



## TÜRKÇE KONUŞMADA DUYGU TANIMA İÇİN MAKİNE ÖĞRENME YÖNTEMLERİ VE DERİN ÖĞRENME TABANLI MODELLERİN KARŞILAŞTIRILMASI

Zekeriya Anıl GÜVEN<sup>1\*</sup>

İzmir Bakırçay Üniversitesi, Mühendislik ve Mimarlık Fakültesi, Bilgisayar Mühendisliği Bölümü, İzmir, Türkiye

### Anahtar Kelimeler

*Duygu analizi,  
Duygu tanıma,  
Ses analizi,  
Makine öğrenmesi,  
Derin öğrenme modelleri.*

### Öz

Son zamanlarda veri miktarına bağlı olarak sağlık, eğitim, pazarlama gibi birçok alanda analizlere ihtiyaç duyulmaktadır. Duygu analizi ise bu alanlarda kişilerin yorumlarını analiz etme, duygularını çıkarma için oldukça popüler bir alandır. Bu çalışmada kızgın, mutlu, sakin ve üzgün duygu etiketleri içeren Türkçe konuşma veri seti üzerinde, ses karakteristik özellikleri ve spektrogramlardan yararlanarak duyguların tespit edilmesi amaçlanmaktadır. Analiz aşamasında Librosa kütüphanesi ile çıkarılan sayısal özellikler ile makine öğrenme yöntemleri ve derin sinir ağları eğitilerek başarıları ölçülmüştür. Ayrıca düşük varyans filtreleme, geri yönlü özellik eleme, ki-kare ve temel bileşen analizi yöntemleri ile özellik azaltım işlemi uygulanarak elde edilen yeni özellikler ile makine öğrenme yöntemlerinin başarısındaki değişiklikler de araştırılmıştır. Görsel veri olan spektrogramlar ise CNN, RNN, LSTM, EfficientNet, ResNet, MobileNet ve DenseNet derin öğrenme tabanlı modellerin eğitilmesi için kullanılmıştır. Modellerin eğitim aşamasında veri seti ile beraber modellere ince ayar işlemi uygulanmıştır. Deneysel çalışmaların sonucunda makine öğrenme yöntemlerinden Ekstrem Gradient Artırma %87.03 doğruluk değeri verirken, ResNet modeli ise %79.23 doğruluk değeri vermiştir.

## COMPARISON OF MACHINE LEARNING METHODS AND DEEP LEARNING-BASED MODELS FOR EMOTION RECOGNITION IN TURKISH SPEECH

### Keywords

*Sentiment analysis,  
Emotion recognition,  
Voice analysis,  
Machine learning,  
Deep learning models.*

### Abstract

Recently, depending on the amount of data, analyzes are needed in many areas such as health, education and marketing. Sentiment analysis is a very popular area for analyzing people's comments and extracting their emotions in these areas. This study aims to determine the emotions by using voice characteristics and spectrograms on the Turkish speech dataset containing angry, happy, calm and sad emotion labels. In the analysis phase, machine learning methods and deep neural networks were trained, and their success was measured with the numerical features extracted from Librosa library. In addition, feature reduction was applied with low variance filtering, backward feature elimination, chi-square and principal component analysis methods, and the changes in the success of machine learning methods were also investigated with the new features obtained. The spectrograms, which are visual data, were used to train CNN, RNN, LSTM, EfficientNet, ResNet, MobileNet and DenseNet deep learning-based models. During the training phase of these models, fine-tuning process was applied to these models together with the dataset. As a result of the experimental studies, Extreme Gradient Boosting, one of the machine learning methods, gave an accuracy value of 87.03%, while the ResNet model gave an accuracy value of 79.23%.

### Alıntı / Cite

Güven Z.A. (2024). Türkçe Konuşmada Duygu Tanıma İçin Makine Öğrenme Yöntemleri Ve Derin Öğrenme Tabanlı Modellerin Karşılaştırılması, Mühendislik Bilimleri ve Tasarım Dergisi, 12(2), 285-297.

