Arabic Music Classification Using Machine Learning Algorithms with Combined Robust Features

M. E. ElAlami[1], S. M. K. Tobar[2], S. M. Khater[1], Eman A. Esmaeil[1,*]

[1]Computer Science Department, Faculty of Specific Education, Mansoura University, Mansoura, Egypt

[2]Musical Education Department, Faculty of Specific Education, Mansoura University, Mansoura, Egypt

Emails: moh_elalmi@mans.edu.eg; sahartobar@mans.edu.eg; shim_khater@mans.edu.eg; emanatya92@mans.edu.eg

Abstract

Machine learning (ML) is the most up-to-date approach for classifying music genres. Due to technological ML advancements, its technologies can help in music genre recognition best. In machine learning, effective fusion of different features could improve recognition performance. Hence, this paper presents a new robust method for Arabic music classification based on the fusion of different sets of features. Frequency-domain, time-domain, and cepstral domain features have been combined and compared with other state-of-the-art approaches. Four machine-learning models that categorize music into its appropriate genre have been created: support vector machines (SVM), K-nearest neighbors (KNN), naïve Bayes (NB), and random forest (RF) classifiers were utilized in a comparative analysis of other ML algorithms, and the accuracy of these models has been assessed and derives the appropriate conclusions. To assess the performance of our method, two various datasets are used: the collected dataset, namely Zekrayati, which was collected by authors in favor of this paper, and the global GTZAN dataset, which was used to compare with previous studies. The experimental findings indicated that the SVM exhibited a higher optimal accuracy of 99.2% and has proven that the fusion proposed features will help to classify music in different fields.

Keywords: Machine learning (ML); Musical classification; Music information retrieval (MIR)

1. Introduction

Music plays a significant role in our lives; it is a kind of information, like any other available data, that must be retrieved and analyzed in order to obtain knowledge [1]. Music genre classification (MGC) has emerged as one of the most important Music Information Retrieval (MIR) methods. Musical genres are difficult to detect due to the illusive nature of auditory musical data [2]. Combining several features improves the performance of real-world categorization systems as a number of features are merged to provide more information, and the information complements one another in favor of the classification process [3]. Recently, the literature has employed a variety of attributes and methods to classify musical sound depending on the fusion between distinguishable qualities, and the process of building a framework that depended on fusioned features for the classification process became an effective method of improving performance [4].

Therefore, this paper suggests a novel fusion method of feature extraction where frequency-domain, time-domain, and cepstral domain features are combined to enhance the performance of Arabic music classification. The introduced feature set has been tested using several ML methods in a comparative analysis of algorithms for choosing the best classifier.

DOI: https://doi.org/10.54216/FPA.170119
Received: November 28, 2023 Revised: March 22, 2024 Accepted: July 28, 2024

The following are the principal contributions of this paper:

- In this paper, the collected dataset has been created and became available for free at: https://www.kaggle.com/datasets/emanatyaesmaeil/zekrayati-dataset.
- Covering the absence of Arabic Music dataset by gathering, structuring, and annotating a large corpus of Arabic audio clips, since the majority of the literary works related to western genre music utilizes the GTZAN dataset. Although this dataset considered a baseline dataset for MGC, but as stated by Strum in [5] it has some limitations, such as mislabeling, distortions, and copies.
- Robust various features have been fusioned to achieve optimal performance in feature extraction.
- The proposed system was applied to the popular GTZAN dataset, and the best accuracy had resulted compared to previous studies.

2. Related Work

Considerable research has been devoted to the music classification in recent years. These works have been supported by recent advancements in machine learning (ML) and deep learning (DL), with many fusion methods for the music information retrieval task. For example, some researchers combined Mel Frequency Cepstral Coefficients (MFCC) with sonograms. While others integrated convolutional neural networks (CNN) with NetVLAD. On the other hand, others had fusioned CNN with Mel-spectrogram and Modgd-gram. On the contrary, timbral feature, chroma feature, and source separation-based features have been combined by many authors. Finally, many works depended on combining the energy feature, time, and frequency domain variance features. This section gives a methodical overview of these fusioned methods for the music industry and examines the prospects of AI in this domain, as shown in Table 1.

