ResearchGate

See discussions, stats, and author profiles for this publication at: https://www.researchgate.net/publication/357156277

MAKİNE ÖĞRENMESİNE DAYALI OTOMATİK MÜZİK TÜRÜ TANIMA MODELİ

Conference Paper · June 2021

CITATIONS
0

READS
24

1 author:



Recep Sinan Arslan
Kayseri University
22 PUBLICATIONS **37** CITATIONS

SEE PROFILE

All content following this page was uploaded by Recep Sinan Arslan on 18 December 2021.

The user has requested enhancement of the downloaded file.

MAKİNE ÖĞRENMESİNE DAYALI OTOMATİK MÜZİK TÜRÜ TANIMA MODELİ

AUTOMATIC MUSIC GENRE RECOGNITION MODEL BASED ON MACHINE LEARNING

Recep Sinan Arslan

Kayseri Üniversitesi, Bilgisayar Mühendisliği Bölümü, Kayseri, Türkiye

sinanarslanemail@gmail.com

Özet: Dijital müzik kaydı son yıllarda hızlı bir şekilde artmaktadır. Büyük veri tabanlarında düzeni sağlamak bu hızla birlikte kolay olmamaktadır. Bu kadar çok bilginin manuel olarak analiz edilmesi, sıralanması ve sınıflandırılması çoğu halde mümkün değildir. Bu sebeple yeni kayıtların işlenmesi, ses içeriğine göre analiz edilmesi ve sınıflandırılması için otomatik sistemlerin geliştirilmesi oldukça kritiktir. Bu amaçla geliştirilen otomatik müzik türü tanıma sistemleri akıllı bilgi eğlence sistemleri ve çevrimiçi müzik akış hizmetlerinde kullanılmaktadırlar. Bu sayede müzik önerileri, müzik türüne göre arama gibi bir takım kişileştirme yapıları oluşturulabilmektedir. Bu modellerin tasarımında müziklerden alınan sinyallerin ön işlem süreçlerine dahil edilmesi, özellik vektörlerinin üretimi için çeşitli algoritmalardan yararlanılması (MFCC, LPC, PCA vb.), ortaya çıkan özellik vektörlerinin optimizasyonu (kodlama, veri standardizasyonu, sentetik veri üretimi veya azaltımı) ve sınıflandırıcı model tasarımı gibi bir dizi süreç yürütülmesi gerekmektedir. Bu çalışmanın amacı ses sinyalinin işlenerek makine öğrenmesi teknikleri ile müzik türlerinin otomatik olarak sınıflandırılmasıdır. Bu amaçla oluşturulan modelde kullanılan özellik vektörü mel frekans kepstral katsayıları (MFKK)'nı içerirken, sınıflandırıcı olarak RandomForest, ExtraTree ve XGBoost algoritmaları kullanılmıştır. GTZAN veri setinde bulunan 10 farklı müzik türünün tanınması için işlevsel bir sistem geliştirilmiştir. Deneysel sonuçlara göre en başarılı performans değeri XGBoost algoritması ile yakalanmış olup doğruluk, duyarlılık, kesinlik ve F-skoru değerleri sırasıyla %90,5, %90,6, %90,4, %90,9 olmuştur. Deneylerde elde edilen umut verici sonuçlar önerilen modelin çevrimiçi bir sürümünün gelecekte geliştirilerek müzik akış platformlarında kullanılması mümkün kılacaktır.

Anahtar Kelimeler: Müzik türü, otomatik tanıma, mel frekans kepstral Katsayısı, özellik çıkarımı