### Yazar Kimliği / Author ID (ORCID Number)

Z.A. Güven, 0000-0002-7025-2815

### Makale Süreci / Article Process

<b>Başvuru Tarihi / Submission Date</b>	28.08.2023
<b>Revizyon Tarihi / Revision Date</b>	09.01.2024
<b>Kabul Tarihi / Accepted Date</b>	18.03.2024
<b>Yayın Tarihi / Published Date</b>	30.06.2024

\* İlgili yazar / Corresponding author: anilguven1055@gmail.com, +90 506 562 3572

# COMPARISON OF MACHINE LEARNING METHODS AND DEEP LEARNING-BASED MODELS FOR EMOTION RECOGNITION IN TURKISH SPEECH

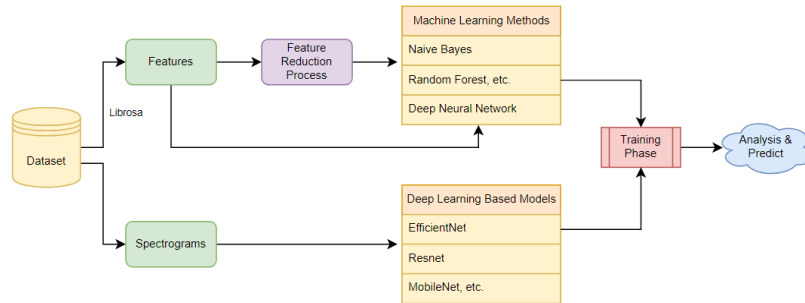
Zekeriya Anıl GÜVEN<sup>1†</sup>

İzmir Bakırçay Üniversitesi, Mühendislik ve Mimarlık Fakültesi, Bilgisayar Mühendisliği Bölümü, İzmir, Türkiye

## Highlights

- It fills the gap in the literature on voice analysis in terms of Turkish language.
- By using the current deep learning-based models in the training-test phase, it is provided to compare with traditional models for Turkish.
- The effect of applying feature reduction methods on the methods was investigated.

## Graphical Abstract



Şekil./Figure. Çalışmanın metodolojisi (The methodology of this study)

## Purpose and Scope

Due to the scarcity of studies on the Turkish language, this study aims to determine the emotions with methods and models by using sound characteristics and spectrograms on the Turkish speech dataset containing angry, happy, calm and sad emotion labels.

## Design/methodology/approach

When the methodology is examined, in the first stage, sound characteristics are extracted by reading the sound files through the Librosa library. Extracted features are trained and tested by applying machine learning methods to deep neural networks. Models, on the other hand, are trained with extracted spectrogram image features specific to audio recordings. Then, the system is re-trained and re-tested with methods as a result of the selection of more relevant features by using feature reduction methods. As a result of these processes, a structure that can extract the emotion of the audio recording for Turkish is presented.

## Findings

Two different feature extractions were used during the training phase, namely the numerical features obtained with the Librosa library and the spectrogram image features. In the first stage, the numerical features obtained with Librosa were trained with machine learning methods and deep neural networks, and the success of the methods was measured. As a result of this analysis, an accuracy value of approximately 87.03% was obtained with the Extreme Gradient Boosting (XGB) method. Since many features are used in the training phase, the effect of these methods on success was analyzed by using feature reduction methods in the next step. As a result of the analysis, as a result of the decrease in the features with the Low Variance Filtering method, the XGB method gave an accuracy value of 86.74%, while there was a decrease of approximately 0.3% in success. The reason for this decrease can be stated as the deletion of the features that have a positive effect on success.

The spectrogram features, in which the frequency spectrum of the audio recordings is displayed visually, were used in the training of deep learning-based models. As a result of the use of normalized spectrograms, ResNet among the models reached an accuracy of 79.23%. Here, the reason for the low success of the model compared to machine learning methods can be shown as the insufficient dataset to fine-tune the model with millions of parameters. The most unsuccessful among the models was MobileNetV2.

## Originality

This study fills the gap in the literature on emotion analysis with the voice in terms of the Turkish language. It also uses the current deep learning-based models in the training-test phase and provides to comparison with traditional models for Turkish.

