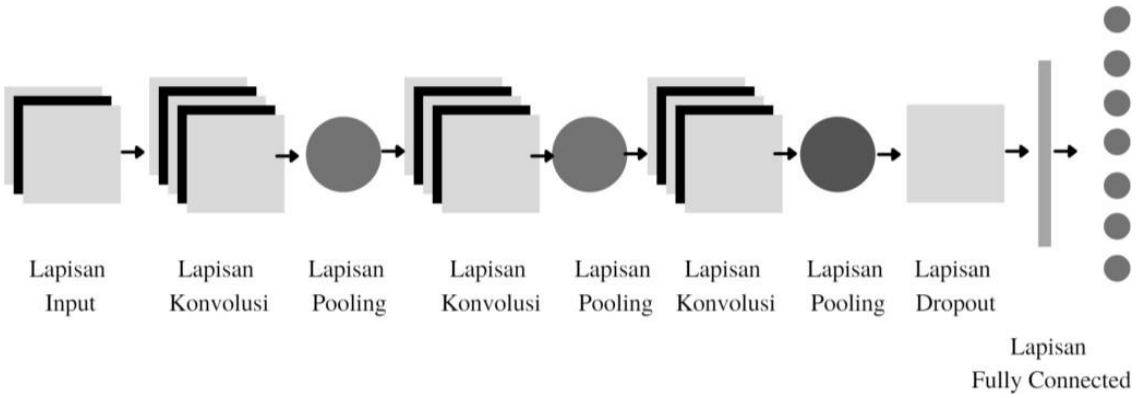

Figure 3. CNN model

Table 3. CNN configuration

Hyperparameter	Best Value
Convolution layer filter size 1	64
Kernel size of convolution layer 1	7
Kernel size of pooling layer 1	2
Dropout rate 1	0.15
Convolution layer filter size 2	64
Kernel size of pooling layer 2	3
Ukuran kernel lapisan pooling tiga	2
Dropout rate 2	0.2
Convolution layer filter size 3	96
Kernel size of convolution layer 3	3
Kernel size of pooling layer 3	2
Dropout rate 3	0.15
Fully-connected layer size	96
Learning Rate	0.001

2.5 Model Prediction

The results of the classification system are obtained by looking for the most predicted labels from existing data pieces. The most prediction labels will be considered as classification labels from user input audio data.

3. RESULTS AND DISCUSSION

3.1 Result

Evaluation of the classification system is carried out by testing test data on the classification system that has been created. In the evaluation process, comparison matrix creation and calculation of classification system evaluation metrics are carried out. The comparison matrix is the comparison matrix of the original label of data with the predictionlabel of the data. The comparison matrix will be used to calculate the evaluation metrics of the classification system. The evaluation metrics used in the study were accuracy, recall, precision, and F1-score.

Classification system testing is carried out with test data, the test data used in the evaluation of the classification system amounts to 242 test data. The test data will be enteredinto the classification system to predict the classification label. The test data prediction label will be used to calculate the performance of the created classification system. The amount oftest data for each region can be diverted in Table 4.

Table 4 Test data

Label	Amount of data
Riau	30
Papua	42
Special Capital District of Jakarta	31
Special Region of Yogyakarta	41
North Sumatra	32
West Java	30
South Sulawesi	36

Test data testing will obtain a comparison matrix. In the comparison matrix it can be seen that most labels are correctly predicted. The comparison matrix of test data shows that there are some labels that tend to be predicted to others. 16 test data from the South Sulawesi region is predicted to come from the Yogyakarta area. In addition to southern Sulawesi, there is also the Yogyakarta area where 15 data is predicted to come from the Jakarta area. The test data comparison matrix can be viewed in Table 5.

Table 5. Comparison matrix

		Prediction						
		Riau	Sumut	Sulsel	Papua	Jakarta	Jabar	Yogya
Real Label	Riau	40	9	0	0	3	2	0
	Sumut	8	52	0	0	0	0	0
	Sulsel	0	0	36	1	2	0	16
	Papua	5	0	0	32	0	2	1
	Jakarta	0	0	0	1	32	11	1
	Jabar	0	8	0	0	1	32	1
	Yogya	0	5	0	0	15	2	44

From the existing comparison matrix, it can be calculated the probability rate of certain labels to be predicted as other labels. All labels have the highest percentage to predict as actual labels. In addition to the original label, the South Sulawesi area has the highest percentage to predict as Yogyakarta with a percentage of 29.1%. Conversely, no other area is predicted as South Sulawesi than South Sulawesi. This shows that the South Sulawesi area has its own uniqueness in traditional music, as other regions are not predicted as South Sulawesi.

Table 6. Possibilities table

		Prediction						
		Riau	Sumut	Sulsel	Papua	Jakarta	Jabar	Yogya
Real Label	Riau	74.1	16.7	0	0	5.5	3.7	0
	Sumut	13.3	86.7	0	0	0	0	0
	Sulsel	0	0	65.5	1.8	3.6	0	29.1
	Papua	12.5	0	0	80	0	5	2.5
	Jakarta	0	0	0	2.2	71.1	24.5	2.2
	Jabar	0	19	0	0	2.4	76.2	2.4
	Yogya	0	7.6	0	0	22.7	3	66.7

The comparison matrix will be used to calculate the evaluation metrics of the existing classification system. The metrics used in this study were the total accuracy of test data, the recall of each classification label, the precision of each classification label, and the F1-score of each classification label. The total accuracy of the classification system to the test data is 74.03%. Other metric results can be seen in Table 7.

