International Research Journal of Advanced Engineering and Science
ISSN (Online): 2455-9024

Testing the Accuracy of Chords in Piano Music Using the Convolutional Neural Network (CNN) Method

Muhammad Kamal Yasya[1], Onny Marleen[2]
[1, 2]Business Information System, Gunadarma University, Depok, West Java, Indonesia-16424
Email Address: [1]kamalyasya@gmail.com, [2]onny_marleen@staff.gunadarma.ac.id

Abstract**—***The implementation of artificial intelligence in the world of discussion has been classified as many, but many of these implementations are still less accurate in terms of accuracy such as Chord Recognition. Chord Recognition is a system or application for knowing or measuring chords correctly. To build the system, methods are needed so that the accuracy generated is high. Some methods that are often used are Convolutional Neural Network (CNN) methods.*

This study aims to measure the accuracy of chord recognition from piano and bass audio by using the Convolutional Neural Network (CNN) method and utilizing chromagrams in the transformation of audio to image form. The data used in this study consisted of 12 major classes and 12 minor classes. Based on the results of tests conducted, the accuracy of each class gives a pretty good F1 value. The average F1 value of the whole chord recognition test based on the root tone is 87%.

Keywords**—** *Convolutional Neural Network (CNN), Chromagram, Chord Recognition.*

I. INTRODUCTION

The implementation of artificial intelligence in the discussion world has been classified as many, but there are still many shortcomings in terms of accuracy as in the case of Chord Recognition. Chord Recognition is a system or application for knowing or measuring chords correctly. To build the system, methods are needed so that the accuracy generated is high. Some methods that are often used are Convolutional Neural Network (CNN) method.

Convolutional Neural Network (CNN) is one of the Machine Learning methods of developing Multi Layer Perceptron (MLP) which is designed to process two-dimensional data. CNN is said to be a further development of MLP because CNN uses a method similar to MLP, but uses more dimensions. CNN is also often used as a powerful feature extractor of data that has been put together, such as images. This method can be extended to various audio signal classification tasks by representing the input signal in the time frequency domain.

This research will measure the accuracy of audio chord recognition from piano and bass instruments. The method used is the Convolutional Neural Network (CNN) method and to change the form of audio into an image will be used chromagram. Chords to be used in this study consisted of 12 major classes and 12 minor classes.

II. THEORY REVIEW AND REFERENCES

A. Audio

Audio is a time series, where the y axis is the amplitude of the current corresponding to a loudspeaker membrane and the x axis is the axis that corresponds to the time unit of the data (Ingo Miersawa & Katharina Morik, 2005). The human ear can only hear sounds with a frequency range between 20 Hz to 20 KHz (20,000Hz). The 20 Hz number is the lowest sound frequency that can be heard, while 20 KHz is the highest frequency that can be heard. Sound waves contain a number of important components, such as amplitude, wavelength, and frequency. These components are able to make one sound different from the others.

B. Akor

Chords are a series of basic notes that are arranged regularly from a scale and can represent these scales (Tito Galit Permana, Dr. Ir. Bambang Hidayat, and Eko Susatio S.T, M.T., 2014). Three-note chords that are three scales apart on the root until the third note are called triads. To facilitate the musician in determining whether a combination of three particular tones is a triad or not, the musician can form a circle of thirds diagram. Scale or tone scale is a distance that indicates the location of a note with a certain starting point. There are four types of triads that are distinguished based on the tone scale, including :

1) Major Triads: Major triads are three notes that sound with a 1-2-1.5 tone scale. The major chords show the identity of a part of the song closer to the root because there is one major chord that occupies Scale I.

2) Minor Triads: Minor triads are three notes that sound with a scale 1-1,5-2.

3) Diminished Triad: Diminished triad is a three-tone sound with a scale 1-1,5-1,5.

4) Augmented Triad: Augmented Triad is a three-tone sound with a scale 1-1,5-2,5.

Muhammad Kamal Yasya, Onny Marleen, "Testing the Accuracy of Chords in Piano Music Using the Convolutional Neural Network (CNN) Method", *International Research Journal of Advanced Engineering and Science*, Volume 1, Issue 1, pp. xx-xx, 2020.

