Online Journal of Music Sciences, 2025, 10(2), 235-247
https://doi.org/10.31811/ojomus.1624182

Research Article ♪ Araştırma Makalesi

Subgenre classification in hip hop music an analysis of machine learning architectures

Hip hop müzikte alt tür sınıflandırması: Makine öğrenimi mimarilerinin analizi

Can Paşa[1], Abdurrahman Tarikci[2]

[1] İnönü Üniversitesi, Güzel Sanatlar ve Tasarım Fakültesi, Müzikoloji Bölümü, Müzik Teknolojisi Pr., Malatya, Türkiye.

[2] Ankara Müzik ve Güzel Sanatlar Üniversitesi, Müzik Bilimleri ve Teknolojileri Fakültesi, Müzik Teknolojisi Bölümü, Ankara, Türkiye.

ABSTRACT

Digitalisation and the proliferation of online music listening platforms have led to the exponential growth of music data on the Internet, thus necessitating the development of automated systems for data organisation and analysis. In this context, automatic genre classification practices have become a significant approach for the efficiency of music discovery and recommendation processes. While significant progress has been made in genre classification, subgenre classification remains an under-researched area, despite its potential to provide more personalised listening experiences. This study aims to address this gap by focusing on the classification of hip-hop music subgenres, namely boombap, jazzrap and trap, utilising a comprehensive dataset comprising 750 audio files. The study extracts a total of 31 features, encompassing both spectral and psychoacoustic characteristics. Machine learning models such as Logistic Regression, K-Nearest Neighbours, Decision Tree and Random Forest are employed, along with the Artificial Neural Network, which attains the highest accuracy of 85%. The findings reveal that subgenre classification poses challenges, especially for categories such as jazzrap and boombap, which share overlapping musical characteristics. In contrast, trap with different timbral characteristics was classified with higher accuracy. This study contributes to the scant research on subgenre classification by underscoring the viability of employing deep learning techniques to enhance the precision of comprehensive datasets and intricate subgenre categorisations. Moreover, this research underscores the pivotal role of subgenre classification within the ambit of digital music platforms. The accurate identification of subgenres not only elevates the overall auditory experience for users but also facilitates the discovery of music selections that resonate closely with their individual preferences.

Keywords: machine learning, deep learning, music subgenre classification, music information retrieval, music genre classification

ÖZ

Dijitalleşme ve çevrimiçi müzik dinleme platformlarının yaygınlaşması, internet ortamında müzik verilerinin katlanarak büyümesine yol açmıştır. Bu durum veri organizasyonu ve analizi için otomatik sistemlerin gerekliliğini ortaya koymuştur. Bu bağlamda, otomatik tür sınıflandırma pratikleri, müzik keşif ve tavsiye süreçlerinin verimliliği adına önemli bir yaklaşım haline gelmiştir. Tür sınıflandırmasında önemli ilerlemeler kaydedilmiş olsa da, daha kişiselleştirilmiş dinleme deneyimleri sunma potansiyeline rağmen alt tür sınıflandırması daha az araştırılmış bir alan olmaya devam etmektedir. Çalışma, 750 ses dosyasından oluşan bir veri kümesi kullanarak hip hop müzik alt türlerinin -boombap, jazzrap ve trap- sınıflandırılmasını ele almaktadır. Çalışmada, spektral ve psikoakustik öznitelikler de dahil olmak üzere toplam 31 özellik çıkarılmıştır. Lojistik Regresyon, K-En Yakın Komşular, Karar Ağacı ve Rastgele Orman gibi makine öğrenimi modellerinin yanı sıra %85'lik

Can Paşa — can.pasa@inonu.edu.tr
Geliş tarihi/Received: 21.01.2025 — Kabul tarihi/Accepted: 14.03.2025 — Yayın tarihi/Published: 30.04.2025

en yüksek doğruluğa ulaşan Yapay Sinir Ağı kullanılmıştır. Bulgular, alt tür sınıflandırmasının, özellikle örtüşen müzikal özellikleri paylaşan jazzrap ve boombap gibi kategoriler için zorluklar yarattığını ortaya koymaktadır. Buna karşılık, farklı tınısal özelliklere sahip trap daha yüksek doğrulukla sınıflandırılmıştır. Çalışma, alt tür sınıflandırması konusundaki sınırlı araştırmayı geliştirmekte ve derin öğrenme tekniklerinin kapsamlı veri kümelerinde ve karmaşık alt tür sınıflandırmalarında hassasiyeti artırmada umut verici şekilde uygulanabilirliğini vurgulamaktadır. Ayrıca bu araştırma, dijital müzik platformları bağlamında alt tür sınıflandırmasının hayati önemini vurgulamaktadır. Alt türlerin doğru bir şekilde tanımlanması, kullanıcılar için genel işitsel deneyimi geliştirmekle kalmaz, aynı zamanda bireysel tercihleriyle yakından uyumlu müzik seçimlerini keşfetmelerini de kolaylaştırmaktadır.

Anahtar kelimeler: makine öğrenmesi, derin öğrenme, müzikte alt tür sınıflandırma, müzik sorgulama sistemleri, müzikte tür sınıflandırma

1. INTRODUCTION

Digitalization has had a profound impact on music consumption and production practices. The widespread use of the Internet at the beginning of this century has led to a proliferation of music-related data, which has necessitated the development of automated systems to organize and process theese data for various purposes (Tzanetakis & Cook, 2002). In this context, the use of machine learning and deep learning architectures for music data organization and analysis has become inevitable. These architectures are now widely applied in various practices within the music industry, including automatic genre classification applications and music recommendation systems.

