Implementation of the Hybrid K-Nearest Neighbour Algorithm for Dangdut Music Sub-Genre Classification

Tria Hikmah Fratiwi [1]*, I Gede Harsemadi [2]*, Putu Tjintia Kencana Dewi [3]*, Luh Rediasih [4]*, M. Alvinnur Filardi [5]*, I Dewa Made Dharma Putra Santika [6]*
* Institut Teknologi dan Bisnis STIKOM Bali
triahikmahfratiwi@stikom-bali.ac.id [1], harsemadi@stikom-bali.ac.id [2], tjintiakencana@stikom.bali.ac.id [3], 230050019@stikom.bali.ac.id [4], 230030642@stikom.bali.ac.id [5], 230040026@stikom.bali.ac.id [6]

Article Info

Article history:

Received 2025-06-02
Revised 2025-07-04
Accepted 2025-07-09

Keyword:

Dangdut,
Genetic Algorithm,
Classification,
K-Nearest Neighbor,
Music Information Retrieval.

ABSTRACT

This research focuses on the classification of dangdut sub-genres — classical, rock, and koplo — by collecting 136 songs from Ellya Khadam, Rhoma Irama, and Denny Caknan, each representing distinct eras of dangdut music. From these, 483 music segments of 30 seconds each were extracted and labelled with expert assistance to ensure accuracy. Six spectral features (centroid, skewness, rolloff, kurtosis, spread, and flatness) were computed and stored in a dataset divided into 70% training and 30% testing sets. The Hybrid K-NN algorithm, integrating Genetic Algorithm (GA) to optimize the k parameter, was applied and evaluated through 5-fold cross-validation. GA parameters were set to a population size of 10, 15 generations, 4-bit chromosome length, and 3-fold cross-validation during optimization. Hybrid K-NN achieved the highest accuracy of 74.31% at k=4 with a processing time of 4.86 seconds, outperforming conventional K-NN (68.75% at k=4, 0.04 seconds), Decision Tree (61.11%, 0.42 seconds), and SVM with ECOC (54.86%, 1.99 seconds). The Hybrid K-NN also demonstrated stable performance with an average accuracy of 72.04% and a standard deviation of 2.22 percent, while the average precision, recall, and F1-score were each around 0.72. Confusion matrix analysis revealed frequent misclassification of class 2 as class 1, highlighting a classification challenge. Overall, this research shows that Hybrid K-NN is more effective than the other methods in capturing data patterns, optimizing parameters, and generalizing to unseen data, though at the cost of longer computation time due to GA's iterative optimization and validation processes.

I. Introduction

Since the recognition of dangdut music as an Intangible Cultural Heritage (ICH) by UNESCO in August 2023, dangdut music is no longer underestimated. Initially perceived as the music of the grassroots, dangdut has evolved into a representation of Indonesia that resonates across various social strata. Major national events often feature dangdut performances as entertainment. For example, during the commemoration of Indonesia's Independence Day at the State Palace, the solemn atmosphere traditionally observed over generations shifts into a lively celebration when dangdut music is performed after the ceremony. Government officials, public figures, and citizens alike sway to the rhythm of dangdut music.

The close connection between dangdut and the public is the result of relentless efforts by dangdut musicians and practitioners to adapt to the times and technological advancements. To remain relevant across generations, dangdut has undergone significant transformations, particularly in instrumentation, moving towards electronic and digital formats. This shift has produced distinct audio characteristics, giving rise to new sub-genres within the dangdut genre. Initially, dangdut was heavily influenced by Bollywood and Indian music styles, known as Malay rhythm dangdut or classic dangdut. It then transformed into a band formation with a rock nuance, popularized by Rhoma Irama and Soneta. In the era of remix music with vibrant beats, dangdut evolved to align with market demands, culminating in a new form known as *dangdut koplo* [1], [2].

Identifying sub-genres within dangdut can be done through auditory perception, processed by the human brain based on experience or knowledge. However, distinguishing between them requires classification methods to simplify the identification and analysis of sub-genre variations, enabling more accurate music retrieval [3]. Despite the popularity and richness of dangdut music, no prior research has specifically attempted to classify its sub-genres using machine learning approaches, making this study a novel contribution to the intersection of musicology and computational analysis.

