Çukurova Üniversitesi Mühendislik Fakültesi Dergisi, 38(3), ss. 739-750, Eylül 2023
Cukurova University Journal of the Faculty of Engineering, 38(3), pp. 739-750, September 2023

Music Genre Recognition Based on Hybrid Feature Vector with Machine Learning Methods

Serhat HIZLISOY[*1] **ORCID** *0000-0001-8440-5539*
Recep Sinan ARSLAN[1] **ORCID** *0000-0002-3028-0416*
Emel ÇOLAKOĞLU[2] **ORCID** *0000-0003-1755-3130*

[1]Kayseri University, Faculty of Engineering, Architecture and Design, Department of Computer Engineering, Kayseri, Türkiye
[2]Kayseri University, Calculated Sciences and Engineering, Kayseri, Türkiye

Geliş tarihi: 16.08.2023 *Kabul tarihi: 29.09.2023*

Atıf şekli/ How to cite: HIZLISOY, S., ARSLAN, R.S., ÇOLAKOĞLU, E., (2023). Music Genre Recognition Based on Hybrid Feature Vector with Machine Learning Methods. Cukurova University, Journal of the Faculty of Engineering, 38(3), 739-750.

Abstract

Music genre recognition is one of the main problems in infotainment tools and music streaming service providers for different tasks such as music selection, classification, recommendation, and personal list creation. Automatic genre recognition systems can be useful for different music-based systems, especially different music platforms. Therefore, this study aimed to classify music genres using machine learning. In this context, GTZAN dataset consisting of 10 classes was used. In this dataset, data augmentation was applied by segmentation. Each record of 30 seconds was divided into 10 parts, increasing the number of samples in the dataset by a factor of 10. Then, features were extracted from the audio signals. The resulting features are chroma, harmony, mel frequency cepstral coefficients, perceptr, root mean square, roll-off, spectral centroid, tempo, and zero crossing rate. The types, variances, and averages of the obtained features were used. Thus, 57 features were obtained. This feature set was pre-processed by delimiting the decimal part, standardization, and label encoding. In the last step, classification was made with different machine learning methods and the results were compared. As a result of hyperparameter optimization in the Extra Tree model, 92.3% performance was achieved. Precision recall and f-score values are 92.4%, 92.3%, and 92.3%, respectively. As a result, an efficient and high-performance model in music genre recognition was created.

Keywords: Machine learning, Music genre recognition, Extra tree classifier, Segmentation

Makine Öğrenimi Yöntemleriyle Hibrit Özellik Vektörüne Dayalı Müzik Türü Tanıma

Öz

Müzik türü tanıma, müzik seçimi, sınıflandırma, öneri ve kişisel liste oluşturma gibi farklı görevler için bilgi-eğlence araçlarında ve müzik akışı servis sağlayıcılarında ana sorunlardan biridir. Otomatik tür tanıma sistemleri, farklı müzik tabanlı sistemler, özellikle farklı müzik platformları için yararlı olabilir. Bu sebeple,

[*]Sorumlu yazar (Corresponding Author): Serhat HIZLISOY, *serhathizlisoy@kayseri.edu.tr*

bu çalışmada makine öğrenmesi kullanılarak müzik türlerinin sınıflandırılması amaçlanmıştır. Bu kapsamda 10 sınıftan oluşan GTZAN veri seti kullanılmıştır. Bu veri setinde, segmentasyon ile veri büyütme uygulanmıştır. 30 saniyelik her kayıt 10 parçaya bölünerek veri kümesindeki örnek sayısı 10 kat artırılmıştır. Daha sonra da ses sinyallerinden öznitelikler çıkarılmıştır. Ortaya çıkan öznitelikler, renk, uyum, mel frekansı kepstral katsayıları, algılayıcı, kök kare ortalama, yuvarlama, spektral merkez, tempo ve sıfır geçiş oranıdır. Elde edilen özniteliklerin türleri, varyansları ve ortalamaları kullanılmıştır. Böylece 57 öznitelik elde edilmiştir. Bu öznitelik seti, ondalık bölümün sınırlandırılması, standardizasyon ve etiket kodlaması ile önceden işlenmiştir. Son adımda ise farklı makine öğrenmesi yöntemleri ile sınıflandırma yapılmış ve sonuçlar karşılaştırılmıştır. Extra Tree modelinde hiperparametre optimizasyonu sonucunda %92,3 performans elde edilmiştir. Kesinlik, hatırlama ve f-skoru değerleri sırasıyla %92,4, %92,3 ve %92,3'tür. Sonuçta, müzik türü tanımada verimli ve yüksek başarıma sahip bir model ortaya çıkarılmıştır.

Anahtar Kelimeler: Makine öğrenmesi, Müzik tür tanıma, Ekstra ağaç sınıflandırma, Bölünme

1. INTRODUCTION

Another way that people use to convey their feelings, feelings, or thoughts to the other party other than speech is music. Studies on music have been attracting a lot of attention recently.