The emergence of studies on the automatic classification of music genres has significantly improved data organization efficiency, particularly on online music listening platforms. This advancement has also made it easier for individuals to access genres and subgenres for consumption. However, while listeners can generally distinguish between genres with relative ease, identifying subgenres often requires a certain level of experience and knowledge (Quinto et al., 2017). The prevalence of music listening platforms has underscored the necessity for subgenre classification studies, as well as genre classification studies. Nevertheless, studies on automatic subgenre classification have received comparatively less attention, as they concentrate on a more circumscribed context (Caparrini et al., 2020).

In the existing literature, one of the earliest studies on automatic genre classification is that by Tzanetakis and Cook (2002). The dataset used in this study comprised 10 distinct music genres. Gaussian Mixture Model (GMM) and K-Nearest Neighbour (K-NN) architectures were utilized in the research, achieving a noteworthy success rate of 61%. Another early study was conducted by Scaringella and Zoia (2005) using a dataset consisting of 1,400 songs and 7 genres. In this study, the following architectures were utilized: Support Vector Machines (SVM), Support Vector Machines with delayed input, Explicit Time Modeling Neural Network (ETM-NN), Recurrent Neural Network (RNN), and Hidden Markov Model (HMM), achieving a success rate of 69.98%. In the study conducted by Karatana and Yıldız (2017), K-Nearest Neighbour (K-NN), Random Forest, Support Vector Machine (SVM), and Artificial Neural Network (ANN) architectures were utilized over data from six different music genres, resulting in an 88.9% success rate. In a subsequent study by Bahuleyan (2018), the use of Deep Neural Network (DNN) and Convolutional Neural Network (CNN) architectures over a total of 40,540 datasets belonging to seven different species yielded in a 64% success rate.

In the context of subgenre classification studies, Quinto et al. (2017) undertook a classification study utilizing a comprehensive dataset comprising 640 minutes of audio material, encompassing three distinct subgenres of jazz music (namely, acid jazz, bebop, and swing). The study incorporated a range of algorithms, including Multi-Layer Perceptron (MLP), Support Vector Machine (SVM), K-Nearest Neighbours (KNN), and Long Short-Term Memory (LSTM) algorithms, culminating in a 90% success rate. In a separate study by Caparrini et al. (2020), Decision Tree, Random Forest, Extremely Randomized Trees, and Gradient Tree Boosting architectures were utilized over a dataset consisting of more than 20 subgenres of electronic dance music and more than 2,000 tracks in total, yielding a 59% success rate.

The present study focuses on classification of three subgenres of hip hop music: boombap, jazz rap and trap. In the course of determining the subgenres, timbral similarities and differences arising from periodic differences and/or musical relationships were taken into consideration.

• Boombap

Boombap, a highly related practice to sampling, has gained a subgenre identity through the technique of direct production. Although the term was first used by T La Rock in 1984 in the song 'It's Yours', it gained popularity with the album 'Return of the Boom Bap' released by KRS-One in 1991 (Mlynar, 2013). The term 'boombap' is an onomatopoetic term based on the sounds of powerful kick (boom) and hard snare (bap) beats in the production process of music. This genre, which typically features minimal instrumentation, is predicated on rhythmic elements (Exarchos, 2019). The genre's distinctive harmonic structure is rooted in the utilisation of samples derived from jazz, soul, and funk recordings of that era, a practice that was significantly influenced by the pervasive use of sampling in the music production process (Exarchos, 2019). The genre has been defined by D'Errico (2015, p. 281) as follows;

During the late 1980s and early 1990s, advancements in sampling technologies resulted in increased accessibility, leading to a period referred to as the 'golden age' of hip hop. The distinctive 'boombap' sound, characterised by the synchronised use of turntables and samplers in the production process, became a hallmark of this era.

• Jazzrap

Jazzrap represents an endeavour to amalgamate African-American musical traditions from bygone eras to the present day. This genre has adopted the rhythmic elements of hip hop and the harmonic elements of jazz and soul (AllMusic, n.d.). Originating from the fusion of dance music, hip hop and jazz, these genres emerged as products of urban African-American creativity. A parallel can be drawn between the rhythmic dominance of hard-bop jazz music in the 1950s and 60s and that of hip hop music (Williams, 2010). The existence of elements in jazz and hip hop music that are both historically and sonically related has contributed to the formation of jazzrap, a crossover genre that emerged from the club performances of DJs from the 1980s onwards.

