



# Ses analizi ile müzik türlerinin sınıflandırılmasına yönelik kapsamlı bir çalışma

## A comprehensive study of classification of music genres with voice analysis

Zekeriya Anıl Güven<sup>1,\*</sup> 

<sup>1</sup> İzmir Bakırçay Üniversitesi, Bilgisayar Mühendisliği Bölümü, 35665, İzmir Türkiye

### Öz

Ses analizi; hastalık tespiti, duygu analizi, müzik türleri sınıflandırma gibi birçok alanda uygulanabilen bir analiz türüdür. Ses karakteristik özellikleri ile probleme göre görüntü analizi, veri analizi uygulanarak sınıflandırma yapılabilmektedir. Bu çalışmada, ses karakteristik özellikleri çıkarılarak müzik türlerinin sınıflandırması gerçekleştirilmiştir. Veri seti olarak 10 müzik türü etiketinden oluşan GTZAN kullanılmıştır. Analiz aşamasında, ses bölütleme işlemi ve özellik seçme yöntemlerinin de müzik türünü sınıflandırmaya etkisi araştırılmıştır. Sınıflandırma için makine öğrenme yöntemleri ve derin sinir ağlarından yararlanılmıştır. Analizin ilk aşamasında, sadece ses bölütleme işlemi uygulanmasıyla doğruluk değeri %10.58 artış göstermiştir. Çalışmanın sonucunda, ses bölütleme işlemi ve özellik seçme yöntemleri uygulanması sonrasında öğrenme yöntemlerinden Derin Sinir Ağları yöntemi %95.19 doğruluk değeriyle %14.19 başarı artışı sağlamıştır.

**Anahtar kelimeler:** Ses analizi, Makine öğrenmesi, Derin sinir ağları, Özellik seçme, Ses bölütleme

### 1 Giriş

Günümüzde ses analizi ile alakalı sağlık, müzik, konuşma (aksan), güvenlik gibi birçok alanda çalışmalar yapılmaktadır. Ses (konuşma) analizi, ses dosyaları üzerinden ses karakteristik özellikleri çıkarılarak yapılan analiz türüdür. Ses analizi ile çıkarılan bu özellikler eğitime veri (mfcc, tempo, ...) olarak verilebilmektedir veya sesin görüntüye (spectrogram, chromogram, ...) dönüştürüldüğü özellikler olarak görüntü işleme yoluyla sistemde eğitilebilmektedir. Her iki özellik türü düşünüldüğünde, ses analizinin çalışma alanı oldukça geniştir.

Ses analizi aracılığıyla ses dosyaları sayısal verilere dönüştürülmektedir. Bu sayısal veriler ile özellik çıkarımı işlemi uygulanmaktadır. Elde edilen bu özellikler, sesin karakteristik özelliklerini oluşturmaktadır. Ses karakteristik özelliklerinin eğitim aşamasında kullanılması ile oluşturulan sistemin başarısı test verileri ile ölçülmektedir.

Ses analizi kullanılması müzik türlerinin tespiti için oldukça yaygındır. Birçok çevrimiçi müzik platformu, otomatik olarak müzik türlerinin belirlenmesini sağlayan sistemlerin faydalı olacağını düşünmektedir. Böylece yüklenen müzikler, daha rahat bir şekilde otomatik olarak sınıflandırılacaktır. Ancak direk ses özellikleri ile bunu yapabilmek zordur, spektrogram gibi görüntüler ile

### Abstract

Voice analysis is a type of analysis that can be applied in many areas such as disease detection, emotion analysis, and classification of music genres. Classification can be made by applying image or data analysis according to the problem with voice characteristics. In this study, the classification of music genres was performed by extracting the voice characteristics. GTZAN, which consists of 10 music genre tags, was used as the dataset. In the analysis phase, the effects of audio segmentation and feature selection methods on the classification of music genres were investigated. Machine learning methods and deep neural networks were used for classification. In the first stage of the analysis, the accuracy value increased by 10.58% by applying only the audio segmentation. As a result of this study, after the application of audio segmentation and feature selection methods, Deep Neural Network provided an increase of 14.19% with an accuracy of 95.19%.

**Keywords:** Voice analysis, Machine learning, Deep neural networks, Feature selection, Audio segmentation

kullanılabilen özelliklerde ses karakteristik özellikleri ile sisteme verildiğinde belli başarı sağlamaktadır.

Ses analizinde ses dosyasının süresinin uzun olabildesinden dolayı ses dosyasının ayrıştırılarak ses bölütleme işlemi ile analiz gerçekleştirilebilmektedir. Böylece hem veri sayısında artış hem de daha doğru analiz sağlanabilmektedir. Ayrıca elde edilen birçok özellik ile beraber analiz aşamasında fayda sağlamayan özellikler bulunabilmektedir. Bundan dolayı özellik seçme yöntemleri kullanılarak özellik sayısı azaltılmaktadır. Geleneksel yöntemlerin kullanımı yaygın olsa da, son zamanlarda bu görev için literatürde oldukça fazla yöntem önerilmiştir [1-2].

Literatür incelendiğinde müzik alanı ile alakalı birçok çalışma bulunmaktadır. Bu çalışmalar genel olarak görüntü işlemeyle ses analizinin kullanılmasına dayanmaktadır. Pelchat ve Gelowitz [3], şarkıları kendi müzik türlerine göre sınıflandırmak için yapay sinir ağlarına girdi olarak şarkıların zaman dilimlerinden oluşturulan spektrogramların görüntülerini kullanmışlardır. Elbir ve Aydın [4], melspectogramların girdi olarak verildiği yeni bir derin sinir ağı modeli ile elde edilen akustik özelliklerin çıkarılmasına odaklanan bir müzik türü sınıflandırma sistemi ve müzik öneri motoru önermişlerdir. Ghildiyal vd. [5], ses

\* Sorumlu yazar / Corresponding author, e-posta / e-mail: anilguven1055@gmail.com (Z.A. Güven)  
Geliş / Received: 17.08.2023 Kabul / Accepted: 11.12.2023 Yayınlanma / Published: 15.01.2024  
doi: 10.28948/ngumuh.1344605