Table 1: Previous studies related to Fusion features for music classification

REF.	Year	Task	Algorithm	Dataset	Accuracy
[6]	2020	Recognition of Music Type by Multiple levels Local Feature Coding Fusion	CNN with NetVLAD have integrated	Extended Ballroom, ISMIR2004, and GTZAN datasets	According to the results, the suggested method outperforms other cutting-edge models in terms of accuracy.
[7]	2021	Combining Various Audio Representations to Classify Music Genres	Multi-Feature Fusion Networks (MUFFN)	GTZAN, Homburg, Extended Ballroom and Small subset of FMA	Results show that Multi_Feature Fusion Networks regularly increase the classification accuracy.
[8]	2022	Instrument Classification in Polyphonic Music in Music Using Spectro/Modgd-gram Fusion.	CNN, Mel-spectrogram, and modgd-gram have fusioned.	IRMAS dataset	The micro and macro F1 scores are 0.69 and 0.62, correspondingly.
[9]	2023	genre categorizing music using the fusion feature	Timbral, chroma and source separation-based features have combined	GTZAN dataset	95.8% overall accuracy and 95.82% F1 score, respectively
[10]	2019	Music Identification Method Employing Feature Fusion.	The energy, time, and frequency domain variance features have been combined.	Different genres of music were chosen as test.	The rate of classification accuracy surpasses 85%.
[11]	2020	Musical Instrument Classification	MFCC + (KNN) and (SVM)	Sixteen musical instruments were collected from various sources	SVM=99.29% KNN=98.17%

DOI: https://doi.org/10.54216/FPA.170119
Received: November 28, 2023 Revised: March 22, 2024 Accepted: July 28, 2024

[12]	2021	Classification of Musical Instrument Sound	SVM+ GoogleNet + KNN	IRMAS online musical database	SVM=99.37% KNN=98.16%
[13]	2022	Music Genre Analysis and Classification	Time and Frequency domain features (NB), (DT) , logistic regression, (RF)	GTZAN dataset	RF=90%
[14]	2022	Music Genre Classification	MFCC RASTA-PLP (Relative Spectral-Perceptual Linear Prediction + SVM	801 music samples for Five genres of: Rock, Folk, Jazz, Lyric, and Dance.	SVM = 81.27%

Reviewing previous studies displayed that the differentiable features play an important role in the categorization of music signals. This work proposes a new fusion way of feature extraction from music signals where frequency-domain, time-domain, and cepstral domain features have been combined. The fusioned features have been tested with several ML classifiers to determine the best-suited model in music recognition.

3. Materials and Methods

The proposed methodology for classifying music signal using our fusioned method is depicted in Figure 1.There are four main stages: dataset building process, pre-processing, Feature extraction, and music signal Classification. The following three subsections will describe each of these stages in detail.

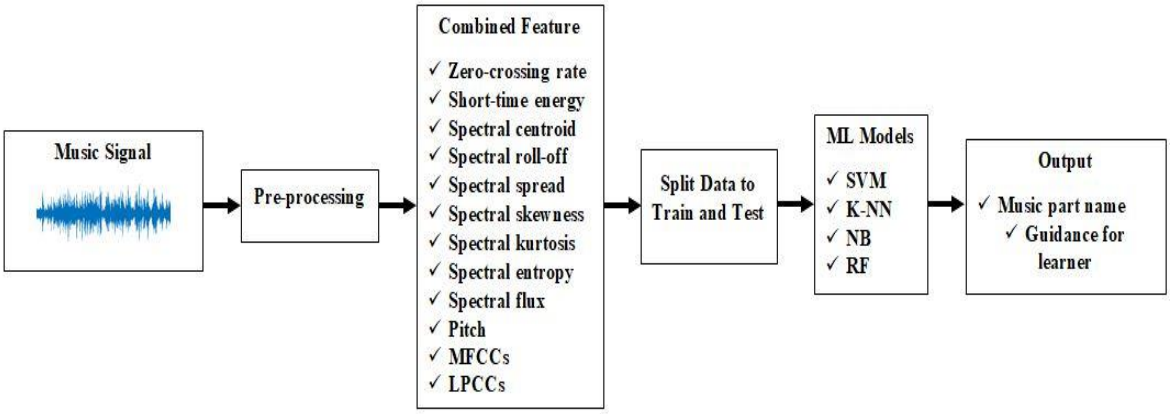


Figure 1. The proposed system overview.

3.1. Collected dataset

- The dataset used in this research was sourced from the Zekrayati Arabic song melody, played using the lute instrument in favor of this research.
- The study dataset consists of 4 classes (Zekrayati_P_1, Zekrayati_P_2, Zekrayati_P_3, and Zekrayati_P_4) with total 440 music track.
- The division has based on the closing tone (cadences) at the end of lines of music.
- The number of wav files for every category is as follows: 120 for Zekrayati_P_1, 120 for Zekrayati_P_2, 80 for Zekrayati_P_3, and 120 for Zekrayati_P_4.
- The number of music tracks in classes isn't equal, as it represents a number of recordings under a variety of conditions.
- As shown in Table 2, the duration of each musical segment varies in seconds, and each segment is saved as an audio file in the .wav format.
- Many features were extracted from the audio files during model-building due to each file's 16-bit mono resolution and 44100 hertz (Hz) sampling rate.

