J Inno Sci Eng, 2026, 10(1):172-188

DOI:https://doi.org/10.38088/jise.1809289

Research Article

Comparative Analysis of SVM, k-NN and Logistic Regression Methods in Classifying Turkish Music Genres

Mehmet Cüneyt Özbalcı [1*], Turgay Tugay Bilgin [1]

[1] *Bursa Technical University, Department of Computer Engineering, Bursa, 16310, Türkiye.*

ARTICLE INFO

Received Date: 23/10/2025
Accepted Date: 7/01/2026

Cite this paper as:
Özbalcı, M.C. & Bilgin, T.T. (2026) Comparative Analysis of SVM, k-NN and Logistic Regression Methods in Classifying Turkish Music Genres. *Journal of Innovative Science and Engineering*. 10(1), 172-188.

*Corresponding author: Mehmet Cüneyt Özbalcı
E-mail:cuneytozbalci@gmail.com

Keywords:
k-NN
Logistic Regression
Music Classification
Signal Processing
SVM

ABSTRACT

Music genre classification represents a fundamental challenge within the field of Music Information Retrieval (MIR). The analysis of audio signals plays a pivotal role in the process of music genre classification, facilitating the extraction of pertinent information from the frequency-based data of the auditory content. In this study, diverse acoustic characteristics were derived through the utilization of the librosa library, and subsequent classification procedures were executed employing machine learning algorithms. For the purpose of this study, a dataset comprising a total of 600 music files in WAV format was meticulously curated. This dataset encompassed six distinct genres, all rooted in Turkish musical traditions. Subsequently, classification tasks were undertaken using Support Vector Machines (SVM), k-Nearest Neighbors (k-NN), and Logistic Regression algorithms. A series of experiments was conducted, varying the kernel functions and distance metrics employed. The findings of this investigation reveal the highest achieved accuracy rates, which amounted to 71.88% with k-NN, 73.44% with Logistic Regression, and 78.65% with the SVM algorithm. Notably, the SVM algorithm demonstrated superior performance in comparison to all other methodologies explored in this study.

1. Introduction

Music serves as a potent medium for the expression of human emotions and thoughts. It comprises intricate audio signals, rich in a diverse array of components, with time and frequency information being particularly salient (Er and Çiğ, 2020). Within the realm of music, two crucial elements are harmony and melody. Harmony pertains to the exploration and application of actual or implied pitch synchronization and chords, while melody involves a sequence of events characterized by distinct pitch variations (Scaringella et al., 2006). The breadth and diversity of musical cultures across various nations and communities have given rise to the intricate challenge of music categorization. In essence, music categorization is a perceptual endeavor. As technologies that mimic human perception continue to advance, the integration of artificial intelligence into music processing has gained traction. Given the fundamental role of music in individuals' lives, it represents a core component of human experience. Consequently, the convergence of artificial intelligence and music has emerged as a significant and influential area of academic research.

The application of artificial intelligence and machine learning in music processing can be broadly categorized into two domains: music synthesis and music analysis. Music synthesis encompasses the generation of music, while music analysis involves the evaluation of music-related information through the processing of existing musical data, yielding meaningful insights. Music analysis encompasses several prominent areas of study, including emotion analysis, instrument analysis, lyric analysis, timbre analysis, and melody analysis (Dhanapala and Samarasinghe, 2024). In music classification, the selection of relevant parameters can vary depending on the specific element under consideration. Parameters such as timbre, melody, lyrics, tempo, pitch, and chord are often integral to classification endeavors. Recent years have witnessed a growing emphasis on classification based on emotional content, instrument identification, timbral characteristics, artist recognition, and pitch analysis. The rapid expansion of data sources and the development of systematic data storage and management applications have accelerated research in the field of music classification (McKinney and Breebaart 2003).

The domain of music classification leverages a spectrum of deep learning and machine learning algorithms. In the existing literature, several algorithms have demonstrated success in music analysis. These algorithms include Support Vector Machines, Artificial Neural Networks (ANN), k-Nearest Neighbors, Logistic Regression, Deep Neural Networks (DNN), Convolutional Neural Networks (CNN), Recurrent Neural Networks (RNN), and Long Short-Term Memory networks (LSTM). The choice of the classification algorithm is contingent on the specific parameters being considered.

Mel-Frequency Cepstral Coefficients (MFCCs) are a prevalent feature extraction technique in music classification. MFCCs exhibit remarkable efficacy in music and sound analysis, owing to their capacity to faithfully capture perceptual characteristics akin to those perceived by the human ear. Furthermore, various filtering and enhancement methods have been devised to enhance the performance of music classification. These methods, aimed at improving audio signal quality and optimizing efficiency, encompass a spectrum of underlying mathematical operations.

Several studies have been conducted in the realm of music classification, each employing different algorithms and features. Here, we provide an overview of these studies:

Vishnupriya and Meenakshi (2018) utilized the CNN algorithm for music classification, relying on Mel Frequency Cepstral Coefficients (MFCCs) as features. They used the GTZAN dataset and achieved an accuracy rate of 76%.

Durdağ and Erdoğmuş (2019) pursued music genre classification by extracting visual and music color information from music. They transformed the extracted data into color images through preprocessing and employed CNN for classification. Their study, trained with Turkish music from the GTZAN dataset, achieved an accuracy rate of 60%.

Jawaherlalnehru and Jothilakshmi (2018) divided music genre classification into four subcategories using the Deep Neural Network algorithm. They extracted MFCC properties from audio signals as features and achieved a remarkable success rate of 97.8% by using 60-second segments of 400 audio files.

Kızrak and Bolat (2015) focused on classifying six Classical Turkish Music maqams. They utilized DNN, incorporating MFCCs and delta MFCCs as features. The highest performance, reaching 92.70%, was achieved when using MFCCs.

Karatana and Yildiz (2017) incorporated signal processing and feature extraction on separate segments of music data. They employed various classification algorithms, including SVM, k-NN, Random Forest, and Artificial Neural Networks, based on timbral features. Their study, utilizing the GTZAN dataset, attained the highest performance (88.9%) with SVM.

Thiruvengatanadhan employed MFCC features for classifying three types of music genres: pop, rock, and classical, using the SVM algorithm. They achieved accuracy rates of 88% with the polynomial kernel function, 91% with the Gaussian kernel function, and 87% with the sigmoidal kernel function (Thiruvengatanadhan, 2018).

Aguiar et al. conducted automatic music genre classification in ten different categories, using the CNN algorithm. They transformed digital audio signals into spectrograms, segmented and resized resulting images, and applied various data processing techniques. The highest accuracy achieved was 89.45% with CNN and single-tone pitch shifting (Aguiar et al., 2018).

McKay and Fujinaga (2004) employed the k-NN algorithm to classify music data in MIDI format into multiple subcategories (classical, jazz, pop). They achieved an accuracy rate of approximately 98% in the initial three subcategorizations (root nodes) but observed a drop in accuracy to about 90% in the subsequent subcategorization (leaf nodes).

(Kotsifakos et al., 2013) employed symbolic music data classification into four music genres: Classical, Rock, Pop, and Blues. Only MIDI-type data was used. A dataset of 100 pieces from four different genres, obtained free of charge from the internet, was selected. SMBGT (Subsequence Matching with Bounded Gaps and Tolerances) was used for similarity detection. The highest performance was achieved with the Logistic Regression method, yielding an AUC value of 0.999.

(Shi, 2025) conducted emotion classification in music. A CNN-based method was used as the approach. The Maestro dataset was used as the data set. The proposed method was compared with CRNN and PCRNN methods. Experimental results showed that the highest emotion recognition accuracy rate was observed with the proposed method at 0.93.

Guo and Wan (2025) conducted a music emotion classification study. A heterogeneous graph neural network (HGN) model was proposed. Instead of focusing solely on song lyrics, the multifaceted relationships between the song, singer, composer, and listener were considered together. The Douban Music and Amazon Music datasets were used. The proposed model achieved the highest accuracy rate of 85.69%, outperforming methods such as KNN, MLP, DeepWalk, GCN, GAT, metapath2vec, and HAN.

Tavares and Ayres (2025) conducted a Sentence-BERT (sBERT)-based approach is proposed for classifying eight different music genres with multiple tags from song lyrics. The data was obtained from the website http://vagalume.com.br/. The results show that the proposed method provides a significant performance improvement over the Bag-of-Words-based approach. In cross-language classification, the average F1-score was observed to increase to approximately 0.69.