melspektrogramı ile birlikte çeşitli modeller için çoklu girdileri CNN ve makine öğrenme yöntemleri için kullanarak GTZAN veri seti üzerinde müzik türlerini sınıflandırmışlardır. Senac vd. [6], müzik türü tespiti için CNN girdileri olarak, dinamikler, tını ve tonalite üç ana müzik boyutundan seçilen sekiz müzik özelliğinden oluşan küçük bir dizi kullanımını önermişlerdir ve GTZAN veri setinin başarısını ölçmüşlerdir. Bahuleyan [7], CNN modelinin yalnızca spektrogramını kullanarak bir ses sinyalinin tür etiketini tahmin etmek için uçtan uca eğitildiği bir derin öğrenme yaklaşımı model ile hem zaman alanından hem de frekans alanından el yapımı özellikleri kullanan modeli makine öğrenme yöntemleri ile eğiterek performanslarını karşılaştırmışlardır. Cheng vd. [8], bir müzik türü sınıflandırma modeli uygulamak için ön işlemede, orijinal ses dosyalarını karşılık gelen Mel spektrumlarına dönüştürmüşler ve eğitim için bu spektrumları önerilen CNN modeline vermişlerdir. Nanni vd. [9], hem görsel hem de akustik olmak üzere farklı özellik gruplarının birleşimine dayanan otomatik ses sınıflandırması yaklaşımı önermişler ve Latin Müzik, ISMIR 2004 ve GTZAN veri seti üzerinde analiz gerçekleştirmişlerdir. Barros vd. [10], karmaşık ses tanımanın genelleştirme yönlerini geliştirmek için kanallar arası bir mimari kullanmayı önermişler ve analiz için SAVEE, GTZAN ve EmotiW veri setlerini kullanmışlardır. Prabhakar ve Lee [11], müzik türü sınıflandırması için beş yeni metodoloji önermişlerdir. Bu metodolojileri, derin öğrenme, seyrek temsil ve transfer öğrenme çerçeveleri, geleneksel makine öğrenimi teknikleriyle karşılaştırıldığında oldukça başarılı ve çok yönlü olmasına dayandırmışlardır. Panagakis ve Kotropoulos [12], yeni müzik sınıflandırma çerçevesi sunmuşlardır. Bu çerçevede, müzik temsili için kortikal temsillere başvururken, tür sınıflandırması için seyrek temsile dayalı sınıflandırma kullanılmışlardır. Ses analizi müzik haricinde duygu analizi ([13]), aktivite ([14]), hastalık ([15-17]), aksan ([18]) tespiti ve robotik ([19]) gibi birçok alanda da uygulanmaktadır.

Çalışmada spektrogram gibi görüntüden yararlanan özellikler kullanılmadan, ses karakteristik özellikleri ile

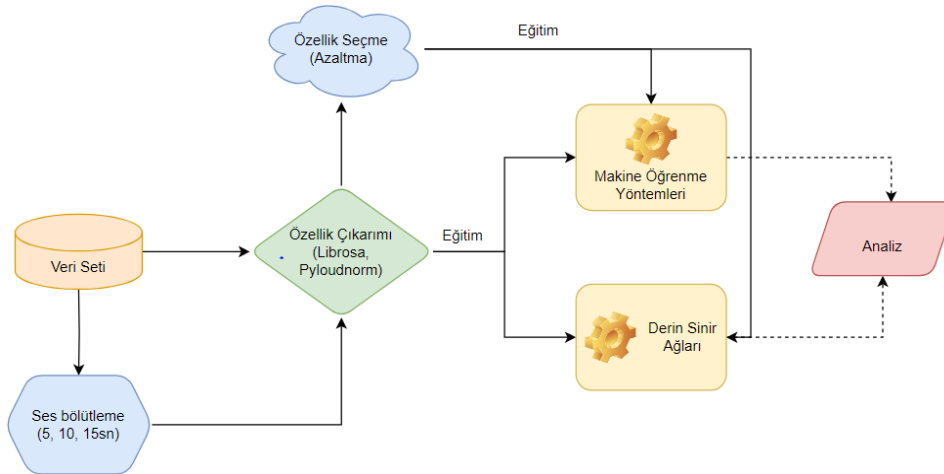
müzik türlerinin tespit edilmesi amaçlanmaktadır. GTZAN veri setine ait müzik seslerinden çıkarılan ses karakteristik özellikleri ile makine öğrenme yöntemleri ve derin sinir ağı (deep neural network - DNN) modeli eğitilerek müzik türünü sınıflandırmanın başarısı ölçülmektedir. Analiz aşamasında, 3, 5, 10, 15, 30sn'lik ses bölütleme işlemi ve çıkarılan özellikler üzerinde özellik seçme yöntemleri uygulamanın etkisi de araştırılmaktadır. Çalışmanın katkıları şunlardır:

- Sadece ses karakteristik özellikleri ile analiz gerçekleştirilmiştir.
- Ses bölütlemenin analize olumlu etkisi gösterilmiştir.
- Özellik seçme yöntemleri kıyaslanarak ses bölütleme ile etkisi araştırılmıştır.

Çalışmanın akışı düşünüldüğünde, veri seti ve makine öğrenme yöntemleri gibi kullanılan tüm materyal ve metotlar ikinci başlık altında anlatılmaktadır. “Bulgular” başlığı altında ise uygulanan tüm analizler tablo ve şekiller ile beraber açıklanmaktadır. Çalışmanın sonuçlarının değerlendirilip tartışılması ise dördüncü başlık altında gerçekleştirilmektedir. “Sonuç” başlığı altında ise çalışmanın özetiyle beraber sonucu aktarılmaktadır.

## 2 Materyal ve metot

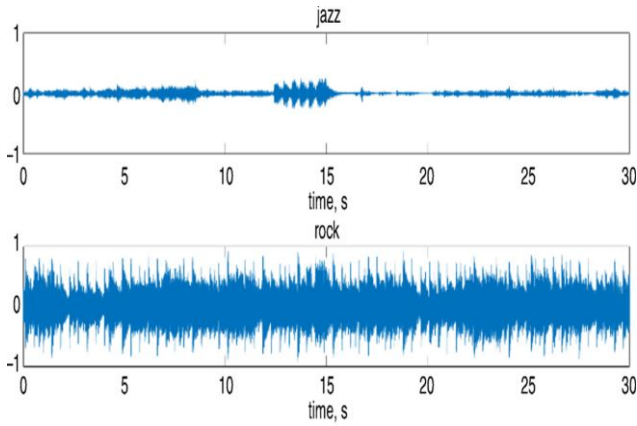
Veri seti, özellik seçme teknikleri, makine öğrenme yöntemleri ve DNN algoritması bu başlık altında anlatılmaktadır. İlk aşamada veri setindeki ses dosyaları Librosa kütüphanesi aracılığıyla sayısal verilere dönüştürülerek ses karakteristik özellikleri çıkarılmaktadır. Ses dosyalarının normal özellik çıkarımına ek olarak hem ses bölütleme (3, 5, 10, 15, 30sn), hem de özellik seçme yöntemi uygulanarak bu özellikler ile yeni eğitim dosyaları elde edilmiştir. Tüm bu varyasyonlar ile Naive Bayes (NB), k-en Yakın Komşu (K-NN), Rastgele Orman (RO), Destek Vektör Makineleri (DVM), Lojistik Regresyon (LR), Ekstrem Gradient Artırma (XGB) makine öğrenme ve DNN yöntemleri eğitilerek sistemlerin başarısı analiz edilmiştir. Sistemin işleyişine ait metodoloji Şekil 1’de gösterilmiştir.