There are various types of music that allow different people to have different musical preferences. The concept called music genre is actually the definition of a piece of music. It can be edited based on historical and cultural roots and also indirectly evokes the types of techniques and instruments used in the piece [1]. Each culture has its own melodies, unique sayings, and many works that have survived from the past to the present day using its own instruments.

With people being able to listen to all kinds of music and have easy access to them, the authors have started to be more influenced by different cultures. These works, which were produced from different cultures over time, resembled each other and formed certain musical genres. These types of music differ from each other in terms of rhythm, timbre, tempo, and melody.

In classification studies about music, it is seen that classification studies are carried out in breakdowns such as music genre, artist, user tags, and mood [2]. The structure of such works is based on melodic content.

People can group the type of music with their senses and feelings [3]. However, species classification with computer systems is a more complex process than with humans. In order to manage this complex process and to use computer systems in the field of music, a branch of science called Music Information Retrieval (MIR) has emerged. Music genre recognition (MGR) is one of the most important MIR field of study such as emotion recognition from music [4,5].

Especially nowadays, digital platforms want to serve their users better by offering similar recommendations to the genre of songs they like. For this purpose, they use machine learning and deep learning methods to increase the success of music genre recognition.

With the increasing number of digital systems and the data they produce, the data size is growing exponentially in many areas. This situation makes it difficult to work in many areas such as organizing, cleaning, and classifying content in different data types [6]. Based on this, a lot of work is being done to automate some of these processes. The increasing number of music makes such classification problems more difficult. Since it is not possible to manually classify and tag millions of music manually, this problem is tried to be solved by making users tag online.

Musical works can also be divided into different subgroups and it is processually appropriate to classify them with automated systems [7]. In this case, a precise and efficient MGR system is required to better retrieve music content.

Basically, an MGR system consists of two main parts: [7] The process of extraction of features and classification. As with many other classification problems, the features used in the classification of musical genres have a direct impact on the performance of the study. As for the classifier, there are studies using both machine learning and deep learning algorithms. The biggest advantage of deep learning is that it is possible to proceed through a single model without considering the feature extraction and classification processes separately.

When previous studies were reviewed, Gwardys and Grzywczak [8] achieved 78% accuracy on the GTZAN dataset for music genre recognition using Convolutional Neural Network (CNN) with Support Vector Machine (SVM).

On the other hand, Durdag and Erdogmus [9] transformed music into images in their study and tried to make recognition by establishing a relationship between the colors in the image and the music. They achieved recognition of 60% in the data set consisting of 5 different types of Turkish music and 54% in the GTZAN data set containing 10 different genres of music.

Arslan [10] made music genre recognition on the GTZAN dataset. Mel Frequency Cepstral Coefficients (MFCC) are used as the feature and Random Forest, Extra Tree, and Extreme Gradient Boosting (XGBoost) algorithms are used as a classifier. The highest accuracy rate was 90.5% with the XGBoost algorithm.

Le Thuy et. al. [11] achieved a 98.97% recognition rate by augmenting data on the Small Free Music Archive (FMA) dataset in their study and using mel spectrogram features as input to the DenseNet121 model.

Sharma et. al. [12], using the GTZAN dataset, studied both 4 music genres and 10 music genres. According to the method they proposed, they made 72.7% and 95.8% recognition for 10 classes with Ensemble Bagged Tree and Wide Neural Network, respectively.
Ashraf et. al. [13] used 4 different hybrid models including CNN and Recurrent Neural Network (RNN) varieties on the GTZAN dataset for music genre recognition in their study. They chose Mel-spectrogram and MFCCs as features. For MFCC, the hybrid model using CNN and LSTM achieved the highest accuracy with 76.40%, while the model using CNN and Bi-directional gated recurrent unit neural network (Bi-GRU) for the mel-spectrogram achieved the highest accuracy with 89.30%.

Yin [14] preferred GTZAN, FMA, and JUNO datasets. In his proposed method, the images were entered into CNN as mel spectrograms obtained over music signals and the classification process was completed with fully connected layers. With the use proposed MR-DCNN model, 63%, 78% and 89.7% recognition was achieved for the FMA, JUNO and GTZAN datasets, respectively.

Zhang [15], compared CNN and RNN with machine learning methods such as decision trees, random forest, achieved the most successful recognition rate with 78% using RNN.

Prabhakar et. al. [16] used GTZAN, ISMIR 2004, and MagnaTagATune as datasets. In this study, 5 different models were examined. Both machine learning algorithms and deep learning algorithms were analyzed. On the machine learning side, voice properties including LLDs, HSFs, pitch, and NMF properties were used, while on the deep learning side, spectrogram was used as a feature. Among these 5 analyses, the best result is the BAG model, which consists of a combination of Bi-LSTM (Bidirectional Long Short-Term Memory), Attention and GCN (Graphical Convolution Network). The highest success rate 93.51%, was achieved with the BAG model in the GTZAN dataset.

