Journal of Political Science and Governance
Volume 10 Issue 1, January-March 2022
ISSN: 2995-4193
Impact Factor: 6.41
https://kloverjournals.org/journals/index.php/psg

ANALYZING AUDIO AESTHETICS: AN EMPIRICAL EXAMINATION OF MUSIC GENRE CLASSIFICATION USING MULTILAYER PERCEPTRON AND FEATURE EXTRACTION

Dr. Ahmad Yusril Abdul Gani
Faculty of Computer Science and Information Technology, Gunadarma University, Jakarta, Indonesia

Abstract
The exponential growth of music databases has led to the challenge of manual music categorization, making it difficult to search for specific music genres in vast collections. Digital music development, particularly in genre classification, has facilitated the study and retrieval of songs. Consequently, there is a need for a convenient and efficient genre classification method that optimizes the learning process and ensures accurate results. This study explores the comparison between two music genre classification approaches: one using the Multilayer Perceptron (MLP) model with Chroma Feature extraction, and the other with Mel Frequency Cepstral Coefficients (MFCC) extraction. The dataset utilized in this research comprises audio data from songs, drawn from the GTZAN music dataset available through http://opihi.cs.uvic.ca/.
Machine Learning, a branch of computer science, provides the theoretical foundation for automating the classification process. In particular, Deep Learning (DL), a subset of Machine Learning, is employed to process inexact data such as language, sound, or images. By applying Artificial Intelligence (AI), computers can learn from patterns and store acquired knowledge. The study focuses on MLP, a machine learning implementation method, to build the classification models.
The primary objective is to determine which extraction features yield better accuracy in classifying song genres. Both Chroma Feature and MFCC extraction features are evaluated, and the classification results obtained from the MLP models are compared. The dataset consists of 1000 sample songs, encompassing ten distinct music genres, each with 100 songs in WAV format. The dataset is further divided into three sections of equal duration (10 seconds), resulting in a 3000-sample dataset comprising 300 songs for each genre.
To evaluate the models, 80% of the dataset (2400 songs) is used for training, and the remaining 20% (600 songs) is reserved for testing. The study demonstrates the potential of AI and deep learning techniques to effectively classify music genres, enabling more efficient music retrieval and analysis.

Keywords: Music genre classification, Multilayer Perceptron (MLP), Chroma Feature, Mel Frequency Cepstral Coefficients (MFCC).

1. Introduction

The growth of music databases is growing rapidly, making it difficult to do the grouping of music in certain categories manually, so it can lead to the difficulty of searching for a music category in large

numbers and large scale. The development of digital music, especially in the genre classification has helped in the ease of studying and searching for a song.

This encourages the creation of convenience in the variation of genre classification that can optimize the learning process that can be done easily, simply and has good quality in the accuracy of the classification of a song. As the times, now various methods have been developed so that an audio file can be recognized automatically. In the world of information technology classifying song genres can be done by applying Artificial Intelligence where computers can be made to learn on an experience (pattern) and are able to save the results of the learning process into some knowledge. In Artificial Intelligence there is a sub-field called Machine Learning.

Machine learning as a branch of computer science that examines how a machine can solve problems without being explicitly programmed. Machine Learning has an implementation method called Deep Learning (DL). Deep Learning is used to calculate inexact data, such as languages, sounds or images [1-3].

Based on the description above, the study discusses the comparison of the results of two music genre classification programs using the Multilayer Perceptron (MLP) model with the *Chroma Feature* extraction feature and the model MLP with the *Mel Frequency Cepstral Coefficients* (MFCC) extraction feature [5-7]. The data set used is in the form of song or music audio data, which is taken based on the GTZAN music dataset downloaded via the site http://opihi.cs.uvic.ca/ [4].

The issues raised in this study is the model with which extraction features that have better accuracy in classifying song genres as seen from the classification results of song genres obtained from MLP models that have been built with *Chroma Feature* and MFCC extraction features [1, 5, 6]. In order that the discussion in the study does not widen, some limitations are given as follows:

1. The dataset used is the GTZAN music dataset in the form of a 30 second sample song. This dataset contains 1000 sample songs, consisting of ten music genres. Each genre has 100 songs in WAV format which are hosted through http://opihi.cs.uvic.ca and can be downloaded.
2. Datasets that have been obtained will be split into three sections with the same duration of 10 seconds. So, we get the 3000-sample dataset consisting of ten tracks of music genres namely: blues, classic, country, disco, hip-hop, jazz, metal, pop, reggae, and rock. Each genre has 300 songs in WAV format.
3. The model that was built in this study uses the Multilayer Perceptron method.
4. Extraction features used in the model are *Mel Frequency Cepstral Coefficients* (MFCC) and *Chroma Feature*. The training data used 2400 songs from all genres (240 songs were taken from each genre) or as much as 80 % of the dataset and the testing data (training) were 600 songs from all genres (60 songs were taken from each genre) or 20 % from the dataset.