DOI: https://doi.org/10.54216/FPA.170119
Received: November 28, 2023 Revised: March 22, 2024 Accepted: July 28, 2024

Table 2: Description the dataset

Class Name	No of file	Minimum duration (s)	Maximum duration (s)
Zekrayati_P_1	**120**	**12**	**60**
Zekrayati_P_2	**120**	**26**	**41**
Zekrayati_P_3	**80**	**12**	**35**
Zekrayati_P_4	**120**	**46**	**91**

3.2. Pre-processing

Pre-processing generally refers to all operations performed on the signal in the frequency or time domain. On the other hand, the pre-processing for music analysis is limited to a specific temporal domain and includes the following steps.

Down-mixing refers to the process of reducing multi-channel signals to a mono channel. Before feature extraction, a stereophonic signal is typically converted to a monophonic signal [15]. in this work linear down-mixing has been used, in which the samples of every channel are averaged linearly to produce a singular channel x_0 :

$$x_0(n) = \frac{1}{C}\sum_{c=1}^{c} x_c(n) \qquad (1)$$

The process of windowing involves the selection of a segment that is sufficiently representative in order to analyze lengthy sound signals. This procedure is employed to eliminate noise from a signal that is contaminated by noise spanning a broad frequency range:

$$x(nt) = y(nt)W(nt), 0 \le nt \le N-1 \qquad (2)$$

In this context, "x(nt)" denotes the result of convoluting the 'input signal' with the window function, "y(nt) " represents the signal to be convolved by the window function, and " W(nt) " typically employs window hamming.

The pre-emphasis process involves signal filtering to reduce the intensity of frequency bands that transport critical information. In speech analysis, it is customary to apply a first-order "high-pass filter" to a signal denoted as "x(n)" to Emphasize details about the formants. [15]. This results in the generation of the pre-emphasis signal xp(n):

$$x_p(n) = x(n) - kx(n-1) \qquad (3)$$

The pre-emphasis coefficient, denoted as $k \in [0;1]$, governs the magnitude of the pre-emphasis (1 signifies the maximum, and 0 signifies the minimum). Speech processing values typically fall within the range of 0.9 to 0.97. For instance, a bandpass filter may be utilized in audio processing to reduce sub-bass effects or emphasize particular octaves. Furthermore, the high-pass pre-emphasis filter of the first order described above could be inverted to create "low-pass de-emphasis filter":

$$x_d(n) = x(n) - k_d x(n-1) \qquad (4)$$

3.3. Feature extraction

In music audio processing, feature extraction involves the numerical derivation or quantification of attributes present in an audio file's specific segment or frame. This is conducted to facilitate applying ML, statistical, and other algorithmic approaches to the signal-processing task. The present study extracted features pertaining to the time and frequency domains, which are elaborated upon in this section. The musical signals are initially partitioned into static frames by implementing a hamming window function. In this paper, the implemented frame size was 1024 ms with overlapping 512 ms, and all the following features were extracted.

3.3.1. Time-domain approaches

a) Zero-crossing rate (ZCR) algorithm

Assignable to the rate of change of signal sign throughout the frame, the ZCR of an audio frame is defined. The mathematical expression represents this as the frequency, a ratio of the signal's reversed sign to the frame's duration. Briefly, ZCR represents the number of occurrences during a one-second interval where a signal crosses zero. As defined and explained in [16], the ZCR for the ith frame of length N is as follows:

DOI: https://doi.org/10.54216/FPA.170119
Received: November 28, 2023 Revised: March 22, 2024 Accepted: July 28, 2024

$$Z(i)=\frac{1}{2N}\sum_{n1-1}^{N}\left|\text{sgn}\left[x_i(n1)\right]-\text{sgn}\left[x_i(n1-1)\right]\right| \quad (5)$$

As; $'sgn'(.)$ is a sign function

$$\text{sgn}\ \left[x_i(n1)\right]=\begin{cases}1, x_i(n1)\geq 0\\ 0, x_i(n1)<0\end{cases}$$

Short-time energy (STE) algorithm

Non-stationary characteristics characterize the auditory signals. The framing/windowing method [16] can convert a non-stationary signal into miniature portions of quasi-stationary signals. Due to the variable nature of the energy accompanying the signal, value prediction is not possible. The energy from a frame, or brief time energy, is computed for this purpose. Energy per frame is the definition of STE. Unvoiced segments have a low STE, while voiced segments have a high STE. From a discrete sound signal *x(n)*, STE is determined as shown in the following Equation

$$E=\frac{1}{N}\sum_{n=1}^{N}\left|x(n)\right|^2 \quad (6)$$

3.3.2. Frequency-domain approaches

a) Spectral centroid algorithm

The spectral centroid is intimately associated with this characteristic. It denotes the median frequency that is inherent in the spectrum of the signal. Because it is the median frequency, the higher and lower energies are in equilibrium. Tracking cadence in musical signals makes use of this function [17].