<sup>†</sup> İlgili yazar / Corresponding author: anilguven1055@gmail.com, +90 506 562 3572

## 1. Giriş (Introduction)

Günümüzde veri sayısının oldukça artmasından dolayı birçok alanda konuya özgü analiz yapılmaktadır. Bilgisayar bilimleri alanında doğal dil işleme, görüntü işleme, ses ve sinyal işleme alanlarına özgü analizler ile çevreye fayda sağlanmaktadır. Bir ürüne özgü yorumları içeren duygu analizi, kameradan belli nesneyi tanımayı sağlayan nesne tespiti, konuşan kişinin aksanını tespit etme gibi konular bu faydalara örnek olarak verilebilmektedir.

Ses analizi alanı sağlık, ticaret gibi birçok sektörde yararlanılan alanlardandır. Canlılara ait ses karakteristik özellikleri ile yapılan analizler ses analizi olarak bilinmektedir. Kişilerin hasta olup olmadığını, kuşların ötme sesine göre türünü, kişilerin aksanını veya duygusunu, müzik türlerini, vs. tespit ederken kullanılmaktadır. Ses karakteristik özellikleri, canlıya özgü olan özelliklerdir. Tempo, shimmer, jitter, spektrogram, chromogram, spektral flux gibi frekans veya zaman domaini için özellikler bulunmaktadır. Bu özelliklerden bazıları (tempo, shimmer, vs.) sayısal veri olarak analiz edilirken, bazıları (spektrogram, vs.) ise görsel olarak analiz edilmektedir. Bu özellikler kullanılacak yöntemlere veya modellere eğitim için verildikten sonra test işlemiyle sistemin başarısı ölçülmektedir.

Duygu analizi veya tanıma, kişilerin verilerden duygusunun çıkarıldığı analiz türüdür. Olumlu, olumsuz, nötr gibi duygu analizine ek olarak kızgın, mutlu, üzgün gibi duyguya özgü analizler de yapılmaktadır. Ses analizi ile de kişinin duygusu tespit edilebilmektedir. Kişinin sesindeki vurgudan, tempodan yararlanarak elde edilen kişinin sesine özgü karakteristik özellikler duyguyu belirleyebilmektedir. Genel olarak bu tür çalışmalar İngilizce dili üzerine yapılmaktadır. Literatürde Türkçe diline katkı sağlamak amacıyla çalışmada Türkçe konuşma ses kayıtları kullanılarak ses karakteristik özellikleri ile kişinin duygu çıkarımı yapılması amaçlanmaktadır. Bunun için kızgın, sakin, mutlu ve üzgün olmak üzere 4 duygu içeren TurEv-DB<sup>‡</sup> veri seti kullanılmıştır. Veri setine ait çıkarılan sayısal özellikler makine öğrenme yöntemleri, derin sinir ağları ile eğitim-test aşamasına verilirken, spektrogram içeren görsel özellikler ise CNN, LSTM, RNN, EfficientNet, Resnet gibi derin öğrenme tabanlı modellere eğitim-test için verilmiştir. Böylece hem sayısal ve görsel özelliklerin hem de yöntemler ve modellerin sınıflandırmaya etkisi araştırılmıştır. Ayrıca sayısal özellikler için özellik azaltım yöntemleri kullanılarak sistemin başarısına etkisi de analiz edilmiştir.

Çalışmanın katkısı incelendiğinde;

- Türkçe dili açısından ses analizi üzerine literatürdeki eksikliği gidermektedir.
- Güncel derin öğrenme tabanlı modellerin eğitim-test aşamasında kullanılmasıyla geleneksel modeller ile karşılaştırılması sağlanmıştır.
- Özellik azaltım yöntemleri uygulamanın yöntemler üzerindeki etkisi araştırılmıştır.
- Geçmiş çalışmalara ek olarak Türkçe konuşmada duygu tanıma için derin öğrenme tabanlı modeller kullanılmasına öncü olmuştur.