Jakubec et. al. [17] used GTZAN as the dataset in their study. They aimed to recognize an automatic type of music for in-car infotainment. Gaussian Mixture Model (GMM), SVM and K-Nearest Neighbour (k-NN) were used as classifiers. The features are MFCC, zero crossing rate (ZCR), spectral centroid, spectral flux, spectral roll-off, and chroma. The highest accuracy rate 69.7%, was obtained with the k-NN classifier.

Hongdan et. al. [18] used the MSD-I, GTZAN and ISMIR 2004 datasets in their study. SVM, k-NN and VGG-16 Net classifiers were studied. Feature vectors were created using BiLSTM. With the VGG-16 Net classifier, the highest accuracy was achieved in the GTZAN dataset with 97.8%.

In Singh et. al [19] studies, chromagram, MFCC, spectral contrast, swaragram, tonnetz, STFT (Short Time Fourier Transform), CQT (Constant-Q Transform), mel-spectrogram and harmonically separated mel-spectrogram were used as features. It was analyzed with 8 different deep learning models. Carnatic, GTZAN, Hindustani, and Homburg are the datasets used. The highest recognition rate of the M8 model was achieved with 98.28% in the GTZAN dataset and 98.62% in the Hindustani dataset. (feature mel-spectrogram).

The rest of the article is organized as follows: the proposed methodology and materials used are mentioned in the Section 2. The results obtained with the proposed model are given in Section 3. Section 4 presents discussions. In the last section, a general evaluation of the study is made, and suggestions for future studies are given.

2. MATERIAL AND METHOD

In this study, a method for automatic classification of music genres is proposed and tested. The architectural structure of the study is shown in Figure 1.

The first step of the model proposed in this study is the selection of the dataset. GTZAN dataset, which is one of the most widely used benchmark datasets for music classification, was selected and used for this study.

The GTZAN dataset consists of 1000 recordings of 10 classes. Each class covers 100 audio files, each with a 30-second audio clip in 22,050 Hz – 16-bit, .wav format, covering different genres of music: Blues, Classical, Country, Disco, Hip-hop, Jazz, Metal, Reggae, Rock, and Pop. This value can cause excessive learning, poor model performance, generalization failure, accuracy, and reliability problems in the analysis. The number of class-based samples should be increased to prevent this situation. To do this, each 30-second recording is divided into 10 parts. Thus, the data set was increased by 10 times. As a result, a dataset containing 1000 samples for each class was obtained.

Later in the process, features were extracted from the audio signals. The resulting features are chroma, harmony, MFCC, perceptr, root mean square (RMS), rolloff, spectral centroid, tempo, and zero crossing rate.

The types, variances, and averages of the obtained features were used. Using the GTZAN dataset, a compact set of 57 features has been identified that experimentally have proven to be efficient for classifying songs by different music genres. This set includes handcrafted song signal features, first and second-order feature derivatives, and lower and higher order feature statistics.

This feature set was pre-processed by delimiting the decimal part, standardization, and label encoding. Decimal parts increase up to 10 digits after feature extraction. This sitution prevents both data processing and learning stages from being efficient. For this reason, decimal parts have been adjusted to a maximum of 2 digits in the pre-processing stage. Thus, while the classification performance is not affected, the amount of data processed is seriously reduced. During the data standardization phase, the data were scaled to fit the normal distribution (mean 0 and standard deviation 1). Then, data were into training and testing data in a 90%-10%.

In this study, machine learning models were used as classifiers. During the creation of a model, different choices can be made in the design of the architecture. It is not possible to directly decide what the most appropriate model is. For this reason, the tuning process was used for hyper-parameter selection in the learning phase.

As the last step, classification was made with different machine learning algorithms, and the obtained values were compared.