b) Spectral roll-off algorithm

When 95% of the energy of a signal is held below a specific frequency, it is referred to as the spectral roll-off point [18].

c) Spectral spread algorithm

This characteristic is intricately linked to the bandwidth of the signal [19]. It is the mean deviation of the rate map concerning its centroid. The spectral spread of noise-like signals is significantly larger than that of pure tonal sounds.

d) Spectral skewness algorithm

Third-order statistical measure spectral skewness illustrates the degree of symmetry exhibited by a given spectrum concerning its arithmetic mean. This attribute would take on a zero value for phases of silence and one for portions of speech. Three-order statistical characteristics comprise the skewness. Skewness with a value of zero signifies a symmetrical distribution, while skewness values less than zero and greater than zero indicate an excess of energy components on the right side of the spectrum and left side, respectively [19].

e) Spectral kurtosis algorithm

Conversely, kurtosis denotes the flattertness of the spectrum concerning its mean value and is a statistical measure of the fourth order. Spectral kurtosis is a parameter associated with Gaussian distributions. A uniform distribution is indicated by a kurtosis value of zero, while a steep peak is detected in the spectral kurtosis value of one. Stylistic classification of music genres also employs spectral kurtosis, similar to spectral skewness [19].

f) Spectral entropy algorithm

Like Shannon's entropy or Renyi's entropy, entropy is additionally a metric for determining the uniformity of flatness. For automatic speech recognition, this function has been implemented [20].

Shannon's entropy can be determined for a signal by employing the formula *sum(P_i log P_i)*, where P_i represents the probabilities of the sample classes.

g) Spectral flux algorithm

DOI: https://doi.org/10.54216/FPA.170119
Received: November 28, 2023 Revised: March 22, 2024 Accepted: July 28, 2024

The two-norm of the difference vector of spectral amplitudes between frames characterizes the spectral flux which detects abrupt shifts in the frequency energy distribution of noises [17].

h) Pitch algorithm

Pitch is a signal's fundamental period, which is used to describe differences in a waveform's periodicity among different audio sources. The average magnitude deference function (AMDF) is one of the techniques that is chosen to determine each frame's pitch. The calculation of AMDF [21] for a discrete signal is as follows:

$$AMDF(m) = \frac{1}{N-m-1} \sum_{n=0}^{N-m-1} |x(n) - x(n+m)| \quad (7)$$

The latency number is denoted by m and *x(n)* is the product of the audio sample sequence and a rectangular window of length N. 0 to *N-1* constitutes the range of *m*.

3.3.3. Cepstral domain features

a) Mel frequency cepstral coefficients (MFCCs) algorithm

The cepstral representation of music provides the framework for MFCCs. The discrete cosine transform of the logarithmic power spectrum on a nonlinear measurand scale is used to generate MFCCs, representing music's short-time power spectrum Figure 13. MFCCs are essential in a wide range of audio signal processing applications since their frequency bands are evenly spaced on the mel-scale, a characteristic that closely resembles the human auditory system [22].

b) Linear prediction cepstral coefficients (LPCCs) algorithm

Numerous benefits are associated with cepstrum, including separation from the source filter, orthogonality, and compactness. Cepstrum coefficients are robust and well-suited for ML due to these characteristics. Conversely, transforming the LPC to the cepstral domain is preferable due to the excessive sensitivity of linear prediction coefficients (LPC) to numerical precision. LPCCs refer to the transformed coefficients that are obtained. LPCCs can be considered as a derivative of LPCs [23].

Following the extraction process, the final feature vector was generated by computing information regarding the level of variance and dispersion, including the (mean, skewness, kurtosis, standard deviation, and minimum and maximum values).

3.4. Music classification

By utilizing solely, the test data features, a classifier attempts to generate a model that can forecast the target value of the test data. To arrive at the most accurate classifier in this work, we first compared the outcomes of widely recognized classifiers as documented in the literature. The various classifiers that are examined in the paper are detailed in the subsequent subsections. The classifier that produced the most accurate results was selected for the proposed music classification.

Four ML approaches were utilized in this article to classify the various stages of music classification: (NB), (KNN), (SVM), and (RF). The selected classifiers are discussed in the subsections that follow.