Madushan and Weerasinghe (2025) conducted a music genre classification study using song lyrics. The features used for classification were Mel Spectrograms and MFCCs. A hybrid-fusion-based deep learning multi-modal model was proposed. The Music4All dataset was used. By processing the data, country, jazz, metal, and pop genres were obtained. An F1-score of 0.72 was achieved with the proposed model.

(Kaushik et al., 2025) presents a study on the classification of Indian music genres using a hybrid DL approach that combines the strengths of deep learning models. It involves collecting various Indian music datasets, preprocessing the data, creating annotations, and training hybrid models. Melodic and rhythmic features are captured using a combination of CNNs and RNNs. Compared to existing studies, it has achieved a higher performance with an accuracy rate of 87.70%.

Music classification performed using the method of creating spectrograms from sound waves. SVM, CNN, ResNet-50, and AlexNet methods were used. GTZAN was used as the data set. The highest accuracy rate was obtained when using the AlexNet method, at 95.38% (Murugan et al., 2024).

(Kilickaya, 2024) conducted music genre classification performed using SVM, Random Forest, and XGBoost methods. The librosa library is used for feature extraction. Features such as MFCCs, spectral centroid, chroma features, zero crossing rate, spectral bandwidth, spectral rolloff, and tempo are extracted. The highest accuracy rate of 70.77% is achieved with the XGBoost method.

These studies showcase the diverse approaches and techniques employed in the field of music classification. Further highlighting the effectiveness of various algorithms and feature extraction methods in achieving notable accuracy rates for genre classification.

This study is based on the graduate dissertation of the corresponding author, which is titled "Classification of Turkish music genres with probabilistic models" (Özbalcı, 2022).

2. Material and Methods

2.1 Classification Algorithms

2.1.1 SVM (Support Vector Machines)

Support Vector Machines is a versatile algorithm applicable to both classification and regression problems. In SVM, classification is achieved using hyperplanes (Jiang et al., 2005). The optimal hyperplane, which effectively separates the data, is referred to as the optimal hyperplane (Osowski et al., 2004). SVM finds extensive utility in music genre classification, owing to several advantages. Notably, SVM offers a distinct advantage in terms of computational efficiency during the learning phase, resulting in reduced time complexity. This makes the algorithm particularly well-suited for handling large datasets, aligning with the demands of big data applications.

SVM encompasses both linear and nonlinear variants, allowing it to accommodate diverse data patterns (Basthikodi et al., 2024). Multidimensional SVM is particularly valuable for making fine-grained distinctions within extensive datasets. To enhance its flexibility in handling different types of data, various kernel functions have been developed. Among these, the Radial Basis Kernel Function (RBF) stands out as a high-performing option in the context of music classification (Özbalcı, 2022). Additionally, polynomial kernel functions and linear kernel functions have also demonstrated their efficacy in delivering successful results (Battineni et al., 2019).

The Radial Basis Function (RBF) kernel is a popular choice for SVMs, particularly when dealing with non-linear problems. The RBF kernel is defined as:

$$K\left(x_i,x_j\right)=\exp\left(-\gamma\|x_i-x_j\|^2\right) \tag{1}$$

where x_i and x_j are two data points, $\|x_i-x_j\|^2$ represents the squared Euclidean distance between them, and γ is a tunable parameter that controls the influence of individual data points on the kernel. The SVC (Support Vector Classification) objective function with the RBF kernel can be formulated as:

$$\min_{w,b}\frac{1}{2}\|w\|^2+C\sum_{i=1}^{n}\xi_i \tag{2}$$

subject to the constraints:

$$y_i(w^T\phi(x_i)+b)\geq 1-\xi_i \tag{3}$$

and $\xi_i\geq 0$. Here, w and b are the weights and bias of the hyperplane, respectively, $\phi(x_i)$ is the non-linear mapping function induced by the RBF kernel. y_i is the class label, ξ_i are the excursive parameters that allow to some of the classification errors. The function of the C parameter is to maximize the margin on one hand. On the other hand, it is to minimize the classification error. The RBF kernel introduces non-linearity into the SVM model, allowing it to handle complex, non-linear decision boundaries effectively (Daviran et al., 2025). The parameter γ plays a crucial role in controlling the behavior of the RBF kernel, influencing the smoothness of the decision boundary and the model's ability to generalize to unseen data. In summary, the SVM with the RBF kernel provides a powerful framework for non-linear classification and regression tasks by leveraging the kernel trick and optimizing the margin between classes in the higher-dimensional feature space induced by the kernel.

2.1.2 k-NN (k-Nearest Neighbors)

In k-NN, a supervised algorithm, classification is performed based on the nearest neighbors. The class of a data point is determined according to the majority class. Data points with the highest similarity are considered to belong to the same class (Atalan, 2025). It is an effective method for labeled data.

Consequently, when the distance between two data points is minimized, these data points are construed as bearing the highest degree of resemblance (Shivashankara and Srinath, 2019). Among the prominent distance metrics are the Manhattan, Euclidean, and Chebyshev distances. The mathematical expression for the Euclidean distance is presented below.

$$d(x,y)=\sqrt{\sum_{i=1}^{n}\left(x_i-y_i\right)^2} \tag{4}$$

The algebraic formula for the Manhattan distance, also known as the taxicab norm or L1 norm, is as follows:

$$d(x,y)=\sum_{i=1}^{n}\left|x_i-y_i\right| \tag{5}$$

The algebraic formula for the Chebyshev distance, which is a special case of the L∞ norm, is as follows:

$$d(x,y)=\max_{1\leq i\leq n}\left|x_i-y_i\right| \tag{6}$$

The variable k in the k-Nearest Neighbors algorithm signifies the number of neighboring data points to be taken into account when making classification decisions (Prabavathy et al., 2020). It is important that selecting an appropriate value for k is critical, as setting it excessively high can detrimentally impact classification performance (Asha et al., 2024), while choosing a value that is too small can lead to overfitting issues. Given the absence of a universally applicable method for determining the optimal value of k, it is typically arrived at through a process of experimentation and iterative evaluation, often employing techniques such as cross-validation or grid search.

2.1.3 Logistic Regression

Logistic Regression stands as a robust statistical technique utilized for analyzing datasets comprising one or more independent variables, particularly apt for linear classification tasks. Renowned for its simplicity and flexibility in both deployment and comprehension, Logistic Regression has proven effective in various domains, (Dey et al., 2025) including music genre classification.

Logistic Regression serves as a fundamental statistical method for modeling the relationship between one or more independent variables and a binary outcome (Visutsak et al., 2025). It finds widespread application in diverse fields due to its interpretability and computational efficiency (Şenel and Alatli, 2014). In recent years, Logistic Regression has emerged as a viable tool for music genre classification, demonstrating promising results (Özbalcı, 2022).

Consider a dataset comprising n samples, each characterized by a set of p features denoted by $X=\{x_1,x_2,\ldots,x_p\}$. Let y represent the binary response variable indicating the presence or absence of a particular music genre. The Logistic Regression model is formulated as:

$$P(y=1 \mid X)=\frac{1}{1+e^{-\left(\beta_0+\beta_1 x_1+\beta_2 x_2+\ldots+\beta_p x_p\right)}} \quad (7)$$

Where

- $P(y=1|X)$ denotes the probability of y being 1 given the features X.
- $\beta_0,\beta_1,\ldots,\beta_p$ represent the coefficients corresponding to each feature.
- e denotes the base of the natural logarithm.

The logistic function $\frac{1}{1+e^{-z}}$ ensures that the predicted probabilities lie within the range [0, 1], facilitating interpretation as probabilities. In the context of music genre classification, Logistic Regression proves valuable for its ability to provide interpretable results and handle high-dimensional feature spaces efficiently. By training on labeled datasets containing audio features extracted from various music tracks, Logistic Regression models can learn to distinguish between different genres effectively. Notably, studies have shown that Logistic Regression outperforms certain other classification algorithms in music genre classification tasks (Özbalcı, 2022).

2.2 Implementation

The schematic representation delineating all stages from data acquisition to the classification phase is illustrated in the diagram depicted below.