2. Literature Review

2.1. Music Genre

Musical genres are categories that have arisen through a complex interplay of cultures, artists, and market forces to characterize similarities between musicians or compositions and organize music

Journal of Political Science and Governance
Volume 10 Issue 1, January-March 2022
ISSN: 2995-4193
Impact Factor: 6.41
https://kloverjournals.org/journals/index.php/psg

collections [8]. Nowadays, music genres are often used to categorize music on radio, television and especially internet [1][3][5-7].

There is not any agreement on musical genre taxonomy. Therefore, most music industries and internet music stores use different genre taxonomies when categorizing music pieces into logical groups within a hierarchical structure. For example, allmusic.com uses 531 genres, mp3.com uses 430 genres, and amazon.com uses 719 genres in their database. Pachet and Cazaly [9] tried to define a general taxonomy of musical genres but they eventually gave up and used self-defined two-level genre taxonomy of 20 genres and 250 subgenres in their Cuidado music browser [10].

There are studies to identify human ability to classify music into a genre. One of them is a study conducted by R.O. Gjerdigen and D. Perrot [11] that uses ten different genres, namely Blues, Classical, Country, Dance, Jazz, Latin, Pop, R&B, Rap, and Rock. The subjects of the study were 52 college students enrolled in their first year of psychology. The accuracy of the genre prediction for the 3 s samples was around 70%. The accuracy for the 2.5 s samples was around 40%, and the average between the 2.5 s classification and the 3 s classification was around 44% **2.2. Automatic Music Genre Classification**

Musical genre classification is a classification problem, and such a task consists of two basic steps that must be performed: feature extraction and classification. The goal of the first step, feature extraction, is to get the essential information out of the input data. The second step is to find what combinations of feature values correspond to what categories, which is done in the classification part. The two steps can be clearly separated: the output of the feature extraction step is the input for the classification step [12]. The standard approach of music genre classification task can be seen in Figure 1.

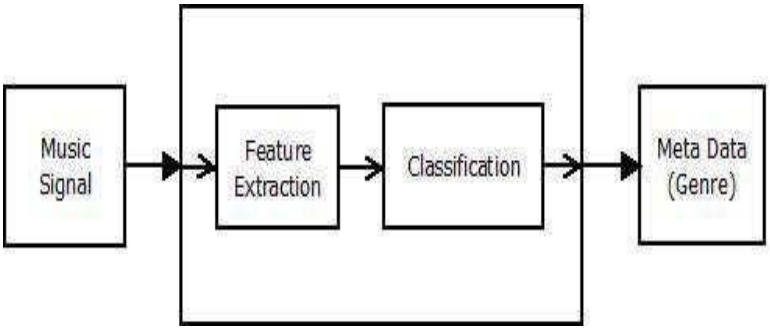

Fig 1. Music genre classification standard approach

3. Research Methods

The research method in this study consists of five stages, including data collection in the form of audio, audio pre-processing, then proceed with Neural Network model design with MLP architecture, then do model training and finally test the model. The model will study the data to determine the optimal features in classifying the music genre and validate each operation to determine the accuracy of the model created.

The stages of the research process in the MLP model with the *Chroma Feature* and MFCC extraction features can be seen in Figure 2 and Figure 3.