Şekil 1. Sistemin metodolojisi

## 2.1 Veri seti

Müzik türlerini sınıflandırma açısından değerlendirmek için GTZAN veri seti [20] kullanılmıştır. Veri seti; toplam 10 sınıf tan, her biri için 100 tane olmak üzere 1000 müzik içermektedir. Her biri 30sn olan müziklerin örnekleme frekansı 22.500 Hz, çözünürlüğü ise 16 bittir. GTZAN'daki müzik türleri blues, klasik, country, disko, hip-hop, jazz, metal, pop, reggae ve rock'tır [20]. GTZAN veri setine ait müzik türlerinin zaman domaininde gösterimine ait bir örnek Şekil 2'de gösterilmiştir. Şekil incelendiğinde jazz ve rock müzik türünün görüntüsünün farklı olduğu gözükmektedir.



Şekil 2. Zaman domaininde örnek veri gösterimi [5]

## 2.2 Özellik çıkarım kütüphaneleri

Ses karakteristik özelliklerinin çıkarılması için librosa<sup>1</sup> ve pyloudnorm<sup>2</sup> kütüphanesi kullanılmıştır. Librosa, müzik ve ses analizi için bir python paketidir. Müzik bilgisi alma sistemleri oluşturmak için gerekli yapı taşlarını, özelliklerini sağlamaktadır [21]. Pyloudnorm ise ses yüksekliğinin ölçülmesini sağlayan bir Python kütüphanesidir [22]. Bu kütüphaneler aracılığıyla çıkarılan özellikler Tablo 1'de verilmiştir. Tablo incelendiğinde frekans domaininde olan spectral özelliklerin maksimum, minimum, ortalama, medyan, standart sapma, uzunluk (range), basıklık (kurtosis), çarpıklık (skew) değerleri de özellik olarak çıkarılmaktadır. Mel-frequency cepstral coefficients (MFCC) özelliğinin ise normal ve birinci türevinin değerleri alınarak yine minimum, maksimum, vs. değerleri elde edilmektedir. Buradaki varyasyonların hepsi göz önüne alındığında toplam çıkarılan özellik sayısı her bir ses kaydı için 242'dir. Tüm özellikler üzerinde eğitim ve test işlemi makine öğrenme yöntemleri ve derin sinir ağları ile gerçekleştirildikten sonra, özellik seçme yöntemleri uygulanarak özellik sayısı azaltılmaktadır. Seçilerek elde edilen yeni özellikler ile makine öğrenme ve derin sinir ağları yöntemlerinde eğitim ve test işlemi gerçekleştirilmektedir.

Tablo 1. Kütüphaneler aracılığı ile çıkarılan özellikler

| Kütüphane | Çıkarılan Özellik          |
|-----------|----------------------------|
| Librosa   | Spectral centroid          |
|           | Spectral flux              |
|           | Spectral bandwidth         |
|           | Spectral flatness          |
|           | Spectral rolloff           |
|           | Zero crossing rate         |
|           | Root mean squared error    |
|           | Chromagram                 |
|           | Harmonic                   |
|           | Tempo                      |
|           | Mel-frequency coefficients |
|           | cepstral                   |
|           | Pyloudnorm                 |

## 2.3 Özellik Seçme Yöntemleri

Çıkarılan özellikler arasında en uygunları seçebilmek ve özellik sayısını azaltabilmek için özellik seçme yöntemleri kullanılmaktadır. 242 olan özellik sayısı yöntemlere göre 50-100 arasında özelliğe düşürülmektedir. Çalışmada kullanılan özellik seçme yöntemleri şunlardır:

- Geri Yönlü Eleme (Backward Feature Elimination - BFE): Prosedür, tüm özelliklerle başlamaktadır. Her adımda, kümede kalan en kötü özellik kaldırılmaktadır [23].
- Düşük Varyans Filtresi (Low Variance Filter - LVF): Her bir sütuna ait varyans değeri hesaplanarak varyans değeri belirli bir eşikten düşük olan özellikleri kaldırmaktadır. Sadece sayısal sütunlar için bu yöntem kullanılmaktadır [24].
- Ki-kare (Chi Square - CS): Sınıflandırma modellerinin performansını artıran özellikleri seçmek için "Ki-kare" istatistiklerini kullanmaktadır. Yöntem özelliklerin hedef sınıfa bağımlılığı veya bağımsızlığı kavramına dayanmaktadır [25].
- Özyinelemeli Özellik Eleme (Recursive Feature Elimination - RFE): RFE, sarmalayıcı tipi bir özellik seçim algoritmasıdır. Bu, yöntemin özünde farklı bir makine öğrenimi algoritmasının verildiği (SVM, RF, vs.) ve kullanıldığı, RFE tarafından sarıldığı ve özelliklerin seçilmesine yardımcı olmak için kullanıldığı anlamına gelmektedir [26].

## 2.4 Makine öğrenme yöntemleri

Özellik çıkarım, ses bölütleme ve özellik seçme işlemleri uygulanması ile eğitim ve test aşamasında makine öğrenme yöntemleri kullanılmaktadır. Kullanılan makine öğrenme yöntemleri:

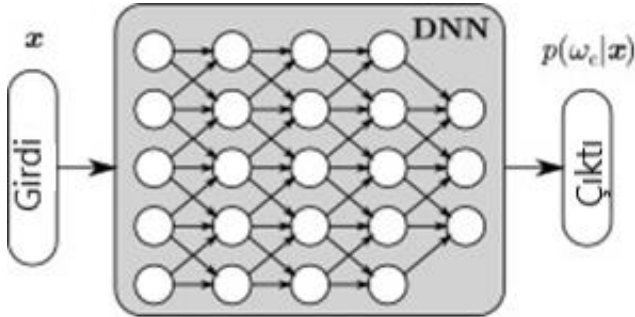
<sup>1</sup> <https://librosa.org/doc/latest/index.html>

<sup>2</sup> <https://github.com/csteinmetz1/pyloudnorm>

- Naïve Bayes: Olasılık prensiplerine göre hesaplama yapılarak, verilerin sınıfını olasılıksal tespit etmeyi amaçlamaktadır [27].
- Rastgele Orman: Veri setinden birden fazla eğitim veri seti elde ederek her bir önyüklemeli eğitim veri seti için bir ağaç üretmektedir. Test aşamasında ağaçların çoğunluk oyu ile sınıflandırma yapılmaktadır [28].
- k-En Yakın Komşu: Belirlenen k sayıda yakın komşunun mesafesini inceleyerek test edilecek bir verinin hangi sınıfa atanacağını belirlemektedir [29].
- Lojistik Regresyon: Sınıflandırılmış bağımlı bir değişken ile birden çok bağımsız değişken arasındaki ilişkiyi, en az değişkeni kullanarak en iyi uyum ile tanımlamaya çalışan bir algoritmadır [30].
- Extreme Gradient Boosting: Karar ağacı altyapısını kullanarak eğitim boosting yöntemini hesaplamalara dahil eden algoritmadır [31].
- Destek Vektör Makineleri: İki veya fazla sınıflı verileri birbirinden ayırabilecek en uygun olan n-boyutlu hiper düzlemi bulma esasına dayanmaktadır [32].

