Journal of Applied Informatics and Computing (JAIC)
Vol.10, No.1, February 2026, pp. 309~315
e-ISSN: 2548-6861

Performance Comparison of Naive Bayes and Support Vector Machine Methods in Music Genre Classification Based on Audio Signal Feature Extraction Using Mel-Frequency Cepstral Coefficients (MFCC)

Kevin Putrayudha Naserwan [1*], [+]Kanda Januar Miraswan [2*], Meylani Utari [3]**
[*] Informatics Engineering, Faculty of Computer Science, Universitas Sriwijaya
[**] Computerized Accounting, Faculty of Computer Science, Universitas Sriwijaya
kevinnaserwan@gmail.com [1], kandajm@ilkom.unsri.ac.id [2], meylani@ilkom.unsri.ac.id [3]
[+]corresponding author : kandajm@ilkom.unsri.ac.id

Article Info

Article history:

Received 2025-11-13
Revised 2026-01-07
Accepted 2026-01-17

Keyword:

Naïve Bayes,
Support Vector Machine,
Music Genre Classification,
Mel-Frequency Cepstral Coefficients,
Audio Signal Processing.

ABSTRACT

Music genre classification has gained increasing attention with the emergence of digital music platforms. One of the relevant features extracted from audio signals is Mel-Frequency Cepstral Coefficients (MFCC), which is widely recognized as an effective technique. MFCC features are extracted at the frame level and aggregated at the clip level to represent each music track, making them suitable for audio-based classification tasks. This study applies Naïve Bayes and Support Vector Machine (SVM) algorithms for classification using the GTZAN dataset consisting of 1,000 audio files from 10 music genres, each with a duration of 30 seconds. The performance of these methods is evaluated using accuracy, precision, recall, and F1-score. The results show that SVM demonstrates superior performance, achieving an accuracy of 95.25% compared to 50.37% for Naïve Bayes. This performance gap can be attributed to SVM's ability to model non-linear decision boundaries and effectively handle high-dimensional MFCC feature spaces. The main contribution of this study lies in the systematic evaluation of multiple SVM kernel configurations and parameter settings, providing empirical insights into the robustness of classical machine learning methods for MFCC-based music genre classification. This study concludes that SVM is better than Naive Bayes in music genre classification with MFCC feature extraction.

I. INTRODUCTION

Music classification, particularly in the digital age, presents significant challenges due to numerous music platforms utilizing technology to recommend genres based on user preferences [1] [2]. Each platform categorizes music into various genres, making it easier for users to choose suitable types of music [3] [4]. However, classifying music genres based on audio signals remains a complex problem requiring advanced technological approaches [5] [6]. Audio signal extraction is crucial in music genre classification processes [7]. These challenges arise from overlapping acoustic characteristics between genres, non-linear relationships between audio features and genre labels, and the high dimensionality of extracted features [8].

Two commonly employed methods are Mel-Frequency Cepstral Coefficients (MFCC) and spectrograms. While spectrograms offer a visual representation of frequency spectrum changes over time, MFCC converts spectral frequencies into a scale more aligned with human auditory perception [7]. Despite having advantages, MFCC has proven more effective in many audio processing applications, including music genre classification [3]. Studies like those conducted by Tzanetakis and Cook show that MFCC outperforms other spectral features in music genre classification tasks [9] [10]. However, MFCC primarily captures short-term spectral information and may not fully represent long-term temporal or rhythmic characteristics of music signals [5]. In contrast, spectrogram-based features provide richer time–frequency information but often require higher computational complexity and more sophisticated classification models [11].

MFCC is widely popular in audio signal processing because it captures acoustic features similar to how humans perceive sound [9]. This technique is frequently used in speech recognition and music identification systems, making it an excellent choice for our research focusing on audio-based music genre classification [12]. In this study, MFCC features are extracted at the frame level and aggregated at the clip level to represent each music track, enabling efficient and compact feature representation for classification tasks.