Çalışmanın akışı düşünüldüğünde; "Kaynak Araştırması" başlığı altında ses ile duygu analizine özgü literatürde yer alan çalışmalar anlatılmaktadır. Veri seti, kullanılan makine öğrenme yöntemleri ve derin öğrenme tabanlı modeller, özellik azaltım teknikleri gibi kullanılan tüm materyal ve metotlar üçüncü başlık altında anlatılmaktadır. "DeneySEL Sonuçlar" başlığı altında ise uygulanan tüm analizler tablo ve şekiller ile açıklanmaktadır. Çalışmanın sonuçlarının değerlendirilip tartışılması ise beşinci başlık altında gerçekleştirilmektedir.

## 2. Kaynak Araştırması (Literature Survey)

Singh (2021), duygu tespiti yapabilmek için kullanıcının sesi üzerinde Librosa kütüphanesinden yararlanarak bir analiz gerçekleştirmiştir. Analiz aşamasında İngilizce veri seti kullanılmış ve sistem makine öğrenme yöntemleriyle test edilmiştir. Murwati ve Aldianto (2022), tüketicilerin Twitter'da kullanıcı tarafından oluşturulan içeriğe dayalı olarak müşteri hizmetleri için sohbet robotlarına yönelik duygularını belirleyen sözlük tabanlı TextBlob ve lojistik regresyon tekniklerini içeren hibrit bir yöntem kullanmışlardır. Sonuçlar, insanların genel olarak chatbotları olumlu karşıladığını göstermişlerdir. Oflazoglu ve Yildirim (2013), Türk filmlerinden konuşma duygu veri seti oluşturmuşlar ve akustik özellikler kullanılarak nötr, üzgün, mutlu, korkunç duyguları tespit etmişlerdir. Koren ve Stipancic (2021), konuşmanın akustik ve dilbilimsel özelliklerine dayanan bir duygu tanıma yöntemi önermişlerdir. Ses, akustik alt yöntemle bağlıyken, dilbilimsel alt yöntem, özel bir doğal dil işleme modelinin kullanıldığı metin mesajlarıyla ilişkilendirilmiştir. Ren vd. (2014), büyük ölçekli ağ bağlantılı ses verilerinden duygu tahmini belirlemeyi amaçlamışlardır. İlk olarak insan-mobil sesli iletişimdeki veri gözlemlerini ve altta yatan duygu kalıplarını araştırmışlardır. Sonraki aşamada duyguları otomatik olarak tahmin etmek için akustik özellikleri, içerik bilgilerini ve coğrafi bilgileri birleştirmek için derin bir sığ sinir ağı (DSNN) modeli önermişlerdir. Canpolat vd. (2020), Türkçe için ses kayıtlarından duygu veri seti oluşturmuşlardır. Veri setini

<sup>‡</sup> <https://github.com/Xeonen/TurEV-DB>

frekans aralıklarında analiz ederek Konvolüsyonel Sinir Ağı (CNN) ve DVM yöntemleri üzerinde eğitim gerçekleştirmişlerdir. Özönmez vd. (2021), akustik özellik çıkarım yöntemleri ile elde edilen özelliklere Temel Bileşenler Analizi yöntemini uygulayarak özellik azaltımını sağlamışlar ve yapay sinir ağı modellerinin mimarisini duygu sınıflandırma için Türkçe konuşma veri setleri üzerinde uygulamışlardır. Anand ve Patra (2022), ses tabanlı duygu analiz sistemini iki farklı yaklaşım ile değerlendirmişlerdir. İlk olarak, sesi metne dönüştürerek metni analiz etmişlerdir. Diğer yaklaşımda ise sesi girdi olarak kullanıp ses özellikleriyle duygu analizi yapmışlardır. Alu vd. (2017), ses örneklerini işleyen ve sınıflandıran derin bir CNN mimarisi ile sesten duygu algılamanın uygulanmasını sunmuşlardır. Satyanarayana vd. (2020), Amazon Comprehend hizmetini kullanarak, görüntü, ses vb. yapılandırılmamış verileri çıkarmışlar ve konuşmanın olumlu, olumsuz, nötr veya karışık duygu olup olmadığı konusunda çıktı elde etmişlerdir.