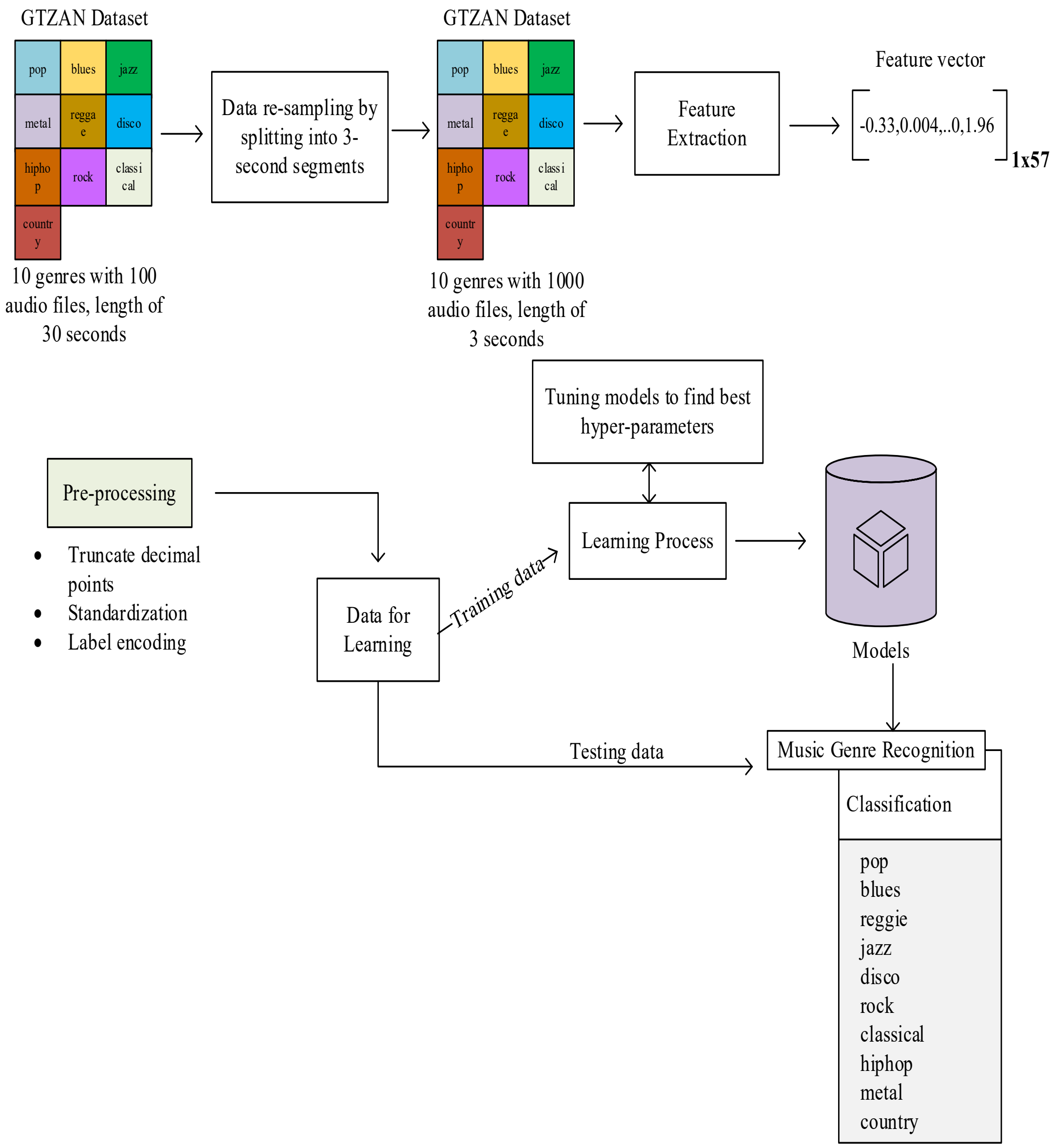

Figure 1. Architectural structure

In Table 1, the feature extraction method and the number of features obtained for each audio are given.

Table 1. Features used in this study

Feature names	Types	Count
Chroma_Stft	chroma_stft_mean,chroma_stft_var	2
Harmony	harmony_mean, harmony_var	2
MFCC	mfcc1_mean,mfcc1_var,mfcc10_mean,mfcc10_var, mfcc11_mean,mfcc11_var,mfcc12_mean,mfcc12_var, mfcc13_mean,mfcc13_var,mfcc14_mean,mfcc14_var, mfcc15_mean,mfcc15_var,mfcc16_mean,mfcc16_var, mfcc17_mean,mfcc17_var,mfcc18_mean,mfcc18_var, mfcc19_mean,mfcc19_var,mfcc2_mean,mfcc2_var, mfcc20_mean,mfcc20_var,mfcc3_mean,mfcc3_var, mfcc4_mean,mfcc4_var,mfcc5_mean,mfcc5_var, mfcc6_mean,mfcc6_var,mfcc7_mean,mfcc7_var, mfcc8_mean,mfcc8_var, mfcc9_mean,mfcc9_var	40
Zero Crossing Rate	zero_crossing_rate_mean, zero_crossing_rate_var	2
Perceptron	perceptr_mean, perceptr_var	2
RMS	rms_mean_rms_var	2
Roll-off	rolloff_mean, rolloff_var	2
Spectral Bandwidth	spectral_bandwidth_mean,spectral_bandwidth_var	2
Spectral Centroid	spectral_centroid_mean,spectral_centroid_var	2
Tempo	tempo	1

2.1. Pre-Processing

The limitation of decimal parts, standardization, and label encoding have been applied to the features. Limiting the decimal part means cutting or shortening the decimal part. That is, it refers to converting the decimal part of a number to a less precise value by cutting or rounding it to a certain number of digits.