• Trap

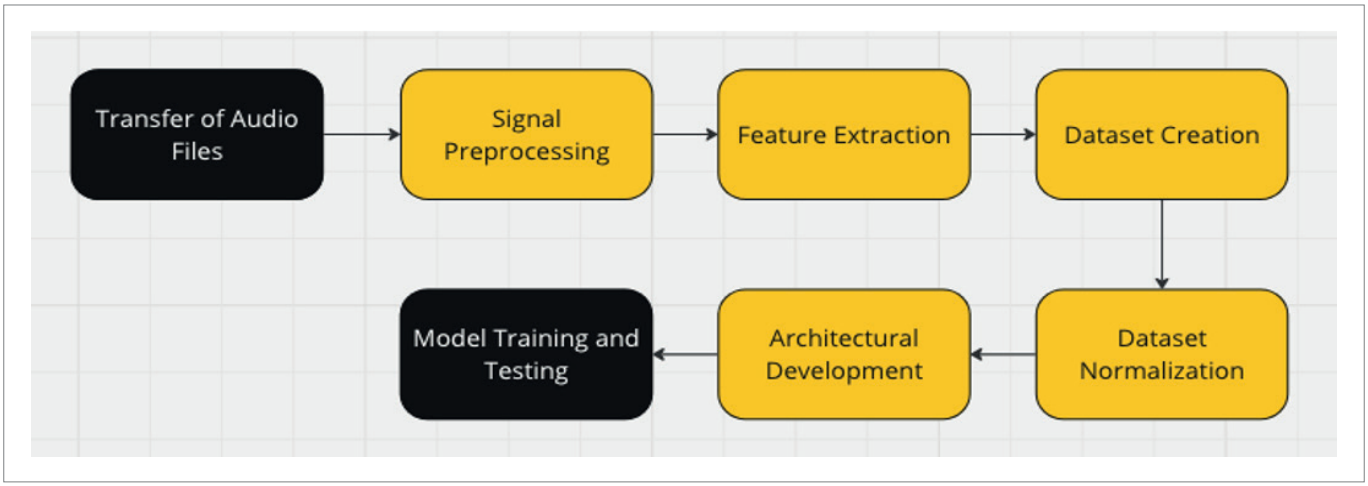
Trap, a hip-hop subgenre that emerged in the southern United States in the early 2000s, finds its origins in Atlanta slang. The most significant sonic characteristic of trap is the utilisation of three-time hi-hat which is based on the use of synthesised drums and bass (Reese, 2022). In addition, the use of intense and aggressive bass elements constitutes the main sonic characteristic of the genre.

The dataset used for the classification was meticulously prepared for this study and comprised 30-second samples extracted from a total of 750 songs, with 250 songs assigned to each genre. Initially, Logistic Regression, KNN, Decision Tree and Random Forest were employed as classical machine learning architectures. Subsequently, the ANN architecture, which forms the basis of deep learning architectures, was utilized. Following the application of hyperparameter optimization to each architecture, training and testing were performed, and the success rates of the architectures were compared. The analysis revealed that the ANN architecture emerged as the most effective, achieving a success rate of 85%. This classification study on the subgenres of hip hop music aims to make a meaningful contribution to the literature in the field.

2. MATERİAL AND METHOD

Figure 1

Flow chart of the method



2.1. Dataset

As shown in Table 1, the dataset under consideration contains a total of 750 audio files, 250 of which categorized into three subgenres. Each audio file has a sampling frequency of 44.1 kHz and a dynamic range of 16 bits. The sound files are in stereo, stored in .wav format and were specifically prepared for this study.

Table 1

Number of audio files in the dataset

Genre	Quantities
Boombap	250
Jazzrap	250
Trap	250
Total	750

2.2. Signal Processing

In order to facilitate the utilization of the audio files in the dataset as an input to machine learning and/or deep learning architectures, it is imperative to convert them into meaningful numerical representations. During this process, the audio data is transformed into a series of numerical values, with the extraction of features performed over the time and/or frequency axis representations.

Analyzing an audio signal as a whole cannot provide information about the instantaneous changes in the signal. Consequently, it is necessary to analyze the signal during specific periods. At this juncture, the audio signal is analyzed over short-term frames, a method referred to as framing (Eyben, 2016; Karatana & Yıldız, 2017). Subsequent to this, a process referred to as 'windowing' is applied to each frame. The purpose of this process is twofold; firstly, to prevent the distortion of data from the frame's end and start points, and secondly, to prevent spectral leakage between frames. Thereafter, the signal in each frame is transformed using the Fourier transform, moving it from the time axis to the frequency axis (Li & Song, 2021).

2.3. Feature Extraction

In this study, a total of 31 features were extracted from the audio files in the dataset. Initially, features commonly used in studies on genre classification in music were analyzed. Thereafter, the features in Table 2 were utilized to create a comprehensive dataset of timbral features of the three subgenres examined in the study. Although the songs included in the dataset were prepared with equal durations, differences in sampling rates and rounding errors during feature extraction may result in variations in the number of

features between tracks. To address this, the averages of the features of each frame in the tracks were taken to obtain an equal-length feature vector from each track (Elbir et al., 2018).

Table 2

Number of features in the dataset

Feature	Statistical Function	Number of Features
RMS	Mean	1
Chroma STFT	Mean	1
Spectral Centroid	Mean	1
Spectral Bandwidth	Mean	1
Spectral Rolloff	Mean	1
MFCC (13 coefficients)	Mean	13
MFCC Delta (13 coefficients)	Mean	13
TOTAL		31