Abstract: Digital music recordings have been increasing rapidly in recent years. Keep order in large databases is not easy with this speed. It is often not possible to analyze, sort and classify this much information manually. For this reason, it is very critical to develop automated systems for processing, analyzing and classifying new recordings according to sound content. Automatic music genre recognition systems developed for this purpose are used in smart infotainment systems and online music streaming services. In this way, some personalization structures such as music suggestions, searching by music genre can be created. In the design of these models, a series of processes must be carried out, such as the inclusion of the signals received from the music in the pre-processing steps, the use of various algorithms for the production of feature vectors (MFCC, LPC, PCA etc.), the optimization of the feature vectors (encoding, standardization, under-sampling or over-sampling) and classifier model design. The aim of this study is to process the audio signal and automatically classify music genres with machine learning techniques. While the feature vector used in the model created for this

purpose includes Mel-Frequency Cepstral Coefficients (MFCC), Random Forest, Extra Tree and XGBoost algorithms were used as classifier. A functional system has been developed to recognize 10 different music genres in the GTZAN dataset. According to the experimental results, the most successful performance value was achieved with the XGBoost algorithm, and the ACC, Recall, Precision and F-score values were 90.5%, 90.6%, 90.4% and 90.9%, respectively. The promising results obtained in the experiments will enable an online version of the proposed model to be developed and used in music streaming platforms in the future.

Keywords: Music genre, automatic classification, mfcc, feature extraction

1. Giriş

Son yıllarda, dijital ses tekniklerinin gelişimi olgunlaşmıştır. İnternet aracılığıyla, tüketicilere bol miktarda müzik hizmeti sunan ve telif hakkıyla korunan müzik indirmelerini sağlamayan 7digital ve Amazon mp3 gibi birçok çevrimiçi platform geliştirilmiştir. Bu platformlardan sağlanan müzik türleri çok çeşitli olabilmektedir ve bu çeşitlilik gün geçtikte hızlı bir şekilde yenileri ile birlikte artmaktadır (Jakubec, 2019). Genel olarak müzik aramalarını kolaylaştırmak için, bu sistemlerde genellikle blues, classic, country gibi birkaç etiket kullanarak müzikleri kategorize ederler. Bununla birlikte bu sistemlerde büyük miktarda müzik depolandığından, müziği manuel olarak etiketlemek zaman alıcıdır. Dahası, çoğu insan farklı müzik anlayışına sahiptir. Bu nedenle, etiketleme sürecinin daha etkin ve standart hale gelmesi için otomatik bir müzik tür sınıflandırma sistemi geliştirilmesi gerekmektedir (Hang, 2014).

Müzik türü tanıma araçları bir müzik içeriğini belirli veya daha fazla müzik türü ile ilişkilendirebilen algoritmalardır. Günümüzde müzik koleksiyonlarına ve video yayın servisleri tarafından talep edilmektedirler. Spotify, iTunes benzeri servisler müzik içeriği aramayı düzenlemek ve kolaylaştırmak için büyük ölçüde müzik türü kullanımına güvenirler (Campobello, 2020). Bunun yanında müzik türlerinin neler olduğuna dair hala kesin bir tanım yapılamamaktadır. Bugüne kadar geliştirilen birçok müzik türü sınıflandırıcısı tek bir türe özgü olarak yüksek tanıma başarısına odaklanmıştır. Deneysel sonuçlar, müzik içeriğinin belirli bir türe ait olduğuna dair insanlarla yapılan çalışmalarda başarı düzeyinin %80 ve altında olduğunu göstermektedir (Syerlehner, 2010; Lippens, 2004). Bunun yanında belirli bir şarkının birden fazla türe uygun olabilme durumu da bulunmaktadır (Correa, 2016). Bu çerçevede, müzik tür belirleme araçları, müziklerin sınıflandırılması için ortak bir anlayış geliştirmek üzere yeni modeller önerilmesi önemli bir konudur. Bu çalışmada müzik türü sınıflandırma için makine öğrenmesi temelli bir yaklaşım önerilmiştir.