The standardization process [20] involves transforming each feature so that it averages to 0 and its standard deviation to 1. In this way, the features have the same scales, and it is ensured that the features of different scales contribute to the models in a balanced way.

Label encoding [21] is used to convert categorical data into numeric values. This method allows machine learning models to process categorical data better. But it is suitable for sequential categorical data. If there is no ranking or rating between categories, this method should not be used. In such cases, other conversion techniques such as One-Hot Encoding may be preferable.

2.2. Hyper-parameter Tuning

In this study, ExtraTree, which has the highest performance, was used as a classifier. Hyper-parameter optimization was performed for the selected classifier, and the results shown in Figure 2 a,b,c, and d graphs were obtained.

The 'Min_samples_leaf' [22] extra tree is a hyperparameter used in machine learning models like decision trees. This value determines the minimum number of instances under a node. If the number of samples under a node falls below this value, then this node is considered a leaf node (does not make branches), and no further divisions are made in this node. According to Figure 2-a chart, the highest accuracy value is reached when we take min_samples_leaf i 1. Therefore, this value was taken as 1 in the model and classified.

The Figure 2-b chart examines the effects of min_samples_split values on model accuracy values. "Min_samples_split" [22]. like the previous hyperparameter is also a parameter used in machine learning models such as the extra tree, decision trees, etc. This hyperparameter determines how many minimum samples must be under decision trees before it can split a node. If the number of samples under a node falls below this value, that node is not divided and is considered a leaf node.

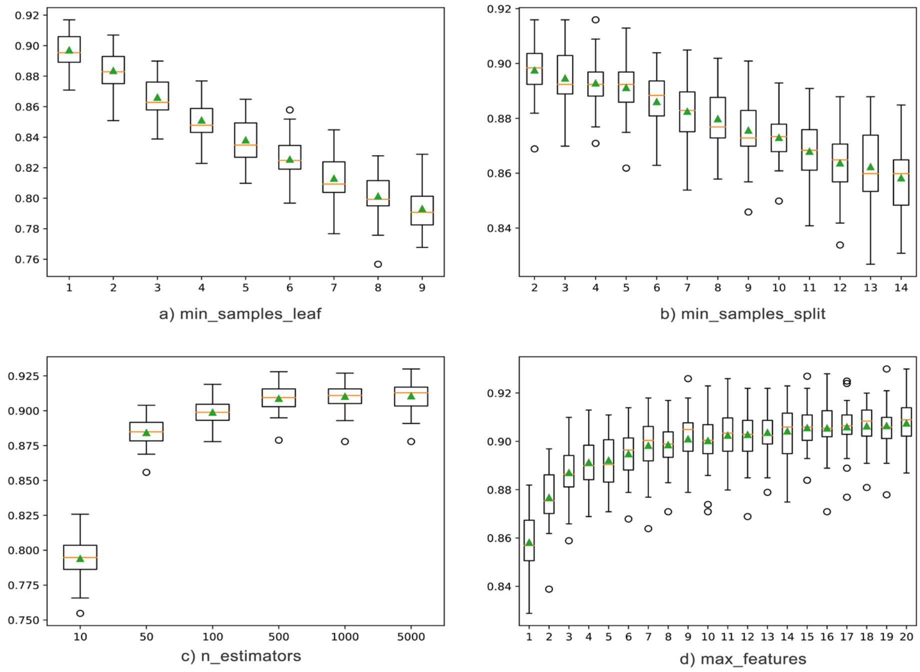

Figure 2. Hyper-parameter tuning

When the chart Figure 2-b was examined, the highest accuracy value was reached when this parameter value was 2. Therefore, this parameter was determined as 2 in the model. In the Figure 2-c chart, the total number of trees to be included in the model is compared with the model accuracy value. n_estimators' Extra Tree is a hyperparameter used in the Random Forest or Gradient Boosting algorithms. This hyperparameter specifies the total number of trees to include in the model. According to the Figure 2-c chart, the highest success rate was obtained with 5000 trees. Therefore, this parameter was given a value of 5000 in the model. The Figure 2-d chart compares the effects of max_features values on model accuracy. Max_features refers to the number of characteristics to consider when looking for the best distinction. When the chart was examined, it was seen that the value selected as 20 offered the highest contribution to the model accuracy.

2.3. Feature Analysis

The importance of 57 Features used in the study was compared in Figure 3.