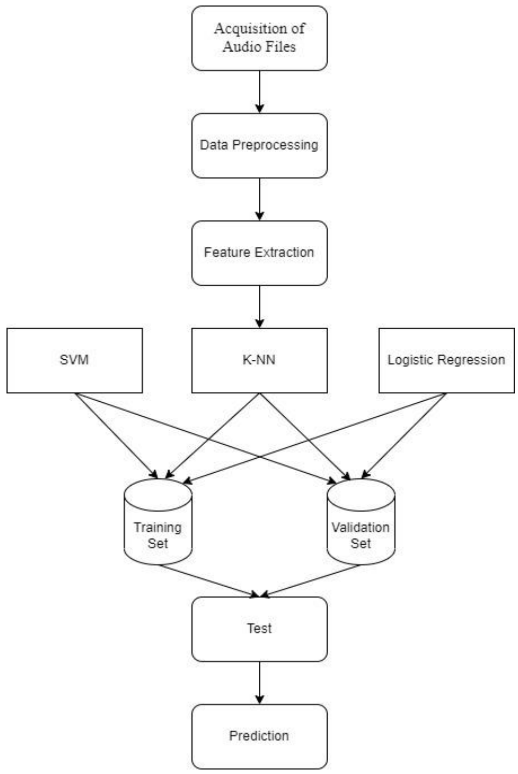

Figure 1: Implementation phases of algorithms

During the implementation phase, the initial step involved acquiring the dataset. Subsequently, the data underwent a processing phase where pertinent sections were trimmed or selected. Following data processing, feature extraction procedures were conducted on the preprocessed dataset. Finally, various algorithms were applied, and predictions or results were derived from these computational processes.

2.3 Features

The features used are MFCCs, ZCR, spectral centroid, spectral bandwidth, spectral rolloff, Chroma STFT and RMSE. They are effectively used in fields such as music classification, music recognition, and signal processing in the literature (Al-Shoshan, 2020). These features accurately represent the temporal, spectral, and harmonic characteristics of audio signals (Jiang et al., 2002). They are prominent features in music information retrieval studies.

2.3.1 MFCCs (Mel Frequency Cepstral Coefficients)

This method involves algebraic calculations and transformations to produce a vector output. It specializes in extracting spectral-based features and is commonly used in voice recognition (Han et al., 2006). It is well-regarded for its high performance and is frequently employed in the field of music information retrieval (MIR) (Sahidullah and Saha, 2012). MFCCs involve six stages: pre-emphasis, framing, windowing, FFT, Mel filter bank, and DCT (Nagawade and Ratnaparkhe, 2017).

2.3.2 ZCR (Zero Crossing Rate)

A zero-crossing denotes a point in a mathematical function where the sign of the function changes, transitioning from positive to negative or from negative to positive. This transition is visually represented by an intersection with the axis, resulting in a zero value on the graph of the function. The frequency of zero crossings serves as a metric for measuring the frequency of a signal (Jalil et al., 2013). To mitigate the impact of noise-induced zero-crossings and enhance the prominence of signal components, a threshold value can be introduced. The algebraic representation of the Zero Crossing function can be expressed as follows:

$$zcr=\frac{1}{T-1}\sum_{t=1}^{T-1} 1_{R<0}\left(s_t s_{t-1}\right) \quad (8)$$

2.3.3 Spectral Centroid

The spectral centroid is a fundamental metric employed in digital signal processing for spectrum characterization. It serves as an indicator of the spectral center of mass, signifying where the bulk of energy in the spectrum is concentrated. The spectral centroid provides insights into the timbral attributes of a signal and serves as a gauge of the dominance of higher harmonics relative to lower harmonics. This feature holds particular significance in discerning the tonal characteristics of musical instruments, with a pronounced relevance in the context of acoustic instruments (Wun et al., 2014).

$$\frac{\sum_{n=0}^{n-1} f(n)x(n)}{\sum_{n=0}^{n-1} x(n)} \quad (9)$$

In equation 9, x(n) denotes the weighted frequency value, while f(n) signifies the center frequency.

2.3.4 Spectral Bandwidth

Spectral bandwidth is a measure that quantifies the spread of spectral components around the spectral center based on the weighted average of their distances (Ramalingam and Krishnan, 2005). This measure is particularly valuable as a timbral feature, as it provides insights into the spread or dispersion of spectral components within a signal's spectrum. It characterizes the distribution of frequencies and how they are distributed around the spectral center.

2.3.5 Spectral Rolloff

Spectral rolloff is a parameter that defines the frequency value below which the total spectral energy remains below a certain percentage (D'Amico et al., 1999). By default, the librosa library uses a threshold of 85% to determine this frequency.

2.3.6 Chroma STFT

Chroma STFT, a feature based on chroma, is categorized as one of the pitch class profiles and holds extensive utility in various aspects of music analysis. Notably, it excels in the classification of pitch classes (Korzeniowski and Widmer, 2016). The Chroma vector, a fundamental component of Chroma STFT, consists of 12 elements, with each element representing the energy associated with a specific pitch class (Misha and Akhtar, 2025). Due to its capacity to capture pitch-related information effectively, Chroma STFT finds application in chord recognition and other music analysis tasks (Kattel et al., 2019).

2.3.7 RMSE (Root Mean Square Error)

This method quantifies the global energy of the audio signal by computing the root mean square (RMS) value for each square derived from the audio samples. The algebraic representation of this calculation is as follows:

$$x_{rms}=\sqrt{\frac{1}{n}\sum_{i=1}^{n} x_i^2} \quad (10)$$

In equation 10, x_i represents the signal amplitude at frame i while n denotes the total number of frames in the sample length (Sandhya et al., 2020).

2.4 Data Preprocessing

The dataset used in this study consists of a total of 600 WAV-format music files representing six different music genres. The dataset was constructed with a balanced distribution, including 100 tracks for each genre. The genres represented in the dataset are rock, pop, rap, arabesque, classical, and religious.

All music tracks were obtained from the YouTube platform, in compliance with copyright regulations. The interval between the 1st minute and the 1.5th minute of each music track was trimmed, resulting in half-minute segments. This duration range was selected after numerous experimental trials, as it was found to best capture the essential musical characteristics of each piece. Moreover, this segment length tends to include at least a portion of the chorus section, which is often a significant component of the musical structure.

All files were stored in uncompressed WAV format. The WAV format was chosen due to its lossless audio quality, which ensures the most accurate preservation of the sound signal's frequency components. This allows for the extraction of critical audio features such as Mel-Frequency Cepstral Coefficients (MFCCs), tempo, harmonic structure, and spectral distribution without any loss of fidelity. To ensure compatibility with the model input, all music files were standardized to a sampling frequency of 44.1 kHz, a commonly adopted rate in audio classification studies as it effectively represents the human auditory range. During the feature extraction process, a total of 20 Mel-Frequency Cepstral Coefficients (MFCCs) were computed for each segment. These coefficients capture the timbral and perceptual aspects of the audio signal and are widely recognized as one of the most effective features in both speech and music classification tasks.

The genre labels were manually assigned. This manual labeling process minimized potential errors that could occur with automated labeling systems and ensured that each track was accurately categorized according to its respective genre.

The total size of the dataset is approximately 3.5 GB. Due to the large file size, the dataset was divided into segments while preserving representational integrity, in order to facilitate efficient data management. This segmentation process not only improved storage efficiency but also enabled the machine learning models to train on a larger number of representative samples.

3. Results

In this study, several machine learning algorithms were employed, including Support Vector Machine, k-Nearest Neighbors, and Logistic Regression. For SVM, different kernel functions were considered, namely polynomial, linear, and Radial Basis Function (RBF). k-NN utilized various distance functions, including Manhattan, Euclidean, and Chebyshev. To assess the impact of different features on performance, experiments were conducted both with individual features excluded and with all features combined.

This approach helped evaluate the significance of each attribute on model performance. For the k-NN algorithm, the optimal values of k were determined through experimentation, resulting in k values of 10 and 7. In Logistic Regression, hyperparameter C was set to 0.1, and the maximum number of iterations was restricted to 100 to optimize model performance. To assess model performance, various metrics were calculated, including F1 Score, precision, and sensitivity. Additionally, confusion matrices were employed to evaluate the classification results, and accuracy rates were computed. The highest accuracy rate achieved in this study was 78.65%, and this result was obtained using the RBF kernel function with the SVM algorithm.

3.1 SVM Results

The graphs of the results obtained with RBF, polynomial and linear kernel functions are given in the figures below.