This study aims to develop a classification system that identifies the class or membership of dangdut music data based on audio signal characteristics, contributing to the field of Music Information Retrieval (MIR). There are significant differences in musical styles across cultures and regions, which also pose specific challenges to building a robust global classification model [4]. The classification method employs the K-NN (K-Nearest Neighbour) algorithm and GA (Genetic Algorithm). The K-NN algorithm classifies new data by comparing its proximity to existing data within a defined neighbourhood value (k) based on characteristic features of the training dataset. The optimal value of k determines the accuracy of the K-NN classification system. Finding the optimal k value requires iterative experimentation, which can be automated using GA to improve efficiency. K-NN and GA complement each other due to their effectiveness in processing non-linear data, such as audio signals. Moreover, GA excels in identifying infinite solution possibilities, including the optimal k value for K-NN.

Previous studies on music genre classification compared the performance of K-NN and SVM algorithms for classifying 13 types of Balinese gamelan music genres based on five spectral analyses: skewness, kurtosis, rolloff, centroid, and spread. The K-NN algorithm achieved an accuracy of 85.38% with k=3, while the SVM algorithm reached 66.9% [5]. The hybrid classification method combining K-NN and GA has been applied to classify moods evoked by music snippets using six spectral features: centroid, skewness, rolloff, kurtosis, spread, and flatness. With the same dataset, K-NN achieved a peak accuracy of 87.84% (k=14), while the K-NN optimized by GA yielded 89.19% (k=3). Although conventional K-NN performs well, the optimal k value remains a critical factor. Optimized K-NN with GA uses a smaller neighbourhood range, reducing ambiguity in classification [6].

This study will classify sub-genres of dangdut music into three labels: classic dangdut, rock dangdut, and *koplo* dangdut. A dataset related to these sub-genres will be collected, segmented, and prepared for classification system requirements. The resulting dataset will be used for training and testing to evaluate the performance of the classification system implementing Hybrid K-NN with GA.

This research contributes to MIR by advancing classification methods for non-linear audio data and to the field of Data Mining through the application of K-NN optimized with GA. For the public, the proposed classification model enhances the accessibility and management of dangdut music, potentially boosting interest and appreciation for its diversity. An accurate classification system enables dangdut audiences to discover similar songs based on audio characteristics, enriching their listening experience and satisfaction.

II. METHODS

This research aims to develop a classification system for dangdut music sub-genres using the Hybrid K-NN algorithm combined with GA (Genetic Algorithm). The music data is classified based on audio signal characteristics obtained through spectral analysis. The research methodology is broadly visualized in the following diagram:

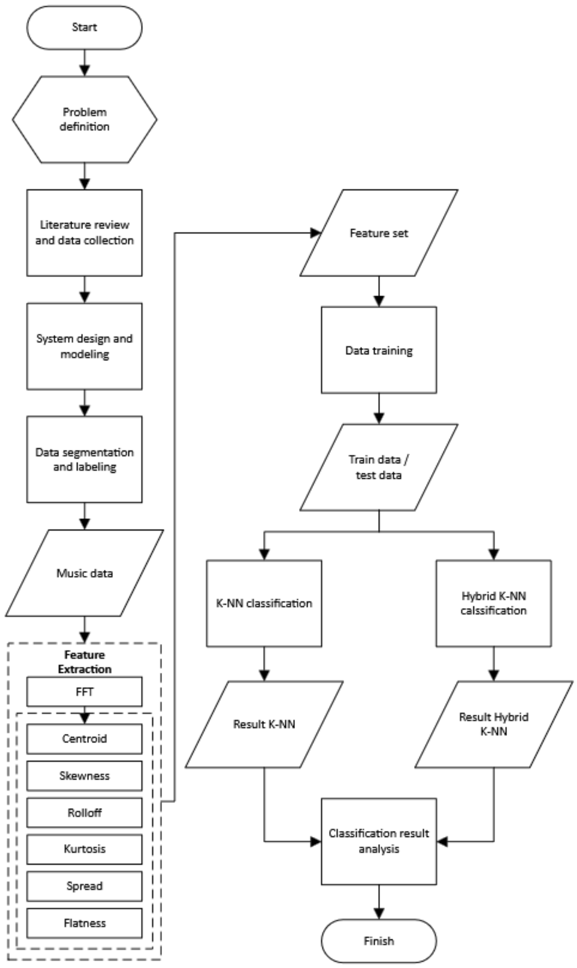

Figure 1. Research Method Flow

A. Problem Definition

The problem definition outlines the background and the issues addressed by the system, namely classifying dangdut music sub-genres using the Hybrid K-NN algorithm with Genetic Algorithm to automatically obtain the optimal k value, thereby streamlining the classification process and generating information about the type of dangdut sub-genre automatically through the system.