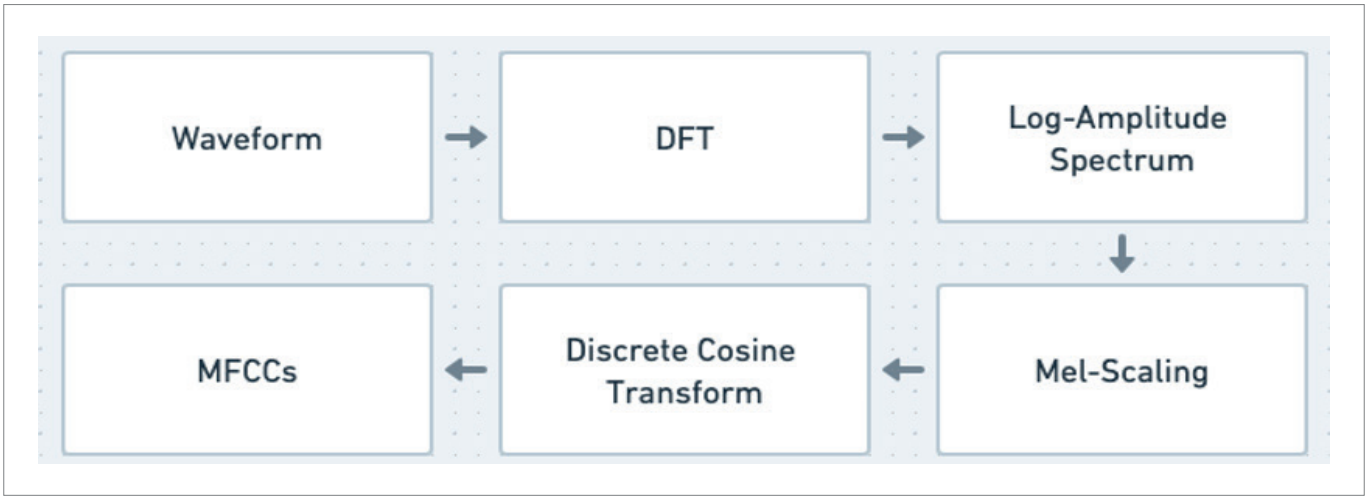
The extraction of features was conducted utilizing the Python[1] programming language and Librosa[2], a library designed for the analysis of music and audio. The following list presents the features that were extracted;

Mel-Frequency Cepstral Coefficients (MFCC): Mel-frequency Cepstral Coefficients are derived from the spectrogram of the signal. Following the framing and windowing processes applied to the signal, Mel-scale filters are applied to the signal carried to the frequency domain as a result of the Fourier transform performed in each frame (Knees & Schedl, 2016). The rationale behind this choice is that human pitch perception is more sensitive at low frequencies than at high frequencies, and the mel scale is a suitable simulation of this (Li & Song, 2021). As shown in Figure 2, the next step in the process is to take the logarithm of the weighted frequency components, apply the discrete cosine transform and move these back to the time domain. This process allows the relevant coefficients to be obtained (Sönmez & Varol, 2019).

$$M(f) = 2595 \cdot log_{10}\left(1 + \frac{f}{700}\right) \quad (1)$$

Figure 2

MFCC extraction scheme



Delta Mel-Frequency Cepstral Coefficients (Delta MFCC): These represent the successive changes between the mel-frequency cepstral coefficients that have been obtained.

[1] https://www.python.org/

[2] https://librosa.org/doc/latest/index.html

Spectral Centroid: This is defined as the point at which the centre of gravity of the frequency content of a signal is located. That is to say, it is the point on the frequency axis where the energy of the signal is concentrated. This representation calculates the mean of the amplitude spectrum of the sound along the frequency axis. As a result, this feature provides information about the 'brightness' level of the sound (Knees & Schedl, 2016).

$$C_t = \frac{\sum_{n=1}^{N} M_t[n] \cdot n}{\sum_{n=1}^{N} M_t[n]} \tag{2}$$

In the formula M_t[n], is the amplitude value at the designated time *t* and frequency point *n* (Tzanetakis & Cook, 2002). It has been established that low spectral centre values indicate 'darker' sounds, whereas high spectral centre values are indicative of 'brighter' sounds and higher frequency energy.

Spectral Rolloff: This is the representation of the frequency value corresponding to below 85% of the total energy obtained over the frequency spectrum of the signal (Tzanetakis & Cook, 2002). This feature insights into how much of the signal's frequency energy is concentrated at lower frequencies.

$$\sum_{\mathrm{n=1}}^{\mathrm{R_t}} \mathrm{M_t[n]} = 0.85 \cdot \sum_{\mathrm{n=1}}^{\mathrm{N}} \mathrm{M_t[n]} \tag{3}$$

Spectral Bandwidth: This metric, which is calculated based on the spectral center of gravity, provides information about the range of frequencies around the spectral center of gravity. In other words, it can be interpreted as the deviation from the signal's center frequency (Knees & Schedl, 2016).

$$BW_t = \frac{\sum_{n=1}^{N} |n - SC_t| \cdot m_t(n)}{\sum_{n=1}^{N} m_t(n)} \tag{4}$$

In the formula, $|n\text{-}SC_t|$ denotes the distance of each frequency point from the spectral centre.

Chroma STFT: The pitch class vector represents the energy distribution of music signals across 12 different pitch classes on the frequency axis (Atahan et al., 2021). In the obtained pitch class vector, the vertical axis represents 12 different notes, while the horizontal axis represents the energy levels of the frequency components (Ashraf et al., 2021).

Root Mean Square Energy– RMSE: The square root of the mean of the squared signal, which fluctuates in amplitude over time, is particularly significant in this context. This metric provides information about the average level of the signal. The formula indicates that *K* represents the number of samples within a window, while *s(k)* denotes the amplitude value of the signal at point *k*

$$RMS_t = \sqrt{\frac{1}{K} \sum_{k=t}^{(t+1)K-1} s(k)^2} \tag{5}$$

2.4. Classification Architectures

Logistic Regression: This machine learning architecture performs classification based on pre-labeled classes, with the features in the dataset forming the basis of this process. In this context, logistic regression is categorized under supervised learning models and is applicable to both binary and multi-class classification tasks. A notable feature of this model is the use of sigmoid activation function for binary classification and a softmax activation function for multiple classification (Jurafsky & Martin, 2024).

Decision Tree: The decision tree architecture utilised in classification and regression analyses exhibits a hierarchical structure, with the generation of output being facilitated through a set of decision rules (Jijo & Abdulazeez, 2021). Each node of the decision tree signifies a decision point based on specific features. The branches from each node represent the probabilities of the decision, with each leaf representing the classification or decision output (Babar, 2024).

