Implementation and Performance Analysis of Mel Frequency Cepstral Coefficient Features in Dangdut Music Sub-Genre Classification

I Nyoman Surya Jaya [1], Tria Hikmah Fratiwi [2], I Gede Harsemadi [3], I Dewa Made Dharma Putra Santika [4], Ni Putu Nanda Maharani [5], Putu Andiny Julia Putri Rapayana [6]
Institut Teknologi dan Bisnis STIKOM Bali
surya_jaya@stikom-bali.ac.id [1], triahikmahfratiwi@stikom-bali.ac.id [2], harsemadi@stikom-bali.ac.id [3]

Article Info

Article history:

Received 2026-04-14
Revised 2026-06-02
Accepted 2026-06-11

Keyword:

Genetic Algorithm,
Hybrid K-NN,
Music Classification,
Dangdut.

ABSTRACT

This study investigates the classification of dangdut music sub-genres using Mel-Frequency Cepstral Coefficients (MFCC) and machine learning approaches. The objective is to evaluate the effectiveness of MFCC in representing audio characteristics and to compare the performance of several classification algorithms, including K-Nearest Neighbor (K-NN), hybrid K-NN optimized with Genetic Algorithm (GA), Support Vector Machine (SVM), and Decision Tree. The dataset consists of 730 audio samples with a duration of 30 seconds each, categorized into three sub-genres: classic dangdut, rock dangdut, and koplo dangdut. The research process includes audio segmentation, extraction of 13 MFCC coefficients, data normalization, train-test splitting (70:30), and performance evaluation using accuracy, precision, recall, F1-score, confusion matrix, and cross-validation. The results indicate that MFCC provides discriminative feature representations, as demonstrated by improved cluster separation in PCA after normalization. Among the evaluated models, hybrid K-NN with GA achieved the highest accuracy of 98.90%, outperforming conventional K-NN, SVM, and Decision Tree. Confusion matrix analysis showed that most samples were correctly classified, with only minor misclassifications between sub-genres sharing similar audio characteristics. Furthermore, consistently high cross-validation accuracies and low standard deviation values confirmed good generalization capability and suggested the absence of significant overfitting. Overall, the findings demonstrate that MFCC is an effective feature for dangdut music sub-genre classification, while normalization and GA-based optimization significantly improve classification performance and model robustness.

I. INTRODUCTION

In the field of Music Information Retrieval (MIR), the feature extraction process is a crucial stage that greatly determines the quality of classification results and audio data analysis. The more features that are successfully extracted, and the more unique the characteristics of each feature, the richer the information that can be obtained from the audio data [1]. These features serve as mathematical representations of musical characteristics, ranging from rhythm, pitch, to timbre, which computationally enable systems to distinguish between different types of music more effectively.

Previous research on the classification of dangdut music sub-genres [2] shows that the use of spectral analysis-based features still has limitations. The values generated across features tend to be highly similar, reducing the system's ability to effectively distinguish between sub-genre characteristics. This condition impacts the classification accuracy, which remains suboptimal, even though the algorithms used are relatively robust in processing audio data. Therefore, a more representative feature extraction approach is needed one that can capture fundamental differences in music signals and improve the performance of dangdut sub-genre classification systems in a more accurate and consistent manner.

As a solution to these limitations, this study employs Mel Frequency Cepstral Coefficients (MFCC) as the primary feature. MFCC is a method for representing the unique characteristics of sound by analyzing the power spectrum over very short time frames. The process involves converting the power spectrum into a logarithmic scale and then transforming it using a linear cosine function on the Mel frequency scale, which approximates how the human ear perceives sound [3]. MFCC is widely used in audio signal processing because it can mimic the human auditory system in capturing the frequency spectrum of sound [4]. Therefore, MFCC is considered more representative in describing the acoustic characteristics of music, particularly in distinguishing differences in timbre and unique sound patterns across sub-genres.

Several previous studies have shown that Mel-Frequency Cepstral Coefficients (MFCC) are effective features in various audio classification tasks. In a gender-based voice identification study, MFCC combined with the K-NN algorithm demonstrated better performance compared to spectral features, achieving an accuracy of 84.18% on training data and 74.71% on testing data, indicating its ability to represent audio signal characteristics more discriminatively [5]. In the context of music genre classification, the use of MFCC on the GTZAN dataset achieved an accuracy of 70%, outperforming Filterbank Energies features, which only reached 56%, thereby highlighting the superiority of MFCC in capturing perceptually relevant spectral information [6]. Additionally, another study that combined MFCC with spectral features such as Chroma, Spectral Contrast, and Tonnetz using the Random Forest algorithm also demonstrated good performance, achieving an accuracy of 76% along with stable precision, recall, and F1-score values [7]. Overall, these studies reinforce that MFCC is an effective and superior feature for audio classification, both as a standalone feature and when combined with other features, making it a suitable foundation for developing music sub-genre classification systems.

The research workflow begins with data collection in the form of a set of songs from three dangdut music sub-genres: classic dangdut, rock dangdut, and koplo dangdut. The collected dataset then undergoes an audio segmentation process to divide each song into smaller segments suitable for analysis. Each segment is subsequently processed through MFCC feature extraction, which represents the frequency spectrum characteristics of the audio signal. The extracted features are then used as input for the classification process by applying a hybrid K-Nearest Neighbor (K-NN) algorithm to evaluate the system’s performance in distinguishing dangdut sub-genres. The selection of the K-NN algorithm is based on its simplicity and effectiveness in performing classification based on distance similarity between data points, making it suitable for audio feature data such as MFCC, which has numerical representations and measurable distances in feature space. In addition, K-NN does not require any specific assumptions about data distribution and can adapt classification results based on the available training data patterns [8]. Through these stages, this study is expected not only to achieve higher classification accuracy compared to previous research but also to contribute to the development of Music Information Retrieval (MIR) systems, particularly by supporting music classification and retrieval based on audio characteristics rather than relying solely on metadata. Such an approach may facilitate more effective music organization, search, and recommendation in digital music collections.