C. Beat Tracking

Beat tracking is used to determine the sample time in an audio recording, where human listeners tend to knock their feet to music (Tavish Srivastava, 2019). Beat tracking the design using the dynamic programming method. Dynamic programming or dynamic optimization is not a mathematical formula that can provide answers by just providing input. Conversely, dynamic programming is a combination of structured thinking and analytical thinking to solve a problem. To do beat tracking, the thing to do first is look for signals in the form of onset. Onset is a location where the signal is at a high value suddenly.

D. Chromagram

Chromagram is the achievement of a tone in a certain duration (Eka Angga Laksana & Feri Sulianta, 2017). Chromagram is the transformation of the signal's time-frequency property into a pitch precursor that changes temporarily. This transformation is based on observing perceptions about the auditory system and has been shown to have some interesting mathematical properties. Chromagrams expand the concept of chroma to include time dimensions. Just as we use a spectrogram to infer properties about the distribution of signal energy from frequency and time, a chromagram can be used to infer properties about the distribution of signal energy over chroma and time.

E. Windowing

Windowing functions to minimize the signal that is not continuous at the beginning and end of each frame (Marwa A. Nasr, 2018). Filter length values are obtained using the formula below. This long value will be used to make impulse responses in simulation and implementation. The filter length value will also affect the roll-off factor or slope factor of a filter.

F. Convolutional Neural Network (CNN)

Convolutional Neural Network (CNN) is a type of neural network commonly used in image data (Samuel Sena, 2019). CNN can be used to detect and recognize objects in an image. CNN consists of neurons that have weight, bias, and activation functions. The way CNN works is similar to MLP, but in CNN each neuron is presented in two dimensions, unlike MLP where each neuron is only one dimensional in size. On CNN, the data propagated by the network are two-dimensional data, so the linear operation and weight parameters on CNN are different. In CNN linear operations use convolution operations, whereas weights are no longer just one dimension, but are in the form of four dimensions which are a collection of convolution kernels. Due to the nature of the convolution process, CNN can only be used on data that has a two-dimensional structure such as image and sound.

G. Music Information Retrieval

Music Information Retrieval (MIR) mainly deals with the extraction and inference of meaningful features of music such as audio signals, symbolic representations or external sources such as web pages. The following are examples of MIR tasks, including fingerprinting, cover song detection, genre recognition, key detection, beat tracking, symbolic melodic similarity, source separation, pitch tracking, tempo estimation, and many more. MIR is used in the system to prepare raw data into a form of data that can be observed by the model. The following are some of the features of the MIR that are used by the system, namely Beat Tracking and Harmonic Separation. Beat tracking feature provided by librosa library which produces output in list form. The value generated is in units of seconds. Harmonic separation feature functions to separate input data in the form of harmonic and percussive. Examples of harmonics are notes that are sounded by the piano and vocal sounds, whereas percussive is a collection of sound data that is played by being hit, swiped, or shaken.

III. RESEARCH METHOD

The framework of thought in this study has the final result of knowing the accuracy of the CNN method in recognizing a chord by using a Chromagram. The first step that will be taken is to accept an audio file input that has the format. Wav. Next will be beat tracking. In the beat tracking process, the uploaded audio file will be cut every second and the waveform will be analyzed to form an onset. After getting the onset of the beat tracking process, then the onset will be changed to chromagram form. In the chromagram process, the onset that has been obtained will be formed into the form of images. The picture will be trained using the CNN method to find the appropriate chord. After finding the appropriate chord, then do the directory setup process. This process is needed so that the placement of the parent and child packages backend and frontend parts can communicate with each other. And finally there is the backend approach process. The backend approach is needed to make it easier to access the functions that exist in the Python script.

IV. DISCUSSION

The results of the research framework are measuring the accuracy of chord data generated in Chord Recognition using the CNN method, then drawing conclusions from the results that have been obtained. This chapter presents several things related to the process, results, and discussion of research results, including, the results of testing the model predictions using the Confusion Matrix, and the results of chord accuracy testing and their discussion.