K-Nearest Neighbour: The K-nearest Neighbours algorithm is a machine learning architecture used for classification, clustering, and regression analysis. In the context of classification, this architecture is employed to store the dataset that has been provided as input and to assign it to one of the previously labeled classes, according to the distance between it and other examples in the dataset when a new dataset is provided. In this particular context, the algorithm uses metrics such as the Euclidean distance and cosine similarity (Boyko & Mykhailyshyn, 2023).

Random Forest: The machine learning architecture under discussion is constructed from multiple decision trees. Each tree is generated from different subsets and features of the training data, and the output of the architecture is determined by averaging the output of each tree or by majority voting (Segal, 2004).

Artificial Neural Network - ANN: The term 'artificial neural networks' refers to a specific architectural configuration that facilitates the learning process through the modeling of the information processing capabilities of the human brain. The functionality of neurons in the human brain is replicated using various mathematical models, thereby simulating their operation. The architecture of the neural network is characterized by its extensive interconnectedness, with each neuron processing data through its activation (output) function, resulting in the generation of its own output. The specific numbers and layers of these neurons vary according to the architectural design. These networks are categorized as either single-layer or multi-layer neural networks (Mehrotra et al., 1997; Wu & Feng, 2018).

In this study, the Python library Scikit-Learn[3] was utilized for the implementation of logistic regression, K-NN, decision tree, and random forest architectures. The dataset was scaled to improve the efficiency of the architecture; in this context, standard scaling was applied to the dataset. Standard scaling is a method of scaling the mean of the features of the dataset as 0 and the standard deviation as 1 (Baladram, 2024). This scaling prevents classification architectures from being biased against features with large scales. Moreover, standard scaling has been shown to be highly efficient for architectures that utilise gradient-based optimisation methods, such as artificial neural networks (ANN), and that function on the basis of vectorial distance between features, such as k-nearest neighbours (KNN) (James et al., 2013).

$$x_{scaled} = \frac{x - \mu}{\sigma} \tag{6}$$

In equation 6, x denotes the features contained within the dataset while μ and σ represent the mean and standard deviation of said features, respectively. Following this, grid search[4], a method of hyperparameter optimization, was employed to enhance the performance of the architectures. In addition, to achieve a more precise measurement of the model's accuracy, the cross-validation method was employed.

The artificial neural network architecture utilized in the present study incorporated the TensorFlow[5] library. In accordance with prior architectures, standard scaling was applied to the dataset values. The architecture consists of two fully connected layers: one hidden layer, and one output layer. The hidden layer employs the ReLU activation function, whereas the output layer utilizes the Softmax activation function, given the execution of multiclass classification. As shown in Figure 3 to enhance the architecture's performance, regularization is applied between the hidden layer and the output layer using a Dropout[6] layer.

[3] https://scikit-learn.org/stable/index.html

[4] Grid search is a widely utilised approach for the optimisation of hyperparameters in machine learning models. This method involves the construction of a predefined grid comprising the values and all possible combinations of hyperparameters, which are then systematically evaluated.

[5] https://www.tensorflow.org/

[6] The term "dropout" refers to dropping out the nodes (input and hidden layer) in a neural network. All the forward and backwards connections with a dropped node are temporarily removed, thus creating a new network architecture out of the parent network. The nodes are dropped by a dropout probability of p (Yadav, 2022).

Figure 3

ANN architecture summary

Layer (type)	Output Shape	Param #
dense_2 (Dense)	(None, 32)	1,024
dropout_1 (Dropout)	(None, 32)	0
dense_3 (Dense)	(None, 3)	99

Total params: 1,123 (4.39 KB)
Trainable params: 1,123 (4.39 KB)
Non-trainable params: 0 (0.00 B)

3. DISCUSSION

In this study, the accuracy score, F1 score, and confusion matrix were utilized to evaluate the applied classification architectures. These metrics, widely employed in the evaluation process of classification architectures, are described in Table 3.

An analysis of Table 4 reveals that the F1 and accuracy scores are closely aligned, indicating that the class distributions within the training data are balanced. This suggests that the number of data points belonging to the three subgenres in the training data is approximately equal or very close to each other. The study aimed to ensure objectivity in comparing different architectures by using an equal number of samples from each subcategory. Based on the literature review, it was observed that the average data duration for each subcategory in classification studies ranged from 250 to 12,000 minutes. In the dataset specifically prepared for this study, the data duration for each subcategory was set at 7,500 minutes. Additionally, the computational cost increases with the size of the dataset. In this context, the number of data samples obtained from each subcategory reflects the diversity of the species while maintaining a reasonable level of analysis cost compared to similar studies.

Table 3

Accuracy metrics used

Metric	Definition	Purpose
Accuracy	Measures the proportion of correctly predicted instances.	(TP + TN) / (TP + TN + FP + FN)
F1 Score	The harmonic mean of Precision and Recall.	2 × (Precision × Recall) / (Precision + Recall)
Confusion Matrix	Displays the counts of true positives, true negatives, false positives, and false negatives.	Used for detailed analysis of prediction errors.