1) (SVM) algorithm

SVM is a supervised ML algorithm that categorizes music into diverse groups. Initially, SVM was designed for binary classification tasks [24]. While addressing multi-class problems, however, numerous scholars employ these prevalent approaches.

We have to construct a multiple-SVM (MSVM) to accommodate multi-class situations, given that SVMs are frequently used for binary data classification. Multiclass SVM employs the one-vs-one method to construct N(N-1) classifiers. Multi-class music classification can be accomplished using the linear SVM and the following formula:

$$C = \arg\max(W_m T * X + b) \quad (8)$$

b represents The term for bias, W_m represents The class weight vector m, and T signifies the transpose operator. In this context, the predicted class for the signal X is denoted as C.

DOI: https://doi.org/10.54216/FPA.170119
Received: November 28, 2023 Revised: March 22, 2024 Accepted: July 28, 2024

2) (K-NN) algorithm

The k-NN classifier assigns the input signal to the category having a higher frequency among its k neighbors [25] by identifying the k training instances in the training set that are in closest physical proximity to the input signal. The k-NN classifier operates based on the following Equation to classify music into numerous categories:

$$C = \arg\max\left(\sum_{m=1}^{n} l(y_m = C_k) \right) \quad (9)$$

The variable C represents the predicted class for the signal X. The number of neighbors to be considered is *n*. In the classification task, y_m represents the label of the m^{th} neighbor. C_k denotes the k^{th} class. The indicator function l accepts a value of 1 if the condition $y_m = C_k$ is satisfied, and 0 otherwise.

Numerous distance metrics, such as cosine similarity, Euclidean distance, and Jaccard similarity, can be employed by the k-NN classifier to compare music signals. The subsequent equations presents the formulations corresponding to the Manhattan, Euclidean, and Minkowski distance metrics.

$$Euclidean = \sqrt{\sum_{i=1}^{n} (p_i - q_i)^2} \quad (10)$$

$$Manhattan = \sum_{i=1}^{n} |p_i - q_i| \quad (11)$$

$$Minkowski = \left(\sum_{i=1}^{n} (|p_i - q_i|)^r \right)^{\frac{1}{r}} \quad (12)$$

3) (NB) algorithm

As shown in Equation (13), the NB classifier is a simple probabilistic classifier that assumes all features are equally independent [26]. It is founded upon Bayes' theorem and rigorous independence assumptions. The probability that the feature is allocated to a class of prior probability will be utilized to determine which class of posterior probability the feature is allotted to via the NB classifier. The class that obtains the greatest probability's is the effect for calculation. In addition to performing well in multiclass prediction, this classifier predicts the class of the test data set with speed and precision.

$$p(c \mid x) = \frac{p(x \mid c)\, p(c)}{p(x)} \quad (13)$$

As: p(c|x): represents the posterior probability of class", (c, target): given predictor (x, attributes), p(x|c): signifies the likelihood, which is the probability of the predictor given class, p(c): denotes the prior probability of the predictor, and p(x) signifies the probabilistic likelihood.

4) (RF) algorithm

RF employs multiple decision trees to enhance the classification's accuracy and robustness [27]. Many decision trees are constructed to generate the ultimate classification utilizing distinct training data and attribute subsets. Each decision tree is constructed using the RF algorithm, which recursively partitions the training data based on the feature values. RF's principal objective is the generation of random decision trees. The subsequent Equation may be applied to implement RF for the classification of music into multiple classes:

$$C = \arg max\left(\sum_{k=1}^{n} T_k(C_m) \right) \quad (14)$$

4. Evaluation Metrics

DOI: https://doi.org/10.54216/FPA.170119
Received: November 28, 2023 Revised: March 22, 2024 Accepted: July 28, 2024

In the following section, various evaluation criteria that assess the predictive performance of the models are detailed. Assessing the accuracy of an ML classification algorithm is one method for determining the frequency with which it correctly classifies a given data point [28]. The accuracy is determined by dividing the combined count of true positives and true negatives by the increased count of false positives, false_ negatives, true_ positives, and true_ negatives. Listed below is the formula for accuracy, precision, and recall:

$$Accuracy = \frac{Numberofcorrectpredi\,ctions}{Totalnumberofpredictions} \quad (15)$$