Journal of Political Science and Governance
Volume 10 Issue 1, January-March 2022
ISSN: 2995-4193
Impact Factor: 6.41
https://kloverjournals.org/journals/index.php/psg

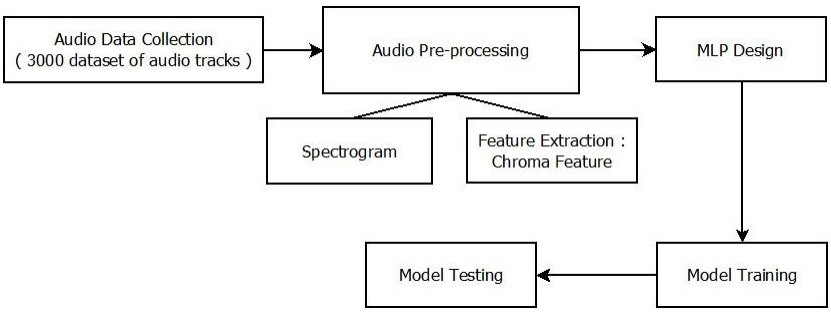

Fig 2. Research Steps with *Chroma Feature Extraction*

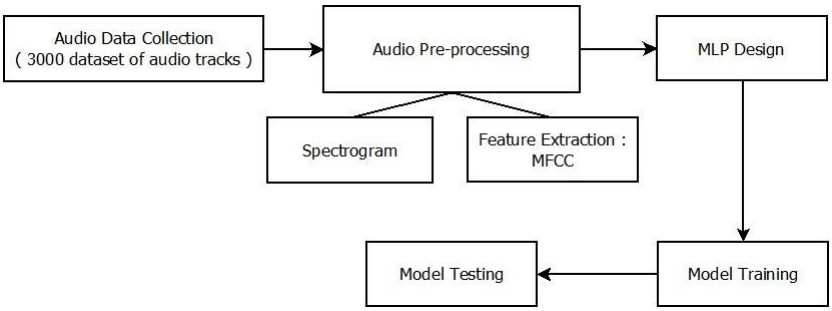

Fig 3. Research Steps with *MFCC Feature Extraction*

3.1. Data Colleting

The collection of data in the form of audio dataset based on the GTZAN music dataset is hosted via http://opihi.cs.uvic.ca and can be downloaded. The dataset consists of 1000 audio tracks from ten different types of genres, each genre is represented by 100 songs with 30 seconds each. All audio tracks are in WAV file format with a sample rate of 22050Hz Mono 16-bit. This dataset consists of ten folders of different music genres. Then the dataset that has been obtained will be split (using Audacity) into 3 parts of tracks with the same duration (10 seconds) so that the dataset is used to be 3000 audio tracks, details of the number of audio tracks used can be seen in Table 1.

Table 1. Details of the Amount and Audio Content

No.	Genre	Duration(second) per Track	Amount
1.	Blues	10	300
2.	Classical	10	300
3.	Country	10	300
4.	Disco	10	300
5.	Hiphop	10	300
6.	Jazz	10	300
7.	Metal	10	300
8.	Pop	10	300
9.	Reggae	10	300
10.	Rock	10	300
∑	-	-	3000

3.2. Audio Pre-Processing

In this stage, audio track will be visualized into a spectrogram. Then all the audio of each genre will become an object for extraction feature using two types of different extraction features, which is Mel

Journal of Political Science and Governance
Volume 10 Issue 1, January-March 2022
ISSN: 2995-4193
Impact Factor: 6.41
https://kloverjournals.org/journals/index.php/psg

Frequency Cepstral Coefficients (MFCC) and Chroma Feature [5, 6]. The extraction feature process will be done separately.

The steps of pre-processing stage with Chroma Feature can be discovered in Figure 4 and for the steps of pre-processing stage with MFCC can be discovered in Figure 4.