### 2.5 Derin sinir ağları

Bir DNN, nöronların önceki katmandan nöron aktivasyonlarını girdi olarak aldığı ve basit bir hesaplama gerçekleştirdiği (örneğin, girdinin ağırlıklı bir toplamı ve ardından doğrusal olmayan bir aktivasyon) birden çok katman dizisinde düzenlenmiş bir nöron koleksiyonudur. Ağın nöronları, girdiden çıktıya karmaşık doğrusal olmayan



Şekil 3. DNN mimarisine bir örnek [33]

bir eşlemeyi (mapping) ortaklaşa uygulamaktadır. Bu eşleme, hata geri yayılımı yöntemi kullanılarak her bir nöronun ağırlıklarının uyarlanmasıyla verilerden öğrenilmektedir [33]. Örnek bir DNN yapısı Şekil 3'te gösterilmiştir.

### 3 Bulgular ve tartışma

Librosa ve pyloudnorm kütüphaneleri ile çıkarılan tüm özellikler üzerinde ilk olarak normalizasyon işlemi uygulanmıştır. Bunun nedeni verilerin belli bir standarta getirilmesidir. Normalizasyon tekniği olarak standartscaler (standart ölçeklendirici) kullanılmıştır. Bu yöntem ile veri setinin her bir öznelik değerinin ortalaması sıfır ve standart sapması bir olacak şekilde yeniden ölçeklendirilmektedir.

Her bir verisi 30sn olan normal veri setine ait normalize edilen tüm özellikler makine öğrenme ve derin sinir ağları

yöntemlerine verilmiştir. Eğitim esnasında ayrıca 5-katmanlı çapraz doğrulama yönteminden de yararlanılmıştır. Tüm aşamalarda veri setinin %80'i eğitim, %20'si ise test aşaması için kullanılmıştır. 30sn'lik ses kayıtlarından oluşan 10 sınıflı veri setine ait yöntemlerin başarısı Tablo 2'de verilmiştir. DNN için epoch sayısı 50 belirlenerek öğrenme oranına göre erken durdurma uygulanmıştır. Tablodaki sonuçlara göre, en yüksek doğruluk ve F1-skor değerini normal analiz sonucunda XGB yöntemi vermiştir.

Tablo 2. 30sn'lik veri setinin yöntemlerdeki başarısı (%)

| Yöntem | Normal Analiz |              | 5-Çapraz Doğrulama |          |
|--------|---------------|--------------|--------------------|----------|
|        | F1-skor       | Doğruluk     | F1-skor            | Doğruluk |
| NB     | 67.48         | 67.50        | 64.61              | 65.16    |
| K-NN   | 42.27         | 43.50        | 72.01              | 72.07    |
| RO     | 79.95         | 80.00        | 74.21              | 74.67    |
| DVM    | 36.56         | 39.50        | 76.80              | 77.17    |
| LR     | 50.11         | 51.00        | 78.40              | 78.67    |
| XGB    | <b>80.87</b>  | <b>81.00</b> | 78.02              | 78.37    |
| DNN    | 76.86         | 76.99        | 80.50              | 80.25    |

Sonraki aşamada literatürde sık kullanılan ses bölütleme işlemi 30sn'lik veri setine uygulanmıştır. 3'er, 5'er, 10'ar ve 15'er sn olmak üzere 30sn'lik ses kaydı parçalara ayrılarak daha fazla ses kaydı elde edilmiştir (3'er sn'lik yaklaşık 10 kat, 5'er sn'lik yaklaşık 6 kat, 10'ar sn'lik yaklaşık 3'er kat, 15'er sn'lik yaklaşık 2 kat fazla eğitim-test verisi). Ses bölütleme işlemi sonrası ses kaydı ve özellik sayısı Tablo 3'te verilmiştir. Tablo incelendiğinde bölütleme yapılacak saniye sayısı azaldıkça, ses kaydı ve özellik sayısının arttığı gözükmektedir.

Tablo 3. Ses bölütleme sonucu ses kaydı ve özellik sayısı

| Ses bölütleme (sn) | Toplam ses kaydı | Özellik sayısı |
|--------------------|------------------|----------------|
| 3                  | 9980             | 9980 x 242     |
| 5                  | 5984             | 5984 x 242     |
| 10                 | 2987             | 2987 x 242     |
| 15                 | 1988             | 1988 x 242     |
| 30                 | 999              | 999 x 242      |

Ses bölütleme işlemi sonrası tüm çıkarılan özelliklere ait dosyalar, makine öğrenme ve derin sinir ağlarına eğitim ve test aşaması için verilmiştir. Ayrıca 5-katmanlı çapraz doğrulamanın da analize etkisi araştırılmıştır. Ses bölütleme işlemine göre tüm yöntemlerin başarısı Tablo 4'te gösterilmektedir. Tablodaki sonuçlar gösteriyor ki en yüksek F1-skor ve doğruluk değeri 3sn'lik ses kaydı içeren veriler ile XGB yönteminde elde edilmiş ve Tablo 3'e göre doğruluk değeri artışı sağlanmıştır.