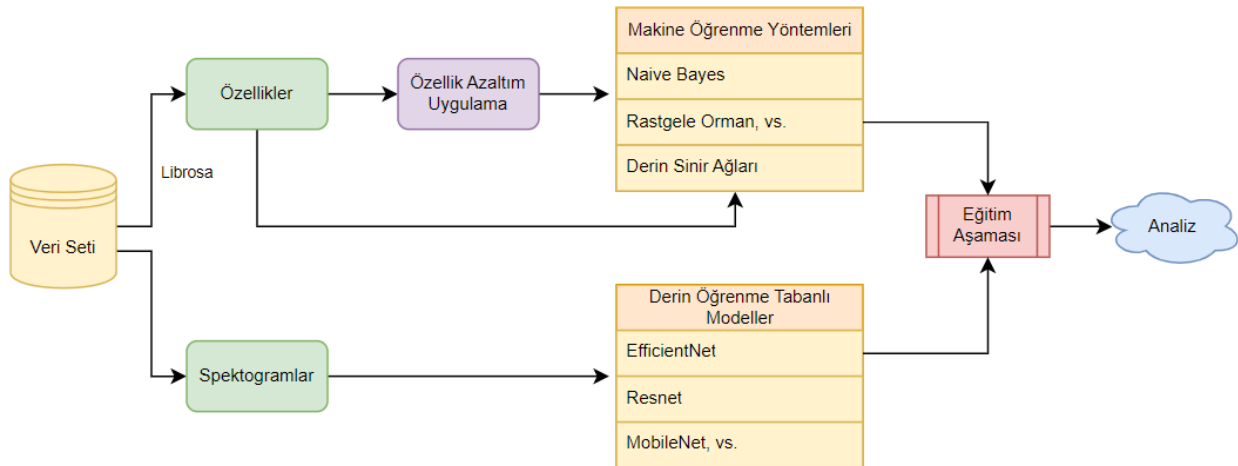
Duygu analizi haricinde ses analizi birçok konu için kullanılmaktadır. Aksan sınıflandırma (Wu vd., 2019; Najafian ve Russell, 2020), kişinin hasta olup olmadığını tespit etme (Tracy vd., 2020; Ali vd., 2019), müzik türlerini belirleme (Pelchat ve Gelowitz, 2020; Elbir ve Aydın, 2020) bu konulara örnek olarak verilebilir.

Literatür incelendiğinde hem Türkçe dili, hem de günümüzde popüler olan derin öğrenme tabanlı modellerin geleneksel yöntemlerle kıyaslanması açısından çalışmalar sınırlıdır. Ayrıca sayısal özellikler ile görsel özelliklerin sınıflandırmaya etkisini araştıran çalışmalar da oldukça azdır. Bundan dolayı, bu çalışma belirtilen boşlukları literatür için doldurmaktadır.

### 3. Materyal ve Yöntem (Material and Method)

Çalışmada kullanılan Türkçe konuşma içeren duygu veri seti, makine öğrenme yöntemleri, derin sinir ağları, derin öğrenme tabanlı Keras uygulama modelleri<sup>§</sup> ve özellik azaltım yöntemleri bu başlık altında anlatılmaktadır.

Metodoloji incelendiğinde, ilk aşamada ses dosyaları Librosa\*\* kütüphanesi aracılığıyla okunarak ses karakteristik özellikleri çıkarılmaktadır. Çıkarılan sayısal özellikler makine öğrenme yöntemleri, derin sinir ağlarına verilerek eğitilip test edilmektedir. Modeller ise ses kayıtlarına özgü çıkarılan görsel spektrogram özellikleri ile eğitilmektedir. Ardından sadece sayısal özellikler için özellik azaltım yöntemleri kullanılarak daha ilgili özelliklerin seçilmesi sonucunda sistem yeniden makine öğrenme yöntemleri ile eğitilip test edilmektedir. Bu işlemler sonucunda Türkçe için ses kaydına ait duyguyu çıkarabilen bir yapı sunulmaktadır. Oluşturulan yapıya ait metodoloji Şekil 1'de gösterilmektedir.