II. METHOD

The overall research methodology is illustrated in the figure1.

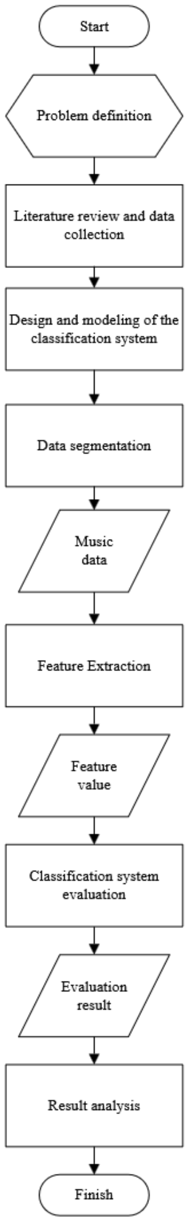

Figure 1. Research methodology flowchart.

This study aims to develop a classification system for dangdut music sub-genres using a Hybrid K-NN algorithm with a Genetic Algorithm (GA). The music data is classified based on audio signal characteristics obtained through the extraction of Mel Frequency Cepstral Coefficient (MFCC) features. The dataset consists of three dangdut sub-genres: classic dangdut, rock dangdut, and koplo dangdut. Through this approach, the system is expected to achieve a higher classification accuracy compared to previous studies that relied on spectral analysis.

A. *Problem Definition*

This study focuses on the need to classify dangdut music sub-genres, namely classic dangdut, rock dangdut, and koplo dangdut. Previous research using spectral analysis features has shown suboptimal accuracy due to the similarity of values across features. Therefore, this study proposes the use of Mel Frequency Cepstral Coefficients (MFCC) as a more representative primary feature, along with the implementation of a Hybrid K-NN algorithm.

B. *Literature Review and Data Collection*

Research on music classification has been widely conducted by analyzing audio signals. One commonly used approach is Mel-Frequency Cepstral Coefficients (MFCC), as it is capable of representing acoustic characteristics that closely resemble human auditory perception. Several previous studies have shown that MFCC is effective in extracting key features from audio signals, both for speech recognition and music genre classification [5], [6], [7]. In dangdut music classification research, the K-Nearest Neighbor (K-NN) method is often used due to its simplicity in determining classes based on the distance between data points [9]. However, a limitation of K-NN is its sensitivity to the selection of the value of k, which significantly affects the system's accuracy. To address this issue, a Hybrid K-NN approach with Genetic Algorithm (GA) is introduced, where GA is used to optimize the value of k to achieve more accurate classification results [2], [10].

Specifically, in the context of dangdut music, Simatupang divides the development of dangdut into three main periods. First, the Malay-style dangdut period (1950s–1960s), which was heavily influenced by Indian music and represented by musicians such as Ellya Khadam. Second, the Rhoma Irama dangdut period (1970s–1990s), which introduced rock elements and brought Soneta to the peak of its popularity. Third, the koplo dangdut period (1990s–present), characterized by faster tempos, dense rhythms, and the use of digital instruments in band formats. These shifts in characteristics have led to the emergence of various dangdut sub-genres with notable differences in tempo, drum (kendang) patterns, and instrumentation formats [11], [12]. Based on these differences, this study collects dangdut music data categorized into three main classes: classic dangdut, rock dangdut, and koplo dangdut.

The dataset used in this study was collected from various digital music databases and online audio distribution platforms through a manual downloading process to ensure sub-genre suitability, recording quality, and audio format consistency. Although there were differences in production periods among the sub-genres, particularly between classic dangdut and modern koplo dangdut, all selected recordings met acceptable audio quality standards. The recordings were characterized by clear audio, minimal noise, the absence of excessive distortion, and sufficient spectral quality to accurately represent their musical characteristics. Therefore, differences in production era did not have a significant impact on audio quality or on the MFCC feature extraction process.

C. *Design and Modeling of the Classification System*

At this stage, the design and modeling of the classification system are carried out to evaluate the extent to which MFCC features can capture and represent important characteristics of dangdut music audio data. This process includes designing the system workflow, starting from feature extraction, selection of classification methods, to model testing in order to determine the level of accuracy and performance in distinguishing characteristic patterns found in dangdut music. Thus, the results of this stage are expected to provide an overview of the effectiveness of using MFCC as the main feature in the analysis and classification of dangdut music audio.

D. *Data Segmentation*

The segmentation process of dangdut audio data is carried out to identify the most representative segments that reflect the intended sub-genre. The duration is standardized across all samples to 30 seconds, and the audio files are converted into mono `.wav` format to meet the requirements of the feature extraction method. The segmentation process resulted in a total of 730 audio samples, consisting of 306 rock dangdut samples, 263 classic dangdut samples, and 161 koplo dangdut samples. These data are subsequently used to obtain feature values that serve as parameters for the classification system.