In the context of machine learning algorithms, Naive Bayes and Support Vector Machines (SVM) are two often-used methods. Naive Bayes operates effectively on simple classification problems and excels in terms of simplicity and implementation speed, although it assumes feature independence [13] [12]. When applied to music genre classification, this independence assumption can become a limitation, as MFCC coefficients are often correlated across frequency bands and time frames [3]. On the other hand, SVM is more complex yet highly efficient for high-dimensional datasets by identifying hyperplanes maximizing separation margins [6] [14] [15]. This characteristic makes SVM particularly suitable for MFCC-based music genre classification, where non-linear decision boundaries frequently occur [16].

Given the importance of comparing these algorithms' performance in real-world scenarios, understanding their strengths and weaknesses becomes essential [17]. Previous studies indicate varied outcomes depending on dataset complexity - Naive Bayes performs well with smaller datasets whereas SVM tends to excel with larger ones [16]. Therefore, evaluating the effectiveness of both methods using MFCC-extracted features is critical for determining the best algorithmic choice for automatic music classification systems. Nevertheless, many existing studies focus on introducing new features or advanced deep learning models, while systematic comparisons of classical machine learning algorithms under multiple parameter configurations using the same feature representation remain limited.

Therefore, this study aims to address this gap by providing a focused and systematic comparison between Naive Bayes and Support Vector Machine for music genre classification using MFCC features. Experiments are conducted on the GTZAN dataset, a widely used benchmark in music genre classification research. The main contributions of this work include: (1) an empirical evaluation of multiple SVM kernel functions and parameter configurations, (2) a direct performance comparison with Naive Bayes to assess robustness and stability, and (3) an analysis of genre-level misclassification patterns using confusion matrices.

II. METHOD

This section outlines the methodology and workflow employed in conducting the comparative performance analysis of the Naive Bayes and Support Vector Machine (SVM) methods. It begins with an explanation of the research framework that provides an overview of the structure and flow of this comparative study. Subsequently, this chapter details the systematic steps taken from start to finish, including feature extraction using Mel-Frequency Cepstral Coefficients (MFCC), implementation of both classification methods, and evaluation and comparison techniques used in music genre classification based on audio signals.