A. Model Prediction Testing Using Confusion Matrix

Table-shaped confusion matrix that is often used to provide a description of the performance of the classification model for a collection of test data. In the system application there are 1192 training data and 247 validation data with 24 classes. There are three variables to calculate the accuracy of the classification made by the model. The three variables are precision, recall, and F1-score. To get the values of these three

Muhammad Kamal Yasya, Onny Marleen, "Testing the Accuracy of Chords in Piano Music Using the Convolutional Neural Network (CNN) Method", *International Research Journal of Advanced Engineering and Science*, Volume 1, Issue 1, pp. xx-xx, 2020.

variables, a confusion matrix is needed which has the following values:[1]

1. *True Negative* (TN) : TN is the number of true predictions for wrong data.
2. *True Positive* (TP) : TTP is the number of correct predictions for true data.
3. *False Negative* (FN) : FN is the number of wrong predictions for incorrect data.
4. *False Positive* (FP) : FP is the number of wrong predictions for true data

B. Model Testing Results

The results of the confusion matrix in the test can be seen in Table 1. Based on the value of the confusion matrix, the values of precision, recall, and F1-score can be seen in Table 2. The results of the test, the accuracy value obtained by 89% of the 72 data tested . In Table 2 it can be seen that the prediction of chord differences can be easily done, while the prediction of chord types is more difficult to distinguish. For example for A # m chord class there is a prediction error for A # chord class. There is also a prediction error in the G # chord class. The model predicts the G # chord class being the G chord class and the G # m chord class. This is because the major and minor chords only have a second pitch difference of half a step. Chords A # (A # major) consist of notes A #, D, and F, while chords A # m (A # minor) consist of notes A #, C #, and F. Not too significant difference in tone causes the system to difficult to distinguish minor and major chords.

As one of the results of chord testing, for C Major chords there are 30 visualizations obtained from C chord input song data with an F1 value obtained by 95 percent. There are several other chord classes predicted by the model namely, C minor chord classes and F major chord classes. This error is because the C major chord and the C minor chord have only one different note. The C major chord consists of C, E, and G. Tones while the C minor chord consists of C, D #, and G. Tones The frequency of E tones is 329.63 Hz while the frequency of the D # tones is 311.13 Hz. The two frequencies are close together so the model incorrectly recognizes the C major chord as a C minor chord twice. Details of the test results on the major C chord classes can be seen in Table 3.

Table 1. Confusion Matrix Results

	A	A#	A#m	Am	B	Bm	C	C#	C#m	Cm	D	D#m	Dm	E	Em	F	F#	F#m	Fm	G	G#	G#m	Gm
A	3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
A#	0	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
A#m	0	1	3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Am	0	0	0	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
B	0	0	0	0	6	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Bm	0	0	0	0	0	3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
C	0	0	0	0	0	0	3	0	0	1	0	0	0	1	0	0	0	0	0	0	2	0	0
C#	0	0	0	0	0	0	0	3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
C#m	0	0	0	0	0	0	0	0	3	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Cm	0	0	0	0	0	0	0	0	0	2	0	0	0	0	0	0	0	0	0	0	0	0	0
D	0	0	0	0	0	0	0	0	0	0	3	0	0	0	0	0	0	0	0	0	0	0	0
D#m	0	0	0	0	0	0	0	0	0	0	0	3	0	0	0	0	0	0	0	0	0	0	0
Dm	0	0	0	0	0	0	0	0	0	0	0	0	3	0	0	0	0	0	0	0	0	0	0
E	0	0	0	0	0	0	0	0	0	0	0	0	0	2	0	0	0	0	0	0	0	0	0
Em	0	0	0	0	0	0	0	0	0	0	0	0	0	0	3	0	0	0	0	0	0	0	0
F	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	3	0	0	1	0	0	0	0
F#	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	3	0	0	0	0	0	0
F#m	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	3	0	0	0	0	0
Fm	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2	0	0	0	0
G	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	3	0	0	0
G#	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	2	0
G#m	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	3	0
Gm	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	3