**Tablo 4.** Ses bölütleme uygulanan veri setinin yöntemlerdeki başarısı (%)

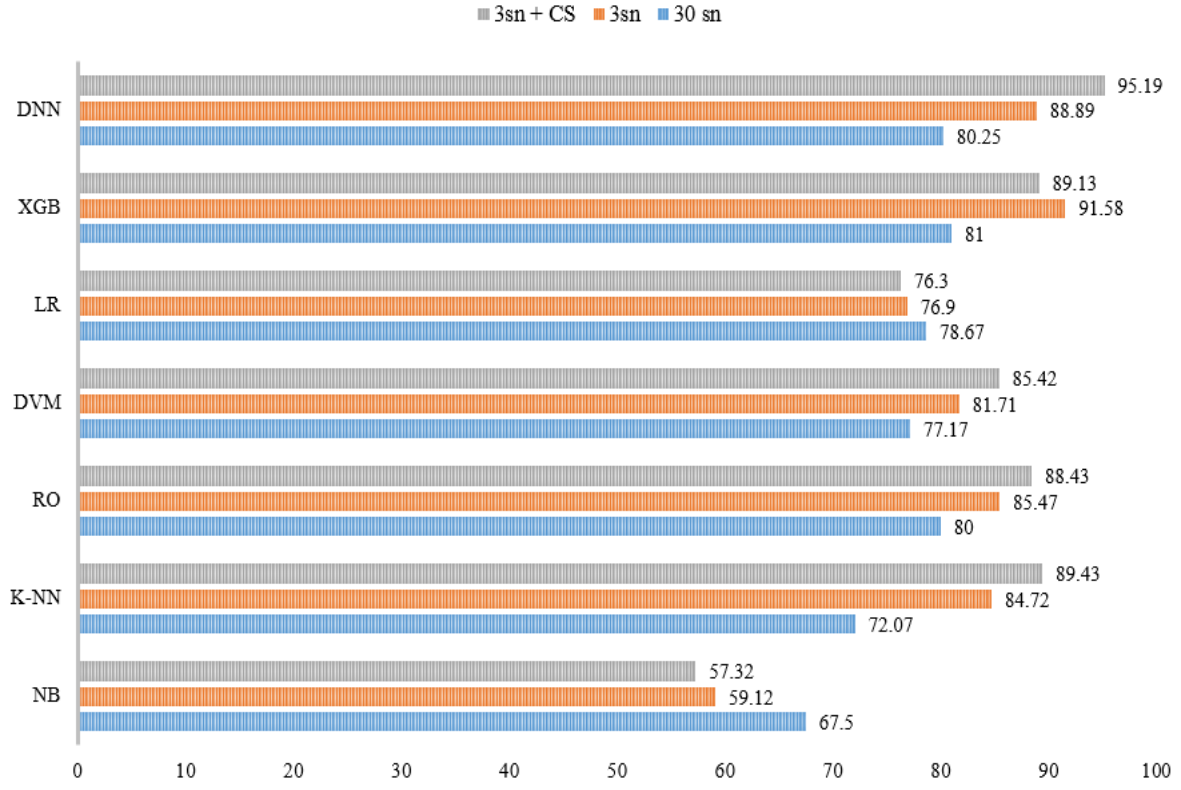
|      |                 | 5-Çapraz Doğrulama = Yok |       |       |       |       |              |       |
|------|-----------------|--------------------------|-------|-------|-------|-------|--------------|-------|
|      |                 | NB                       | K-NN  | RO    | DVM   | LR    | XGB          | DNN   |
| 3sn  | <i>F1-skor</i>  | 58.54                    | 84.77 | 85.40 | 81.60 | 76.81 | <b>91.58</b> | 88.98 |
|      | <i>Doğruluk</i> | 59.12                    | 84.72 | 85.47 | 81.71 | 76.90 | <b>91.58</b> | 88.89 |
| 5sn  | <i>F1-skor</i>  | 60.08                    | 49.29 | 84.87 | 38.13 | 45.48 | 89.68        | 88.91 |
|      | <i>Doğruluk</i> | 60.65                    | 49.88 | 84.96 | 41.02 | 46.95 | 89.72        | 88.47 |
| 10sn | <i>F1-skor</i>  | 64.10                    | 47.47 | 82.39 | 41.56 | 49.52 | 85.88        | 84.59 |
|      | <i>Doğruluk</i> | 64.72                    | 49.16 | 82.61 | 44.31 | 50.84 | 85.95        | 84.28 |
| 15sn | <i>F1-skor</i>  | 61.33                    | 44.26 | 77.75 | 37.42 | 45.24 | 84.36        | 83.64 |
|      | <i>Doğruluk</i> | 61.81                    | 45.48 | 77.89 | 40.45 | 46.23 | 84.42        | 83.67 |
|      |                 | 5-Çapraz Doğrulama = Var |       |       |       |       |              |       |
| 3sn  | <i>F1-skor</i>  | 55.81                    | 69.91 | 70.83 | 74.33 | 71.73 | 78.14        | 89.49 |
|      | <i>Doğruluk</i> | 56.68                    | 69.99 | 71.34 | 74.53 | 71.95 | 78.26        | 89.44 |
| 5sn  | <i>F1-skor</i>  | 58.90                    | 71.30 | 72.19 | 76.23 | 74.85 | 77.22        | 88.54 |
|      | <i>Doğruluk</i> | 59.42                    | 71.02 | 72.64 | 76.27 | 75.00 | 77.29        | 88.82 |
| 10sn | <i>F1-skor</i>  | 61.31                    | 73.18 | 73.87 | 76.93 | 76.72 | 77.97        | 83.68 |
|      | <i>Doğruluk</i> | 61.57                    | 72.85 | 74.39 | 76.17 | 76.80 | 78.07        | 83.72 |
| 15sn | <i>F1-skor</i>  | 62.87                    | 73.62 | 74.89 | 77.98 | 78.39 | 78.11        | 80.60 |
|      | <i>Doğruluk</i> | 63.33                    | 73.29 | 75.30 | 78.07 | 78.42 | 78.22        | 80.67 |

Ses bölütleme yapıma işleminin ardından 3, 5, 10, 15, 30sn'lik oluşturulan tüm dosyalar üzerinde BFE, LVF, CS, RFE özellik seçme yöntemleri kullanılarak çıkarılan özellikler azaltılmıştır. Azaltılan özellikler üzerinde tüm sistem yeniden eğitilerek test edilmiştir. Özellik azaltılmanın makine öğrenme ve derin sinir ağları yöntemleri üzerindeki etkisi [Tablo 5](#)'te gösterilmiştir. Tablo incelendiğinde özellik seçme uygulanması ve ses bölütleme işlemi ile daha başarılı sınıflandırma yapılabildiği gösterilmiştir. Her sn veri seti için normal veri setine göre başarı artışı sağlanmıştır. Sırasıyla en yüksek F1-skor ve doğruluk değerleri, 3sn'lik ses kayıtları ile CS özellik seçme yöntemleri kullanılarak %95.37, %95.19 ile elde edilmiştir. Çalışmanın her aşaması için en iyi F1-skor ve doğruluk değerleri kullanılması sonucunda analizlerin etkileri [Şekil 4](#)'te gösterilmiştir. Ses bölütleme işlemi ve özellik seçme yöntemleri uygulanmasıyla başarı oranında artış sağlanmıştır.

En son aşamada 10 müzik türü için en başarılı yöntemlerin sınıflandırma raporları çıkarılmıştır. Bunun sonucunda hangi müzik türlerinin her aşamada daha doğru tespit edildiği analiz edilmiştir. İlk olarak 30sn için en başarılı olan XGB yöntemine ait sınıflandırma raporu [Tablo 6](#)'da gösterilmiştir. Sınıflandırma raporuna göre rock etiketine sahip müzik türleri en başarısız olarak sınıflandırılırken, klasik etiketli müzik türü en başarılı olarak sınıflandırılmıştır. Bir diğer analiz için ise ses bölütleme ve özellik seçme işlemleri sonrası en başarılı olan 3sn'lik ses kayıtları için CS özellik seçme yöntemi seçilmiştir. Bu yöntemle ait sınıflandırma raporu [Tablo 7](#)'de verilmiştir. Tablo incelendiğinde klasik ve hiphop müzik türü en başarılı sınıflandırılırken, rock müzik türü en başarısız sınıflandırılmıştır. Bunun nedeni olarak rock müzik türünün reggae, hip-hop, metal gibi temposu benzer müzik türleriyle karışması belirtilebilir.