Table 7. Classification system performance

	Recall	**Presisi**	**F1-Score**
Riau	74.07%	75.47%	74.77%
Sumatera Utara	**86.67%**	70.27%	77.61%
Sulawesi Selatan	65.45%	**100%**	79.12%
Papua	80%	94.12%	**86.49%**
Jakarta	71.11%	60.38%	65.31%
Jawa Barat	76.19%	65.31%	70.33%
Yogyakarta	66.66%	69.84%	68.22%
Rata-rata	74.31%	76.48%	74.55%

The recall value for each classification label is obtained by dividing the number of test data of each label predicted to be correct with the amount of data with the label Based on these calculations, the highest recall value on the North Sumatra label with a recall of 86.67% and the lowest recall value on the South Sulawesi label with a recall of 65.45%. The average recall value of all test data is 74.31%. The recall average shows that the classification system works quite well in classifying specific area data to actual areas. From the existing classification labels, there are 2 classification labels that have recall values far from the average such as South Sulawesi and Yogyakarta. Low recall values are due to the similarity of certain area data with other regions so that there is an error of classification.

The precision value for each classification label is obtained by dividing the amount of test data of each label predicted correctly by the amount of data predicted to have that label. Based on these calculations, the highest precision value was obtained on the South Sulawesi label with 100% precision and the lowest precision value on the Jakarta label with a precision

of60.38%. The average precision value of all test data is 76.48%. The average value of precision indicates that the classification system works well in determining a particular area. However, there is a long range between the highest precision value and the lowest precision value, this shows that the south Sulawesi regional data has the highest uniqueness of other regions.

The F1-Score value for each classification label is obtained by the formula of each classification label. Based on these calculations, the highest F1-Score score was obtained on the Papua label with an F1-Score of 86.49% and the lowest F1-Score value on the Jakarta label with an F1-Score of 65.31%. The average F1-score score is 74.55%, this average value indicates the classification system is working well. From the results of this study, it can be concluded that the classification system designed works well in classifying the area of origin of traditional music.

4. CONCLUSIONS

In this study it can be concluded that the creation of a traditional music classification system based on the region of origin has been successfully carried out. The traditional music classification system produced in the study worked well. The combination of the Convolutional Neural Network (CNN) classification model and the Mel Frequency Cepstrum Coefficients (MFCC) works well in the case of traditional music classification. The classification system produced an accuracy of 74.03%.

REFERENCES

[1] Rahmawati, "Makalah Pengaruh Globalisasi Terhadap Kebudayaan," 2016, [Online]. Available: https://www.academia.edu/28879985/MAKALAH_PENGARUH_GLOBALISASI_TER HADAP_KEBUDAYAAN.

[2] C. N. Silla Jr., A. L. Koerich, and C. A. A. Kaestner, "A machine learning approach toautomatic music genre classification," *J. Brazilian Comput. Soc.*, vol. 14, no. 3, pp. 7–18, 2008, doi: 10.1590/s0104-65002008000300002.

[3] B. McFee *et al.*, "librosa: Audio and Music Signal Analysis in Python," *Proc. 14th Python Sci. Conf.*, no. August 2020, pp. 18–24, 2015, doi: 10.25080/majora-7b98e3ed-003.

[4] K. Choi, G. Fazekas, and M. Sandler, "Explaining Deep Convolutional Neural Networks on Music Classification," 2016, [Online]. Available: http://arxiv.org/abs/1607.02444.

[5] A. L.-C. Wang, "An Industrial Strength Audio Search Algorithm.," *Proc. 4th Int. Soc.Music Inf. Retr. Conf. (ISMIR 203), Balt. Maryl. (USA), 26-30 Oct. 2003*, no. January 2003, pp. 7–13, 2003, doi: 10.1109/IITAW.2009.110.

[6] J. S. Downie and S. J. Cunningham, "Toward a theory of music information retrieval queries: System design implications," *Proc. 3rd Int. Soc. Music Inf. Retr. Conf.*, pp. 13–17, 2002.

[7] G. Tzanetakis and P. Cook, "Musical genre classification of audio signals," *IEEE Trans. Speech Audio Process.*, vol. 10, no. 5, pp. 293–302, 2002, doi: 10.1109/TSA.2002.800560.

[8] J. H. Lee and J. S. Downie, "Survey of Music Information Needs, Uses, and Seeking Behaviours: Preliminary Findings," *Ismir 2004*, pp. 441–446, 2004, [Online]. Available: http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.98.8649%7B&%7Damp;rep=r ep1%7B&%7Damp;type=pdf.

[9] A. Hossan, S. Memon, and M. A. Gregory, "A Novel Approach for MFCC FeatureExtraction," no. July 2014, 2011, doi: 10.1109/ICSPCS.2010.5709752.
[10] S. Davis and P. Mermelstein, "Experiments in syllable-based recognition of continuous speech," *IEEE Trans. Speech Audio Process.*, vol. 28, no. May 1980, pp. 357–366, 1980,doi: 10.1109/ICASSP.1980.1170934.
[11] A. P. Viswanathan, "Music Genre Classification," *Int. J. Eng. Comput. Sci.*, vol. 7, no. 1,pp. 8–13, 2016, doi: 10.18535/ijecs/v4i10.38.