$$Precision = \frac{TruePositive}{TruePositive + FalsePositive} \quad (16)$$

$$Recall = \frac{TruePositive}{TruePositive + FalseNegative} \quad (17)$$

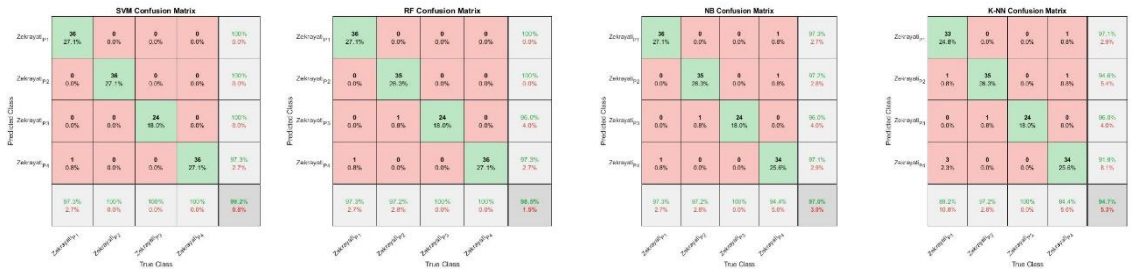
$$F1\_Score = 2 * \frac{Precision * Recall}{Precision + Recall} \quad (18)$$

5. Results and discussion

Each ML model was assessed with the comparison procedure. The sole purpose of the ML experiment was to verify the execution of the code and the configuration of the parameters, as detailed in Table 3. With an Intel 'i5' processor and '8' GB of RAM, we conducted our experiment in MATLAB 2021a. The suggested approach has combined the time-domain features (ZCR and STE) and frequency-domain features (spectral centroid, spectral roll-off, spectral spread, spectral skewness, spectral kurtosis, MFCCs, and LPCCs) for feature extraction for music part classification into Zekrayati_P_1, Zekrayati_P_2, Zekrayati_P_3, and Zekrayati_P_4. The dataset is partitioned into numerous sets for efficient model evaluation and training. At first, a training set comprising 70% of the dataset is designated, with the test set comprising the remaining 30%. A significant proportion of the data is utilized in the model's training process due to this division. Figure 2 displays confusion matrices for (a) SVM (b) RF (c) NB and (d) K-NN ML models.

Table 3: ML Models and Hyper parameters

Model	Parameters	Tested values
SVMs classifier	**Kernel**	**linear, rbf, sigmoid**
RF classifier	**min_samples_leaf**	**4, 5**
k-Nearest Neighbor classifier	**N_Neighbors**	**5, 7**
NB classifier	**Method**	**Gaussian**



(a) (b) (C) (D)

Figure 2. Confusion matrices for (a) SVM (b) RF (c) NB and (d) K-NN ML models

Table 4 illustrates the performance metrics pertaining to multi-class classification using SVM, RF, NB, and K-NN. The SVM model exhibits-superior performance in the categorization of music when compared to alternative classifiers. The SVM with the rbf kernel achieved a high performance rate of 99.20% for accuracy, 97.30% for recall, 99.20% for precision, and 98.26% for the F1 score, followed by RF with min_samples_leaf (4), then k-nn with n_neighbors (5), and finally NB with Gaussian.

DOI: https://doi.org/10.54216/FPA.170119
Received: November 28, 2023 Revised: March 22, 2024 Accepted: July 28, 2024

Table 4: The Performance Metrics for the Multi-Class Classification

ML algorithm	Precision	Recall	F1-Score	Accuracy
SVM	**99.20**	**97.30**	**98.26**	**99.20**
RF	**98.50**	**97.30**	**97.89**	**98.50**
NB	**97.00**	**97.30**	**97.14**	**97.00**
K-NN	**94.70**	**89.19**	**91.88**	**94.70**

- **GTZAN dataset for affirming the efficiency of the proposed system**

GTZAN-dataset results are discussed in this section. Downloading and accessing the dataset is possible via a public hyperlink: ("https://www.kaggle.com/datasets/andradaolteanu/gtzan-dataset-music-genre-classification"). GTZAN Dataset contains 1000 audio tracks with 30-second duration. It is organized into ten genres, each including 100 tracks. The dataset includes the following ten classes: (blues (bl)), (classical (cl)), (country (co)), (disco (di)), (hiphop (hi)), (jazz (ja)), (metal (me)), (pop (po)), (reggae (re)) and (rock (ro)). The number of songs for each genre is similar. All the tracks are "22050Hz" Mono "16-bit" audio files in .wav format. Table 5 shows the comparison accuracy of the proposed system with other studies over the GTZAN dataset.

Table 5: Comparison Accuracy with GTZAN Dataset

Reference	Dataset	Accuracy (%)
Tzanetakis et al. [29]	**GTZAN**	**61.00**
Holzapfel et al. [30]	**GTZAN**	**74.00**
Martins de Sousa et al. [31]	**GTZAN**	**79.70**
Kobayashi et al. [32]	**GTZAN**	**81.50**
Salamon et al. [33]	**GTZAN**	**82.00**
Proposed Method	**GTZAN**	**90.33**

With comparison with previous studies that used the GTZAN dataset, as shown in Table 5, it was found that most studies depended on "frequency-domain features," "time-domain features", or "cepstral domain." In this paper, all these features were combined with LPCCs, which are considered an effective technique for extracting musical information.