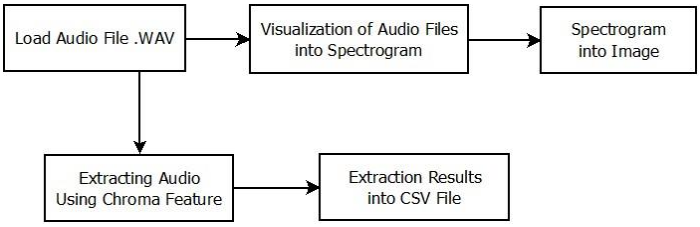

Fig 4. Audio pre-processing with *Chroma Feature Extraction*

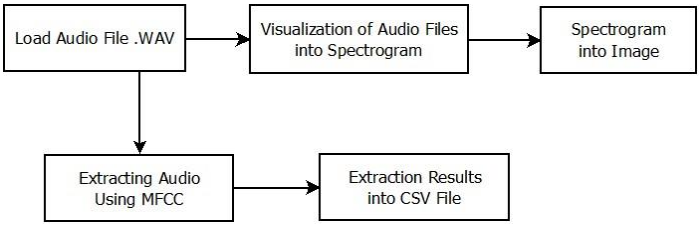

Fig 4. Audio pre-processing with *MFCC Feature Extraction* **3.3. MLP Design**

In this stage, models are built using MLP architecture (Multilayer Perceptron). MLP architecture is used as a classification method, where this method has better ability in audio classification than other architectures. It is used in research because of its simplicity and its learning algorithm is easy to apply and has good accuracy. Figure 6 explains about illustration of MLP architecture. Architecture model that is used in the research can be seen in Figure 7.

In input layer, X train is used as an input data. And then it will be processed in a hidden layer. The hidden layer has three different layers. The first layer has dense 256 with activation function ReLU. The second layer has dense 128 with activation function ReLU. And the third layer has dense 64 with activation function ReLU. After that, the output layer will receive the result from activation process in hidden layer with dense 10 and activation softmax.

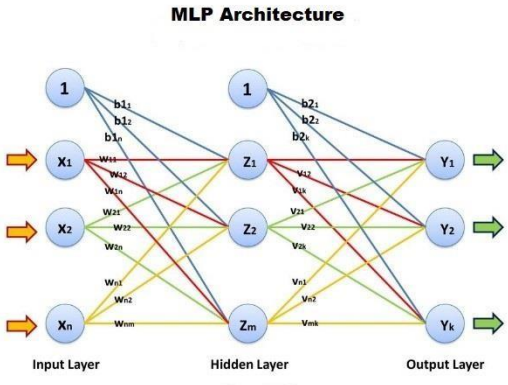

Fig 6. Illustration of MLP Architecture [1]

Journal of Political Science and Governance
Volume 10 Issue 1, January-March 2022
ISSN: 2995-4193
Impact Factor: 6.41
https://kloverjournals.org/journals/index.php/psg

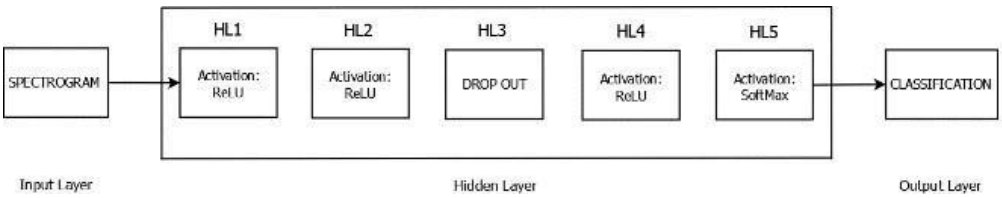

Fig 7. Multilayer Perceptron Model Architecture **3.4. Model Training**

The model that has been built by those features and functions, will be compiled for training. The feature values of the dataset as much as 3000 audios will be saved separately in CSV file. The result of this extraction will be divided in two types of data, training data and testing data. In training data, we used 80% of total extraction data, which is as much as 2400 data. Model is trained using Google Collaboratory to help data processing with specifications:

1. Python 3 Google Compute Engine backend (GPU) as compute engine.
2. RAM of 12.72 GB.
3. Disk storage of 358.27 GB.

Model training will be done with 150 epochs. These epochs are to decide how many models must be executed in order to reach error rates/accuracy that is desired. Samples that used are 32 out of 2400 (batch size) for each training in each epoch until the 150th epoch. Each iteration in the training will use *Adam Optimizer*. After that, the prediction result will be saved. **3.5. Model Testing.**

Model testing is used 20% of total extraction data, or as much as 600 data. Model testing is done by evaluating the comparison of result accuracy between testing result and prediction from training model. Model testing uses data as much as 20% of the total data extraction or as much as 600 data. Model testing is done by evaluating the model by comparing the accuracy of the test results with the correct predictive value of the model training process. Testing is also done using Confusion Matrix.

4. **Results and Discussion 4.1. Model Training Results.**

The graph of the calculation results of the accuracy and loss of training data and validation on the MLP model with the Chroma Feature extraction can be seen in Figures 8 and 9.