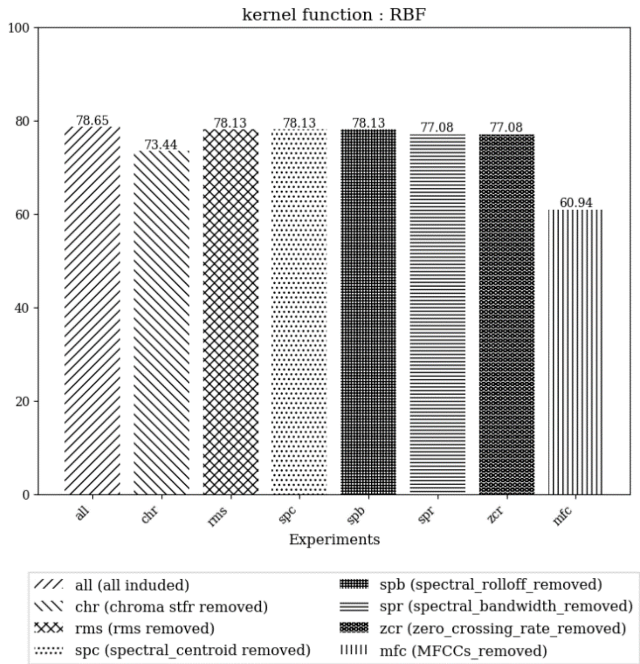

Figure 2: Combinatorial performances of features when using RBF

The figure demonstrates that the RBF kernel function achieves its highest performance when all features are utilized together. This suggests that employing all features collectively has a positive impact on the classification performance. Notably, when the MFCCs attribute is excluded, there is a substantial drop in performance. In essence, this indicates that the MFCCs attribute plays a pivotal role in influencing the model's performance, making it a critical feature for classification in this context.

The confusion matrices obtained when using the SVM algorithm and the RBF kernel function are given below.

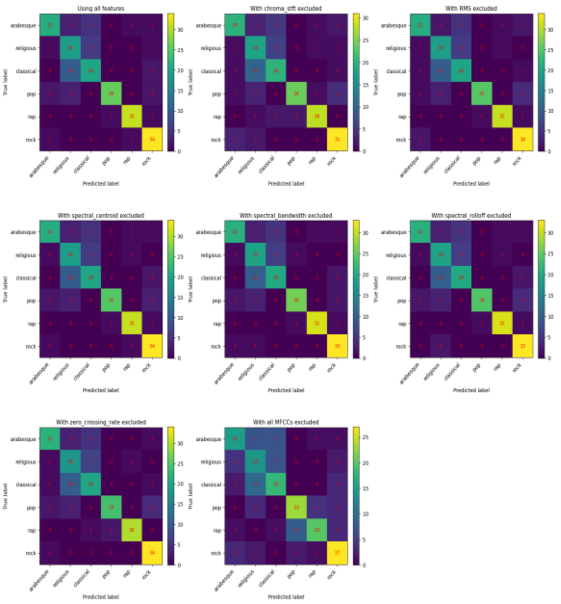

Figure 3: Confusion matrix when RBF is used with SVM

The confusion matrix shows that the highest accuracy rate was achieved when all features were used.

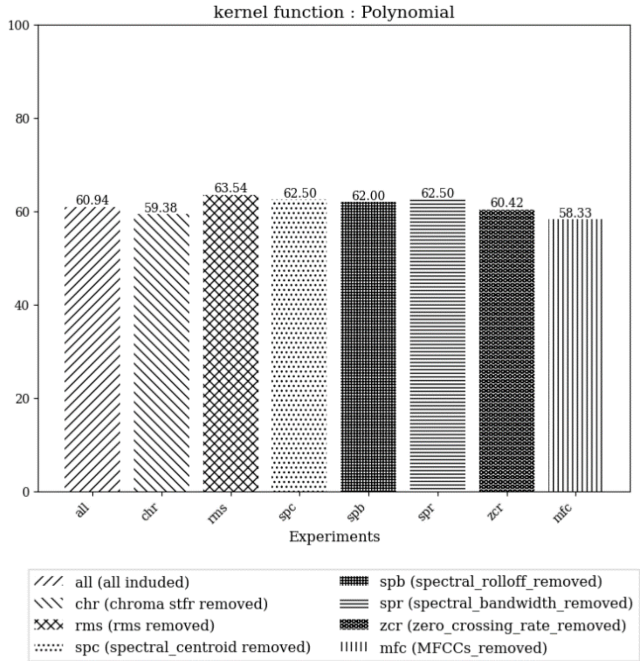

Figure 4: Combinatorial performances of features when using polynomial kernel function

The results indicate that the polynomial kernel function attains its highest performance when the rms feature is omitted. Including both the rms feature and all spectral features leads to a decrease in performance. Moreover, the most influential attribute, as inferred from these results, is once again the MFCCs attribute, emphasizing its significance in determining the model's performance with the polynomial kernel function.

The confusion matrices obtained using the SVM algorithm and polynomial kernel function are given below.

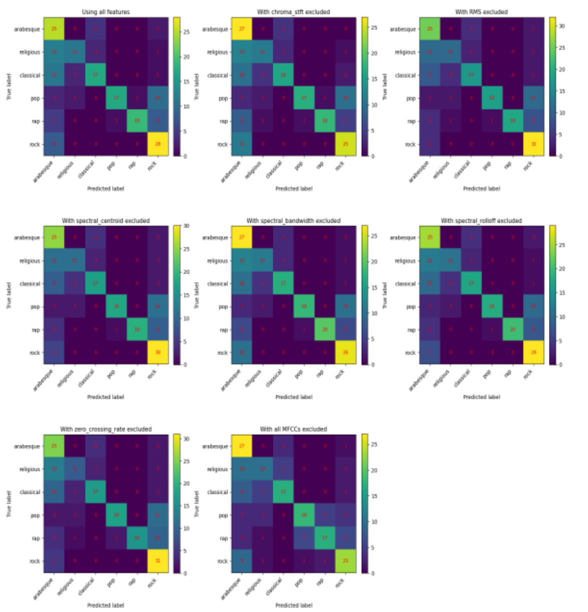

Figure 5: Confusion matrix when polynomial kernel function is used with SVM

The confusion matrix shows that the highest accuracy rate was achieved when the rms feature was removed.

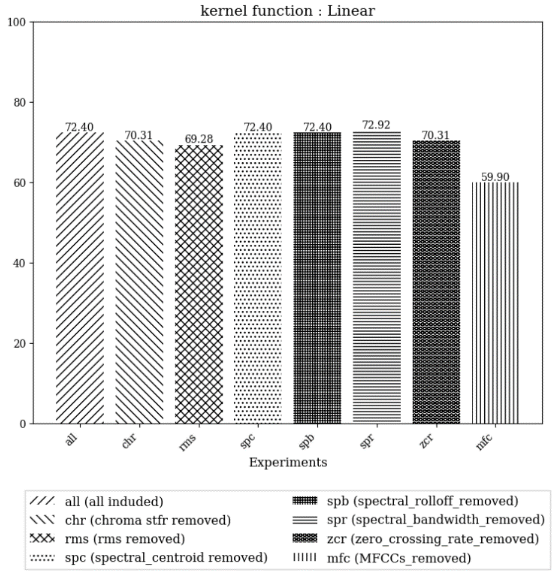

Figure 6: Combinatorial performances of features when using linear kernel function

The outcomes demonstrate that the linear kernel function reaches its peak performance when the spectral_bandwith feature is excluded. Solely the spectral_bandwith feature appears to have a detrimental impact on the classification performance with this kernel function. Once again, the most pivotal attribute in determining performance remains the MFCCs feature, underscoring its paramount importance in the context of the linear kernel function.

The confusion matrices obtained using the SVM algorithm and linear kernel function are given below.

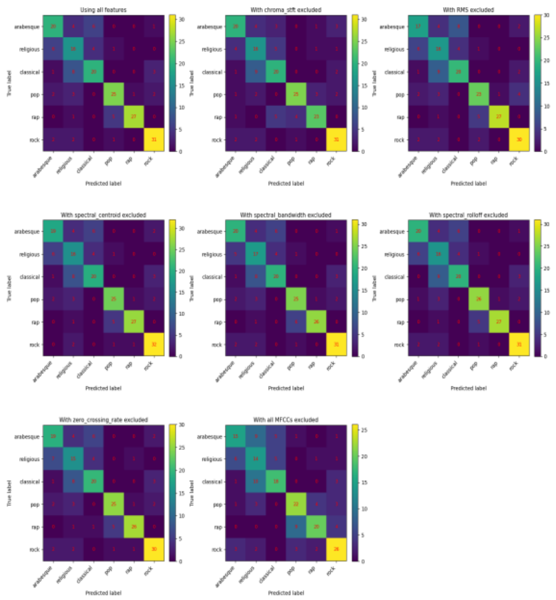

Figure 7: Confusion matrix when linear kernel function is used with SVM

The confusion matrix shows that the highest accuracy rate was achieved when the spectral_bandwidth feature was removed.