E. *Feature Extraction*

The system's basis for performing classification is the values obtained from the feature extraction process. The feature extraction stage is carried out to obtain distinctive characteristics that can differentiate one type of data from another. This study applies the Mel-Frequency Cepstral Coefficients (MFCC) method, which is a feature extraction technique that represents audio signal characteristics based on human auditory perception through a series of stages [13], including:

1) Pre-emphasis, a filtering process used to amplify high-frequency components so that the spectral shape becomes more balanced.

2) Framing, which involves dividing the audio signal into several short segments (frames) so that the signal can be assumed to be stationary over a short period.
3) Windowing, the application of a window function (Hamming Window) to each frame to reduce discontinuities at the edges of the signal.
4) Fast Fourier Transform (FFT), which transforms the signal from the time domain to the frequency domain to obtain the frequency spectrum.
5) Mel-Filterbank, the process of mapping linear frequency to the Mel scale, mimicking the human ear's sensitivity to frequency.
6) Discrete Cosine Transform (DCT), a transformation process used to decorrelate the coefficients, resulting in a more compact and informative set of MFCC features.

In the MFCC feature extraction process, several key parameters are used to ensure accurate representation of the audio signal characteristics. The number of MFCC coefficients is set to 13, representing the most significant cepstral features of the audio signal. The Fast Fourier Transform window size is set to 2048 samples to provide sufficient frequency resolution during spectral analysis, while the hop length is set to 512 samples to determine the shift between consecutive analysis windows. In addition, the sampling rate of each audio file and the audio time-series signal serve as fundamental inputs for the extraction process. These parameters enable MFCC to effectively capture the spectral and perceptual characteristics of dangdut music, which are subsequently used as inputs for the classification models.

F. Classification System Evaluation

The dangdut music sub-genre classification system in this study is developed using the K-Nearest Neighbor (K-NN) algorithm and a hybrid K-NN approach that optimizes the search for the optimal k value using a Genetic Algorithm. The main objective is to evaluate the performance of MFCC features in representing audio characteristics. This process not only focuses on improving accuracy through k parameter optimization but also examines the effectiveness of combining methods to produce a classification model that is more adaptive to variations in dangdut music patterns. Furthermore, the classification results obtained from K-NN and hybrid K-NN are compared with the performance of other algorithms such as Support Vector Machine (SVM) and Decision Tree. This allows for a comprehensive analysis of the strengths and weaknesses of each method in processing MFCC features for dangdut music sub-genre classification tasks.

G. Results Analysis

The analysis of dangdut music sub-genre classification results in this study is conducted using MFCC features processed through several algorithms, namely K-Nearest Neighbor (K-NN), hybrid K-NN with optimization of the k value using a Genetic Algorithm (GA), Support Vector Machine (SVM), and Decision Tree. The performance of each model is evaluated comprehensively by calculating accuracy, applying cross-validation, and measuring precision, recall, and F1-score for each class, in order to assess the model's ability to classify data accurately and consistently [14]. Furthermore, the overall accuracy is calculated as the main indicator for comparing the performance of the classification methods. In addition, the results obtained using MFCC features are compared with classification results based on spectral features to determine the differences in effectiveness between the two feature types in representing the characteristics of dangdut music audio, as well as their impact on improving the overall performance of the classification system.

III. RESULT AND DISCUSSION

In the data segmentation stage, the most representative audio segments with a duration of 30 seconds were selected for each dangdut music sample. This process resulted in a total of 730 data samples divided into three sub-genre classes: 263 samples of classic dangdut, 306 samples of rock dangdut, and 161 samples of koplo dangdut. Subsequently, all data underwent feature extraction using the Mel-Frequency Cepstral Coefficients (MFCC) method, which involves several processes, including pre-emphasis to enhance high frequencies, framing to divide the signal into short frames, windowing using a Hamming window, Fourier Transform (FFT) to convert the signal into the frequency domain, mapping to the Mel scale, and Discrete Cosine Transform (DCT) to generate MFCC coefficients [15]. From this process, 13 MFCC parameters were obtained, representing the main characteristics of the audio signal and used as input for the classification system.

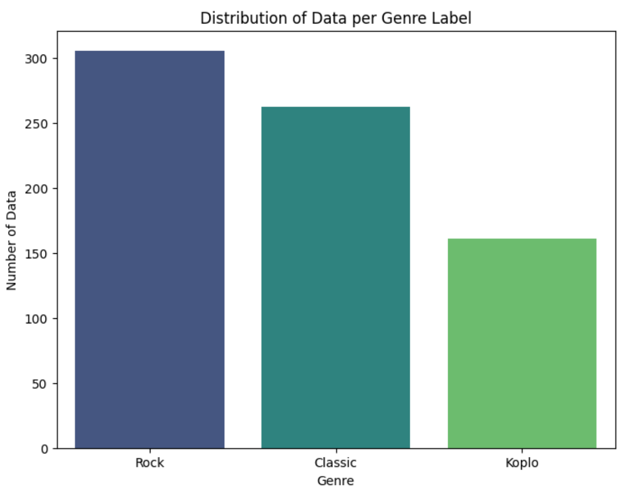

Figure 2. Distribution of dangdut music data by sub-genre label.

To evaluate the robustness of the classification models, two data scenarios were employed: imbalanced and balanced datasets. The imbalanced dataset represents the original

e-ISSN: 2548-6861

distribution of the collected data, consisting of 306 rock dangdut samples, 263 classic dangdut samples, and 161 koplo dangdut samples. Meanwhile, the balanced dataset was created by equalizing the number of samples in each class to 161, corresponding to the smallest class size. The use of both scenarios allows for the analysis of model performance under real-world data distributions as well as under conditions where class representation is evenly distributed.