B. Literature Review and Data Collection

The initial stage involves a literature review on the issue, which in this case is the classification of dangdut music sub-genres divided into three class categories or labels: classic dangdut, rock dangdut, and koplo dangdut. This categorization is based on the periodization of dangdut proposed by Simatupang, an academic.

The first period, Malay rhythm dangdut, spans the 1950s to the 1960s and is characterized by Indian music influences, with representative musicians such as Ellya Khadam. The second period, Rhoma Irama dangdut, marks the genre's peak popularity in the 1970s to the 1990s, spearheaded by Rhoma Irama and the Soneta band with its distinctive rock nuances. The final period, koplo dangdut, emerged in the 1990s and continues to the present day, following the fall of Suharto. Deny Caknan represents the contemporary development of koplo dangdut [1].

The division of dangdut sub-genres results from instrumentation shifts influenced by technological advancements, creating diverse types of dangdut music. The most prominent differences are tempo and drum performance. The tempo in original dangdut music is more relaxed and structured, while digital era dangdut music tends to use faster tempos with a 4/4 rhythm, producing a denser and more lively sound. Drum performances in the digital era incorporate digital tools such as electronic drums or pads and often appear in a band format, contrasting with the original dangdut style, which features a Malay orchestra ambiance and traditional tabla drum instrumentation [2][7].

Based on the differences in the characteristics of dangdut sub-genres, music data is collected to compile training and testing datasets for the classification system. The data is gathered from various digital sources and selected based on specific artists and the distinctive features of their works, aiming to represent three main dangdut sub-genres: classic, rock, and koplo. The classification system uses Hybrid K-NN with GA to address the k-value issue in conventional K-NN. K-NN is inherently non-linear, meaning its decision boundaries are flexible as they are based on the nearest-neighbour distance, making it well-suited for classifying music data (audio signals) [6].

C. System Design and Modeling

At this stage, the design and modeling of two types of classification systems will be carried out. One system utilizes the Hybrid K-NN algorithm with GA, while the other uses the conventional K-NN algorithm. The goal is to compare the performance of these two algorithms in addressing the issue of determining the optimal k value to achieve the highest accuracy. The systems are developed using the Matlab application on a Windows operating system.

D. Data Segmentation and Labeling

The music data segmentation process is carried out using Adobe Audition to extract the most representative section of each song that reflects the characteristics of one of the dangdut sub-genres. Each segment has a duration of 30 seconds and is saved in mono .wav format. Each music segment represents one sub-genre label of dangdut music: classic dangdut, rock dangdut, or koplo dangdut. Specifically, the works of Ellya Khadam are labeled as classic dangdut, the works of Rhoma Irama are labeled as rock dangdut, and the works of Denny Caknan are labeled as koplo dangdut, based on the distinctive musical style and influence of each artist within the respective sub-genre.

E. Feature Extraction

The feature extraction stage is carried out to obtain specific characteristics from the music data. The .wav formatted data is input into the extraction system, starting with converting the audio signal from the time domain to the frequency domain using the FFT (Fast Fourier Transform) method. FFT transforms the sound signal into a digital form in the frequency spectrum, which is then analyzed using six types of spectral feature: centroid, skewness, rolloff, kurtosis, spread, and flatness. Spectral features are a set of audio characteristics that describe the distribution of frequency energy in a sound signal and play a crucial role in music analysis and classification. The main components include spectral centroid, which indicates the center of mass of the spectrum; spectral skewness, which measures the asymmetry of energy distribution; spectral rolloff, which shows the frequency limit where most of the energy is concentrated; spectral kurtosis, which describes the sharpness or peakedness of the distribution; spectral spread, which reflects the width of frequency dispersion around the centroid; and spectral flatness, which assesses the uniformity of the spectrum. Together, these features help distinguish sound characteristics such as brightness, complexity, and texture [8].