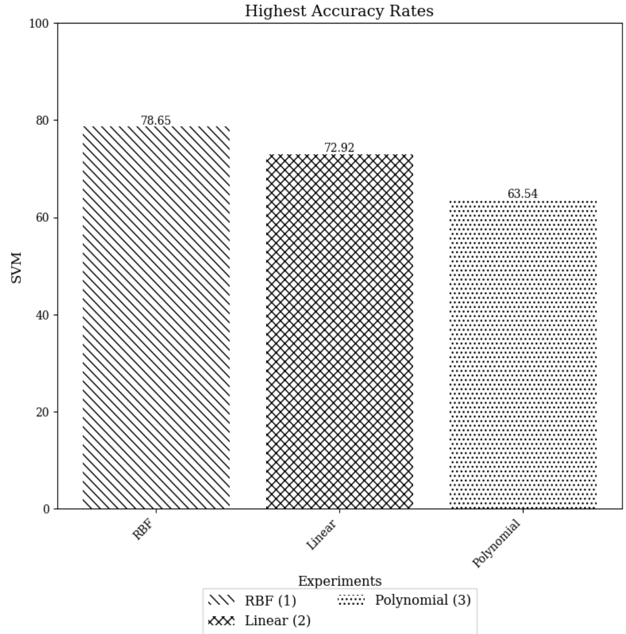

Figure 8: Peak performances of kernel functions with SVM

After conducting numerous experiments, it was noted that the RBF kernel function consistently yielded the highest performance, achieving an accuracy rate of 78.65%.

Table 1: Metrics showing the performance of genres classified by selecting the highest performing features using radial basis kernel function

Metrics	Precision	Recall	F1-Score
arabesque	0.84	0.68	0.75
religious	0.54	0.70	0.61
classical	0.65	0.62	0.63
pop	0.96	0.79	0.87
rap	0.94	0.94	0.94
rock	0.83	0.94	0.88
Weighted Average	0.80	0.79	0.79

When employing the Radial Basis Function (RBF) kernel function in conjunction with the SVM algorithm, the highest precision value was achieved in the pop music genre, with a precision score of 0.96. Additionally, the highest recall values were observed in the rap and rock music genres, both registering a recall score of 0.94. Furthermore, the highest F1-Score was attained in the rap music genre, with an F1-Score of 0.94. Finally, among the various metrics, precision holds the highest weighted average.

Table2: Metrics showing the performance of genres classified by selecting the highest performing feature using polynomial kernel function

Metrics	Precision	Recall	F1-Score
arabesque	0.45	0.81	0.58
religious	0.61	0.41	0.49
classical	0.71	0.53	0.61
pop	0.95	0.55	0.69
rap	1.00	0.58	0.73
rock	0.83	0.94	0.88
Weighted Average	0.72	0.64	0.64

When applying the polynomial kernel function in conjunction with the SVM algorithm, the highest precision value was observed in the rap music genre, with a precision score of 1.00, indicating perfect precision for that genre. Moreover, the highest recall value was recorded for the rock music genre, reaching a recall score of 0.94. Additionally, the highest F1-Score was achieved in the rock music genre, with an F1-Score of 0.88. In terms of overall performance, the metric with the highest weighted average is precision.

Table3: Metrics showing the performance of genres classified by selecting the highest performing feature using linear kernel function

Metrics	Precision	Recall	F1-Score
arabesque	0.69	0.65	0.67
religious	0.46	0.59	0.65
classical	0.67	0.62	0.65
pop	0.79	0.79	0.79
rap	0.96	0.82	0.89
rock	0.84	0.86	0.85
Weighted Average	0.74	0.73	0.73

When applying the linear kernel function in conjunction with the SVM algorithm, the following performance metrics were observed: The highest precision value was achieved in the rap music genre, with a precision score of 0.96. The highest recall value was obtained for the rock music genre, with a recall score of 0.86. The highest F1-Score was attained in the rap music genre, with an F1-Score of 0.89. Consistent with previous observations, among these metrics, precision held the highest weighted average.

3.2 k-NN Results

The graphs of the results obtained with Manhattan, Euclidean and Chebyshev distance functions are given in the figures below.

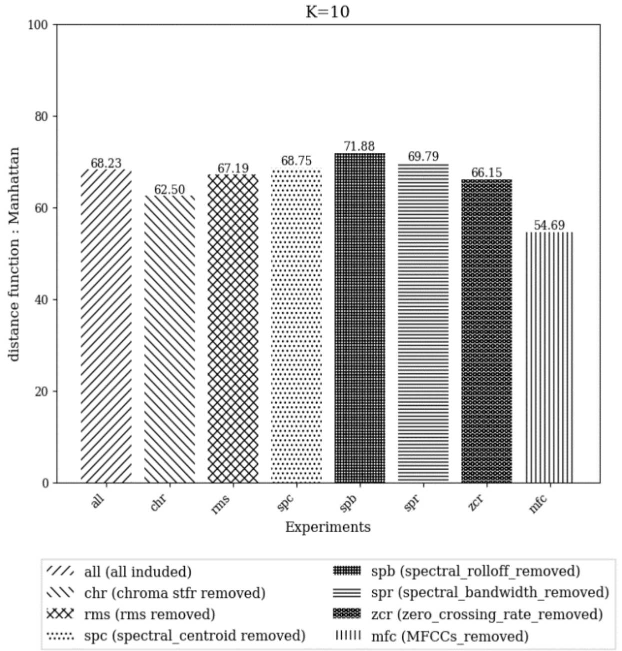

Figure 9: Combinational performances of features using k-NN with k=10 and Manhattan distance function

The results indicate that the Manhattan distance function performs optimally when the spectral_rolloff feature is omitted. Additionally, it was observed that the inclusion of spectral_centroid and spectral_bandwidth features has a detrimental effect on performance. Consistent with previous observations, the MFCCs feature emerged as the most influential and decisive feature in determining the model's performance when using the Manhattan distance function.

The confusion matrices obtained using the k-NN algorithm with the Manhattan distance function are given below.

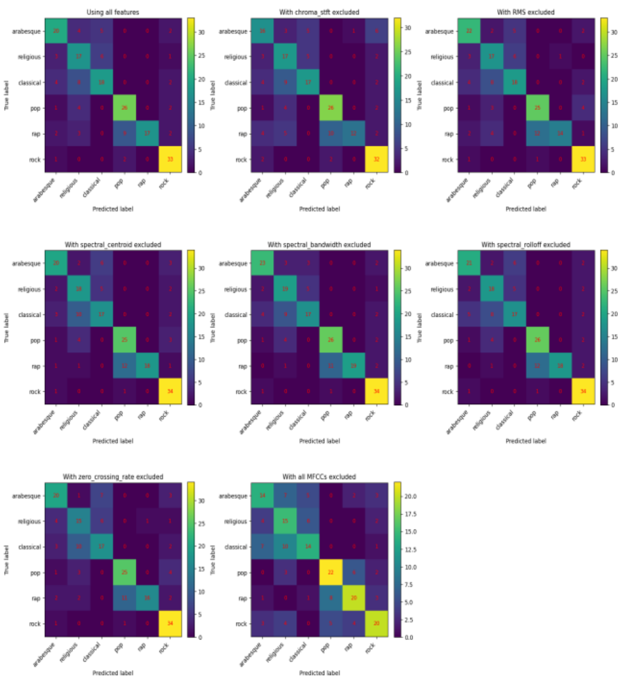

Figure 10: When k=10 and Manhattan distance function is used, the confusion matrix obtained with k-NN

The confusion matrix shows that the highest accuracy rate was achieved when the spectral_rolloff feature was removed.

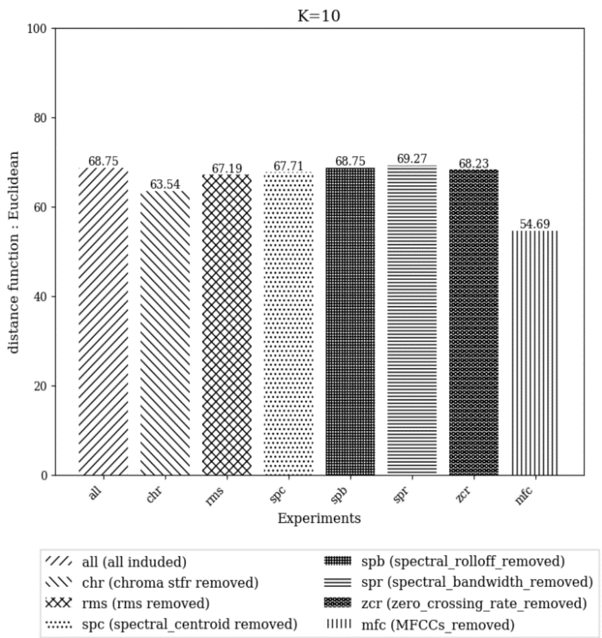

Figure 11: Combinational performances of features using k-NN with k=10 and Euclidean distance function

The results indicate that the Euclidean distance function yields the highest performance when the spectral_bandwidth feature is excluded. Furthermore, it was observed that the inclusion of the spectral_bandwidth feature has a negative impact on performance. As consistently observed, the MFCCs feature emerged as the most influential and decisive feature in influencing the model's performance when using the Euclidean distance function. The confusion matrices obtained using the k-NN algorithm with the Euclidean distance function are given below.