Table 2. Calculation of Precision, Recall, F1-score, and Support

	precision	recall	f1-score	support
A	1.00	1.00	1.00	3
A#	0.67	1.00	0.80	2
A#m	1.00	0.75	0.86	4
Am	0.67	1.00	0.80	2
B	1.00	1.00	1.00	6
Bm	1.00	1.00	1.00	3
C	1.00	0.43	0.60	7
C#	1.00	1.00	1.00	3
C#m	1.00	1.00	1.00	3
Cm	0.67	1.00	0.80	2
D	1.00	1.00	1.00	3
D#m	1.00	1.00	1.00	3
Dm	1.00	1.00	1.00	3
E	0.67	1.00	0.80	2
Em	1.00	1.00	1.00	3
F	1.00	0.60	0.75	5
F#	1.00	1.00	1.00	3
F#m	1.00	1.00	1.00	3
Fm	0.67	1.00	0.80	2
G	1.00	1.00	1.00	3
G#	0.00	0.00	0.00	0
G#m	1.00	0.75	0.86	4
Gm	1.00	1.00	1.00	3
accuracy			0.89	72

Table 3. Test Results on Major C Chords

	Precision	Recall	F1-Score	Support
C	0,90	1,00	0,95	27
Cm	0,00	0,00	0,00	2
F	0,00	0,00	0,00	1
Macro Avg	0,30	0,33	0,32	-
Weighted Avg	0,82	0,90	0,85	-

[1] Muthu, Understanding the Classification Report through Sklearn, https://muthu.co/understanding-the-classification-report-in-sklearn/, 15 Desember 2019.

For major C chord training data can be seen in Figure 1 and for C major validation data can be seen in Figure 2. Can be seen in Figure 1 there are bright lines, namely the chord G and C chord lines. After going through the CNN process which acts as a validator, it was found that the chord from the training data was chord C (Figure 2).

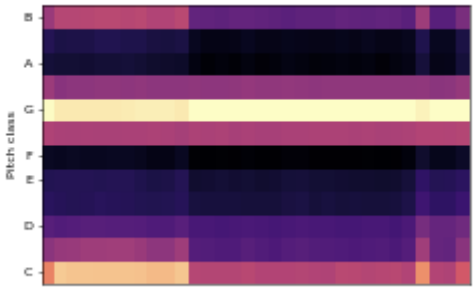

Figure 1. Chord C Major Training Data

Figure 2. Major C Chord Validation Data

C. Detailed Test Results

From the accuracy testing that has been done, it can be concluded that the accuracy testing in each class gives a pretty good F1 value. The value of F1 in each class can be seen in Table 4.

Table 4. Detailed F1 Value Overall Testing for Chord Recognition Based on Root

Nomor	Kelas Akor	Nilai F1
1	C	0,95
2	Cm	0,77
3	C#	0,91
4	C#m	0,80
5	D	0,83
6	Dm	0,75
7	D#	0,81
8	D#m	0,82
9	E	0,89
10	Em	0,84
11	F	0,99
12	Fm	0,85
13	F#	0,84
14	F#m	1,00
15	G	0,83
16	Gm	0,98
17	G#	0,92
18	G#m	0,86
19	A	0,98
20	Am	0,80
21	A#	0,82
22	A#m	1,00
23	B	0,86
24	Bm	0,86
Rata-rata		0,87

V. CONCLUSIONS AND SUGGESTIONS

A. Conclusions

From the accuracy testing that has been done, it can be concluded that the accuracy testing in each class gives a pretty good F1 value. The largest F1 value was obtained in the F # minor and A # minor chord testing that is equal to 100 percent. The smallest F1 value obtained in the D minor chord test is equal to 75 percent. The average F1 value of the overall chord recognition test based on the root tone was 87 percent.

B. Suggestions

Suggestions that can be given to expand this research are as follows:

1. Examine chord recognition from the sounds of musical instruments other than piano and bass.
2. Conduct research using the KNN method and compare the results with the CNN method.

REFERENCES

[1] Abhirawa, H., Jondri, and Arifianto, A., 2017, Face Recognition Using Convolutional Neural Network, Journal E-Proceeding of Engineering, Vol.4, No.3, Pg. 4907 – 4916.

[2] Clarissa, H.J., Hidayat, B., Suhardjo. 2018, Cyst Detection Image Processing Through Periapical Radiography Using Local Binary Pattern Method and Learning Vector Quantization, Journal E-Proceeding of Engineering, Vol.5, No.2, Pg. 2169 – 2177.