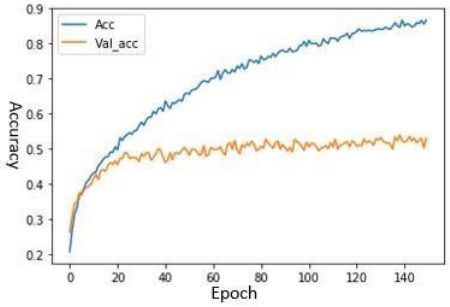

Fig 8. Accuracy Results for Data Train and Validation (*Chroma Feature*)

Journal of Political Science and Governance
Volume 10 Issue 1, January-March 2022
ISSN: 2995-4193
Impact Factor: 6.41
https://kloverjournals.org/journals/index.php/psg

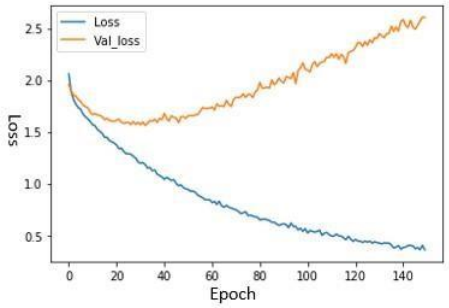

Fig 9. Loss Results for Data Train and Validation (*Chroma Feature*)

In Figure 7 shows that there is an increase in the value of accuracy in training and validation. The increase in accuracy in the validation looks fluctuating, this is because the features obtained from songs in each genre have similar values. This similarity eventually led to the misclassification of genres. In figure 8 shows a decrease in the value of loss in training data. In the validation data the initial loss value decreases, but after passing through epoch the 30th there is an increase in the loss value caused by validation errors and increasingly away from the value of the results of training or can be referred to as overfitting, an event where the value of training results both accuracy and loss far from the validation results.

The graph of the calculation results of the accuracy and loss of training data and validation on the MLP model with the MFCC feature extraction can be seen in Figures 10 and 11.

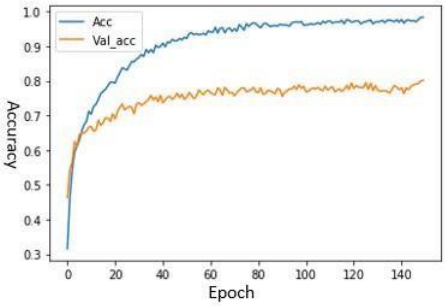

Fig 10. Accuracy Results for Data Train and Validation (*MFCC*)

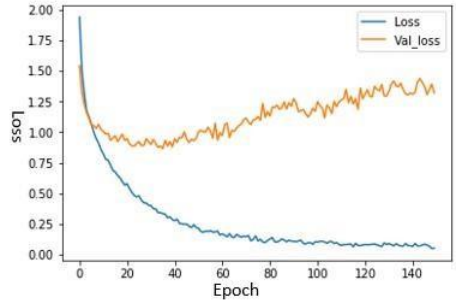

Fig 11. Loss Results for Data Train and Validation (*MFCC*)

In Figure 10 shows that there is an increase in the value of accuracy in training and validation. The increase in accuracy in the validation looks fluctuating, this is because the features obtained from the songs in each genre have similar values. The similarity of feature values makes genre misclassification

Journal of Political Science and Governance
Volume 10 Issue 1, January-March 2022
ISSN: 2995-4193
Impact Factor: 6.41
https://kloverjournals.org/journals/index.php/psg

happen. Even though the MLP model with the MFCC extraction feature is still misclassified, the level of accuracy of the validation data is still quite good. In Figure 11 shows a decrease in the value of loss in training data. In the validation data the initial loss value decreases, but after passing the 20th epoch there is an increase in the loss value caused by validation errors and increasingly away from the value of the training results. Although the increase in the value of loss in the validation data is not too high, it is still classified as overfitting.

4.2. Model Testing Results

Testing on the model training was carried out with test data of 600 audio tracks that had been extracted with the Chroma Feature extraction and with MFCC feature extraction. The test is carried out to test whether the models that have been built can accurately classify the music genre and to determine which models with extraction features have a better level of accuracy. Testing with confusion matrix uses 600 audio tracks from ten types of music genres. Confusion matrix measurement method in this test is useful to get the required values such as accuracy, precision, recall, and F1.

The results of the classification test of each type of music genre using a confusion matrix on model that uses the Chroma Feature extraction can be seen in Figure 12. Classification test results of each type of music genre using the confusion matrix on model that uses MFCC feature extraction can be seen in Figure 13.