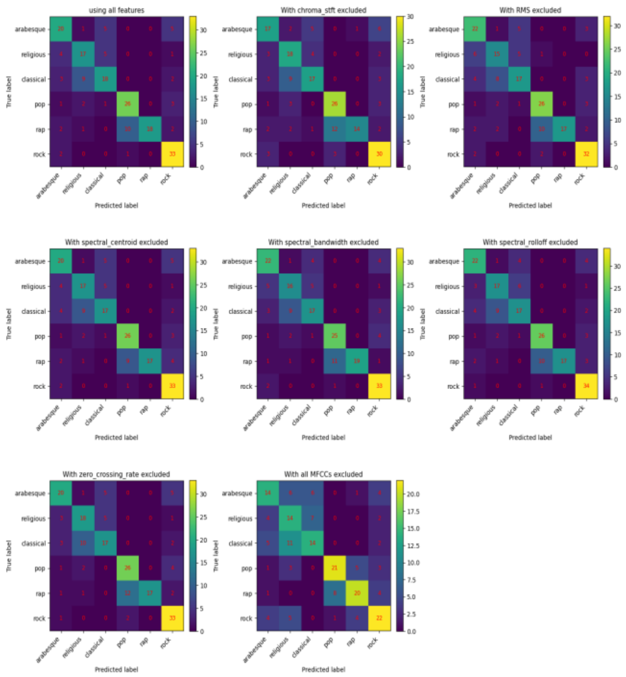

Figure 12: When k=10 and Euclidean distance function is used, the confusion matrix obtained with k-NN

The confusion matrix shows that the highest accuracy rate was achieved when the spectral_bandwidth feature was removed.

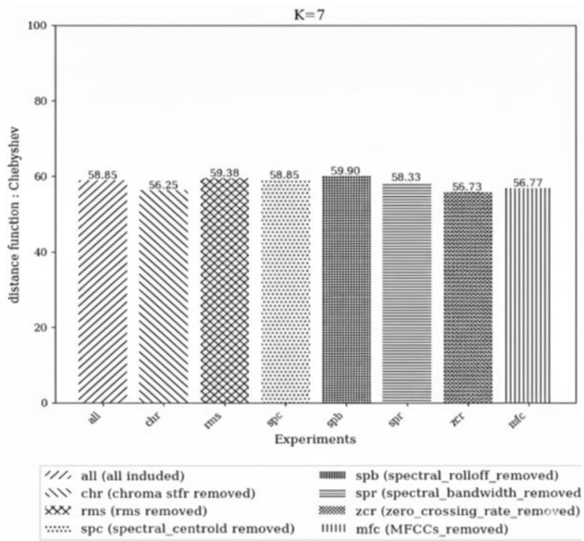

Figure 13: Combinational performances of features using k-NN with k=7 and Chebyshev distance function

The results indicate that the Chebyshev distance function performs optimally when the spectral_rolloff feature is excluded. Additionally, it was observed that including the rms feature has a negative impact on performance. Furthermore, the most decisive feature in influencing the model's performance with the Chebyshev distance function was found to be the zero_crossing_rate. The confusion matrices obtained using the k-NN algorithm with the Chebyshev distance function are given below.

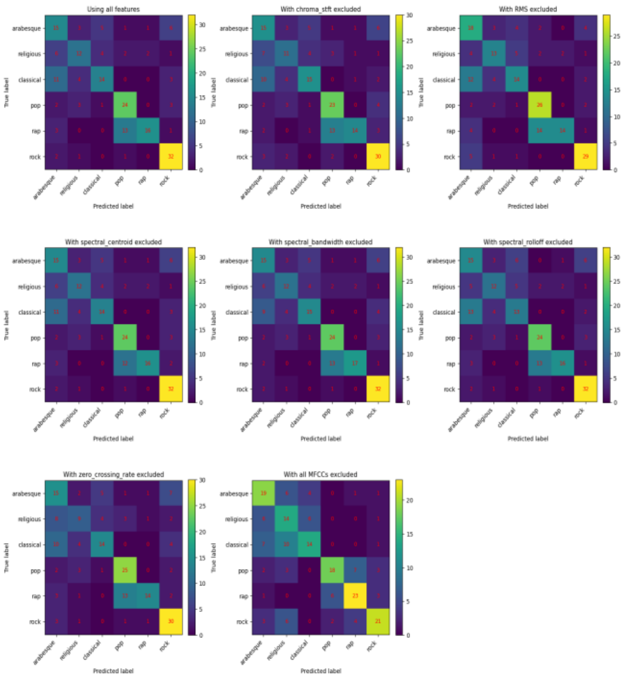

Figure 14: When k=7 and Chebyshev distance function is used, the confusion matrix obtained with k-NN

The confusion matrix shows that the highest accuracy rate was achieved when the spectral_rolloff feature was removed.

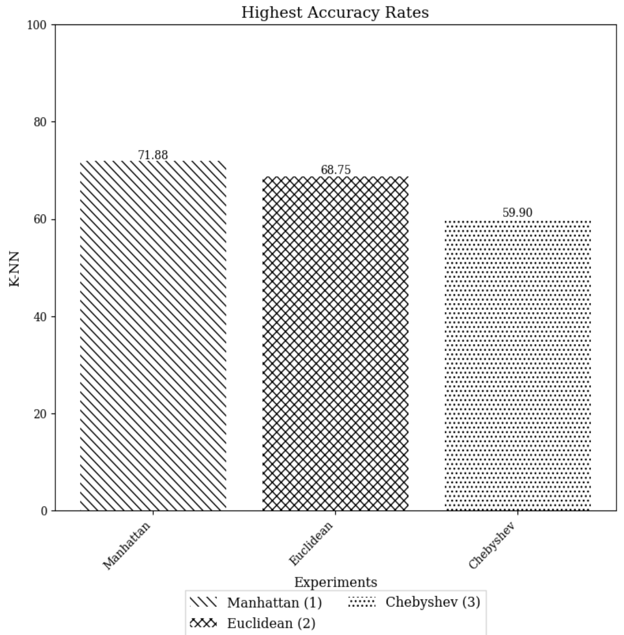

Figure 15: Highest performances of all distance functions with k-NN

As a result of various combinatorial experiments, it is observed that the best result achieved with various distance functions using k-NN is the Manhattan distance function with the accuracy of 71.88%.

Table 4: Highest performance metrics achieved with k-NN using manhattan distance function

Metrics	Precision	Recall	F1-Score
arabesque	0.74	0.74	0.74
religious	0.53	0.70	0.60
classical	0.68	0.53	0.60
pop	0.68	0.79	0.73
rap	1.00	0.58	0.73
rock	0.79	0.94	0.86
Weighted Average	0.75	0.72	0.72

The performance metrics for the k-NN method using the Manhattan distance function and those using the Euclidean distance function are nearly identical. In both cases, the metric with the highest weighted average remains precision. This suggests that both distance functions provide similar classification performance, with precision being the primary performance indicator in both scenarios.

Table 5: Highest performance metrics achieved with k-NN using euclidean distance function

Metrics	Precision	Recall	F1-Score
arabesque	0.67	0.71	0.69
religious	0.57	0.63	0.60
classical	0.61	0.53	0.57
pop	0.70	0.79	0.74
rap	1.00	0.52	0.68
rock	0.72	0.94	0.82
Weighted Average	0.72	0.69	0.69

When the Euclidean distance function is applied with the k-NN algorithm, the highest precision value is obtained in the rap music genre with score of 1.00. The highest recall value was obtained in the rock music genre. The highest F1-Score was also obtained in the rock music genre. The metric with the highest weighted average was precision. The results obtained with the Euclidean distance function were very similar to those achieved with the Manhattan distance function.