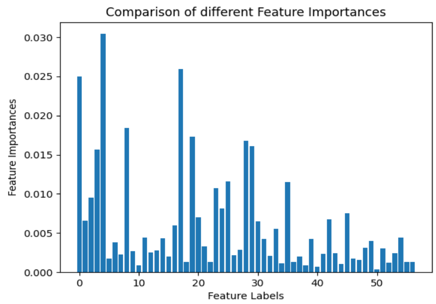

Figure 3. Features in the acope of operation

The impact of the features used when creating a machine learning model on process success varies. This effect determines the degree to which features are cared for by the model. According to the graph in Figure 3, the features in the order 0,4,8,18,20,29,30 (chroma_std_mean, spectral centroid mean, roll-off mean, mfcc1_var, mfcc2_var, mfcc7_mean, mfcc7_var) are more effective in the success of the model. It can also be noted as a general comment that features after feature 30 have lower effects on model success.

2.4. Classifier Selection

Within the scope of this study, machine learning algorithms were preferred as classifiers and accuracy values were compared. The algorithms used are Extra Tree, XGB, Random Forest, Logistic Regression, Decision Trees, Linear Discriminant Analysis, SVM, and k-NN.

3. RESULTS

In this study, music genre recognition was performed by using machine learning algorithms with the features extracted after data augmentation by segmentation in the GTZAN dataset. All obtained results are listed in Table 2.

Table 2. Comparison of all results

Algorithm	Accuracy	Precision	Re-call	F1-Measure
SVM	88%	88.1%	88%	88%
Logistic Regression	73.7%	73.6%	73.7%	73.5%
K Nearest Neighbor	90.3%	90.5%	90.3%	90.3%
Decision Trees	66.1%	65.9%	66.1%	65.9%
Extra Tree	**92.3%**	**92.4%**	**92.3%**	**92.3%**
LDA	70.3%	70.8%	70.3%	70.2%
XGB	89.2%	89.2%	89.2%	89.2%
Random Forest	89.5%	89.6%	89.5%	89.4%

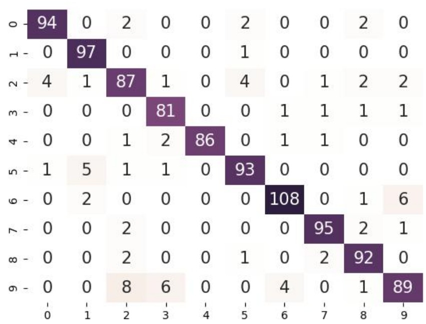

Figure 4. Extra tree confusion matrix

When we examine the results in Table 2, the highest accuracy value was obtained in the Extra Tree classifier and was 92.3%. In addition, SVM, k-NN, XGB and Random Forest classifiers achieved accuracy rates of 88% and above.

The confusion matrices for the Extra Tree, from which the highest accuracy value was obtained, are shown in Figure 4 and Figure 5.

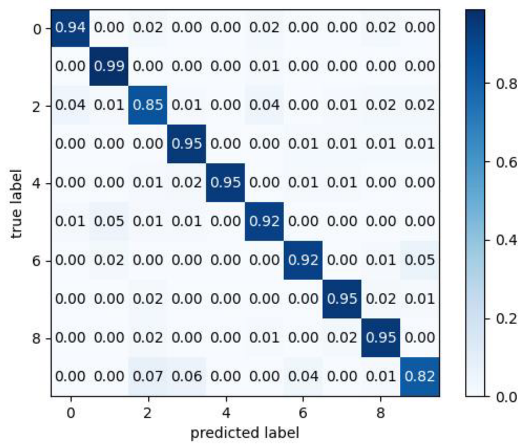

Figure 5. Extra tree confusion matrix percentage

The textual equivalents of the class labels used in Figure 4 and Figure 5 are: 0-Blues, 1-Classical, 2-Country, 3-Disco, 4-Hiphop, 5-Jazz, 6-Metal, 7-Pop, 8-Reggae, 9-Rock. Here, the class with the highest accuracy value is "Classical", and the lowest accuracy value is "Rock".

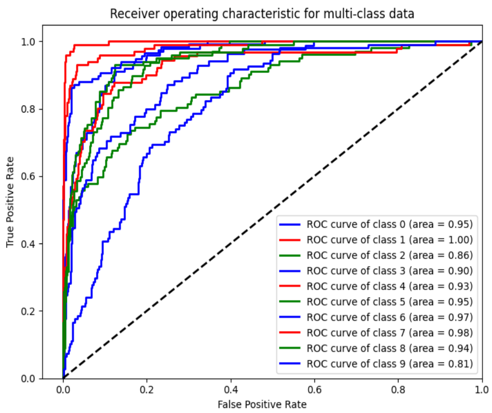

Figure 6. ROC curve

In the ROC curve, the y-axis is the positive ratio of True (TPR). It is also called sensitivity. Sensitivity (TPR) indicates the classification model's ability to accurately predict the positive class. A high TPR means that you are effectively predicting the positive class. If the False Positive Rate is FPR, the ratio of false positives (samples of negative class, that are mistakenly predicted as positive) to the total samples of negative classes. FPR indicates the rate at which the negative class is mistakenly predicted as positive. A low FPR indicates that the model correctly predicted the negative class, minimizing false positive predictions.