Subsequently, the data in each scenario was divided into training and testing sets with a ratio of 70:30. In addition, a comparison was conducted between normalized and non-normalized data to ensure that there was no significant influence of extreme values (outliers) on the evaluation results. The results of implementing MFCC features in the classification system using various algorithms are described as follows:

A. K-NN Classification System

The performance of MFCC features in the classification system in this study was evaluated using the conventional K-Nearest Neighbor (K-NN) algorithm with k = 5 as the number of nearest neighbors, which determines the class of a data point based on distance proximity and the majority label of the five nearest data points. Initial testing was conducted under imbalanced data conditions without normalization, resulting in an accuracy of 92.69%. This result indicates that MFCC features are quite effective in representing audio characteristics for dangdut music sub-genre classification. Cross-validation evaluation under this condition also shows stable performance, with an average accuracy of 0.9370 and a standard deviation of 0.0218 (5-fold), and an average accuracy of 0.9411 with a standard deviation of 0.0319 (10-fold). The low standard deviation indicates that the model has good generalization ability and is not sensitive to data partitioning [16].

TABLE I
K-NN CLASSIFICATION EVALUATION METRICS

Class	Precision	Recall	F1-Score
Classic	0,94	0,91	0,92
Rock	0,94	0,98	0,96
Koplo	0,91	0,91	0,91

Based on the classification evaluation metrics in Table 1, the model is able to consistently classify all classes. However, the high performance observed on imbalanced data without normalization raises the possibility of overfitting [17]. Therefore, further testing was conducted by incorporating a normalization process to improve the model's generalization capability.

TABLE II
K-NN ACCURACY EVALUATION (NORMALIZED DATA)

K	Imbalanced Data %	Balanced Data %
1	96,80	94,48
2	96,35	92,41
3	96,35	95,86
4	95,89	95,17
5	96,35	94,48
6	96,35	91,72
7	96,35	93,10
8	95,43	91,03
9	94,52	91,03
10	95,43	91,72

The test results (Table 2) show that the application of normalization is able to improve the model's performance. The normalized data is divided into two scenarios: imbalanced data as real-world data and balanced data with 161 samples in each class. The accuracy on imbalanced data tends to be higher and more stable compared to balanced data, although both scenarios still demonstrate good performance at several values of k.

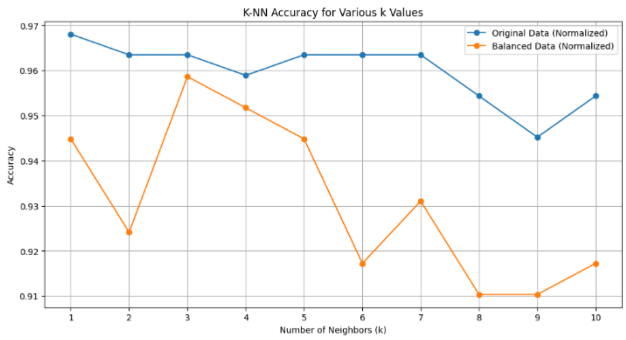

Figure 3. K-NN classification accuracy across various values of k on normalized data, comparing imbalanced and balanced datasets.

Compared to the initial testing, normalization has been proven to improve accuracy while also indicating that the influence of outliers is not a significant issue. Furthermore, cross-validation results on the normalized data show an increase in average accuracy to 0.9671 with a standard deviation of 0.0159 (5-fold), and 0.9726 with a standard deviation of 0.0274 (10-fold). The consistently low standard deviation indicates that the model is more stable, consistent, and has better generalization capability.

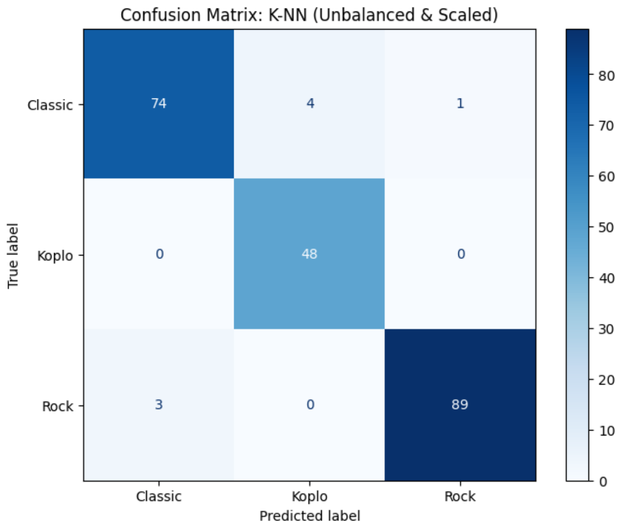

Figure 4. Confusion matrix of K-NN classification in unbalanced and normalized data.

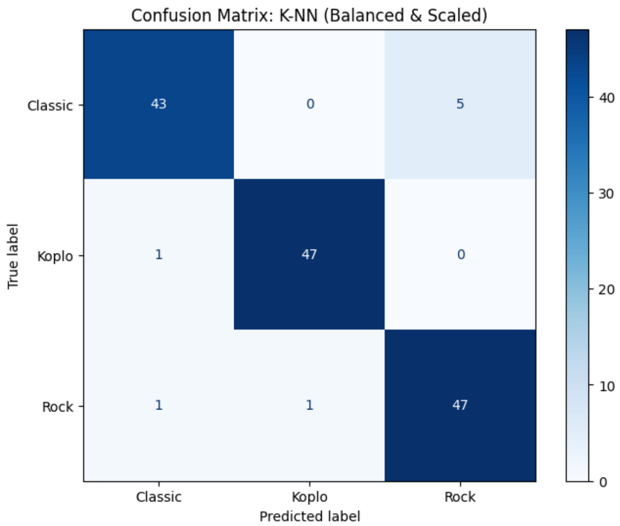

Figure 5. Confusion matrix of K-NN classification in balanced and normalized data.