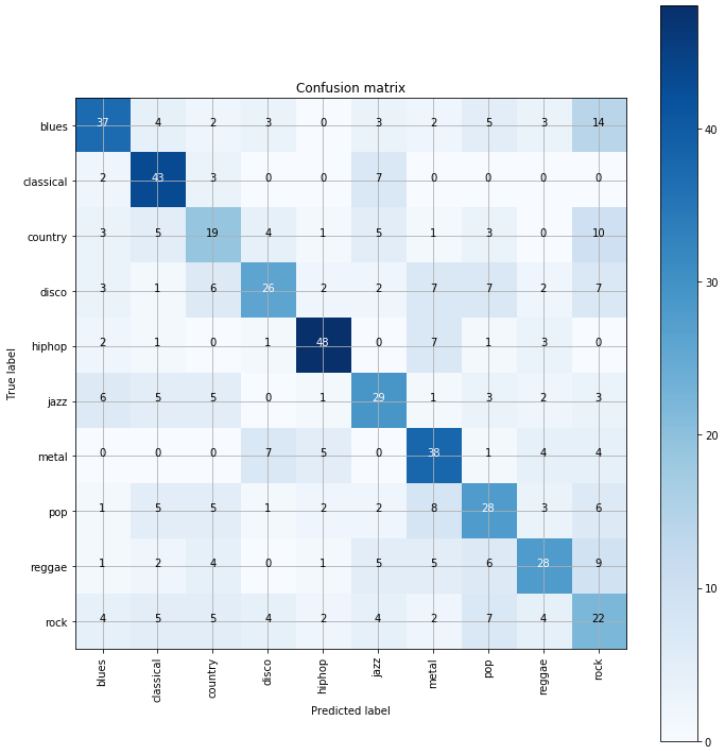

Fig 12. Confusion Matrix of The Classification of Music Genres with *Chroma Feature*

Journal of Political Science and Governance
Volume 10 Issue 1, January-March 2022
ISSN: 2995-4193
Impact Factor: 6.41
https://kloverjournals.org/journals/index.php/psg