6. Conclusion and Future Work

ML techniques are beneficial for classification tasks, especially music genre classification, in which music is classified into different genres concerning its features. In the literature, there are a few works that indicate Arabic music classification. Therefore, this was the basis of this work. The dataset used in this paper has been sourced from the Zekrayati Arabic song melody, played using the lute instrument, in favour of this research. The dataset consisted of 440 music parts of different 4 classes. Four ML classifiers were utilized for supervised ML algorithm comparison and analysis. The experimental findings indicate that the SVM exhibited an optimal accuracy of 99.2%. As future work, authors will attend to combining CNN with mel-spectogram for abest result.

References

[1] Ness, A. J., & Kolczynski, C. A. (2001). Sources of Lute Music. Oxford Music Online. doi:10.1093/gmo/9781561592630.article.26299

[2] Yu, X., Ma, N., Zheng, L., Wang, L., & Wang, K. (2023). Developments and applications of Artificial Intelligence in music education. Technologies, 11(2), 42. doi:10.3390/technologies11020042

[3] Chen, J., Ramanathan, L., & Alazab, M. (2021). Holistic Big Data Integrated Artificial Intelligent Modeling to improve privacy and security in data management of Smart Cities. Microprocessors and Microsystems, 81, 103722. doi:10.1016/j.micpro.2020.103722

[4] Lee, J., Nazki, H., Baek, J., Hong, Y., & Lee, M. (2020). Artificial intelligence approach for Tomato Detection and mass estimation in Precision Agriculture. Sustainability, 12(21), 9138. doi:10.3390/su12219138

[5] Chaudhury, M., Karami, A., & Ghazanfar, M. A. (2022). Large-scale music genre analysis and classification using Machine Learning with apache spark. Electronics, 11(16), 2567. doi:10.3390/electronics11162567

DOI: https://doi.org/10.54216/FPA.170119
Received: November 28, 2023 Revised: March 22, 2024 Accepted: July 28, 2024

[6] Ng, W. W., Zeng, W., & Wang, T. (2020). Multi-level local feature coding fusion for music genre recognition. IEEE Access, 8, 152713–152727. doi:10.1109/access.2020.3017661

[7] Silva, D. F., Silva, M. V., Filho, R. S., & Silva, A. C. (2021). On the fusion of multiple audio representations for music genre classification. Anais Do XVIII Simpósio Brasileiro de Computação Musical (SBCM 2021), 37–44. doi:10.5753/sbcm.2021.19423

[8] Lekshmi, C. R., & Rajeev, R. (2023). Multiple predominant instruments recognition in polyphonic music using Spectro/Modgd-gram fusion. Circuits, Systems, and Signal Processing, 42(6), 3464–3484. doi:10.1007/s00034-022-02278-y

[9] Sharma, D., Taran, S., & Pandey, A. (2023). A fusion way of feature extraction for automatic categorization of music genres. Multimedia Tools and Applications, 82(16), 25015–25038. doi:10.1007/s11042-023-14371-8

[10] Zhao, Z. qi. (2019). Music classification model based on feature fusion. 2019 IEEE/ACIS 18th International Conference on Computer and Information Science (ICIS). doi:10.1109/icis46139.2019.8940205

[11] 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. doi:10.35940/ijitee.g5836.059720

[12] Prabavathy, S., Rathikarani, V., & Dhanalakshmi, P. (2021). Musical Instrument Sound classification using GoogleNet with SVM and KNN Model. Lecture Notes in Networks and Systems, 300, 230–240. doi:10.1007/978-3-030-84760-9_21

[13] Chaudhury, M., Karami, A., & Ghazanfar, M. A. (2022). Large-scale music genre analysis and classification using Machine Learning with apache spark. Electronics, 11(16), 2567. doi:10.3390/electronics11162567

[14] Deng, G., & Ko, Y. C. (2022). Active learning music genre classification based on support vector machine. Advances in Multimedia, 2022, 1–11. doi:10.1155/2022/4705272

[15] Zhang, W. (2022). Music genre classification based on Deep Learning. Mobile Information Systems, 2022, 1–11. doi:10.1155/2022/2376888

[16] syah, A., Yuliadi, B., & Sahara, R. (2018). Music genre classification using naïve Bayes algorithm. International Journal of Computer Trends and Technology, 62(1), 50–57. doi:10.14445/22312803/ijctt-v62p107