The confusion matrix analysis indicates that the K-NN model still produces a number of classification errors on the testing data, particularly between sub-genres that share similar audio characteristics, such as classic dangdut and rock dangdut. In both the balanced and imbalanced normalized datasets, the majority of samples were classified correctly, with no dominant misclassification occurring in any particular class. This finding, together with the consistent cross-validation results and low standard deviation values, suggests that the model has good generalization capability and does not suffer from significant overfitting. Therefore, the high accuracy achieved is more likely to reflect the effectiveness of MFCC features in representing the characteristics of dangdut music sub-genres rather than merely the model's ability to memorize the training data.

B. *Hybrid K-NN Classification System*

The hybrid K-NN classification system in this study integrates a Genetic Algorithm (GA) to optimize the search for the optimal k value, which has long been one of the main challenges in K-NN implementation due to the need for repeated trials across different k values to obtain the best accuracy [18]. By utilizing GA, the parameter search process is no longer performed through trial-and-error but instead through a more efficient evolution-based optimization mechanism.

In general, the GA process begins with the initialization of a population, where a number of candidate solutions (values of k) are generated randomly. This is followed by evaluation and selection, where each individual is assessed based on its fitness (classification accuracy), and the best-performing individuals are selected for retention. The next stage is crossover, which combines two selected individuals to produce new solutions expected to have better quality. After that, mutation is applied to introduce random variations into certain individuals to prevent premature convergence and to expand the solution search space. This iterative process continues until the most optimal k value is obtained [19].

The experiments in this study were conducted using various combinations of Genetic Algorithm (GA) parameters, including population size, number of generations, mutation rate, and the range of k values tested. These parameter variations aim to analyze their influence on optimization performance in finding the optimal k value for the K-NN algorithm. Through this approach, it is expected that the most effective GA parameter configuration can be obtained to improve classification accuracy. An illustration of one of the GA operations applied in this study is presented in Table 3.

TABLE III
GA OPERATION WITH A NUMBER OF GENERATIONS PARAMETER OF 10

Generation	Best K	Accuracy %
1	3	96,28
2	3	96,28
3	3	96,28
4	1	97,31
5	1	97,31
6	1	97,31
7	1	97,31
8	1	97,31
9	1	97,31
10	1	97,31

The Genetic Algorithm (GA) operation with a parameter of 10 generations produces ten generations of the best solutions throughout the evolutionary process. From this set of solutions, one individual with the highest fitness value is selected as the optimal k value to be used in the subsequent K-NN classification stage. The results of the best k selection along with the corresponding accuracy achieved, obtained from various parameter variations and both imbalanced and balanced data scenarios, are presented in Table 4 and Table 5.

TABLE IV
ACCURACY PERFORMANCE OF HYBRID K-NN WITH IMBALANCED DATA

Parameter				**Best K**	**Accuracy %**	
Pop	**Gen**	**Mutase**	**K Range**		**Raw**	**Scaled**
30	15	0,1	1 – 20	1	98,90	96,80
20	10	0,1	1 – 20	1	98,90	96,80
10	5	0,05	1 – 10	1	98,90	96,80

TABLE V
ACCURACY PERFORMANCE OF HYBRID K-NN WITH BALANCED DATA

Parameter				**Best K**	**Accuracy %**	
Pop	**Gen**	**Mutase**	**K Range**		**Raw**	**Scaled**
30	15	0,1	1 – 20	1	97,31	94,48
20	10	0,1	1 – 20	1	97,31	94,48
10	5	0,05	1 – 10	1	97,31	94,48

Based on Tables 4 and 5, it can be concluded that the application of hybrid K-NN with optimization using a Genetic Algorithm (GA) consistently produces an optimal k value of 1 across various parameter settings, both for imbalanced and balanced data scenarios. This indicates that the GA optimization process is capable of finding a stable solution despite variations in configuration parameters such as population size, number of generations, mutation rate, and the range of k values. In terms of performance, the imbalanced data scenario yields an accuracy of 98.90% on non-normalized (raw) data and 96.80% on normalized (scaled) data, while the balanced data scenario achieves accuracies of 97.31% (raw) and 94.48% (scaled). These differences indicate that data distribution affects accuracy levels, with generally higher performance observed in the imbalanced data scenario.

TABLE VI
HYBRID K-NN CLASSIFICATION EVALUATION METRICS IN IMBALANCED AND SCALED DATA

Class	**Precision**	**Recall**	**F1-Score**
Classic	0,97	0,96	0,97
Rock	1	0,96	0,98
Koplo	0,91	1	0,95

TABLE VII
HYBRID K-NN CLASSIFICATION EVALUATION METRICS IN BALANCED AND SCALED DATA

Class	**Precision**	**Recall**	**F1-Score**
Classic	0,97	0,96	0,97
Rock	1	0,96	0,98
Koplo	0,91	1	0,95

Compared to conventional K-NN, the classification results of the hybrid K-NN show higher performance even when using the same neighborhood value, namely k = 1. This is because the Genetic Algorithm not only selects the best k value but also indirectly optimizes the solution search space based on fitness evaluation (accuracy) [2]. The selection, crossover, and mutation processes in GA enable the selection of more adaptive parameter configurations according to the data characteristics, meaning that the resulting k = 1 is not merely a minimum value, but a truly optimal value in the context of the dataset distribution used. In addition, this optimization process helps minimize the influence of less representative data, thereby improving the quality of classification decisions compared to conventional K-NN, which uses a direct approach without optimization.