Şekil 1. Çalışmanın metodolojisi (The methodology of this study)

#### 3.1. TurEV-DB Veri Seti (TurEV-DB Dataset)

Veri seti, belirli insanların dört farklı duyguda (kızgın, sakin, mutlu, üzgün) söylediği 82 kelimeye dayanan 1735 tane ses kaydı içermektedir. 82 kelime, kelimelerin fonolojik özellikleri dikkate alınarak Türkçenin Ses Dizgesi'nden (Ergenç ve Uzun, 2017) seçilmiştir. Seçilen kelimelerde, kelimelerin çeşitli konumlarında kullanılan hem sesli hem de sessiz harflerden oluşan bir dizi fonem bulunmaktadır (Canpolat vd., 2020). Tüm ses kayıtlarında bir

<sup>§</sup> <https://keras.io/api/applications/>

<sup>\*\*</sup> <https://librosa.org/doc/main/index.html>

kelime üzerinde duyguya ait söylem ile kayıt gerçekleştirilmiştir. Tablo 1 korpustaki ses kayıtlarının kayıt alınan kullanıcılara göre sayısını göstermektedir.

**Tablo 1.** Ses kayıtlarının kullanıcıya göre istatistikleri (Canpolat vd., 2020) (Statistics of audio recordings by user)

Etiket	Kullanıcı Id						Toplam
	7895	1984	1234	1358	1157	6783	
<i>Kızgın</i>	82	82	82	82	77	82	487
<i>Sakin</i>	80	82	82	82	0	82	408
<i>Mutlu</i>	29	82	82	82	0	82	357
<i>Üzgün</i>	82	82	82	82	73	82	483
Toplam	273	328	328	328	150	328	1735

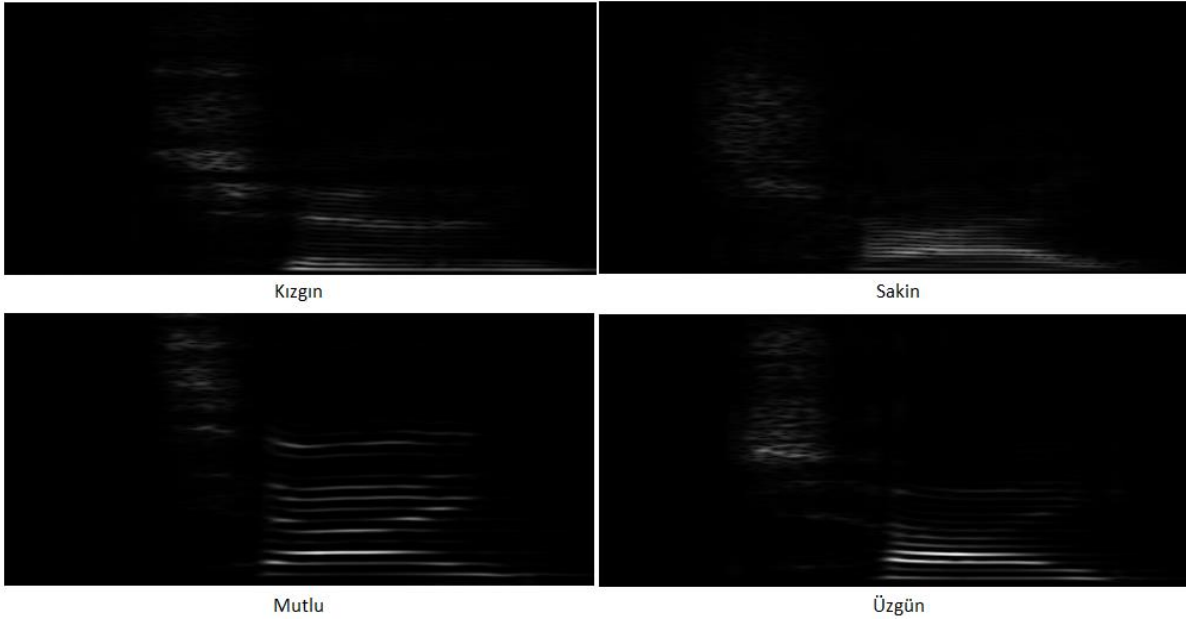
### 3.2. Özellik Çıkarım Kütüphanesi (Feature Extraction Library)