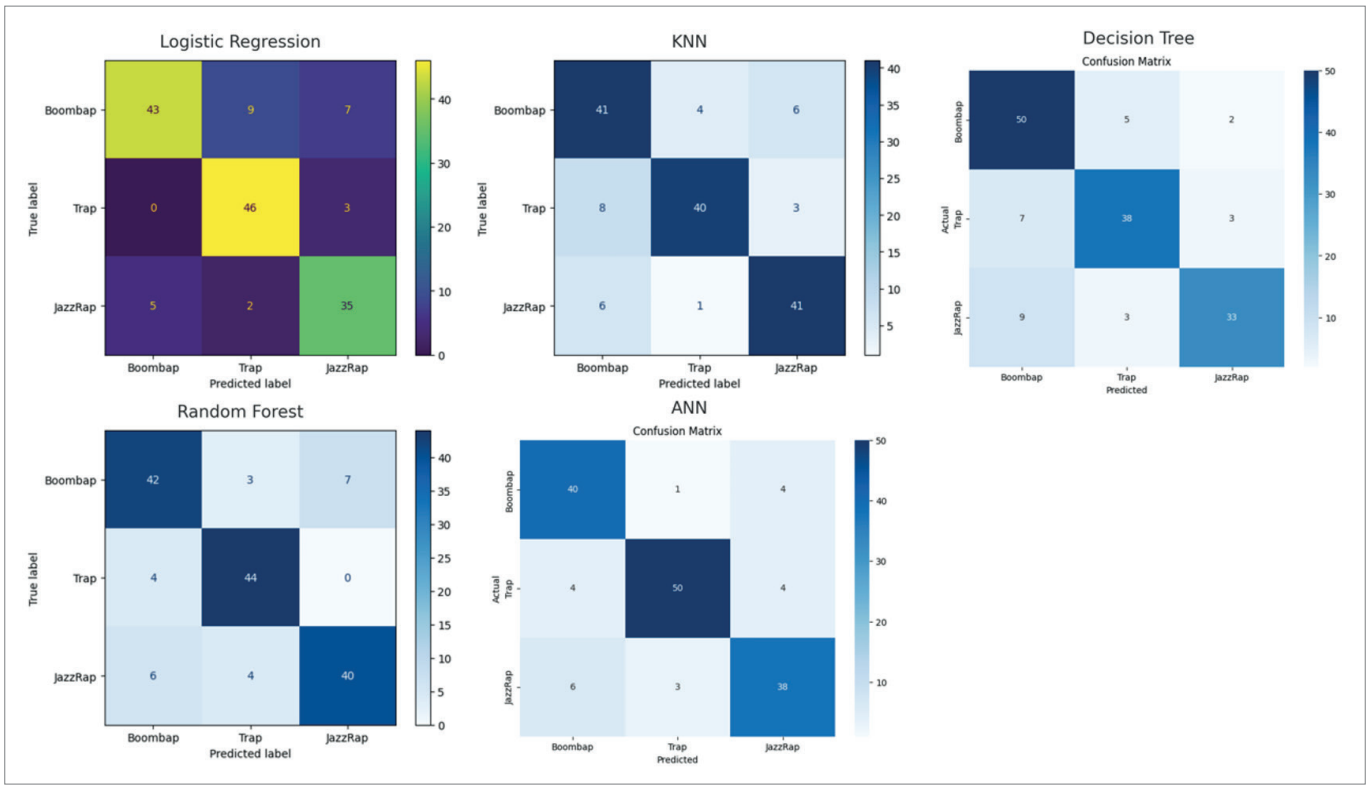
Table 4

Accuracy and F1 scores for architectures

Architecture	Accuracy	F1 Score
Logistic Regression	0.82	0.82
KNN	0.81	0.81
Decision Tree	0.80	0.80
Random Forest	0.84	0.83
ANN	0.85	0.85

Figure 4

Confusion matrix of the architectures



When the confusion matrices of the architectures are analyzed, it is observed that Trap is predicted correctly at a high rate in the ANN and Random Forest architectures, with minimal confusion in other genres. However, greater confusion is noted in Jazzrap and Boombap, particularly in the Decision Tree and KNN architectures.

The confusion of classification architectures in the case of jazzrap and boombap is considered to be due to the fact that these subgenres have similar musical characteristics. In fact, both frequently incorporate jazz elements. Jazzrap directly draws on jazz music codes in terms of rhythmic and harmonic structure, employing instrumentation similar to standard jazz orchestration. Boombap, on the other hand, builds its musical infrastructure directly from sampling. Its rhythmic elements are shaped by samples from acoustic drums and/or electronic sounds via drum machines, while its harmonic structure is typically derived from samples of old jazz, funk, and soul recordings. The shared jazz elements in these subgenres are believed to result in similar spectral features.