F. Data Training

The result of the feature extraction is a set of feature values from the six spectral analyses, which are then assigned label codes to represent one of the three dangdut music sub-genres. Code 1 represents classic dangdut, code 2 represents rock dangdut, and code 3 represents koplo dangdut. This data forms the training dataset, which will serve as the reference for determining the class of music data used to test the performance of the classification system.

G. K-NN Classification

The K-NN classification process begins with selecting the value of k (the number of nearest neighbours). When new data (test data) is provided, the distance between the test data and all the training data is calculated using the Euclidean metric. The test data is then classified into the majority class of its k nearest neighbours. The value of k is chosen manually or through experimentation, and the model's accuracy depends on selecting the appropriate k and the distribution of the data [9], [10]. The K-NN classification system in this study limits the value of k between 1 and 15. An illustration of the K-NN classification system can be seen in the following diagram:

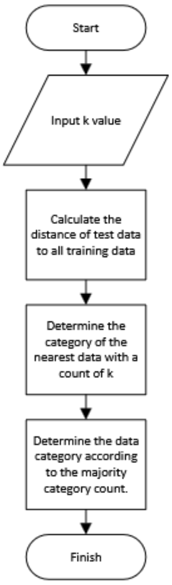

Figure 2. Flow of K-NN classification.

H. Hybrid K-NN Classification

The application of GA in Hybrid K-NN serves as an optimization method for determining the best k value to improve the performance of the classification system. GA operates during the data training process to obtain the optimal k value through the best individuals formed from three main genetic processes: crossover (swapping chromosomes), mutation (replacing one solution to increase population diversity), and selection (using solutions with high fitness values to pass on to the next generation [11]. As a random search method that mimics the principles of natural biological evolution, GA is suitable for use with nonlinear data, just like K-NN. The steps involved in the Hybrid K-NN with GA process are explained in the following diagram:

1. GA starts by creating an initial population, which is a set of chromosomes representing potential solutions from the training data.
2. Each chromosome is evaluated using a fitness function. Fitness in this study is measured based on cross-validation accuracy. The fitness value reflects the performance of the K-NN model on the training data.
3. The best chromosomes are selected to produce the next generation. Selection is done using the Roulette Wheel Selection method, where chromosomes with higher fitness have a greater chance of being selected.
4. After selection, crossover is used to combine pairs of chromosomes (parents) and generate new chromosomes (offspring).

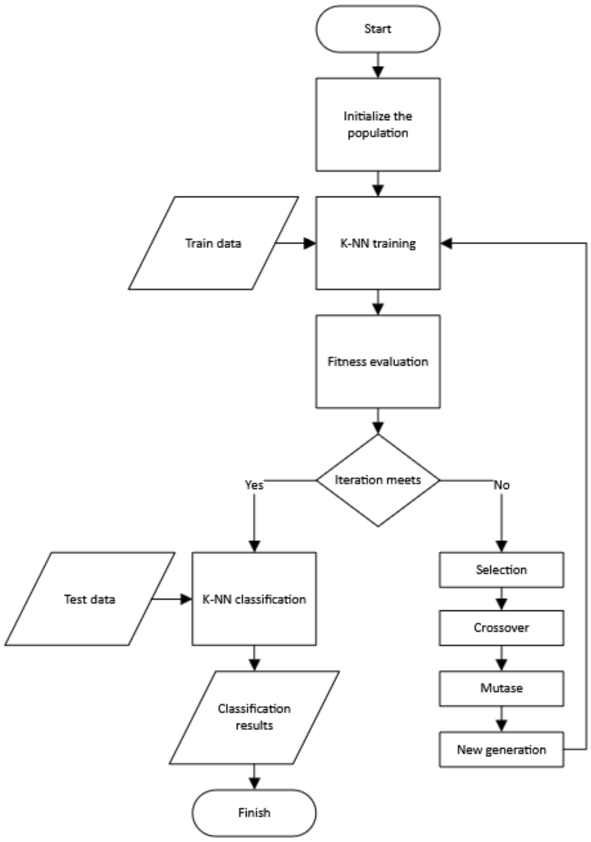

Figure 3. Flow of Hybrid K-NN classification.

5. Each gene (bit) in the chromosome has a small chance of mutation. The mutation probability used is 0.01, which helps prevent the algorithm from getting stuck in a local solution.
6. The new generation replaces the old population, and the selection-crossover-mutation process repeats until the maximum number of generations is reached.
7. After all generations are completed, the best chromosome is selected based on the highest fitness value.
8. K-NN is trained with the optimal k value from GA, and its performance is evaluated on the test data.