**Tablo 5.** Ses bölütleme ve özellik seçme yöntemlerine göre yöntemlerin başarısı (%) (sırasıyla F1-skor, doğruluk)

| Sn   | Yöntem | NB              | K-NN            | RO              | DVM             | LR              | XGB             | DNN                 |
|------|--------|-----------------|-----------------|-----------------|-----------------|-----------------|-----------------|---------------------|
| 3sn  | BFE    | 47.78,<br>48.55 | 67.57,<br>67.79 | 70.35,<br>70.69 | 63.04,<br>63.33 | 54.28,<br>55.11 | 70.92,<br>70.89 | 94.72,94.64         |
|      | CS     | 55.53,<br>57.32 | 89.38,<br>89.43 | 88.39,<br>88.43 | 85.36,<br>85.42 | 76.21,<br>76.30 | 89.17,<br>89.13 | <b>95.37, 95.19</b> |
|      | LVF    | 41.01,<br>43.74 | 67.78,<br>67.89 | 71.29,<br>71.54 | 61.26,<br>61.92 | 51.03,<br>52.86 | 70.91,<br>70.94 | 95.16, 95.14        |
|      | RFE    | 55.53,<br>57.32 | 89.38,<br>89.43 | 88.33,<br>88.38 | 85.36,<br>85.42 | 76.21,<br>76.30 | 89.17,<br>89.13 | 94.97, 94.40        |
| 5sn  | BFE    | 54.25,<br>55.56 | 76.57,<br>76.69 | 75.85,<br>76.19 | 73.43,<br>73.52 | 62.44,<br>63.16 | 78.50,<br>78.53 | 81.17, 80.87        |
|      | CS     | 59.69,<br>60.90 | 86.48,<br>86.55 | 87.66,<br>87.72 | 87.10,<br>87.14 | 78.78,<br>78.86 | 89.05,<br>89.06 | 94.02, 93.82        |
|      | LVF    | 46.87,<br>49.12 | 73.70,<br>73.94 | 78.59,<br>78.53 | 68.78,<br>68.92 | 58.27,<br>59.15 | 80.43,<br>80.45 | 83.81, 83.54        |
|      | RFE    | 59.69,<br>60.90 | 86.48,<br>86.55 | 87.58,<br>87.64 | 87.10,<br>87.14 | 78.78,<br>78.86 | 89.05,<br>89.06 | 93.56, 93.57        |
| 10sn | BFE    | 55.43,<br>56.19 | 69.74,<br>69.90 | 70.83,<br>71.41 | 69.35,<br>69.73 | 61.12,<br>62.21 | 73.32,<br>73.58 | 73.96, 73.71        |
|      | CS     | 63.21,<br>64.38 | 80.43,<br>80.60 | 82.64,<br>82.78 | 84.70,<br>84.78 | 78.38,<br>78.60 | 84.66,<br>84.78 | 90.57, 90.47        |
|      | LVF    | 52.49,<br>54.85 | 70.25,<br>70.40 | 78.79,<br>78.93 | 75.56,<br>75.75 | 65.86,<br>66.39 | 81.02,<br>81.10 | 84.39, 84.28        |
|      | RFE    | 63.21,<br>64.38 | 80.43,<br>80.60 | 85.35,<br>85.45 | 84.70,<br>84.78 | 78.38,<br>78.60 | 84.66,<br>84.78 | 89.99, 89.80        |
| 15sn | BFE    | 50.30,<br>51.51 | 60.49,<br>60.80 | 67.04,<br>67.59 | 62.99,<br>63.32 | 57.61,<br>58.79 | 68.46,<br>68.34 | 69.06, 68.59        |
|      | CS     | 67.11,<br>67.34 | 78.27,<br>78.14 | 81.80,<br>81.91 | 83.28,<br>83.17 | 78.94,<br>79.15 | 84.82,<br>84.93 | 87.92, 87.44        |
|      | LVF    | 50.97,<br>53.02 | 67.34,<br>67.09 | 79.14,<br>79.15 | 73.86,<br>73.87 | 71.12,<br>71.36 | 81.25,<br>81.16 | 82.69, 82.16        |
|      | RFE    | 67.11,<br>67.34 | 78.27,<br>78.14 | 82.35,<br>82.41 | 83.28,<br>83.17 | 78.94,<br>79.15 | 84.82,<br>84.93 | 86.40, 85.93        |
| 30sn | BFE    | 62.95,<br>63.00 | 72.67,<br>72.00 | 76.57,<br>77.00 | 78.67,<br>79.00 | 75.11,<br>75.50 | 74.33,<br>74.50 | 81.35, 79.00        |
|      | CS     | 70.15,<br>70.00 | 75.34,<br>76.00 | 78.49,<br>78.50 | 86.91,<br>87.00 | 82.32,<br>82.50 | 81.06,<br>81.50 | 82.59, 82.00        |
|      | LVF    | 64.10,<br>64.50 | 66.91,<br>67.50 | 75.62,<br>76.00 | 78.46,<br>78.50 | 77.50,<br>77.50 | 81.33,<br>81.50 | 76.55, 78.50        |
|      | RFE    | 70.15,<br>70.00 | 75.34,<br>76.00 | 80.25,<br>80.50 | 86.91,<br>87.00 | 82.32,<br>82.50 | 81.06,<br>81.50 | 81.99, 80.50        |



Şekil 4. Her aşama için en başarılı yöntemlerin doğruluk değerlerini içeren grafik (%)

Tablo 6. 30sn'lik veri seti için XGB yönteminin sınıflandırma raporu

| Etiketler             | Kesinlik | Geri Çağırma | F1-skor     | Destek |
|-----------------------|----------|--------------|-------------|--------|
| Disko                 | 0.73     | 0.76         | 0.74        | 21     |
| Metal                 | 0.75     | 1            | 0.86        | 12     |
| Reggae                | 0.75     | 0.88         | 0.81        | 24     |
| Blues                 | 0.86     | 0.82         | 0.84        | 22     |
| Rock                  | 0.5      | 0.47         | <b>0.48</b> | 15     |
| Klasik                | 0.93     | 0.96         | <b>0.95</b> | 28     |
| Jazz                  | 0.86     | 0.71         | 0.77        | 17     |
| Hiphop                | 0.94     | 0.84         | 0.89        | 19     |
| Country               | 0.8      | 0.73         | 0.76        | 22     |
| Pop                   | 0.89     | 0.85         | 0.87        | 20     |
| <i>Doğruluk</i>       |          |              | <b>0.81</b> | 200    |
| <i>Makro ort.</i>     | 0.8      | 0.8          | 0.8         | 200    |
| <i>Ağırlıklı ort.</i> | 0.81     | 0.81         | 0.81        | 200    |