[17] Zhang, J. (2021). Music feature extraction and classification algorithm based on Deep Learning. Scientific Programming, 2021, 1–9. doi:10.1155/2021/1651560

[18] Atahan, Y., Elbir, A., Enes Keskin, A., Kiraz, O., Kirval, B., & Aydin, N. (2021). Music genre classification using acoustic features and Autoencoders. 2021 Innovations in Intelligent Systems and Applications Conference (ASYU). doi:10.1109/asyu52992.2021.9598979

[19] Prabhakar, S. K., & Lee, S.-W. (2023). Holistic approaches to music genre classification using efficient transfer and deep learning techniques. Expert Systems with Applications, 211, 118636. doi:10.1016/j.eswa.2022.118636

[20] Bhattacharjee, M., Prasanna, S. R., & Guha, P. (2020). Speech/music classification using features from Spectral Peaks. IEEE/ACM Transactions on Audio, Speech, and Language Processing, 28, 1549–1559. doi:10.1109/taslp.2020.2993152

[21] Zhang, K. (2021). Music style classification algorithm based on music feature extraction and Deep Neural Network. Wireless Communications and Mobile Computing, 2021(1). doi:10.1155/2021/9298654

[22] Khasgiwala, Y., & Tailor, J. (2021). Vision transformer for music genre classification using Mel-frequency cepstrum coefficient. 2021 IEEE 4th International Conference on Computing, Power and Communication Technologies (GUCON). doi:10.1109/gucon50781.2021.9573568

[23] Sul, A., Sarda, H., Billa, A., & Mishra, T. K. (2023). LPCC based music Genrefication using hybrid computational model. 2023 7th International Conference On Computing, Communication, Control And Automation (ICCUBEA). doi:10.1109/iccubea58933.2023.10392080

[24] S. Amer, R. (2020). Detection and Classification of Alcoholics Using Electroencephalogram Signal and Support Vector Machine. Fusion: Practice and Applications, 2(1), 14-21. DOI: https://doi.org/10.54216/FPA.020103

[25] Chen, L., Wang, C., Chen, J., Xiang, Z., & Hu, X. (2021). Voice disorder identification by using Hilbert-Huang transform (HHT) and k nearest neighbor (KNN). Journal of Voice, 35(6). doi:10.1016/j.jvoice.2020.03.009

[26] Wankhede, D. S., S., G. V., Manikjade, A., Meher, N., Atkale, T., Ghule, A., & Gujar, D. (2023). Analyzing the performance of naive Bayes, logistic regression, SVM and random forest for identifying

DOI: https://doi.org/10.54216/FPA.170119
Received: November 28, 2023 Revised: March 22, 2024 Accepted: July 28, 2024

hate speech from Twitter Social Media. 2023 7th International Conference On Computing, Communication, Control And Automation (ICCUBEA). doi:10.1109/iccubea58933.2023.10392242

[27] Raghav, A. Gupta, M. (2020). Ensemble Learning for Facial Expression Recognition. Fusion: Practice and Applications, 2(1), 31-41. DOI: https://doi.org/10.54216/FPA.020104

[28] Sharma, N. M., Kumar, V., Mahapatra, P. K., & Gandhi, V. (2023). Comparative analysis of various feature extraction techniques for classification of speech Disfluencies. Speech Communication, 150, 23–31. doi:10.1016/j.specom.2023.04.003

[29] Tzanetakis, G., & Cook, P. (2002). Musical genre classification of Audio Signals. IEEE Transactions on Speech and Audio Processing, 10(5), 293–302. doi:10.1109/tsa.2002.800560

[30] Holzapfel, A., & Stylianou, Y. (2008). Musical genre classification using nonnegative matrix factorization-based features. IEEE Transactions on Audio, Speech, and Language Processing, 16(2), 424–434. doi:10.1109/tasl.2007.909434

[31] Martins de Sousa, J., Torres Pereira, E., & Ribeiro Veloso, L. (2016). A robust music genre classification approach for Global and Regional Music Datasets Evaluation. 2016 IEEE International Conference on Digital Signal Processing (DSP). doi:10.1109/icdsp.2016.7868526

[32] Kobayashi, T., Kubota, A., & Suzuki, Y. (2018). Audio feature extraction based on sub-band signal correlations for music genre classification. 2018 IEEE International Symposium on Multimedia (ISM). doi:10.1109/ism.2018.00-15

[33] Salamon, J., Rocha, B., & Gomez, E. (2012). Musical genre classification using melody features extracted from Polyphonic Music Signals. 2012 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). doi:10.1109/icassp.2012.6287822

DOI: https://doi.org/10.54216/FPA.170119
Received: November 28, 2023 Revised: March 22, 2024 Accepted: July 28, 2024