I. Classification Result Analysis

In this stage, both the Hybrid K-NN classification system and the conventional K-NN classification system will be tested for their accuracy in classifying dangdut music sub-genres. The accuracy level is determined by measuring the comparison between the number of correctly classified test data or correct answers produced by the classification system, and the total number of test data [12]. The system with the highest accuracy will be considered the best classification system.

III. RESULTS AND DISCUSSION

A. Dataset

Based on the findings from the literature review, which divides dangdut music sub-genres into three categories, music data from Ellya Khadam, Rhoma Irama, and Denny Caknan was collected. These three artists each represent different eras of dangdut music with distinct characteristics. Ellya Khadam represents the classical dangdut era, characterized using kendang and music influenced by Indian adaptations. Rhoma Irama represents the rock music experiment era, harmonizing rock elements with traditional dangdut music to create a new sub-genre known as dangdut rock. Denny Caknan represents the contemporary dangdut music with a koplo style that is currently popular among the public.

From the 136 collected dangdut songs obtained from various digital music sources, a total of 483 music segments were extracted. These segments, each with a duration of 30 seconds, were labeled with one of the dangdut sub-genres: classical, rock, or koplo. During the labeling process, the research team was assisted by a music expert to ensure the accuracy of the labels. Accurate labeling is crucial for training the classification model with correct data, enabling it to effectively recognize the characteristic patterns of each sub-genre.

```
9085.9 27.345 12732 836.14 4820.3 0.099297 1
9350.7 28.559 12791 920.67 5105 0.14641 1
3616.2 28.907 10858 902.82 4325.2 0.066041 1
2667.3 27.919 9581.3 842.95 3090.8 0.096488 2
4997.3 26.675 10179 785.39 3705.1 0.1034 2
3063.1 25.084 8679.4 753.65 3231.3 0.13404 2
1700.9 26.4 10881 786.88 3139.5 0.21677 3
4749.7 28.266 11936 899.44 4804.1 0.19527 3
1645.1 26.803 10880 835.24 3254.8 0.22041 3
```

Figure 4. Feature extraction database from six types of spectral analysis along with class codes.

Each label category consists of 161 data entries, which were then extracted to obtain feature values from six types of spectral analysis: centroid, skewness, rolloff, kurtosis, spread, and flatness. These feature values were stored in a database along with their respective class category codes. Code 1 represents the classical class, code 2 represents the rock class, and code 3 represents the koplo class. This dataset was then divided into training and testing sets for the next step of testing the performance of the K-NN and Hybrid K-NN classification algorithms.

The centroid feature values have a wide range, from approximately 714.78 to 21,104, with an average of 4,011.24. Skewness ranges from 19.45 to 33.88, with an average of 27.17. Other features show a normal distribution of data without extreme values. However, the feature values for each class category have significant density, which could potentially affect the performance of the classification system [13].

B. Classification System

The application of the Hybrid K-NN algorithm with GA in the classification system shows that the integration of these two methods can improve the model's performance by automatically selecting the optimal k parameter. This process is carried out by utilizing GA's ability to efficiently explore the search space through selection, crossover, and mutation, thus avoiding dependence on manual k selection. In this study, the dataset was split into a training set and a testing set with a ratio of 7:3, where 70% of the data was used for training the model and 30% for testing its performance. The results of the performance test of the Hybrid K-NN classification system with the configuration of several parameters are shown in Table 1 below.

TABEL I
TEST RESULTS OF HYBRID K-NN PARAMETERS

Pop Size	Max Gen	Amt Gen	Amt Fold	Best K	Accuracy (%)	Time (s)
30	20	4	10	11	70,83	111,47
30	20	4	5	12	69,44	60,70
30	20	4	3	10	70,14	19,23
20	20	4	3	5	71,53	12,55
15	20	4	3	12	69,44	9,61
10	20	4	3	3	72,22	6,80
10	15	4	3	4	74,31	4,86
10	5	4	3	3	72,22	1,63

The choice of parameters in the GA operation is critical to ensure optimal performance and efficiency in the solution search process. The population size (pop size) determines the number of individuals (chromosomes) in each generation. A larger population size expands the search space but increases computation time. The number of generations (max gen) specifies how many times the population evolves through selection, crossover, and mutation. More generations allow for a more optimal solution search but may prolong the process time. The number of genes (amt gen) in each chromosome reflects the bit representation of the value being optimized, such as the k parameter in K-NN. The gene length should be sufficient to cover the parameter value range with adequate precision. In this study, a 4-bit length was applied, providing enough range to select k values between 1 and 15. The number of folds (amt fold) in cross-validation determines how the training data is divided for model evaluation. More folds yield more accurate evaluations but risk longer processing times [14].