Tablo 7. 3sn'lik ses bölütleme veri seti için CS yöntemi sonrası DNN yönteminin sınıflandırma raporu

| Etiketler             | Kesinlik | Geri Çağırma | F1-skor     | Destek |
|-----------------------|----------|--------------|-------------|--------|
| Disko                 | 0.93     | 0.94         | 0.93        | 186    |
| Metal                 | 0.98     | 0.95         | 0.96        | 192    |
| Reggae                | 0.95     | 0.98         | 0.96        | 202    |
| Blues                 | 0.96     | 0.97         | 0.97        | 206    |
| Rock                  | 0.90     | 0.92         | 0.91        | 192    |
| Klasik                | 0.98     | 0.97         | 0.97        | 206    |
| Jazz                  | 0.95     | 0.96         | 0.96        | 189    |
| Hiphop                | 0.96     | 0.97         | 0.97        | 218    |
| Country               | 0.95     | 0.92         | 0.93        | 189    |
| Pop                   | 0.96     | 0.96         | 0.96        | 216    |
| <i>Doğruluk</i>       |          |              | <b>0.95</b> | 1996   |
| <i>Makro ort.</i>     | 0.95     | 0.95         | 0.95        | 1996   |
| <i>Ağırlıklı ort.</i> | 0.95     | 0.95         | 0.95        | 1996   |

Geçmiş çalışmalarda kullanılan GTZAN veri setinin başarı kıyaslaması Tablo 8’de verilmiştir. Ses bölütleme ve özellik seçme yöntemlerinin etkisi analiz edildiğinde çalışma başarı konusunda ön plana çıkmıştır. Diğer çalışmalarda spektrogram ile görüntü analizi de yoğunlukta olduğundan, bu çalışma sadece ses üzerine analiz gerçekleştirme açısından özgündür.

**Tablo 8.** Geçmiş çalışmalar ile başarı kıyaslaması

| Çalışma                       | Doğruluk (%) |
|-------------------------------|--------------|
| Cheng vd. [8]                 | 83.3         |
| Nanni vd. [9]                 | 90.7         |
| Ghildiyal vd. [5]             | 91.0         |
| Senac vd. [6]                 | 91.0         |
| Prabhakar ve Lee [11]         | 93.5         |
| Panagakis ve Kotropoulos [12] | 93.7         |
| <b>Çalışmanın Başarısı</b>    | <b>95.2</b>  |
| Barros vd. [10]               | 96.4         |
| Elbir ve Aydin [4]            | 97.6         |

#### 4 Sonuçlar

Çalışmada GTZAN veri seti kullanılarak müzik türlerinin sınıflandırması amaçlanmıştır. Sınıflandırma için spektrogram gibi görüntü özelliklerinden ziyade sayısal değerlere sahip özelliklerden (mfcc, tempo, flux, vs.) yararlanılmıştır. Analiz aşamasında ses bölütleme ve özellik seçme yöntemlerinin müzik türlerini sınıflandırmaya etkisi de araştırılmıştır. Bunun için 3, 5, 10, 15sn’lik ses bölütleme işlemi ve RFE, LVF, CS, BFE özellik seçme yöntemleri sisteme uygulanmıştır.

İlk aşamada 30sn’lik veri setinin özellikleri çıkarılarak makine öğrenme ve DNN yöntemleriyle analiz edilmiştir. Analiz sonucunda XGB yöntemi %81.00’lik doğruluk değeri elde etmiştir. Sonraki aşamada ses bölütleme işlemi uygulanarak ses kaydı sayısı artırılmış ve daha fazla eğitim-test verisi elde edilmiştir. Ses bölütleme işlemi ile eğitilen sistemde 3sn için %91.58, 5sn için %89.72, 10sn için %85.95, 15sn için ise %84.42’lik doğruluk değerleri XGB yöntemi ile elde edilmiştir. Makine öğrenme yöntemleri incelendiğinde XGB yöntemleri diğer tüm yöntemlerden daha başarılı olmuştur. Bu analizlerin sonucunda, 30sn’lik ses kaydının test sonuçlarına göre %10.58 doğruluk değeri artışı sağlanmıştır. Bunun nedeni olarak ses kaydı sayısının artması ve müzik türlerinin daha belirgin eğitilmesi gösterilebilir.

Ses bölütleme işlemi sonrasında, elde edilen veri setlerine özellik seçme yöntemleri uygulanmıştır. Özellik seçme yöntemleri uygulanması ile elde edilen analizler sonucunda, tüm veri setlerinde başarı artışı sağlanmıştır. En yüksek başarı, 5sn ses kaydı için CS uygulanması ile DNN yöntemi kullanılarak %95.19 doğruluk değeri ile elde edilmiştir. Sonuçlar incelendiğinde özellik seçme yöntemleri uygulandığında en iyi özellikler seçilerek eğitim-test işlemi gerçekleştirilmektedir. Böylece başarı artışı sağlanmıştır. Ses bölütleme işlemi ve özellik seçme uygulanması

sonucunda %14.19’luk doğruluk değeri artışı gerçekleşmiştir.

Müzik türlerinin kendi içerisinde başarısı da analiz edilmiştir. Bunun için, 30sn’lik ses kaydına herhangi işlem uygulanmadan elde edilen en iyi başarıya sahip XGB ve 3sn’lik CS yöntemi uygulanması ile en başarılı olan DNN yönteminin sınıflandırma raporları incelenmiştir. Raporların sonuçları rock müzik türünün daha zor, en kolay ise klasik müzik türünün tespit edilebildiğini göstermiştir.

Gelecek çalışmalar sağlık alanında ses analizi çalışmalarına odaklanması planlanmaktadır. Hastalık evreleri tespiti, hastalığın olup olmadığı ses analizi yoluyla elde edilmeye çalışılacaktır.

#### Çıkar çatışması

Çıkar çatışması yoktur.

**Benzerlik oranı (Turnitin):** %9

#### Kaynaklar

- [1] M. B. Gulmezoglu, V. Dzhabarov, M. Keskin, and A. Barkana, A novel approach to isolated word recognition. *IEEE Transactions on Speech and Audio Processing*, 7(6), 620-628, 1999. <https://doi.org/10.1109/89.799687>
- [2] S. Keser, Improvement of face recognition performance using a new hybrid subspace classifier. *Signal, Image and Video Processing*, 17(5), 2511-2520, 2023.
- [3] N. Pelchat and C. M. Gelowitz, Neural network music genre classification. *Canadian Journal of Electrical and Computer Engineering*, 43(3), 170-173, 2020. <https://doi.org/10.1109/CJECE.2020.2970144>
- [4] A. Elbir and N. Aydin, Music genre classification and music recommendation by using deep learning. *Electronics Letters*, 56(12), 627-629, 2020. <https://doi.org/10.1049/el.2019.4202>
- [5] A. Ghildiyal, K. Singh and S. Sharma, Music genre classification using machine learning. In 2020 4th international conference on electronics, communication and aerospace technology (ICECA), pp. 1368-1372, India, 5-7 November 2020.
- [6] C. Senac, T. Pellegrini, F. Mouret and J. Pinquier, Music feature maps with convolutional neural networks for music genre classification. In Proceedings of the 15th international workshop on content-based multimedia indexing, 1-5, Florence, Italy, 19-21 June 2017. <https://doi.org/10.1145/3095713.3095733>
- [7] H. Bahuleyan, Music genre classification using machine learning techniques. arXiv preprint, 2018. arXiv:1804.01149.
- [8] Y. H. Cheng, P. C. Chang and C. N. Kuo, Convolutional neural networks approach for music genre classification. In 2020 International Symposium on Computer, Consumer and Control (IS3C), 399-403, Taiwan, 13-16 November 2020. <https://doi.org/10.1109/IS3C50286.2020.00109>
- [9] L. Nanni, Y. M. Costa, D. R. Lucio, C. N. Silla Jr, and S. Brahmam, Combining visual and acoustic features for audio classification tasks. *Pattern Recognition*