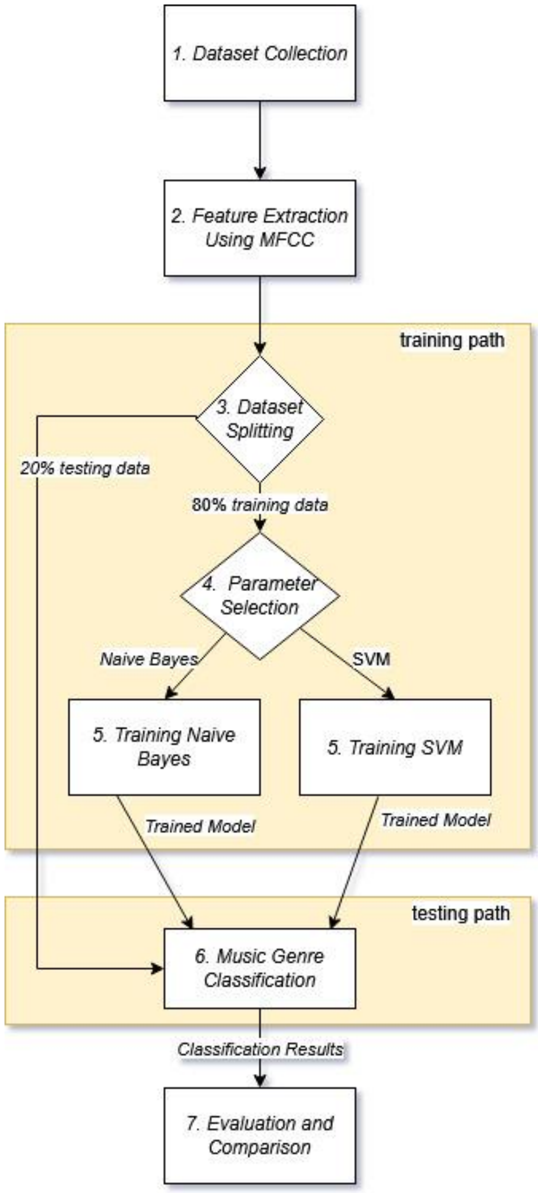

Figure 1. Music Genre Classification Workflow

A. Data Collection

Data collection is a critical stage in this research, where audio data from various music genres is gathered for further analysis and classification.

The data utilized in this research consists of digital audio files with the following specifications:

a. File Format: WAV (Waveform Audio File Format)
b. Duration: 30 seconds per file
c. Total Number of Files: 1000
d. Distribution: 100 files for each of the 10 music genres

The data for this research is sourced from the GTZAN dataset, a standard dataset widely used in various studies for evaluating music genre classification algorithms. This dataset comprises 1000 audio files, each lasting 30 seconds, divided into 10 different music genres, with each genre represented by 100 audio samples. The genres include blues, classical, country, disco, hip-hop, jazz, metal, pop, reggae, and rock. The audio files in this dataset are in WAV format with a sample rate of 22050Hz mono 16-bit. This dataset was chosen due to its public availability and its status as a benchmark standard in music genre classification research.

B. Research Framework

This section details the sequence of research processes comprehensively through a diagram that provides an overview of the steps involved in this study comparing the performance of Naive Bayes and SVM methods in music genre classification based on MFCC feature extraction.

This section details the sequence of research processes comprehensively through a diagram that provides an overview of the steps involved in this study comparing the performance of Naive Bayes and SVM methods in music genre classification based on MFCC feature extraction.

a. Dataset Collection: Gathering the dataset to be used for music genre classification.
b. Feature Extraction Using MFCC: Extracting features using MFCC to obtain a numerical representation suitable for classification.
c. Dataset Splitting: Dividing the extracted dataset into training (80%) and testing (20%) sets to ensure fair evaluation.
d. Parameter Selection: Choosing parameters for both Naive Bayes and SVM models to optimize classification performance.
e. Training Models: Training both models using the training dataset.
f. Music Genre Classification: Using trained models to classify music genres based on test data.
g. Evaluation and Comparison: Evaluating results from both models using metrics such as accuracy, precision, recall, and F1-score.

C. MFCC Feature Extraction

In this study, MFCC features are extracted using a frame-based approach to capture the short-term spectral characteristics of audio signals. Each audio file is segmented into overlapping frames using a window size of 2048 samples and a hop length of 512 samples. From each frame, 13 MFCC coefficients are computed. To incorporate dynamic information, first-order (delta) and second-order (delta-delta) derivatives are also calculated, resulting in a total of 39 MFCC-related features per frame.

The frame-level MFCC features are then aggregated at the clip level by computing the mean value of each coefficient across all frames. This aggregation process produces a fixed-length feature vector for each music track, which is suitable for input into the classification models.

D. Classification Method

This research employs two classification algorithms: Naive Bayes and Support Vector Machine (SVM).

Naive Bayes is a probabilistic classification method based on Bayes' theorem and the assumption of feature independence. In this study, the Gaussian Naive Bayes variant is applied due to its suitability for continuous-valued features such as MFCC. Several values of the var smoothing parameter are evaluated to observe their effect on classification performance.

Support Vector Machine (SVM) is a supervised learning algorithm that constructs an optimal hyperplane to separate data points from different classes. In this research, SVM is evaluated using multiple kernel functions, including radial basis function (RBF), polynomial, and sigmoid kernels, with different parameter configurations (such as C, gamma, and degree) to examine their influence on classification performance.