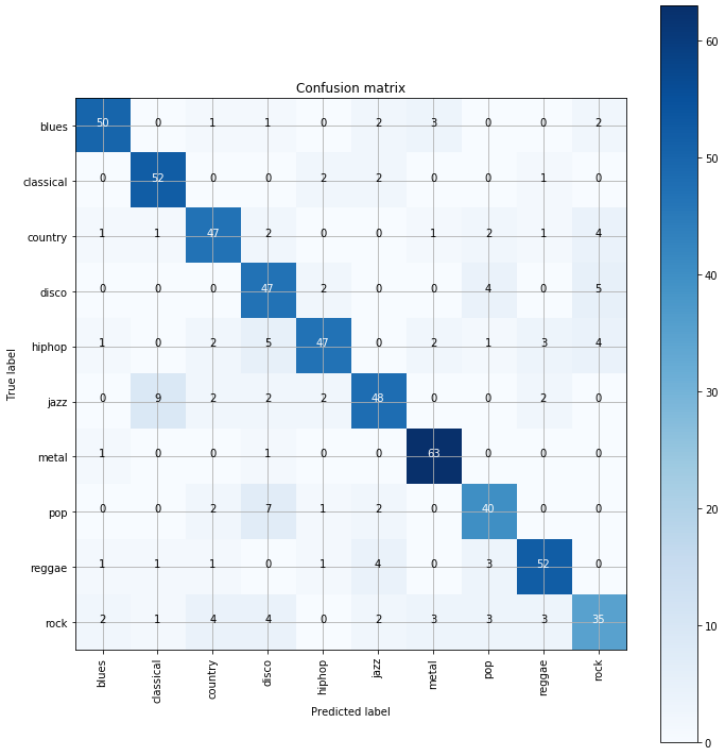

Fig 13. Confusion Matrix of The Classification of Music Genres with *MFCC*

5. Conclusions and Suggestions

5.1. Conclusions

The results of the training of the models built with the Chroma Feature extraction obtained training accuracy values of 98.375% and for testing accuracy obtained by 53.5%. While the training results of the model built with the MFCC feature extraction produced 99,958% training accuracy and 83,333% testing accuracy. Based on the accuracy obtained in the two models-built shows that the model with the MFCC feature extraction has a much better accuracy in identifying the music genre compared to the model built with the Chroma Feature extraction. The loss value obtained from training on models built with the Chroma Feature extraction is 1,625% and for testing loss is 46.50%. As for the value of loss in models built with MFCC feature extraction, 0.0416% for training loss and 16.666% for testing loss. Based on the results of testing with the Confusion Matrix on models with the Chroma Feature extraction, the calculation produces an average recall value of 53.5% and a precision average of 53.6%. As for models with MFCC features extraction, the average recall value was 83.1% and the precision rate was 83.5%. From the average recall and precision mean values produced an F1 value of 53.5% for models with Chroma Feature extraction and 83.3% for models with MFCC features extraction. From the comparison of these results, it can be concluded that the two models have a good harmony between the recall and precision values, but the percentage calculation results on the model with the MFCC feature extraction have better values.

After observing the differences in the results obtained between models with Chroma Feature and MFCC, the authors conclude that the results of music genre classification on models with Chroma

Feature have lower results compared to models with MFCC because models with Chroma Feature only have a maximum number of filter features of 12. While at the MFCC has a filter feature as many as 20. This affects the music genre classification results in a model with MFCC feature extraction is more accurate than the model with Chroma Feature extraction.

5.2. Suggestions

Some suggestions that might be applied for related research in classifying music genres using the Multilayer Perceptron with the Chroma Feature extraction and the Multilayer Perceptron with the MFCC extraction include the use of song or audio datasets that are more numerous than the use of previous audio datasets with a constant duration. Or by using more than one type of extraction feature in each model built to get better audio extraction values.

Acknowledgement

The authors would like to thank the Gunadarma Education Foundation for financial support.

References

R. Rekha, and R.S. Tharani (2021), Speech Emotion Recognition using Multilayer Perceptron Classifier on Ravdess Dataset. ICCAP 2021, December 07-08, Chennai, India, 2021, doi: 10.4108/eai.7-12-2021.2314726.

Sudianto, A.D. Sripamuji, I. Ramadhanti, Risa Riski Amalia, J.Saputra, B. Prihatnowo (2022), Penerapan Algoritma Support Vector Machine Dan Multi-Layer Perceptron Pada Klasifikasi Topik Berita. Jurnal Nasional Pendidikan Teknik Informatika, Volume 11, Nomor 2, pp.84-91.

M. Farooq, F. Hussain, N.K. Baloch, F.R. Raja, H. Yu, Y.B. Zikria (2020), Impact of feature selection algorithm on speech emotion recognition using deep convolutional neural network. Sensors, 20(21), 6008.

GTZAN, http://opihi.cs.uvic.ca/.

M.A. Hossan, S. Memon, and M.A. Gregory (2010), A novel approach for MFCC feature extraction. 2010 4th International Conference on Signal Processing and Communication Systems, Gold Coast, QLD, Australia, pp. 1-5, 2010, doi: 10.1109/ICSPCS.2010.5709752.

P.P. Singh, P. Rani (2014), An Approach to Extract Feature using MFCC. IOSR Journal of Engineering (IOSRJEN), Vol. 04, Issue 08, pp.21-25.

A. Sithara, A. Thomas, D. Mathew (2018), Study of MFCC and IHC Feature Extraction Methods with Probabilistic Acoustic Models for Speaker Biometric Applications. Procedia Computer Science 143, pp.267-276.

N. Scaringella, G. Zoia, and D. Mlynek (2006). Automatic genre classification of music content: a survey. IEEE Signal Process. Mag., vol. 23, no. 2, pp. 133–141.

F. Pachet and D. Cazaly (2000). A taxonomy of musical genres. In Proc. Content-Based Multimedia Information Access (RIAO), Paris, France, 2000.

F. Pachet, J. Aucouturier, A. L. Burthe, A. Zils, and A. Beurive (2004), The cuidado music browser: an end-to-end electronic music distribution system. In Multimedia Tools and Applications, 2004, Special Issue on the CBMI03 Conference, Rennes, France, 2003.

D. Perrot and R. O. Gjerdigen (1999), Scanning the dial: An exploration of factors in the identification of musical style. In Proceedings of the 1999 Society for Music Perception and Cognition.

K. Kosina (2002), Music genre recognition, Master's thesis. Hagenberg Technical University, Hagenberg, Germany, June 2002.