In an ideal situation, the TPR should be high (more true positive predictions), and the FPR should be low (fewer false positive predictions) so that the ROC curve converges to the upper left corner. In Figure 6, the ideal situation is realized. When the AUC values produced by the model for each class are examined, it is seen that it shows high performance for other species except the Rock genre. For some classes, this value reached 1.00. This means that the proposed model has high performance regardless of class.

4. DISCUSSIONS

With the model developed within the scope of this study, the highest accuracy value was obtained with the Extra Tree classifier as a result of the classification made with the features extracted after the data augmentation process through segmentation in the GTZAN dataset. (%92.3). In addition, when examined on a class basis, the class with the highest accuracy value is "Classical" (99%) and the class with the lowest accuracy value is "Rock" (82%).

One of the developers of this study (Arslan, 2021) conducted the study of music genre recognition with machine learning algorithms using the GTZAN data set. His 2021 study used MFCC features and the Random Forest, Extra Tree, and XGBoost algorithms. The best result was obtained with the XGBoost algorithm. (%90.5) The difference of this study from the previous study is that the data augmentation process was performed through segmentation through the GTZAN dataset, 57 feature vectors were included in the process except MFCC, and 3 different pre-processing steps were performed on the data set. As a result, the accuracy rate reached 92.3%.

Table 3 is the resulting table obtained when we compare our study with similar studies conducted recently. When we examine the table, it is seen that the studies with a higher recognition rate than this study are based on deep learning. However, the results in all deep learning models are not more successful than the developed model. Within the scope of the studies examined, one of the best results was obtained with machine learning algorithms.

Table 3. Previous studies

Study	Dataset	Features	Classifier	Accuracy rate (%)
Gwardys and Grzywczak (2014) [8]	GTZAN	Deep Image Features	SVM, CNN	78
Durdağ and Erdoğmuş (2019) [9]	5-Class Dataset GTZAN	Transformed Images	DCNN	60(5 Class) 54 (GTZAN)
Le Thuy et. al. (2022) [11]	FMA	Mel Spectrogram	DenseNet121	98.97
Sharma et. al.(2023) [12]	GTZAN	Q-Wavelet Transform, Teager Energy Operator Timbral, Chroma, and Source Separation Based Features	Ensemble Bagged Tree, Wide Neural Network	72.7(10 class) 95.8(4 class)
Ashraf et. al. (2023) [13]	GTZAN	MFCC, Mel Spectrogram,	CNN, RNN	76.40(MFCC) 89.30(Mel)
Yin (2023) [14]	GTZANFMA JUNO	Mel Spectrogram	CNN	63(FMA) 78(JUNO) 89.7(GTZAN)
Zhang (2023) [15]	GTZAN	DWCH	CNN, RNN, Machine Learning Algorithms	78(RNN)
Yilmaz et. al. (2022)[23]	GTZAN	Harmony, MFCC, Roll-Off, Chroma, Spectral Centroid, Perceptual, Zero Crossing Rate	K-NN, Gaussian Naive Bayes , SVM, Random Forest, Ada Boost ,Gradient Boosting , Logistic Regression,XGBoost	91,7
Liu et. al. (2023) [24]	GTZAN	STFT Spectrogram, Mel Spectrogram, CQT Spectrogram	LGNet (Locally Activated Gated Network), Naive Bayes, SVM, LSTM, ResNet, Bi-LSTM, FCN, FCN-LSTM	82.43
Arslan (2021) [10]	GTZAN	MFCC	Random Forest, Extra Tree, XGBoost	90.5
Proposed Method	**GTZAN**	**Chroma, Harmony, MFCC, Perceptual, RMS, Roll-Off, Spectral Centroid, Tempo, Zero Crossing Rate**	**SVM, Logistic Regression, K-NN, Decision Trees, Extra Tree, LDA, XGB, Random Forest**	**92.3**

5. CONCLUSION

Over the years, different methodologies have been proposed and tested for music genre recognition. While some of these are based on visuals, others are classified as sound-based. The approach of extracting sound-based features, combining them in a hybrid feature vector and classification with machine learning methods for music genre recognition has been adopted in this study.

In the data set used, each 30-second record was divided into 10 parts by segmentation, increasing the number of samples 10 times, and this process increased the performance of the recognition rate.

The model achieved an average classification performance of 92.3% with the Extra Tree classifier after hyperparameter optimization. Obtained results are given comparatively for different classifiers. The advantage of the model is its high performance with an uncomplicated structure.