E. Dataset Splitting and Evaluation

The dataset is divided into training and testing sets using an 80:20 ratio, with stratified sampling applied to ensure that each music genre is proportionally represented in both sets. This approach helps prevent class imbalance and provides a more reliable evaluation of classification performance.

Model performance is evaluated using accuracy, precision, recall, and F1-score. In addition, confusion matrices are generated for selected model configurations to analyze genre-level classification behavior and misclassification patterns.

F. Software Development Methodology

The software development method implemented in this research is the Waterfall Development Methodology due to its linear and step-by-step nature suitable for projects with clear stages from start to finish.

a. Requirements Analysis : Defining system requirements necessary for developing music genre classification software.
b. System Design : Designing software architecture based on analyzed requirements.
c. Implementation : Coding the system according to the design using Python with libraries such as Librosa for feature extraction and Scikit-learn for classification algorithms.

This structured methodology ensures that each aspect of the research is meticulously planned and executed to facilitate accurate comparisons between Naive Bayes and SVM in music genre classification tasks based on audio signal features extracted through MFCC techniques.

III. RESULTS AND DISCUSSION

In this research, nine different experimental configurations were conducted to compare the performance of Naive Bayes and SVM in music genre classification. Each configuration had different parameters.

Before explaining the experimental configuration, it's important to understand the composition and distribution of the dataset used:

a. Total number of samples: 1000
b. Number of music genres (categories): 10
c. Samples per genre: 100
This dataset is divided into a training set and a test set with an 80:20 ratio, using the stratified sampling method:
a. Training set: 800 samples (80 samples per genre)
b. Test set: 200 samples (20 samples per genre)

Stratified sampling was applied to ensure that the proportion of samples for each genre was maintained in both sets. This is important to ensure balanced and representative evaluation for all genres.

TABLE I
EXPERIMENTAL CONFIGURATION

Config.	Model	Parameter	Accuracy	Precision	Recall	F1-score
I	Naive Bayes	*var_smoothing*:1e-9 (default)	49,25%	0,4848	0,4925	0,4828
II	Naive Bayes	*var_smoothing*:1e-2	50,37%	0.4994	0.5037	0.4921
III	Naive Bayes	*var_smoothing*:1e-5	49.62%	0.4892	0.4962	0.4868
Average			49,74%	0,4911	0,4975	0,4873
IV	SVM	*kernel*: rbf , C: 1.0, *gamma*: 'scale' (default)	82.25%	0.8256	0.8225	0.8218
V	SVM	*kernel*: rbf , C: 10, *gamma*: 0.1	94.87%	0.9500	0.9487	0.9488
VI	SVM	*kernel*: 'poly', C:1, degree: 3	78.62%	0.8312	0.7862	0.7956
VII	SVM	*kernel*: 'sigmoid', C: 1, *gamma*: 'auto'	35.5%	0.3390	0.3550	0.3422
VIII	SVM	*kernel*: rbf, C: 10, *gamma*: 0.07	95.25%	0.9531	0.9525	0.9525
IX	SVM	*kernel*: rbf, C: 100, *gamma*: 1	57.37%	0.9167	0.5737	0.6469
Average			73,97%	0,4911	0,7397	0,7513

From the test results, the final results of a series of experiments conducted in this study are presented. Each model configuration for both Naive Bayes and Support Vector Machine (SVM) is evaluated based on four key performance metrics: Accuracy, Precision, Recall, and F1-score.

The table below summarizes the results from the nine configurations that were tested, covering parameter variations for both models. In the Naive Bayes model, the var_smoothing parameter is adjusted to observe the effect of smoothing on model performance. Meanwhile, in the SVM model, various. The combination of kernel and parameters such as C and gamma is optimized to achieve the best performance.

From the experiments conducted, it is clear that SVM consistently outperforms Naive Bayes in the task of music genre classification. The per-configuration analysis provides several important results:

A. Naive Bayes Performance

Configurations I to III show that Naive Bayes with various var_smoothing values yields relatively low accuracy, ranging from 49.25% to 50.37%.