[3] Collins, Nick, 2011, SCMIR: A Supercollider Music Information Retrieval Library, Brighton : University of Sussex.

[4] Falah, Adnan Hassal, and Jondri, 2019, Normal and Abnormal Lung Sound Classification Using Deep Neural Network and Support Vector Machine, Journal E-Proceeding of Engineering, Vol.6, No.1, Pg. 2451 – 2459.

[5] Fatihah, N., Karna, N., and Patmasari, R, 2019, Evaluation Of Dlx Microprocessor Instructions Efficiency For Image Compression. Journal E-Proceeding of Engineering, Vol.6, No.1, Pg. 720 – 725.
[6] Gumilar , T., Suwandi, and Bethaningtyas, H., 2015, Detection of Tone Error in Guitar Strings Using Harmonic Product Spectrum, Journal E-Proceeding of Engineering, Vol.2, No.2, Pg. 3276 – 3283.
[7] Hakim, D.M., and Rainarli, E., 2019, Convolutional Neural Network for Introduction to Music Notation Imagery, Techno.COM, Vol. 18, No. 3, Pg. 214-226.
[8] Ibrahim, H.S., Jondri, and Wlsesty, U.N., 2018, Deep Learning Analysis To Recognize Complex Qrs on Ecg Signals with CNN Method. Journal E-Proceeding of Engineering, Vol.5, No.2, Pg. 3718 – 3725.
[9] Krebs, F., Bock, S., dan Widmer, G., 2013, Rhythmic Pattern Modeling For Beat And Downbeat Tracking In Musical Audio, Austria : Johannes Kepler University.
[10] Kumalasari, D.O., Novamizanti, L., and Ramatryana, I N.A., 2019, Determination of Chorus Location in Mp3 Music Using 2-D Correlation Coefficient on the Frame Based on Mel-Frequency Cepstral Coeffisient (Mfcc), Journal E-Proceeding of Engineering, Vol.6, No.1, Pg. 910 - 918.
[11] Laksana, E.A., and Sulianta, F., 2017, *Analysis and Comparative Study of Music Genre Classification Algorithms*, STMIK AMIKOM Yogyakarta : was published.
[12] Lionel, D., Adipranata, R., and Setyati, E., 2019, *Music Genre Classification Using Deep Learning Convolutional Neural Network and Mel-Spectrogram Method*, Universitas Kristen Petra : was published.
[13] Maharani, M.P., Hidayat, B., and Suhardjo. 2018, Comparison of Pulpitis Detection Through Periapical Radiographic Images with the Extraction of Watershed and Gray Level Co-Occurrence Matrix (GLCM) Classifications with K-Nearest Neighbor (K-NN), Journal E-Proceeding of Engineering, Vol.5, No.3, Pg. 5414 – 5421.
[14] Mahmud, K.H., Adiwijaya, and Al Faraby, S., 2019, *Multi-Class Image Classification Using Convolutional Neural Networks. Journal of E-Proceeding of Engineering*, Vol.6, No.1, Pg. 2127 – 2136.
[15] Parlys, A., Zahra, A.A., and Hidayatno, A., 2018, *Song Classification Based on Spectogram with Convolution Neural Network. Transient*, Vol. 7, No. 1, Pg. 28-33.
[16] Permana, T.G., Hidayat, B., and Susatio, E., 2014, *Guitar Chord Identification Using the Harmonic Product Spectrum Algorithm. Journal of E-Proceeding of Engineering*, Vol.1, No.1, Pg. 162 – 170.
[17] Putra, I.W.S.E., 2016, Image Classification Using Convolutional Neural Network (CNN) on Caltech 101, Surabaya : Institut Teknologi Sepuluh Nopember.
[18] Waskito, T.B., Sony, S.T., and Setianingsih, C., 2019, Wheeled Robot Control With Hand gestures Based on Image Processing, Journal E-Proceeding of Engineering, Vol.6, No.3, Pg. 10052 – 10059.
[19] Zulfikar, M.M.M., 2015, Introduction and Representation of Music Chords Symbols Using the Hidden Markov Model With Doubly Nested Circle Of Fifth Approach, Telkom University : was published.