Çoklu etikete sahip bir sınıflandırıcı geliştirirken karşılaşılan en büyük zorluklardan birisi önerilen modeli doğrulamak için ihtiyaç duyulan verisetine erişimin kısıtlı olmasıdır. GTZAN veriseti birbiriyle örtüşmeyen on müzik türünde sınıflandırılmış 1000 şarkıdan oluşan bir koleksiyondur. Veri kümesinde her bir şarkı kendi müzik içerine göre onu en iyi temsil eden türe göre etiketlemiştir. Daha önce de belirtildiği gibi bazı müziklerin ait olduğu baskın sınıfın belirlenmesi her zaman mümkün olamadığından yanlızca en uygun duruma göre belirlenmiş bir etiketleme yapılmıştır. Her bir müziğin tek bir etiket ile tanımlanmış olması sebebiyle tam olarak doğru kabul edilmesi sakıncalıdır. Bu veriseti ile eğitilen modellerde yanlış öğrenme, gürültü gibi durumlar sebebiyle doğru sınıflandırma oranlarında düşüşler yaşanmaktadır. Ancak GTZAN, en azından özellikle baskın bir türün varlığında ve etiketli

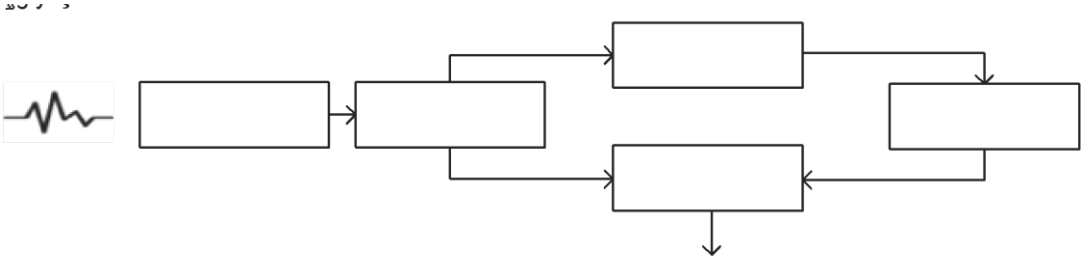
tür baskın olanla eşleştirildiğinden diğer verisetlerine göre daha iyidir (Nanni, 2016). Bu sebeple bu çalışmada bu veriseti kullanılmıştır.
Son yıllarda yayınlanan birçok çalışma, müzik bilgisine erişim senaryosunda müzik türü ile ilgili araştırmaların tükenmediğini göstermektedir. Müzik türü sınıflandırmasına yönelik olarak yapılan güncel çalışmalardan aşağıda verilmiştir:

- Müzik türü tanımada evrişimli sinir ağı ile NetVLAS'ın hibrit olarak kullanıldığı bir model önerilmiştir. Bu modelde amaç geleneksel temsili öğrenme yöntemlerinin aksine müziğe ait küresel ve soyut özelliklere odaklanılmıştır. Böylece seviyeler arası bilgilerin yakalanması ve uzun vadeli bağımlılıkların modellenmesi mümkün olmuştur. Deneysel sonuçlarda GTZAN veri seti ile yapılan testlerde, son teknoloji modellerden daha başarılı sonuçlar elde edilebilmiştir (Ng, 2020).
- Philip ve ark. tarafından yapılan çalışmada altı farklı türde etiketlenmiş toplam 1550 müzik parçasından oluşan bir veri seti ile çalışılarak KNN ve RF ile sınıflandırma yapılmıştır. Farklı müzik türlerini birbirinden ayırmayı sağlayacak akorlar, armoniler, enstrüman tınıları gibi sekiz farklı özellik ile modeller eğitilmiştir. Her bir tür için hangi özelliklerin önemli olduğu ve modele katkısını ölçmek için özellik seçimi yapılmıştır. Sonuçta akor vektörünün diğer özelliklere göre daha iyi performans gösterdiği anlaşılmıştır (Ginsel, 2020).
- Müzik önerisi ve alt tür tanımlama için veri madenciliği araçlarının kullanımı ile şarkı özellikleri ve bunlara karşılık gelen türlerin ilişkisinin incelendiği (Sergio, 2020) numaralı çalışmada iki aşamalı bir geliştirme yapılmıştır. İlk aşamada RF, NN ve LR ile sınıflandırma modelleri oluşturularak karşılaştırılmıştır. Sonrasında veri görselleştirme ve özellik vektörlerinde boyut azaltımı yapılmıştır. Sonuçta RF ile en iyi başarı elde edilmiş olsa da aşırı uyumdan kaçınma sağlanamamıştır.
- Müzik türü tanımada ses sinyali işleme kapsamında önemli bir araştırma alanıdır. Nirmal ve ark. tarafından yapılan çalışmada müzik türünü belirlemek için ses sinyalinin kısa süreli Fourier dönüşümü olan spektrogramların kullanıldığı bir yöntem önerilmiştir. Elde edilen spektrogramlar CNN ağlarının eğitiminde kullanılmıştır. Sonuçlar karışıklık matrisi ile değerlendirmiştir. Testler GTZAN verisetinden seçilen üç müzik türü için gerçekleştirilmiştir (Nirmal, 2020). Benzer şekilde müzik türünü otomatik olarak belirlemek için müzik sinyalinden MFCC yöntemi ile özelliklerin çıkarıldığı ve YSA ile eğitimin yapıldığı bir diğer çalışmada sınıflandırma performansının artırılması amaçlamıştır. Model müzikteki güncel eğilimleri yansıtan ve mevcut veri kümelerinde karşılan soruların ele alan bir veri kümesi ile eğitilmiştir (Mandal, 2020).
- Ses spektrogramlarından alınan alt seviye bilgilerin kullanıldığı ve CNN mimarisinde geliştirilmiş olan çalışmada, müzik klip türünü ayırt etmek amaçlanmıştır. Karar verme katmanında daha uygun anlamsal özelikleri üreten çok ölçekli zaman frekansı bilgileri kullanılmıştır. Deneyler GTZAN veri seti ile gerçekleştirilmiştir. Sonuçta %93,9'luk bir sınıflandırma doğruluğuna ulaşılmıştır. Önerilen ağın mobil cihazlarda ve sınırlı bellek alanlarında kullanılabilmesi mümkündür (Liu, 2021).
- Müzik türünün belirlenmesine yönelik olarak 2020 yılında yapılan bir yüksek lisans tezinde de müzik sinyallerinden elde edilen özelliklerin derin sinir ağları ve gradient boosting(GB) makinelerin kullanıldığı bir model tasarlanmıştır. Amaç genel müzik sınıflandırma taksonomisini öğrenmektir. Sonuçta GB ile en iyi sınıflandırma performansı yakalanmıştır (Bozler, 2015).

Bu çalışmada şarkıların on farklı müzik türüne ait olma derecesini değerlendirebilen müzik içeriğinin çok türlü sınıflandırması için makine öğrenmesi metotlarının kullanıldığı bir yöntem önerilmiştir. Model eğitimi öncesinde eğitim veri setinde sentetik veri üretimi yapılmış olup RandomForest, ExtraTree ve XGBoost algoritmaları kullanılmıştır. Modelleri eğitmek için GTZAN veri seti kullanılmıştır. Önerilen modelin yeniliği büyük ölçüde veriseti özelliklerinin tamamını kullanan ve düşük karmaşıklıkta sınıflandırıcı oluşturulması ve yüksek sınıflandırma performansına sahip olmasıdır. Düşük hesaplama karmaşıklığı sayesinde, model gerçek zamanlı uygulamalar için düşük maliyetli gömülü sistemlere entegre edilmesi için uygundur.
Bu çalışmanın ikinci bölümünde sistem tasarımı gösterilmiştir. Üçüncü bölümde en iyi performans sonuçlarının elde edildiği 3 model için sonuçlar farklı metriklere göre değerlendirilmiş ve önerilen model tartışılmıştır. Son bölümde ise çalışmanın genel bir değerlendirilmesi yapılarak bu çalışmanın genişletilmesi için bazı öneriler getirilmiştir.

2. Sistem Tasarımı (System Design)

Müzik türlerinde otomatik tür tanımlama, dijital sinyalin önceden tanımlanmış sınıflara göre sınıflandırılması problemini ifade eder. Bu nedenle öncelikle analog sinyalin dijital formata dönüştürülmesi gerekir. Sonrasında müzik özelliklerin çıkarılması gereklidir. Müzik türlerini en iyi şekilde tanımlayabilmek için en uygun özellik kümesinin üretilmesi beklenir. Sınıflandırma ise üretilen bu özellik vektörlerinin eğitimi ve test edilmesini gösterir. Şekil-1'de buna ilişkin önerilen sistem gösterilmiştir.