Table6: Highest performance metrics achieved with k-NN using chebyshev distance function

Metrics	Precision	Recall	F1-Score
arabesque	0.42	0.48	0.45
religious	0.55	0.44	0.48
classical	0.60	0.47	0.53
pop	0.60	0.73	0.66
rap	0.81	0.52	0.63
rock	0.68	0.89	0.77
Weighted Average	0.61	0.60	0.59

When the Chebyshev distance function is applied with the k-Nearest Neighbors algorithm, the following performance metrics were observed: The highest precision value of 0.81 was achieved for the rap music genre, indicating the correctness of positive predictions for this genre. The highest recall value was obtained in the rock music genre, with a recall score of 0.89, signifying the successful identification of 89% of situations expected to be predicted positively. The highest F1-Score was also attained in the rock music genre. Among the metrics, the one with the highest weighted average was precision with a score of 0.61.

3.3 Logistic Regression Results

Performance graphs of Logistic Regression are given below.

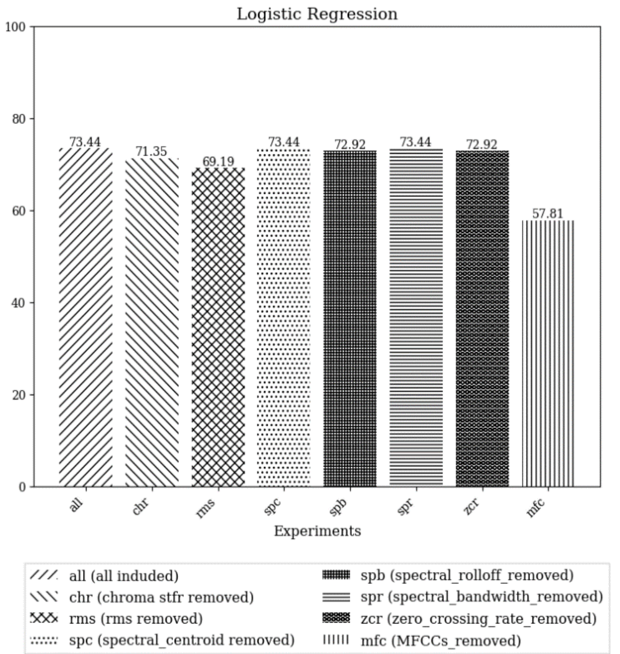

Figure 16: Combinatorial performances of features obtained by Logistic Regression

Experimental studies using Logistic Regression revealed that MFCCs had the most significant positive impact on performance. The highest overall performance was achieved when all features were utilized, resulting in an accuracy rate of 73.44%.

The confusion matrices obtained when the Logistic Regression algorithm is used are given below.

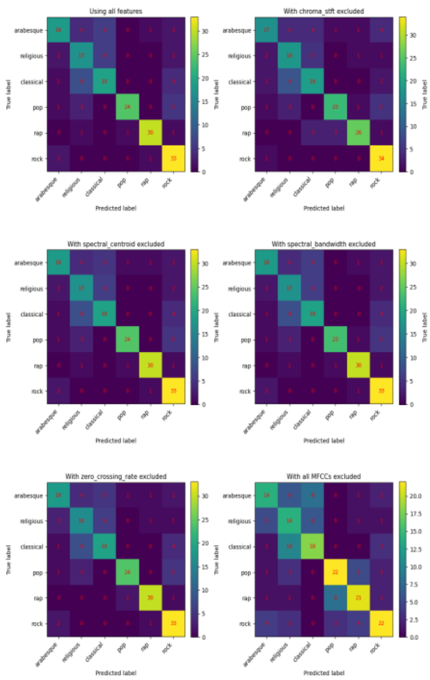

Figure 17: Confusion matrices obtained using the Logistic Regression algorithm

As seen in the confusion matrix, the highest performance was achieved when all features were used.

Table7: Performance metrics with the highest performance obtained with Logistic Regression

Metrics	Precision	Recall	F1-Score
arabesque	0.75	0.58	0.65
religious	0.52	0.63	0.57
classical	0.61	0.59	0.60
pop	0.96	0.73	0.83
rap	0.94	0.91	0.92
rock	0.70	0.92	0.80
Weighted Average	0.75	0.73	0.74

The results from the Logistic Regression algorithm indicate the following performance metrics: The highest precision value was achieved in the pop music genre, with a precision score of 0.94. The highest recall value was obtained in the rock music genre, registering a recall score of 0.92. The highest F1-Score was attained in the rap music genre, with an F1-Score of 0.92. Once again, the metric with the highest weighted average was precision underlining its significance as the primary performance metric in this context.

4. Conclusion and Discussion

In this study, a series of experiments were carried out to evaluate the performance of different machine learning algorithms in the classification of Turkish music genres. Among the tested models, the Support Vector Machine (SVM) algorithm achieved the highest accuracy rate of 78.65%, surpassing both k-Nearest Neighbors (k-NN) with 71.88% and Logistic Regression with 73.44%. This result underscores the robustness and generalization capability of SVM in handling multidimensional and complex feature spaces derived from audio signals.

An important finding of this research is the significant contribution of Mel-Frequency Cepstral Coefficients (MFCCs) to classification performance. The comparative experiments demonstrated that models incorporating MFCC features consistently achieved higher accuracy, confirming their central role in capturing the timbral and spectral characteristics of musical audio. Therefore, MFCCs can be regarded as an indispensable feature for effective music genre classification, particularly in the context of Turkish music, which exhibits rich tonal and rhythmic diversity.

Additionally, the study observed that genres with overlapping acoustic and cultural characteristics such as religious, arabesque and classical tended to produce higher misclassification rates. This suggests that their shared tonal structures and instrumentation may introduce ambiguity in the feature space. Conversely, rap displayed more distinctive rhythmic and vocal patterns, enabling more reliable classification results.

For future research, several directions are proposed. First, expanding and diversifying the dataset could improve model generalization, especially given the unique properties of Turkish music. Second, the integration of deep learning architectures, such as Convolutional Neural Networks (CNNs) or Recurrent Neural Networks (RNNs), may allow for more sophisticated temporal and hierarchical feature extraction. Finally, enhancing data preprocessing techniques, for instance by segmenting tracks into meaningful sections such as choruses or refrains, could lead to a more nuanced understanding of genre-specific characteristics.

Overall, this study contributes to the growing body of literature on music information retrieval (MIR) by emphasizing the importance of feature selection, dataset quality, and algorithmic diversity in improving the accuracy of automatic music genre classification systems.

Article Information

Authors' Contrtibution: The conceptualization of the study was carried out by T.T.B. The study design was developed by M.C.O. and T.T.B. Supervision of the research was conducted by T.T.B. Resources were provided by M.C.O. Data collection was performed by M.C.O. Data analysis was carried out by T.T.B. and M.C.O. The literature search was conducted by M.C.O. The manuscript was written by M.C.O. and T.T.B. Critical revision of the manuscript was performed by T.T.B.

Conflict of Interest/Common Interest: No conflict of interest or common interest has been declared by the authors.

Ethics Committee Approval: Ethics committee approval was not applicable to this study.

Artificial Intelligence Statement: The author(s) bear full responsibility for the content and accuracy of their work, including any use of artificial intelligence (AI) technologies, and confirm that they have read the AI Policy, which is accessible on the journal's website.

References

Aguiar, R. L., Costa, Y. M., & Silla, C. N. (2018). Exploring data augmentation to improve music genre classification with convnets. In 2018 *International Joint Conference on Neural Networks (IJCNN)* (pp. 1–8). IEEE.

Al-Shoshan, A. I. (2020). Classification and Separation of Audio and Music Signals. In *Multimedia Information Retrieval*. IntechOpen.

Asha, P., UT, A. P., Ankayarkanni, B., Praveena, M. A., & Suresh, L. P. (2024, August). Assessment and Evaluation of Music Genre Classification by employing various AI Techniques. In 2024 *7th International Conference on Circuit Power and Computing Technologies (ICCPCT)* (Vol. 1, pp. 112-116). IEEE.

Atalan, Y. A. (2025). Consideration of Environmental, Economic, and Oil Factors for Unit-based Estimation of Consumed Electrical Energy with ML Algorithms: A Case Study of

Çanakkale Region. *Journal of Innovative Science and Engineering*, *9*(2), 247-258.

Basthikodi, M., Chaithrashree, M., Ahamed Shafeeq, B. M., & Gurpur, A. P. (2024). Enhancing multiclass brain tumor diagnosis using SVM and innovative feature extraction techniques. *Scientific Reports*, 14(1), 26023.

Battineni, G., Chintalapudi, N., & Amenta, F. (2019). Machine learning in medicine: Performance calculation of dementia prediction by support vector machines (SVM). *Informatics in Medicine Unlocked*, 16, 100200.