The precision, recall, and F1-score metrics also indicate suboptimal performance, with the highest precision value being only 0.50 and the F1-score around 0.48.

The confusion matrix for Naive Bayes indicates that this model often misclassifies genres with similar acoustic characteristics. Naive Bayes is more suitable for simple classification tasks, but its performance is limited for high-complexity data such as music genres. This limitation is mainly caused by the strong independence assumption of Naive Bayes, which is less appropriate for MFCC features that exhibit correlations across coefficients and time frames.

The average accuracy of the three Naive Bayes configurations is 49.74%, with a precision of 0.4911, recall of 0.4975, and F1-score of 0.4873.

B. SVM Performance

SVM significantly outperformed Naive Bayes, particularly in configurations IV thru IX. The default SVM configuration (RBF kernel, C=1, gamma='scale') yielded an accuracy of 82.25%, which was significantly better than Naive Bayes.

The best configuration was found in configuration VIII with an RBF kernel, C=10, and gamma=0.07, resulting in an accuracy of 95.25%. In addition to accuracy, the precision, recall, and F1-score for this configuration were also very high (above 0.95), indicating SVM's excellent ability to classify genres.

In other configurations, such as configuration VII with a sigmoid kernel, SVM performance dropped drastically with an accuracy of 35.5%. This indicates that the choice of kernel significantly affects SVM performance, and the sigmoid kernel may be less suitable for music genre classification tasks.

The Confusion Matrix from the best configurations (V and VIII) shows minimal error distribution, indicating SVM's ability to effectively distinguish genres with similar characteristics. However, analysis of the confusion matrices reveals that misclassifications still occur between genres with overlapping acoustic characteristics, such as rock and metal, as well as disco and pop. In contrast, genres with more distinctive timbral characteristics, such as classical and jazz, are classified more accurately.

The average accuracy for all SVM configurations is 73.97%, with a precision of 0.4911, recall of 0.7397, and F1-score of 0.7513.

C. *Parameter Influence*

Changes in SVM parameters, such as the values of C and gamma, have a significant impact on performance. Higher C values and appropriate gamma for the RBF kernel provide optimal results. Conversely, the sigmoid kernel (configuration VII) and the polynomial kernel (configuration VI) are unable to produce optimal results, with lower accuracy.

It should be noted that the number of experimental configurations evaluated for SVM is greater than that of Naive Bayes. This design choice reflects the fact that SVM performance is highly sensitive to kernel selection and parameter tuning, whereas Naive Bayes has a more limited set of tunable parameters. While this difference may introduce an imbalance in experimental exploration, the comparison remains valid in highlighting the relative performance potential of both algorithms under optimized conditions.

Selecting the RBF kernel with optimized parameters (C=10 and gamma=0.07) proved to be the most effective combination for the music genre classification task.