Şekil 1. Müzik Türü Tanıma Modeli Blok Diagramı (Block Diagram of Music Genre Recognition)

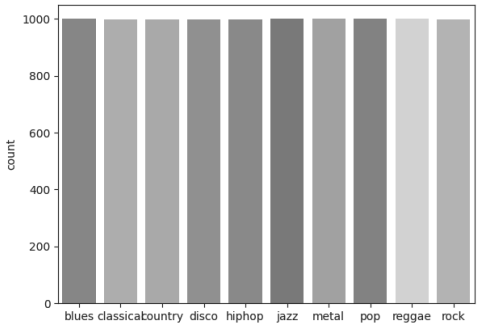
Sınıflandırma, sistemin test edilen her bir örnek için uygunluğun belirlendiği süreçtir. Sınıflandırma işlevinin girdisi, akustik özellik grubu tarafından temsil edilen vektördür. Çıktı ise test edilen örneğin tahmin edildiği sınıftır. Bu çalışmada RandomForest, ExtraTree ve XGB algoritması ile sınıflandırma deneyleri yapıldı ve sonuçlar gözlemlendi.

3. Deneyler ve Sonuçları

3.1. Veriseti (Dataset)

Müzik türlerinin sınıflandırılmasında önerilen modelin test edilebilmesi için GTZAN veri seti kullanılmıştır. Model eğitimi ve performans değerlendirmesinde kullanılmıştır. Tzenatakis ve Cook tarafından oluşturulmuş bu veriseti (Tzanetakis, 2002) MARSYAS yazılım altyapısı sayesinde açık dağıtıma sahiptir. Blues, Klasik müzik, Hip Hop, Jazz gibi 10 müzik türüne sahip ve her bir tür için 100 şarkıdan oluşan bir koleksiyondur (Şekil-2). Her bir

kaydın 30 saniyelik bir uzunluğu için 22050 Hz. örnekleme hızı vardır ve 16 bit .au formatında saklanmaktadır. Veriseti %70-%30 oranında ayrılarak sırasıyla eğitim ve test için kullanılmıştır.



Şekil 2. Müzik Türlerine Göre Örnek Sayısı Dağılımı (Distribution of Samples According to Music Genre)

3.2. Deney Ortamı ve Parametreler

Deneyler herhangi bir GPU kullanılmayan ve aşağıda özellikleri verilen PC üzerinde gerçekleştirildi.

- İşlemci: Intel Xeon @ 3.80 GHz, 12-cores
- Bellek (RAM): 32GB
- İşletim Sistemi: Ubuntu 18.04
- Altyapı: Keras 2.2.4
- Altyapı-arka plan: Engine Tensorflow

Sınıflandırma aşamasında kullanılan XGB algoritmasında hiper parametre seçimi yapılmış olupta sonuçta en iyi performansın elde edildiği parametreler seçilmiştir. Buna göre max depth=4, estimators=50, learning rate = 0.2, min child weight=3, gamma=1, objective= 'multi:softmax', scale pos weight=1 olarak belirlenmiştir.

3.3. Deney Sonuçları (Results)

Bu çalışmada GTZAN veriseti kullanılarak 10 farklı müzik türünün tanınmasına yönelik olarak bir makine öğrenmesi modeli geliştirilmiştir. Model farklı sınıflandırıcılar kullanılarak test edilmiş olup en iyi performansa sahip 3 algoritmanın sonuçları Tablo-1'de gösterilmiştir.