Genetic Algorithm (GA) is used to find the best k value for KNN by optimizing prediction accuracy. The fitness function is calculated as the KNN accuracy on the test data — the higher the accuracy, the better the chromosome (representing k). The algorithm starts with an initial population of candidate solutions and evolves them over several generations. In each generation, crossover combines two chromosomes to produce new ones, while mutation randomly flips bits in a chromosome to maintain diversity and avoid local optima. The proper combination of GA parameters ensures that GA can explore and exploit the solution space efficiently to find the best parameter. The optimal k value was achieved with a combination of a population size of 10, 15 generations, a gene length of 4 bits, and 3-fold cross-validation, yielding the highest accuracy of 74.31% for k=4 and a processing time of 4.86 seconds. This accuracy reflects the model's ability to classify data reasonably well on the test dataset.

The performance of the Hybrid K-NN algorithm was then compared to the conventional K-NN algorithm using the same training and testing datasets. Table 2 provides the classification system performance results for the conventional K-NN algorithm.

TABEL 2
RESULTS OF CONVENTIONAL K-NN TESTING

K	Accuracy (%)	Time (s)
1	59,72	0,02
2	59,72	0,02
3	63,19	0,03
4	68,75	0,04
5	63,89	0,03
6	63,19	0,04
7	63,89	0,04
8	62,50	0,03
9	60,42	0,03
10	63,89	0,04
11	60,42	0,03
12	62,50	0,03
13	61,81	0,03
14	61,11	0,03
15	61,11	0,03

Through several trials using different kk values within the range of 1 to 15, the K-NN classification achieved the highest accuracy of 68.75% for k=4 with a processing time of 0.04 seconds. In a small dataset population, such as the 483 data points used in this study, k=4 is considered effective because it provides relatively sharp decisions compared to larger kk values. With k=4, the algorithm considers the 4 nearest neighbours to determine the class, capturing local patterns effectively, especially if the dataset has uneven distribution. The classification performance on the test data showed low accuracy, indicating potential overfitting, where the model performs well on the training data but struggles to generalize to new, unseen data [15]. Nevertheless, the model's execution time for classification was efficient.

Both the Hybrid K-NN and conventional K-NN algorithms achieved their best performance at k=4, but their accuracy results differed. Despite using the same training and testing datasets, Hybrid K-NN delivered a higher accuracy of 74.31%, compared to 68.75% for conventional K-NN. This discrepancy in accuracy is likely due to the use of Genetic Algorithm (GA) in Hybrid K-NN, which evaluates the model through k-fold cross-validation. This method provides a more comprehensive understanding of the model's performance on the training data, enabling GA to identify solutions that generalize better to test data and thus improve accuracy.

However, the Hybrid K-NN classification required a longer execution time compared to conventional K-NN. This is due to the iterative process of finding the optimal solution in Hybrid K-NN, which involves fitness evaluation, selection, crossover, mutation, and cross-validation.

The performance of Hybrid K-NN was also compared with two multiclass classification methods on the same dataset with a 70% training and 30% testing split. Decision Tree achieved a test accuracy of 61.11% with a very fast execution time (0.42 seconds), indicating high efficiency but moderate generalization ability. Meanwhile, SVM (Support Vector Machine**)** based on ECOC with a linear kernel achieved a test accuracy of 54.86% with an execution time of 1.99 seconds, showing less optimal performance on the test data. These results indicate that the Hybrid K-NN method is more effective in adjusting parameters and recognizing data patterns compared to SVM and Decision Tree.