- Letters, 88, 49-56, 2017. <https://doi.org/10.1016/j.patrec.2017.01.013>
- [10] P. Barros, C. Weber, and S. Wermter, Learning auditory neural representations for emotion recognition. In 2016 International Joint Conference on Neural Networks (IJCNN), 921-928, July 2016.
- [11] S. K. Prabhakar, and S. W. Lee, Holistic approaches to music genre classification using efficient transfer and deep learning techniques. Expert Systems with Applications, 211, 118636, 2023. <https://doi.org/10.1016/j.eswa.2022.118636>
- [12] Y. Panagakis, and C. Kotropoulos, Music genre classification via topology preserving non-negative tensor factorization and sparse representations. In 2010 IEEE International Conference on Acoustics, Speech and Signal Processing, 249-252, March 2010.
- [13] C. Bakir, and M. Yuzkat, Speech emotion classification and recognition with different methods for Turkish language. Balkan Journal of Electrical and Computer Engineering, 6(2), 122-128, 2018. <https://doi.org/10.17694/bajece.419557>
- [14] M. Sharma, S. Joshi, T. Chatterjee, and R. Hamid, A comprehensive empirical review of modern voice activity detection approaches for movies and TV shows. Neurocomputing, 494, 116-131, 2022. <https://doi.org/10.1016/j.neucom.2022.04.084>
- [15] O. Karaman, H. Çakın, A. Alhudhaif and K. Polat, Robust automated Parkinson disease detection based on voice signals with transfer learning. Expert Systems with Applications, 178, 115013, 2021. <https://doi.org/10.1016/j.eswa.2021.115013>
- [16] C. Quan, K. Ren and Z. Luo, A deep learning based method for Parkinson's disease detection using dynamic features of speech. IEEE Access, 9, 10239-10252, 2021. <https://doi.org/10.1109/ACCESS.2021.3051432>
- [17] O. Amir, W.T. Abraham, Z. S. Azzam, G. Berger, S. D. Anker, S. P. Pinney, D. Burkhoff, I. D. Shallom, C. Lotan and E. R. Edelman, Remote speech analysis in the evaluation of hospitalized patients with acute decompensated heart failure. Heart Failure, 10(1), 41-49, 2022. <https://doi.org/10.1016/j.jchf.2021.08.008>
- [18] Z. Al-Jumaili, T. Bassiouny, A. Alanezi, W. Khan, D. Al-Jumeily and A. J., Hussain, Classification of spoken English accents using deep learning and speech analysis. In International Conference on Intelligent Computing, 277-287, Xi'an, China, 7-11 August 2022.
- [19] O. F. Çıplak, ve S. Keser, Gerçek Zamanlı Ses Tanıma ile Robot Kolu Kontrolü. Avrupa Bilim Ve Teknoloji Dergisi, (31), 34-39, 2021. <https://doi.org/10.31590/ejosat.969608>
- [20] G. Tzanetakis, and P. Cook, Musical genre classification of audio signals. IEEE Transactions on speech and audio processing, 10(5), 293-302, 2002. <https://doi.org/10.1109/TSA.2002.800560>
- [21] B. McFee, C. Raffel, D. Liang, D. P. Ellis, M. McVicar, E. Battenberg and O. Nieto, librosa: Audio and music signal analysis in python. In Proceedings of the 14th python in science conference, 18-25, Texas, (8), 6-12 July 2015.
- [22] C. J. Steinmetz, and J. Reiss, pyloudnorm: A simple yet flexible loudness meter in Python. In Audio Engineering Society Convention 150. Audio Engineering Society, 2021.
- [23] J. Han, M., Kamber and J. Pei, Data preprocessing. Data Mining (Third Edition). The Morgan Kaufmann Series in Data Management Systems, pp. 83-124, 2012.
- [24] L. V. Filter, and P. Filter, Seven techniques for dimensionality reduction. Technical report, 2014.
- [25] A. Sikri, N. P. Singh, and S. Dalal, Chi-Square Method of feature selection: impact of pre-processing of data. International Journal of Intelligent Systems and Applications in Engineering, 11(3s), 241-248, 2023.
- [26] I. Guyon, J. Weston, S. Barnhill, and V. Vapnik, Gene selection for cancer classification using support vector machines. Machine learning, 46, 389-422, 2002.
- [27] Ö. Şahinaslan, H. Dalyan, ve E. Şahinaslan. Naive bayes sınıflandırıcısı kullanılarak youtube verileri üzerinden çok dilli duygu analizi. Bilişim Teknolojileri Dergisi, 15 (2), 221-229, 2022. <http://doi.org/10.17671/gazibtd.999960>
- [28] E. Çavuş ve İ. Sancaktar, Batarya sağlık durumunun makine öğrenmesi ile kestirimi. Niğde Ömer Halisdemir Üniversitesi Mühendislik Bilimleri Dergisi, 11 (3), 601-610, 2022. <https://doi.org/10.28948/ngumuh.1112985>.
- [29] N. Akyel ve K. Seçkin. K-en Yakın Komşuluk Algoritmasının Hile Denetiminde Kullanımı. Journal of Accounting and Taxation Studies, 5(1), 21-40, 2012.
- [30] A. C. Başara ve Y. Şişman, Frekans oranı, kanıt ağırlığı ve lojistik regresyon yöntemleri kullanılarak heyelan duyarlılık haritalarının CBS tabanlı karşılaştırılması. Niğde Ömer Halisdemir Üniversitesi Mühendislik Bilimleri Dergisi, 11 (3), 647-660, 2022. <https://doi.org/10.28948/ngumuh.1065284>
- [31] A. C. Kelle ve H. Yüce, MQTT Trafığında DoS Saldırılarının Makine Öğrenmesi ile Sınıflandırılması ve Modelin SHAP ile Yorumlanması. Journal of Materials and Mechatronics: A, 3(1), 50-62, 2022.
- [32] M. Açıkkar, Prediction Of Gross Calorific Value Of Coal From Proximate And Ultimate Analysis Variables Using Support Vector Machines With Feature Selection. Niğde Ömer Halisdemir Üniversitesi Mühendislik Bilimleri Dergisi, 9 (2), 1129-1141, 2020. <https://doi.org/10.28948/ngumuh.585596>.
- [33] G. Montavon, W. Samek, and K. R. Müller, Methods for interpreting and understanding deep neural networks. Digital signal processing, 73, 1-15, 2018.