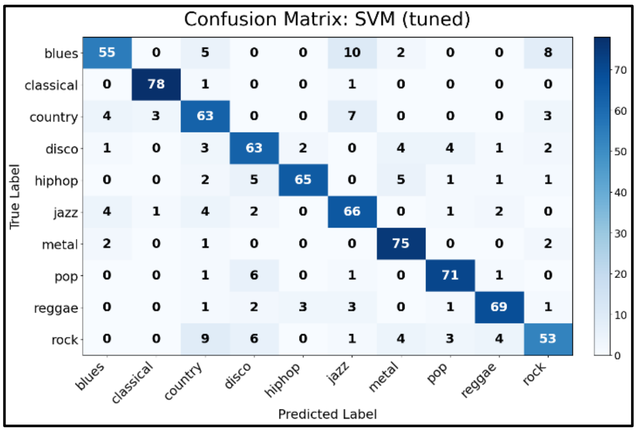

Figure 2. Confusion Matrix Configuration IV

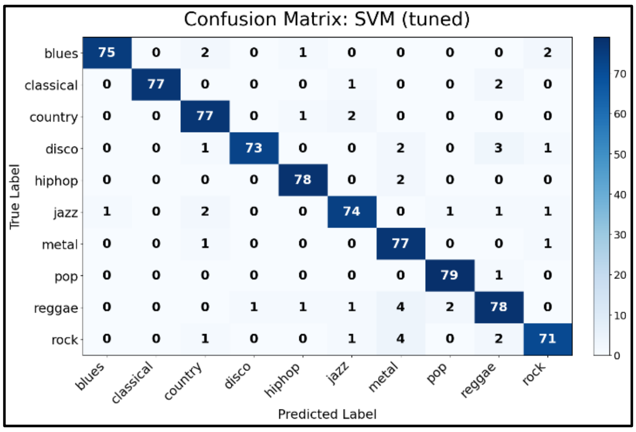

Figure 3. Confusion Matrix Configuration V

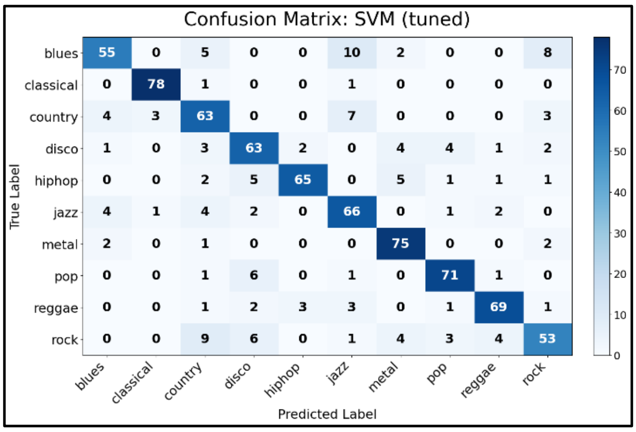

Figure 4. Confusion Matrix Configuration VI

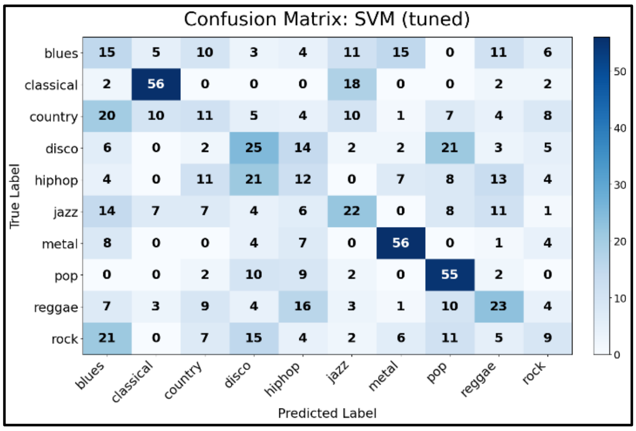

Figure 5. Confusion Matrix Configuration VII

From a scientific perspective, this research reinforces the suitability of margin-based classifiers such as SVM for audio classification tasks where feature correlations and non-linear class boundaries are prevalent. Although deep learning approaches have gained popularity in recent years, this study shows that classical machine learning models remain competitive and relevant when carefully configured and systematically evaluated.