Furthermore, cross-validation results for the model with k=1 show very strong performance. On balanced and normalized data, the model achieves an average accuracy of 97.31% with a standard deviation of 0.0207, while on imbalanced but normalized data, a higher average accuracy of 98.90% is obtained with a standard deviation of 0.0103. The low standard deviation values in both scenarios indicate that the model has high consistency across folds and good generalization capability.

Although theoretically the use of k=1 has limitations due to the absence of a majority voting mechanism and its sensitivity to noise, the cross-validation results in this study show that the model remains stable and robust. This suggests that the dataset distribution is sufficiently representative, making k=1 still effective in producing accurate and consistent classification performance. However, this limitation should still be considered, especially when the model is applied to more complex or noisy datasets.

C. SVM and Decision Tree Classification Systems

Next, the performance of MFCC features was evaluated using other classification algorithms, namely Support Vector Machine (SVM) and Decision Tree, as a comparison to K-NN and hybrid K-NN methods. Based on the results in Table 8, the SVM algorithm shows a significant improvement in performance after normalization, with higher accuracy on scaled data compared to raw data, in both imbalanced and balanced data conditions. This indicates that SVM is quite sensitive to feature scaling and benefits from the normalization process.

TABLE VIII
SVM AND DECISION TREE ACCURACY PERFORMANCE

Algorithm	**Accuracy % (Imbalanced)**		**Accuracy % (Balanced)**	
	Raw	**Scaled**	**Raw**	**Scaled**
SVM	86,30	96,35	83,45	94,48
Decision Tree	85,39	85,39	88,28	88,28

TABLE VI
SVM CLASSIFICATION EVALUATION METRICS IN IMBALANCED AND SCALED DATA

Class	Precision	Recall	F1-Score
Classic	0,96	0,94	0,95
Rock	0,95	0,97	0,96
Koplo	1	1	1

TABLE VII
DECISION TREE CLASSIFICATION EVALUATION METRICS IN IM BALANCED AND SCALED DATA

Class	Precision	Recall	F1-Score
Classic	0,90	0,82	0,86
Rock	0,82	0,87	0,84
Koplo	0,86	0,88	0,87

Meanwhile, the Decision Tree shows relatively stable performance between raw and scaled data, indicating that this algorithm is not significantly affected by feature scaling. However, overall, the accuracy produced by the Decision Tree is still lower than that of SVM under normalized data conditions. When compared with K-NN and hybrid K-NN results, both algorithms (SVM and Decision Tree) demonstrate lower performance. Therefore, it can be concluded that MFCC features are more optimal when combined with distance-based approaches such as K-NN, especially when optimized using a Genetic Algorithm.

D. Feature Extraction Analysis

To analyze the discriminative capability of MFCC features, a visualization was performed using Principal Component Analysis (PCA) under two data scenarios: imbalanced data without normalization and imbalanced data with normalization. In the raw data, the first two principal components explained 73.05% of the total variance. However, the visualization showed that the genre clusters still exhibited considerable overlap and were not clearly separated. The data points were more widely dispersed, and class boundaries were less distinct due to the dominance of several MFCC features with larger value ranges. This condition indicates that the characteristics of each sub-genre were not optimally represented, potentially reducing the model’s ability to distinguish between classes with similar audio characteristics.

In contrast, for the normalized (scaled) data, the first two principal components explained 53.38% of the total variance but produced a more structured distribution pattern. Data points belonging to the same class appeared more concentrated, and the distance between clusters became more apparent, although a small degree of overlap remained in certain areas. This improvement occurred because normalization standardized the scale of all MFCC features, allowing each feature to contribute proportionally to the data representation rather than being dominated by features with larger numerical values. As a result, the distinctive characteristics of each sub-genre were represented more effectively, leading to clearer cluster separation within the feature space.

These findings demonstrate that normalization plays an important role in improving the quality of data representation for classification. The PCA visualization results are also consistent with the improved performance observed in the K-NN, Hybrid K-NN, and SVM classification models, where clearer cluster separation indicates that MFCC features are more capable of distinguishing the characteristics of different dangdut music sub-genres.

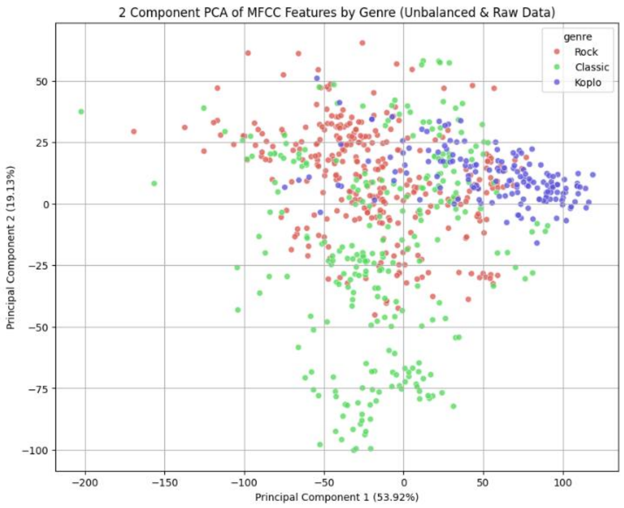

Figure 6. MFCC data distribution using PCA on unnormalized data.