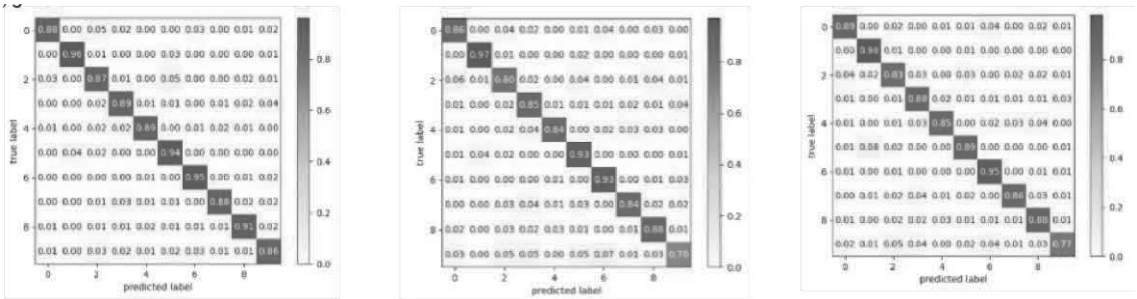
Tablo detaylı olarak incelendiğinde algoritmaların genelde başarılı sonuçlar ürettiği anlaşılmaktadır. Bu başarı sadece sınıflandırma performansında olmayıp kesinlik, duyarlılık ve f-skoru değerlerinde paralel sonuçlar elde edilmiştir. Bu durum bu çalışmada önerilen modelin otomatik müzik tür tanımında kullanılabilir bir yapı olduğunu göstermektedir. %90 ve üzerinde bir doğrulukla doğru tür tahmini yapılmıştır.

Tablo 1. GTZAN veri ile yapılan sınıflandırma test sonuçları

Classification Method	Sonuçlar (Özellik Seçimi Yapılmadan)				Sonuçlar (Özellik Seçimi Yapılarak)			
Algoritma	Doğruluk	Kesinlik	Duyarlılık	F-skoru	Doğruluk	Kesinlik	Duyarlılık	F-skoru
XGB	0,905	0,904	0,906	0,909	0,826	0,827	0,826	0,826
RandomForest	0,861	0,861	0,861	0,860	0,857	0,857	0,857	0,856
ExtraTree	0,886	0,887	0,886	0,885	0,880	0,881	0,880	0,879

Aynı testler için özellik seçimi yapılanrak tekrarlandığında ise sınıflandırma performanslarında dramatik düşüşler yaşanmaktadır. Bu durum MFCC ile elde edilen özelliklerin tümümün sınıflandırma modelinde anlamlı katkılar sunduğunu göstermektedir. F-skorunda %90 ve üzerinde bir değer elde edilmesi de modelin hem gerçek pozitif hem de gerçek negatif sınıflandırmasında başarılı olduğunu yani modelin iki yönlü başarılı sonuçlar ürettiğini göstermektedir.

Bunun yanında ortalama olarak %10 düzeyinde olan hata durumunu gözlemlemek ve hangi sınıfları ayırt etmede sorun olduğu gözlemlemek üzere üretilen karmaşıklık matrisleri her üç algoritma için hazırlanmış olup Şekil-3'te gösterilmiştir.



Şekil 3. Karmaşıklık Matrisi (Confusion Matrix)

Matrisler incelendiğinde modelin blues, country ve hiphop müziklerini ayırt etmede sorun yaşadığı anlaşılmaktadır. Diğer müzik türlerinde daha başarılı bir ayırt edicilik mümkün olmuştur. Bu duruma çözüm üretmek için her bir müziğin birde fazla sınıfa dahil edildiği verisetlerinin oluşturulması faydalı olacaktır. Böylece belirli sınıflarda (jazz, disco, classical) edilen %95 ve üzeri sınıflandırma düzeyi tüm sınıflara uygulanacak ve daha başarılı bir model ortaya çıkacaktır.

4. Sonuç

Bu çalışmanın amacı RandomForest, ExtraTree ve XGBoost yöntemleri kullanarak müzik türlerini tanıma yeteneğini değerlendirmektir. Bu yöntemler için sırasıyla %86,1, %88,6,

%90,50 sınıflandırma sonuçları elde edilmiştir. Konuşma tanımaya benzer hususlar bulundurmak ile birlikte bazı spesifik noktalarıda bulunmaktadır. İnsanın görece daha basit gerçekleştirdiği ancak manuel tür tanımlamanın mümkün olmadığı problemler için kullanılabilir bir sistem tasarlanmıştır. Sonuçta sistemin temel adımları ile birlikte model sonuçları detaylandırılmıştır.