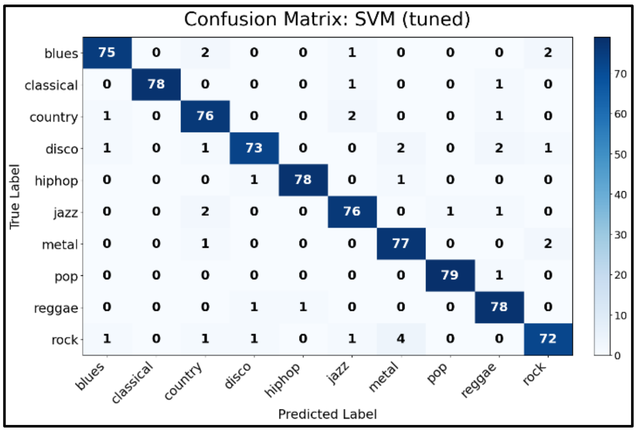

Figure 6. Confusion Matrix Configuration VIII

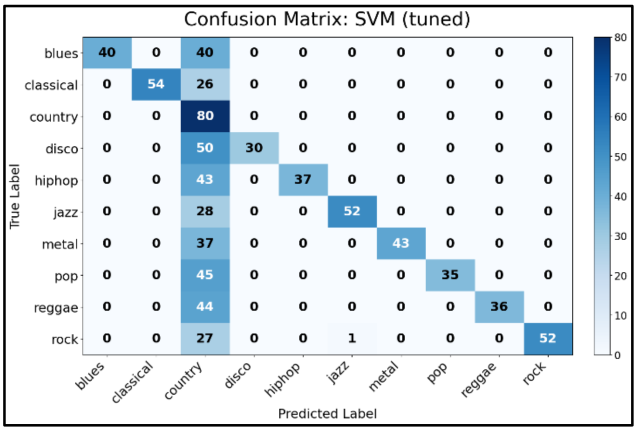

Figure 7. Confusion Matrix Configuration IX

In practical terms, the results of this study suggest that SVM can be effectively applied in real-world music information retrieval systems, such as music recommendation platforms and digital music libraries, where reliable genre classification based on audio content is required.

Despite the promising results, this study has several limitations. The experiments were conducted using a single benchmark dataset (GTZAN), and feature extraction was limited to MFCC without incorporating additional audio descriptors such as rhythmic or harmonic features. Furthermore, the comparison was restricted to classical machine learning algorithms and did not include deep learning-based models.

Future research may address these limitations by exploring additional audio features, evaluating more advanced classification models, and testing the proposed approach on larger and more diverse datasets. Further analysis of genre-level confusion patterns may also contribute to improving feature design and classification performance.

IV. CONCLUSION

Based on the experimental results and analysis conducted in this study, several key conclusions can be drawn regarding music genre classification using Mel-Frequency Cepstral Coefficients (MFCC) features with two classification models: Naive Bayes and Support Vector Machine (SVM).

Overall, SVM consistently outperformed Naive Bayes in MFCC-based music genre classification, both in terms of peak performance and average evaluation metrics. Based on experimental results, SVM with an RBF kernel (C=10, gamma=0.07) provided the best performance with an accuracy of 95.25%. SVM consistently outperformed Naive Bayes, which only achieved a maximum accuracy of 50.37%. These results indicate that SVM is more effective in handling high-dimensional and non-linear audio feature spaces than probabilistic models with strong feature independence assumptions.

Average performance difference between Naive Bayes and SVM. The average accuracy achieved by Naive Bayes across all configurations was only 49.748%, while the average accuracy of SVM across all configurations reached 73.97%. In addition to accuracy, SVM also demonstrated more stable and consistent performance across precision, recall, and F1-score metrics when compared to Naive Bayes. This consistency indicates SVM's superior ability to generalize across different music genres and reduce misclassification errors, particularly for genres with complex acoustic characteristics.

SVM is more consistent in evaluation metrics other than accuracy. SVM showed consistent performance in terms of precision, recall, and F1-score across all of its best configurations, with a precision value of 0.95. SVM can avoid errors in prediction better than Naive Bayes, providing more balanced results in classifying various music genres.

Error distribution in SVM is more concentrated based on Confusion Matrix analysis. SVM shows fewer errors and they are more focused on genres that are difficult to distinguish, while Naive Bayes shows a wider error distribution, especially for genres with similar features.

The influence of SVM parameters on the performance of selecting the correct parameters, such as the values of C and gamma in the RBF kernel, significantly affects SVM performance. The optimal parameter combination (C=10, gamma=0.07) yielded the best results in music genre classification.