Figure 5

Distribution of subgenres in 2D

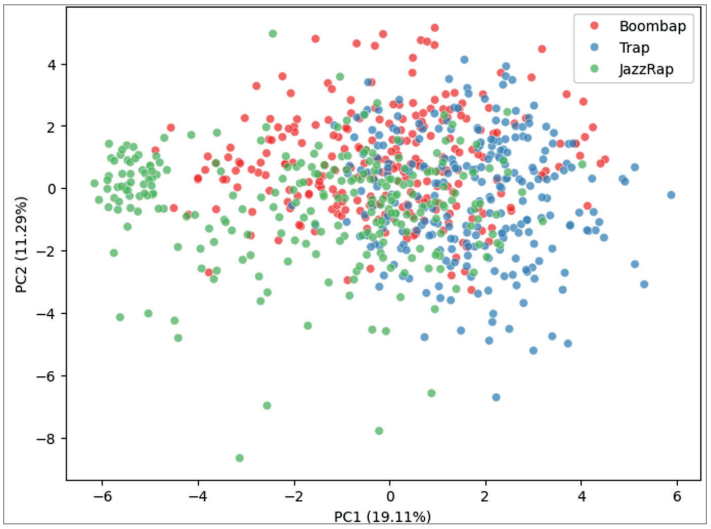


Figure 5 demonstrates that a considerable proportion of the subgenres exhibit a high degree of intermingling in the middle region (PC1 and PC2 are close to zero values). This finding suggests that the subgenres are to a certain extent incapable of differentiating due to the presence of shared musical elements (e.g. the utilisation of jazz samples, comparable rhythmic structures or analogous timbral features). Of particular note is the observation of an intertwining of jazzrap and boombap in the central region, a phenomenon consistent with the presence of jazz-based elements in both subgenres. The overlap of jazzrap and boombap, both in the central region and partially on the left-hand side, is indicative of the similarities between the two subgenres, as evidenced by the shared presence of jazz elements. Consequently, while the PCA graph discloses certain divergences between the genres, particularly with regard to Trap, it also illuminates a substantial degree of intertwinement between jazzrap and boombap, attributable to shared timbral characteristics.

In the study, the ANN architecture proved to be the most successful, achieving an 85% accuracy rate and F1 score. This is due to its efficiency in non-linear decision structures, which minimized the complexity of the relationship between boombap and jazzrap compared to other architectures, despite the ANN being a single-layer neural network. Deeper neural networks were also trained during the testing phase of the ANN architecture, but overfitting was encountered due to the number of datasets, which reduced the generalization performance of the architecture.

4. CONCLUSION

The present study seeks to enhance the existing body of research on subgenre classification by analyzing the trap, boombap, and jazzrap subgenres of hip hop music. It emphasizes the timbral distinctions among these three subgenres, prioritizing frequency domain and psychoacoustic features during the feature extraction process. The classification employs both traditional machine learning models and artificial neural networks, which underpin deep learning. At this juncture, the performance of a shallow artificial neural network and that of traditional machine learning architectures can be observed in conjunction on the same dataset.

The classification architectures faced greater challenges in differentiating between jazzrap and boombap, which share similar musical characteristics, compared to trap, which is marked by more distinct timbral features. This difficulty arises from the jazz-based elements present in both jazzrap and boombap, leading to similarities in their spectral features. In contrast, the unique timbral characteristics of the trap genre allowed the architectures to achieve higher accuracy rates in its classification. This highlights the critical importance of the feature extraction process in subgenre classification.

It has been posited that traditional machine learning architectures and shallow neural networks may be inadequate in handling the intricacies and non-linear patterns inherent in bigger datasets (Najafabadi et al., 2015). As a result, the utilization of deep neural networks with extensive datasets has been shown to improve the effectiveness of differentiating between closely related subgenres in terms of their timbral features. In a similar vein, the integration of deep neural network-driven frameworks within digital music streaming services, distinguished by their expansive data repositories, is anticipated to elevate the accuracy of subgenre classification functionalities. In addition to timbral features, classification processes that include rhythmic features and metadata such as lyrics are predicted to be able to distinguish between genres with similar timbral elements more efficiently. In this case, subgenre classification algorithms that allow users to make more accurate music recommendations can improve the user experience and facilitate the library management of online music listening platforms. Specifically, the demand for discovering new genres or accessing songs in similar genres can be met more effectively with such comprehensive and high-performance subgenre classification systems. In conclusion, the accurate and automatic classification of hip hop subgenres is not only of academic interest, but also of great importance in terms of industrial applications and user experience.

Ethical approval

This study does not require ethics committee approval as it does not involve human, animal or sensitive data.

Author contribution

Study conception and design: CP, AT; data collection: CP; analysis and interpretation of results: CP; draft manuscript preparation: CP, AT. All authors reviewed the results and approved the final version of the article.

Source of funding

The authors declare the study received no funding.

Conflict of interest

The authors declare that there is no conflict of interest.

Etik kurul onayı

Bu çalışma insan, hayvan veya hassas veriler içermediği için etik kurul onayı gerektirmemektedir.

Yazarlık katkısı

Çalışmanın tasarımı ve konsepti: CP, AT; verilerin toplanması: CP; sonuçların analizi ve yorumlanması: CP; çalışmanın yazımı: CP, AT. Tüm yazarlar sonuçları gözden geçirmiş ve makalenin son halini onaylamıştır.

Finansman kaynağı

Yazarlar, çalışmanın herhangi bir finansman almadığını beyan etmektedir.

Çıkar çatışması

Yazarlar, herhangi bir çıkar çatışması olmadığını beyan etmektedir.

REFERENCES

AllMusic. (n.d.). *Jazz rap*. AllMusic. Retrieved February 17, 2025, from https://www.allmusic.com/subgenre/jazz-rap-ma0000012180

Ashraf, M., Ahmad, F., Rauqir, R., Abid, F., Naseer, M., & Haq, E. (2021). Notice of violation of IEEE publication principles: Emotion recognition based on musical instrument using deep neural network. In *2021 International Conference on Frontiers of Information Technology (FIT)* (pp. 323-328). IEEE. https://doi.org/10.1109/FIT53504.2021.00066

Atahan, Y., Elbir, A., Keskin, A. E., Kiraz, O., Kirval, B., & Aydın, N. (2021). Music genre classification using acoustic features and autoencoders. In *2021 Innovations in Intelligent Systems and Applications Conference (ASYU)* (pp. 1-5). IEEE. https://doi.org/10.1109/ASYU52992.2021.9598979

Babar, K. (2024). Performance evaluation of decision trees with machine learning algorithm. *International Journal of Scientific Research in Engineering & Management, 8*(5), 1-5. https://doi.org/10.55041/ijsrem34179