D'Amico, A. A., D'Andrea, A. N., & Mengali, U. (1999). Feedforward joint phase and timing estimation with OQPSK modulation. *IEEE Transactions on Vehicular Technology*, 48(3), 824–832.

Daviran, M., Maghsoudi, A., & Ghezelbash, R. (2025). Optimized AI-MPM: Application of PSO for tuning the hyperparameters of SVM and RF algorithms. *Computers & Geosciences*, *195*, 105785.

Dey, D., Haque, M. S., Islam, M. M., Aishi, U. I., Shammy, S. S., Mayen, M. S. A., ... & Uddin, M. J. (2025). The proper application of logistic regression model in complex survey data: a systematic review. *BMC Medical Research Methodology*, *25*(1), 15.

Dhanapala, S., & Samarasinghe, K. (2024). Music emotion classification: A literature review. *Journal of Research in Music*, 2(2).

Durdağ, Z., & Erdoğmuş, P. (2019). Müzik türlerinin derin öğrenme ağları ile sınıflandırılması. *Sakarya University Journal of Computer and Information Sciences*, 2(1), 53–60.

Er, M. B., & Çiğ, H. (2020). Türk müziği uyaranları kullanılarak insan duygularının makine öğrenmesi yöntemi ile tanınması. *Gazi University Journal of Science Part C: Design and Technology*, 8(2), 458–474.

Guo, J., & Wan, P. (2025). Music Emotion Classification Based on Heterogeneous Graph Neural Networks. *IEEE Access.*

Han, W., Chan, C. F., Choy, C. S., & Pun, K. P. (2006). An efficient MFCC extraction method in speech recognition. 2006 *IEEE International Symposium on Circuits and Systems*, 145–148.

Jalil, M., Butt, F. A., & Malik, A. (2013). Short-time energy, magnitude, zero crossing rate and autocorrelation measurement for discriminating voiced and unvoiced segments of speech signals. 2013 *International Conference on Technological Advances in Electrical, Electronics and Computer Engineering (TAEECE)*, 208–212.

Jawaherlalnehru, G., & Jothilakshmi, S. (2018). Music genre classification using deep neural networks. *International Journal of Scientific Research in Science, Engineering and Technology*, 4(4).

Jiang, D. N., Lu, L., Zhang, H. J., Tao, J. H., & Cai, L. H. (2002, August). *Music type classification by spectral contrast feature.* In Proceedings. IEEE international conference on multimedia and expo (Vol. 1, pp. 113-116). IEEE.

Jiang, H., Bai, J., Zhang, S., & Xu, B. (2005). SVM-based audio scene classification. 2005 *International Conference on Natural Language Processing and Knowledge Engineering*, 131–136. IEEE.

Karatana, A., & Yildiz, O. (2017). Music genre classification with machine learning techniques. In 2017 *25th Signal Processing and Communications Applications Conference (SIU)* (pp. 1–4). IEEE.

Kattel, M., Nepal, A., Shah, A. K., & Shrestha, D. (2019). Chroma feature extraction using Fourier transform. Conference: *Chroma Feature Extraction using Fourier Transform.*

Kaushik, P., Rathore, S. P. S., Chahal, K., Saraf, S., Chauhan, G. S., & Kumar, P. (2024, March). RhythmQuest: Unifying Indian music classification and prediction with hybrid deep learning techniques. *In 2024 IEEE International Conference on Interdisciplinary Approaches in Technology and Management for Social Innovation (IATMSI)* (Vol. 2, pp. 1-6). IEEE.

Kızrak, M. A., & Bolat, B. (2015). Classification of classic Turkish music makams by using deep belief networks. In 2015 *23rd Signal Processing and Communications Applications Conference (SIU)* (pp. 527–530). IEEE.

Kilickaya, O. (2024). Genre Classification and Musical Features Analysis. *International Journal of Latest Engineering Research and Applications (IJLERA)*, *9*(04), 18-33.

Kotsifakos, A., Kotsifakos, E. E., Papapetrou, P., & Athitsos, V. (2013, May). Genre classification of symbolic music with SMBGT. *In Proceedings of the 6th international conference on PErvasive technologies related to assistive environments* (pp. 1-7).

Korzeniowski, F., & Widmer, G. (2016). Feature learning for chord recognition: The deep chroma extractor. *Proceedings of the 17th International Society for Music Information Retrieval Conference*, 37–43.

Madushan, J., & Weerasinghe, R. (2025). Music Genre Classification with Multi-Modal Properties of Lyrics and Spectrograms. *International Journal on Advances in ICT for Emerging Regions (ICTer)*, 18(2).

McKay, C., & Fujinaga, I. (2004). Automatic genre classification using large high-level musical feature sets. *Proceedings of the International Society for Music Information Retrieval (ISMIR)*, 525–530.

McKinney, M., & Breebaart, J. (2003). *Features for audio and music classification.*

Mishra, A., & Akhtar, R. (2025). Music genre classification using machine learning techniques. arXiv preprint arXiv:2509.01762. https://arxiv.org/abs/2509.01762.

Murugan, B. S., Dhanasekaran, S., & Kumar, P. (2024, August). *Utilizing Spectrograms and Deep Learning Techniques for Improved Music Genre Classification.* In 2024 7th International Conference on Circuit Power and Computing Technologies (ICCPCT) (Vol. 1, pp. 1723-1728). IEEE.

Nagawade, M. S., & Ratnaparkhe, V. R. (2017). Musical instrument identification using MFCC. 2017 *2nd IEEE International Conference on Recent Trends in Electronics, Information & Communication Technology (RTEICT)*, 2198–2202.

Osowski, S., Siwek, K., & Markiewicz, T. (2004, June). *MLP and SVM networks-a comparative study.* In Proceedings of the 6th Nordic Signal Processing Symposium, 2004. NORSIG 2004. (pp. 37-40). IEEE.

Özbalcı, M. C. (2022). *Classification of Turkish music genres with probabilistic models* [Master's thesis].

Prabavathy, S., Rathikarani, V., & Dhanalakshmi, P. (2020). Classification of musical instruments using SVM and KNN. *International Journal of Innovative Technology and Exploring Engineering*, 9(7), 1186–1190.

Ramalingam, A., & Krishnan, S. (2005). Gaussian mixture modeling using short time Fourier transform features for audio fingerprinting. 2005 *IEEE International Conference on Multimedia and Expo*, 1146–1149.

Sahidullah, M., & Saha, G. (2012). Design, analysis and experimental evaluation of block-based transformation in MFCC computation for speaker recognition. *Speech Communication*, 54(4), 543–565.

Sandhya, P., Spoorthy, V., Koolagudi, S. G., & Sobhana, N. V. (2020). Spectral features for emotional speaker recognition. 2020 *Third International Conference on Advances in Electronics, Computers and Communications (ICAECC)*, 1–6.

Scaringella, N., Zoia, G., & Mlynek, D. (2006). Automatic genre classification of music content: A survey. *IEEE Signal Processing Magazine*, 23(2), 133–141.

Shi, Y. (2025). *A CNN-Based Approach for Classical Music Recognition and Style Emotion Classification.* IEEE Access.

Shivashankara, S., & Srinath, S. (2019). Shape-based continuous real time hand gesture recognition system of American Sign Language using KNN classifier.

Şenel, S., & Alatli, B. (2014). Lojistik regresyon analizinin kullanıldığı makaleler üzerine bir inceleme. *Journal of Measurement and Evaluation in Education and Psychology*, 5(1), 35–52.

Tavares, T. F., & Ayres, F. J. (2025). *Multi-label cross-lingual automatic music genre classification from lyrics with Sentence BERT* [arXiv preprint]. https://arxiv.org/abs/2501.03769

Thiruvengatanadhan, R. (2018). Music classification using MFCC and SVM. *International Research*

Journal of Engineering and Technology, 5, 922–924.

Vishnupriya, S., & Meenakshi, K. (2018). Automatic music genre classification using convolution neural network. In 2018 *International Conference on Computer Communication and Informatics (ICCCI)* (pp. 1–4). IEEE.

Visutsak, P., Loungna, J., Sopromrat, S., Jantip, C., Soponkittikunchai, P., & Liu, X. (2025). Mood-Based Music Discovery: A System for Generating Personalized Thai Music Playlists Using Emotion Analysis. *Applied System Innovation*, 8(2), 37.

Wun, S., Horner, A., & Wu, B. (2014). Effect of spectral centroid manipulation on discrimination and identification of instrument timbres. *Journal of the Audio Engineering Society*, 62(9), 575–583.