REFERENCES

[1] Y. Zhang and T. Li, "Music genre classification with parallel convolutional neural networks and capuchin search algorithm," *Scientific Reports*, vol. 15, no. 1, p. 9580, 2025.

[2] R. A. Oladejo, A. A. Abayomi-Alli, O. Arogundade, A. A. Adeyanju and M. O. Lawrence, " Artificial Intelligence in Music Genre Classification: A Systematic Review of Techniques, Applications, and Emerging Trends," *Cureus Journal Of Computer Science*, 2025.

[3] A. Alamsyah, F. Ardiansyah and A. Kholiq, "Music Genre Classification Using Mel Frequency Cepstral Coefficients and Artificial Neural Networks: A Novel Approach," *Sceintific Journal of Informatics,* vol. 11, no. 4, pp. 937-948, 2024.

[4] P. Ghosh, S. Mahapatra, S. Jana and R. K. Jha, "A Study on Music Genre Classification using Machine Learning," *International Journal of Engineering Business and Social Science,* vol. 1, no. 4, pp. 257-268, 2023.

[5] T. Pratiwi, A. Sunyoto and D. Ariatmanto, "Music Genre Classification Using K-Nearest Neighbor and Mel-Frequency Cepstral Coefficients," *Sinkron : Jurnal Dan Penelitian Teknik Informatika,* vol. 8, no. 2, pp. 861-867, 2024.

[6] L. Abdoune, M. Fezari and A. Dib, "Indoor Sound Classification with Support Vector Machines: State of the Art and Experimentation," *International Journal of Computational Methods and Experimental Measurements,* vol. 12, no. 3, pp. 269-279, 2024.

[7] M. Müller, Fundamentals of Music Processing : : Audio, Analysis, Algorithms, Applications, Switzerland: Springer, 2015.

[8] N. Asanah and I. Pratama, "Deep Learning Approach for Music Genre Classification using Multi Feature Audio Representations," *Sistemasi: Jurnal Sistem Informasi,* vol. 14, no. 5, pp. 2045-2054, 2025.

[9] G. Tzanetakis and P. R. Cook, "Musical Genre Classification of Audio Signals," *IEEE Transactions on Speech and Audio Processing,* vol. 10, no. 5, pp. 293 - 302, 2002.

[10] S. Aggarwal and N. Aggarwal, "Classification of Audio Data using Support Vector Machine," *International Journal of Computer Science and Technology,* vol. 2, no. 3, pp. 398-405, 2011.

[11] P. Sawaengsawangarom, S. Phongoen and P. Wongchaisuwat, "Deep Learning for Music Genre Classification: A case study of Thai music," in *2025 International Conference on Multimedia Retrieval,* 2025.

[12] Ardiansyah, B. Yuliadi and R. Sahara, "Music Genre Classification using Naïve Bayes Algorithm," *International Journal of Computer Trends and Technology (IJCTT),* vol. 62, no. 1, pp. 50-57, 2018.

[13] J. Han, M. Kamber and J. Pei, Data Mining: Concepts and Techniques (3rd Edition). The Morgan Kaufmann Series in Data Management Systems., Elsevier, 2012.

[14] C.-W. Hsu, C.-C. Chang and C.-J. Lin, "A Practical Guide to Support Vector Classification," 2003.

[15] T. Li, M. Ogihara and Q. Li, "A Comparative Study on Content-Based Music Genre Classification," in *Proceedings of the 26th annual international ACM SIGIR conference on Research and development in informaion retrieval*, 2003.

[16] D. Barchiesi, D. Giannoulis, D. Stowell and M. D. Plumbley, "Acoustic Scene Classification: Classifying environments from the sounds they produce," *IEEE Signal Processing Magazine,* vol. 32, no. 3, pp. 16-34, 2015.

[17] T. C. Ba, T. D. T. Le and L. T. Van, Music genre classification using deep neural networks and data augmentation, Elsevier, 2025.