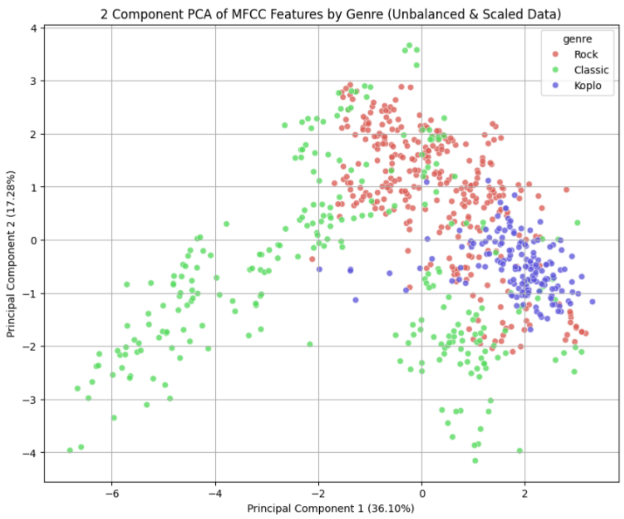

Figure 7. MFCC data distribution using PCA on normalized data.

In addition, the correlation matrix analysis in both scenarios shows that MFCC features have relatively consistent linear relationships, where adjacent features tend to be positively correlated because they capture similar spectral characteristics. Although normalization does not change the correlation pattern between features, it standardizes the feature scales so that each variable contributes more fairly in distance-based computations. Thus, normalization does not alter the underlying structure of feature relationships but improves their effectiveness in distance-based models.

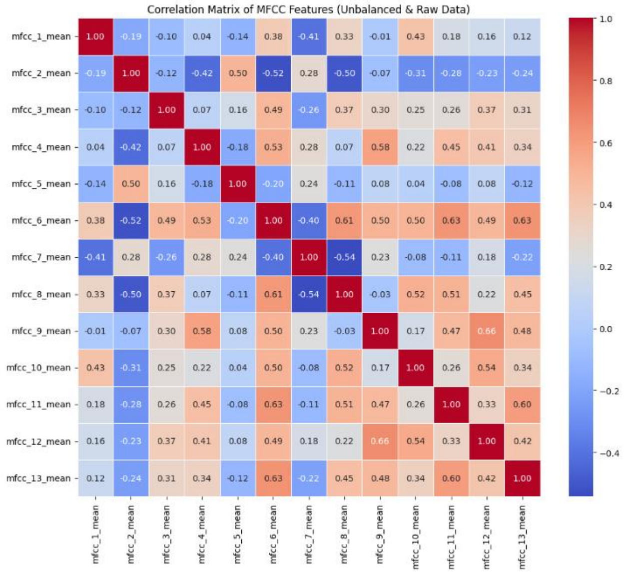

Figure 8. MFCC feature correlation matrix on unnormalized data.

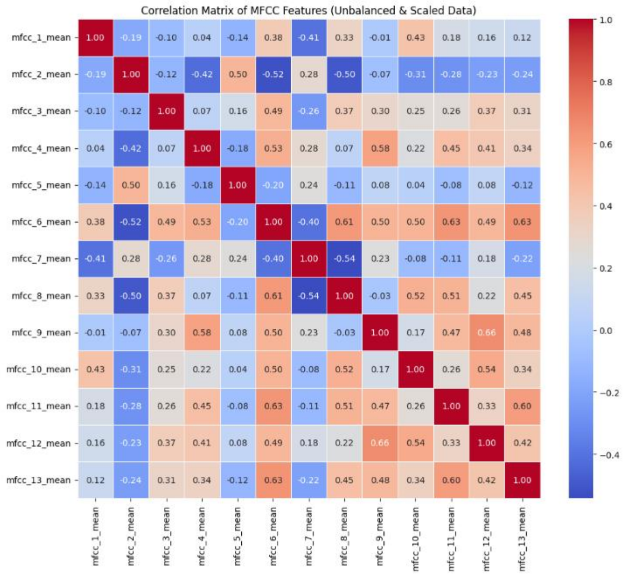

Figure 9. MFCC feature correlation matrix on normalized data.

Overall, the PCA and correlation matrix results confirm that normalization improves the quality of data separation in the feature space, which is consistent with the improved performance observed in the classification system evaluation, particularly for K-NN and SVM algorithms. This finding is also aligned with previous studies, where spectral features showed lower performance because they only represent global frequency characteristics and are less aligned with human auditory perception, resulting in less distinct class separation.

In contrast, MFCC proves to be superior because it maps the signal onto the Mel scale, which is more aligned with human auditory perception, and produces a more discriminative feature representation. Based on the PCA results, correlation matrix, and classification evaluation, MFCC is able to form more separable clusters and achieve higher accuracy compared to spectral features. Therefore, MFCC is more effective for use in dangdut music sub-genre classification systems than spectral features.

IV. Conclusion

Based on the results, MFCC proved effective in representing the audio characteristics of dangdut music for sub-genre classification. Among the evaluated algorithms, hybrid K-NN with Genetic Algorithm optimization achieved the best performance, reaching 98.90% accuracy and outperforming conventional K-NN, SVM, and Decision Tree. Data normalization further improved model stability and classification performance, as evidenced by clearer cluster separation in the PCA visualization and more balanced feature contributions. The confusion matrix analysis showed that the majority of samples were correctly classified, with only a small number of misclassifications occurring between sub-genres with similar audio characteristics. Furthermore, consistently high cross-validation accuracies accompanied by low standard deviation values indicate that the models possess good generalization capability and do not suffer from significant overfitting. Therefore, the high classification performance is more likely attributable to the discriminative power of MFCC features rather than the model simply memorizing the training data. Overall, the results demonstrate that MFCC, combined with normalization and GA-based optimization, provides an effective and robust approach for dangdut music sub-genre classification.