Bahuleyan, E. (2018). *Music genre classification using machine learning techniques*. arXiv. https://arxiv.org/abs/1804.01149

Baladram, S. (2024, Sep 6). *Scaling numerical data explained: A visual guide with code examples for beginners*. Medium. https://medium.com/towards-data-science/scaling-numerical-data-explained-a-visual-guide-with-code-examples-for-beginners-11676cdb45cb

Boyko, N. I., & Mykhaylyshyn, V. Y. (2023). K-NN's nearest neighbors method for classifying text documents by their topics. *Radio Electronics, Computer Science, Control,* 3, 83-96. https://doi.org/10.15588/1607-3274-2023-3-9

Caparrini, A., Arroyo, J., Pérez-Molina, L., & Sánchez-Hernández, J. (2020). Automatic subgenre classification in an electronic dance music taxonomy. *Journal of New Music Research, 49*(3), 269-284. https://doi.org/10.1080/09298215.2020.1761399

D'Errico, M. (2015). Off the grid: Instrumental hip-hop and experimentation after the golden age. In J. A. Williams (Ed.), *The Cambridge companion to hip-hop* (pp. 280-291). Cambridge University Press.

Elbir, A., Çam, H. B., İyican, M. E., Öztürk, B., & Aydın, N. (2018). Music genre classification and recommendation by using machine learning techniques. In *2018 IEEE Signal Processing and Communications Applications Conference (SIU)* (pp. 1-5). IEEE. https://doi.org/10.1109/ASYU.2018.8554016

Exarchos, M. (2019). Boom bap ex machina: Hip-hop aesthetics and the Akai MPC. In *Producing music. Perspectives on music production*. Routledge.

Eyben, F. (2016). *Real-time speech and music classification by large audio feature space extraction*. Springer. https://doi.org/10.1007/978-3-319-27299-3

James, G., Witten, D., Hastie, T., & Tibshirani, R. (2013). *An introduction to statistical learning: with applications in R*. Springer.

Jijo, B. T., & Abdulazeez, A. M. (2021). Classification based on decision tree algorithm for machine learning. *Journal of Applied Science and Technology Trends, 2*(1), 20-28. https://doi.org/10.38094/jastt20165

Jurafsky, D., & Martin, J. H. (2024). *Speech and language processing*. https://web.stanford.edu/~jurafsky/slp3/old_aug24/

Karatana, A., & Yıldız, O. (2017). Music genre classification with machine learning techniques. In *2017 25th Signal Processing and Communications Applications Conference (SIU)* (pp. 1-4). IEEE. https://doi.org/10.1109/SIU.2017.7960694

Knees, P., & Schedl, M. (2016). *Music similarity and retrieval: An introduction to audio- and web-based strategies*. Springer-Verlag Berlin Heidelberg. https://doi.org/10.1007/978-3-662-49722-7

Li, Z., & Song, P. (2021). Audio similarity detection algorithm based on Siamese LSTM network. In *6th International Conference on Intelligent Computing and Signal Processing (ICSP)* (pp. 182-186). IEEE. https://doi.org/10.1109/ICSP51882.2021.9408942

Mehrotra, K., Mohan, C. K., & Ranka, S. (1997). *Elements of artificial neural nets*. MIT Press.

Mlynar, P. (2013). *In search of boom bap*. Red Bull Music Academy Daily. https://daily.redbullmusicacademy.com/2013/11/in-search-of-boom-bap

Najafabadi, M. M., Villanustre, F., Khoshgoftaar, T. M., Seliya, N., Wald, R., & Muharemagic, E. (2015). Deep learning applications and challenges in big data analytics. *Journal of Big Data, 2*(1), 1-21. https://doi.org/10.1186/s40537-014-0007-7

Reese, E. (2022). *The history of trap*. Eric Reese.

Scaringella, N., & Zoia, G. (2005). On the modeling of time information for automatic genre recognition systems in audio signals. *Proceedings of the 6th International Conference on Music Information Retrieval (ISMIR 2005)*. Queen Mary, University of London.

Segal, M. R. (2004). *Machine learning benchmarks and random forest regression*. eScholarship. https://escholarship.org/uc/item/35×3v9t4

Sönmez, Y. Ü., & Varol, A. (2019). New trends in speech emotion recognition. *2019 7th International Symposium on Digital Forensics and Security (ISDFS)* (pp. 1-7). IEEE. https://doi.org/10.1109/ISDFS.2019.8757528

Tzanetakis, G., & Cook, P. (2002). Musical genre classification of audio signals. In *IEEE Transactions on Speech and Audio Processing* (pp. 293-302). https://doi.org/10.1109/TSA.2002.800560

Quinto, R. J. M., Atienza, R. O., & Tiglao, N. M. C. (2017). Jazz music sub-genre classification using deep learning. In *TENCON 2017 - 2017 IEEE Region 10 Conference* (pp. 3111-3116). IEEE. https://doi.org/10.1109/TENCON.2017.8228396

Williams, J. A. (2010). The construction of jazz rap as high art in hip-hop music. *The Journal of Musicology, 27*(4), 435-459.

Wu, Y., & Feng, J. (2018). Development and application of artificial neural network. *Wireless Personal Communications, 102*(4), 1645-1656. https://doi.org/10.1007/s11277-017-5224-x

Yadav, H. (2022, Jul 5). *Dropout in neural networks*. Medium. https://towardsdatascience.com/dropout-in-neural-networks-47a162d621d9