Gelecekte, veri setinde bulunan müzik türlerinin çeşitlendirilmesi, her bir tür için daha fazla müzik içeren bir veri kümesinin toparlanması faydalı olacaktır. Ayrıca her bir müzikten elde edilecek yüksek seviye özelliklerin kullanıldığı derin öğrenme modellerinin tasarlanması da hali hazırda %90 seviyesinde olan sınıflandırma sonucunun artırılmasında katkı sunacaktır.

5. Kaynaklar

Bozler, D. (2015). Machine Learning Techniques Applied to Musical Genre Recognition (M.Sc Thesis). Colorado State University, USA.

Campobello, G, Del'Aquila, D., Russo, M. and Segreto, A. (2020). Neuro-genetic programming for multigenre classification of music content. Applied Soft Computing Journal, 94, 1-13.

Correa, D.C and Francisco, R. (2016). A survey on symbolic data-based music genre classification. Expert Systems with Applications, 60, 190-210.

Ginsel, P., Vatolkin, I. and Rudolph, G. (2020). Analysis of Structural Complexity Features for Music Genre Recognition, Glasgow, 1-8. Available at: https://ieeexplore.ieee.org/xpl/conhome/9178820/proceeding (Accessed 6 May 2021).

Hang, Y.F, Sheng-Min, L., Huan-Yu, W. and Yu-Siou, L. (2014). Music Genre classification based on local feature selection using a self-adaptive harmony search algorithm. Data & Knowledge Engineering, 92, 60-76,2014.

Jakubec, M. and Chmulik, M. (2019). Automatic music genre recognition for in-car infotainment. Transportation Research Procedia, 40, 1364-1371.

Lippens, S., Martens, P. and Mulder, D. (2004). A Comparison of human and automatic musical genre classification, Montreal, 1-4. Available at: https://ieeexplore.ieee.org/xpl/conhome/9248/proceeding (Accessed 6 May 2021).

Liu, C., Nath, L., Feng, N., Liu, G., Wang, H. and Liu, S. (2021). Bottom-up broadcast neural network for music genre classification, Multimedia tools and applications, 80, 7313-7331, 2021.

Mandal, P., Nath, I., Gupta, N. and Jha, M. K. (2020). Automatic Music Genre Detection Using Artificial Neural Networks, Intelligent Computing in Engineering,17-24, 2020.

Nanni, L., Yandre, C., Lumini, A., Moo-Young, K. and Seung-Ryul B. (2016). Combining visual and acoustic features for music genre classification. Expert Systems with Applications, 45, 108-117, 2016.

Ng, W.W.Y., Zeng, W. and Wang, T. (2020). Multi-Level Local Feature Coding Fusion for Music Genre Recognition IEEE Access, 8, 152713-152727, 2020.

Nirmal, M.R. and Mohan, S.(2020). Music Genre Classification using Spectrograms, Thrissur, 1-5. Available at: https://ieeexplore.ieee.org/xpl/conhome/9362340/proceeding (Accessed 6 May 2021).

Sergio, R., Jesus, L.G. and Ortiz, C. (2020). Data-Driven Techniques for Music Genre Recognition, Francisco, 49-63. Available at: https://sites.google.com/view/icaita2020/home (Accessed 6 May 2021).

Syerlehner, K., Widmer, G. and Knees, P. (2010). A Comparison of Human, Automatic and Collaborative Music Genre Classification and User Centric Evaluation of Genre Classification Systems, Linz, 1-15. Available at: http://amr.dke-research.de/amr2010/ (Accessed 6 May 2021).

Tzanetakis, G. and Cook, P. (2002). Marsyas: A framework for audio analysis, Organized Sound, 4(3), 1-16.

View publication stats