Acknowledgements

The author would like to express sincere gratitude to ITB STIKOM Bali for the support and facilities provided, which have made it possible for this research to be completed successfully.

References

[1] P. D. Pamungkasari, *Buku Ajar kecerdasan Artifisial: Teori dan Implementasi Praktis*. Gresik: CV Global Kreatif Media, 2025.

[2] T. H. Fratiwi *et al.*, "Implementation of the Hybrid K-Nearest Neighbour Algorithm for Dangdut Music Sub-Genre Classification," *Journal of Applied Informatics and Computing (JAIC)*, vol. 9, no. 4, pp. 2548–6861, Aug. 2025, doi: https://doi.org/10.30871/jaic.v9i4.9702.

[3] M. Tajuddin, D. Supatmiwati, S. Hidayat, and A. S. Anas, *Digitalisasi: Konsep, Teknologi & Penerapannya (Bunga Rampai Pengetahuan, Gagasan & Rekomendasi Untuk Indonesia)*, I. Malang: Media Nusa Creative, 2022.

[4] H. Rheza Paleva and B. Henryranu Prasetio, "Penerapan Short Time Fourier Transform pada MFCC untuk Sistem Pengenalan Ucapan Tingkat Stres," 2024. [Online]. Available: http://j-ptiik.ub.ac.id

[5] R. Aurelius *et al.*, "Perbandingan Kualitas Pengenalan Suara untuk Ekstraksi Fitur Menggunakan MFCC dan Spectral," *NARATIF: Jurnal Ilmiah Nasional Riset Aplikasi dan Teknik Informatika*, vol. 06, Jun. 2024.

[6] I. Aufa, A. Ramadhan, and A. Qoiriah, "Perbandingan MFCC dan Filterbank Energies dalam Metode HMM untuk Identifikasi Genre Musik," *Journal of Informatics and Computer Science*, vol. 06, 2025.

[7] R. Sefiyanti and Murinto, "Klasifikasi Genre Musik Berdasarkan Fitur Mel Frequency Spectral Coefficient Menggunakan Random Forest," in *CORISINDO*, Sep. 2025.

[8] Suyanto, *Data Mining untuk Klasifikasi dan Klasterisasi Data*, vol. 1. Bandung: Penerbit Informatika Bandung, 2019.

[9] I. G. Harsemadi, "Perbandingan Kinerja Algoritma K-NN dan SVM dalam Sistem Klasifikasi Genre Musik Gamelan Bali," *INSANI Informatics for Educators And Professionals : Journal of Informatics*, vol. 8, no. 1, pp. 1–10, 2023, [Online]. Available: www.joox.com

[10] T. H. Fratiwi, M. Sudarma, and N. Pramaita, "Sistem Klasifikasi Musik Gamelan Angklung Bali Terhadap Suasana Hati Menggunakan Algoritma K-Nearest Neighbor Berbasis Algoritma Genetika," *Majalah Ilmiah Teknologi Elektro*, vol. 20, no. 2, p. 265, Dec. 2021, doi: 10.24843/mite.2021.v20i02.p10.

[11] F. G. Hananta, "Pergeseran Instrumentasi Karakteristik Musik Dangdut Era Digital Di Jawa Timur," *Repertoar*, vol. 1, Jul. 2023.

[12] M. Ikhwan, *Kepikiran Dangdut*. Yogyakarta: Warning Books, 2023.

[13] Z. K. Abdul and A. K. Al-Talabani, "Mel Frequency Cepstral Coefficient and its Applications: A Review," 2022, *Institute of Electrical and Electronics Engineers Inc.* doi: 10.1109/ACCESS.2022.3223444.

[14] N. T. S. Saptadi *et al.*, *Data Mining*. Batam: Yayasan Cendikia Mulia Mandiri, 2024.

[15] K. M. Rezaul *et al.*, "Enhancing Audio Classification Through MFCC Feature Extraction and Data Augmentation with CNN and RNN Models," *International Journal of Advanced Computer Science and Applications*, vol. 15, no. 7, pp. 37–53, 2024, doi: 10.14569/IJACSA.2024.0150704.

[16] T. Abedin, H. Xu, and S. Uddin, "The impact of K selection in K-fold cross-validation on bias and variance in supervised learning models," *Sci. Rep.*, vol. 16, no. 1, Dec. 2026, doi: 10.1038/s41598-026-37247-x.

[17] M. Mujahid *et al.*, "Data oversampling and imbalanced datasets: an investigation of performance for machine learning and feature engineering," *J. Big Data*, vol. 11, no. 1, Dec. 2024, doi: 10.1186/s40537-024-00943-4.

[18] R. K. Halder, M. N. Uddin, M. A. Uddin, S. Aryal, and A. Khraisat, "Enhancing K-nearest neighbor algorithm: a comprehensive review and performance analysis of modifications," *J. Big Data*, vol. 11, no. 1, Dec. 2024, doi: 10.1186/s40537-024-00973-y.

[19] P. Suwirmayanti, R. A. N. Diaz, I. K. G. D. Putra, and K. Budiarta, *Data Mining: Teknik, Implementasi, dan Aplikasi*, Pertama. Yogyakarta: Andi, 2023.