In future studies, we plan to modify the number and types of features and analyze the results using deep learning algorithms. In addition, the success of the proposed model in different datasets and different problems will be examined

6. REFERENCES

1. Farajzadeh, N., Sadeghzadeh, N., Hashemzadeh, M., 2023. PMG-Net: Persian Music Genre Classification Using Deep Neural Networks. Entertainment Computing, 100518.
2. Çoban, Ö., Özyer, G.T., 2016. Music Genre Classification from Turkish Lyrics. In 2016 24th Signal Processing and Communication Application Conference (SIU), 101-104, IEEE.
3. Karatana, A., Yıldız, O., 2017. Music Genre Classification with Machine Learning Techniques. 2017 25th Signal Processing and Communications Applications Conference (SIU), Antalya.
4. Hizlisoy, S., Tufekci, Z., 2021. Derin Öğrenme ile Türkçe Müziklerden Müzik Türü Sınıflandırması. Avrupa Bilim ve Teknoloji Dergisi, (24), 176-183.
5. Hizlisoy, S., Yildirim, S., Tufekci, Z., 2021. Music Emotion Recognition Using Convolutional Long Short Term Memory Deep Neural Networks. Engineering Science and Technology, An International Journal, 24(3), 760-767.
6. Salazar, A.E.C., 2022. Hierarchical Mining with Complex Networks for Music Genre Classification. Digital Signal Processing, 103559.
7. Yu, Y., Luo, S., Liu, S., Qiao, H., Liu, Y., Feng, L., 2020. Deep Attention Based Music Genre Classification. Neurocomputing, 84-91.
8. Gwardys, G., Grzywczak, D., 2014. Deep Image Features in Music Information Retrieval. Intl Journal of Electronics and Telecommunications, 60(4), 321-326.
9. Durdağ, Z., Erdoğmuş, P., 2019. A New Genre Classification with the Colors of Music. Sakarya University Journal of Computer and Information Sciences, 2(1), 53-60.
10. Arslan, R.S., 2021. Automatic Music Genre Recognition Model Based on Machine Learning. Art and Desing-2021, 21-22 June. Niğde: Omer Halisdemir University.
11. Le Thuy, D., Loan, T., Thanh, C., Cuong, N., 2022. Music Genre Classification Using Densenet and Data Augmentation. Computer Systems Science and Engineering, 47(1), 657-674.
12. Sharma, D., Taran, S., Pandey, A., 2023. A Fusion Way of Feature Extraction for Automatic Categorization of Music Genres. Multimedia Tools and Applications (82), 25015-25038.
13. Ashraf, M., Abid, F., Din, I., Rasheed, J., Yesiltepe, M., Yeo, S., Ersoy, M., 2023. A Hybrid Cnn and Rnn Variant Model for Music Classification. Applied Sciences 13(3),1476.
14. Yin, T., 2023. Music Track Recommendation Using Deep-CNN and Mel Spectrograms. Mobile Networks and Applications, 1-8.
15. Zhang, X., 2023. Music Genre Classification by Machine Learning Algorithms. Highlights in Science, Engineering and Technology, 38, 215-219.
16. 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.
17. Jakubec, M., Chmulik, M., 2019. Automatic Music Genre Recognition for In-Car Infotainment. Transportation Research Procedia, 1364-1371.
18. Hongdan, W., SalmiJamali, S., Zhengping, C., Qiaojuan, S., Ren, Le., 2022. An Intelligent Music Genre Analysis Using Feature Extraction and Classification using Deep Learning Techniques. Computers and Electrical Engineering, 100, 107978.
19. Singh, Y., Biswas, A., 2022. Robustness of Musical Features on Deep Learning Models for Music Genre Classification. Expert Systems with Applications, 199, 116879.
20. Çiftler, A.F., 2019. Veri Bilimi Notları 4 – Özellik Ölçeklendirme / Normalizasyon / Standartlaştırma. https://tr.linkedin.com/pulse/veri-bilimi-notlar%C4%B1-4-%C3%B6zellik-%C3%B6l%C3%A7eklendirme-abdullah-faruk-%C3%A7i%CC%87ftler. Access date: August 2023.

21. Tilki, M., 2020. Label Encoder ve Onehotencoder Karşılaştırması. medium: https://medium.com/operations-management-T%C3%Bcrkiye/label-encoder-veonehotencoderkar%C5%9f%C4%B1la%C5%9ft%C4%B1rmas%C4%B1-C0983e884fc5, Access date: August 2023.

22. Scikit Learn, 2023. Sklearn. Ensemble. Extratreesclassifier. Sklearn: https://scikit learn.org/stable/modules/generated/sklearn.ensemble.extratreesclassifier.html, Access date: July 2023.

23. Yılmaz, P., Akçakaya, Ş., Özkaya, Ş.D., Çetin, A., 2022. Machine Learning Based Music Genre Classification and Recommendation System. El-Cezeri, 9(4), 1560-1571.

24. Liu, Z., Bian, T., Yang, M., 2023. Locally Activated Gated Neural Network for Automatic Music Genre Classification. Applied Sciences, 13(8), 5010.