The Hybrid K-NN model was then validated using 5-fold cross-validation. GA was executed with parameters: initial population of 10, 15 generations, and chromosome length of 4 bits. In each fold, the model was trained and tested, then evaluated using accuracy, precision, recall, F1-score, and the confusion matrix to comprehensively assess its performance. The average results showed an accuracy of 72.04% with a standard deviation of 2.22%, while the average precision, recall, and F1-score were all around 0.72, reflecting stable and balanced performance in recognizing the three classes.

	Pred1	Pred2	Pred3
True1	124	24	13
True2	40	93	28
True3	18	12	131

Figure 5. Confusion matrix result.

From the accumulated confusion matrix, the most frequent misclassification occurred in class 2, which was often classified as class 1 as many as 40 times, indicating a challenge in distinguishing between these two classes. Overall, the Hybrid K-NN with GA proved effective in optimizing parameters and producing a reliable classification model.

IV. CONCLUSION

The implementation of Hybrid K-NN in the classification system for dangdut sub-genres demonstrates a significant improvement in accuracy and generalization compared to conventional K-NN, SVM, and Decision Tree. By optimizing the k parameter using a Genetic Algorithm (GA) with parameters (population = 10, generations = 15, chromosome length = 4 bits) and validating the model through 5-fold cross-validation, Hybrid K-NN achieved an average accuracy of 72.04% (up to 74.31% on the test set), outperforming conventional K-NN at k=4 (68.75%), SVM with ECOC (54.86%), and Decision Tree (61.11%) on the same dataset. Although Hybrid K-NN requires longer processing time (4.86 seconds) than the others — particularly conventional K-NN (0.04 s), SVM (1.99 s), and Decision Tree (0.42 s) — its improved precision, recall, F1-score, and ability to effectively capture data patterns make it a more reliable model overall. Despite some challenges in distinguishing certain classes (e.g., class 2 often misclassified as class 1), the integration of GA into K-NN proves highly effective in optimizing parameters and delivering superior classification performance compared to the other methods tested.

ACKNOWLEDGMENTS

Thank you to the Institute of Technology and Business (ITB) STIKOM Bali for its support in facilitating this research process.

REFERENCES

[1] D. Setiaji, “Tinjauan Karakteristik Dangdut Koplo Sebagai Perkembangan Genre Musik Dangdut,” *Handep (Pasca Sarjana institut Seni Indonesia Surakarta)*, vol. 1, pp. 19–34, Dec. 2017.

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

[3] B. R. Ismanto, T. M. Kusuma, and D. Anggraini, “Indonesian Music Classification on Folk and Dangdut Genre Based on Rolloff Spectral Feature Using Support Vector Machine (SVM) Algorithm,” *IJCCS (Indonesian Journal of Computing and Cybernetics Systems)*, vol. 15, no. 1, p. 11, Jan. 2021, doi: 10.22146/ijccs.54646.

[4] X. Li, F. Li, Z. P. Lu, and Z. Yang, “Music Genre Classification: A Comprehensive Study on Feature Fusion with CNN and MLP Architectures,” 2025, doi: 10.54254/2755-2721/132/2024.20632.

[5] I. Gede 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

[6] 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.

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

[8] A. Lerch, *An Intoduction to Audio Content Analysis: Music Information Retrieval, Task and Application*, Second Edition. New Jersey: John Wiley & Sons, Inc., 2023.

[9] W. Andriyani *et al.*, *Matematika Pada Kecerdasan Buatan*, Pertama. Makassar: CV. Tohar Media, 2024.

[10] W. Andriyani *et al.*, *Data Sebagai Fondasi Kecerdasan Buatan*, Pertama. Makassar: CV. Tohar Media, 2024.

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

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

[13] J. Eronen, M. Ptaszynski, F. Masui, G. Leliwa, and M. Wroczynski, “Exploring the Potential of Feature Density in Estimating Machine Learning Classifier Performance with Application to Cyberbullying Detection,” in *CEUR Workshop Proceedings 2935*, 2022. doi: 10.48550/arXiv.2206.01949.

[14] C. Mahlich, T. Vente, and J. Beel, “From Theory to Practice: Implementing and Evaluating e-Fold Cross-Validation,” in *International Conference on Artificial Intelligence and Machine Learning Research (CAIMLR 2024)*, Singapore, 2024.

[15] B. Ghojogh and M. Crowley, “The Theory Behind Overfitting, Cross Validation, Regularization, Bagging, and Boosting: Tutorial,” May 2023, [online]. Available: http://arxiv.org/abs/1905.12787