Librosa kütüphanesi, ses kayıtlarının okunması ve ses karakteristik özelliklerinin çıkarılması için kullanılmaktadır. Librosa, müzik ve ses analizi için bir python paketidir. Müzik bilgisine ait getirim sistemleri elde etmek için gerekli özellikleri sağlamaktadır (McFee vd., 2015). Bu kütüphane aracılığıyla zaman ve frekans domainine ait birçok ses karakteristik özellikleri çıkarılmaktadır. Çıkarılan özellikler Tablo 2’de verilmiştir. Tabloda yer alan frekans domainine ait spektral özelliklerin maksimum, minimum, ortalama, standart sapma, basıklık (kurtosis), çarpıklık (skew) değerleri ek özellik olarak elde edilmektedir. Mel-Frekans Kepstral Katsayıları (MFCC) özelliği ise ses frekanslarındaki değişimi belirlenen sayıda çıkarmaktadır. Ayrıca MFCC’nin birinci türevinin değerleri alınarak yine minimum, maksimum, vs. değerleri çıkarılmaktadır. Tüm çıkarılan özellikler göz önüne alındığında toplam özellik sayısı her bir ses kaydı için 175’tir. Tüm özellikler üzerinde yöntemler ile eğitim-test işlemi gerçekleştirildikten sonra, sonraki aşamalar için özellik seçme yöntemleri uygulanarak özellik sayısı azaltılmaktadır.

**Tablo 2.** Librosa kütüphanesi aracılığıyla çıkarılan özellikler (Extracting features via the Librosa library)

Kütüphane	Çıkarılan Özellik	Özellik Sayısı
Librosa	Spectral centroid	6
	Spectral flux	6
	Spectral bandwidth	6
	Spectral flatness	6
	Spectral rolloff	6
	Spectral contrast	6
	Zero crossing rate	6
	Root mean squared error	6
	Chromagram	6
	Tempo	1
	Mel-Frekans Kepstral Katsayıları	120

Eğitim ve test aşamasında kullanılacak derin öğrenme tabanlı modeller için ise ses kayıtlarının normalize edilmiş görsel spektrogram özellikleri kullanılmıştır. Spektrogram, zamana göre değiştiği için bir sinyalin frekans spektrumunu görsel olarak göstermektedir. Modellerde bu özellik ile eğitilerek frekans üzerindeki görselden ses kayıtlarını duyguya göre sınıflandırmaya çalışmaktadır. Spektrogramların elde edilmesi sürecinde Hamming pencereyi kısa süreli Fourier dönüşümünden (STFT) yararlanılmıştır. STFT algoritmasında %95 örtüşme ile 50 milisaniyeye tekabül eden 2205 örneklilik uzun pencere boyutu kullanılmıştır (Canpolat vd., 2020). Veri setindeki her sınıfa ait spektrogram örnekleri Şekil 2’de verilmiştir.



Şekil 2. Her sınıfa ait normaliz spektrum örnekleri (Examples of normalized spectrograms for each class)

### 3.3. Özellik Azaltım Teknikleri (Feature Reduction Techniques)

Çıkarılan özellikler arasında en uygunları seçebilmek ve özellik sayısını azaltabilmek için özellik azaltım teknikleri kullanılmaktadır. Geri Yönlü Eleme (Backward Feature Elimination - BFE), Düşük Varyans Filtresi (Low Variance Filtering - LVF), Ki-kare (Chi Square - CS) yöntemler var olan özellikler arasında seçim yaparak özellikleri azaltırken, Temel Bileşen Analizi (Principal Component Analysis - PCA) istenen özellik sayısı kadar var olan özelliklerden yararlanarak yeni özellikler elde etmektedir. Teknikleri açıklamak gerekirse;

- Geri Yönlü Eleme (BFE): Tüm aday özelliklerini içeren tam bir modelle başlamaktadır. Önceden belirlenmiş bir durdurma kuralı karşılanana kadar özellikler sistematik olarak modelden kaldırılması sağlanmaktadır. Eleme sürecinin belirli bir aşamasında, özet ölçümde en az düşüşe neden olan değişken elenmektedir (Yılmaz ve Kuvat, 2023).
- Düşük Varyans Filtresi (LVF): Her bir sütuna ait varyans değeri hesaplanarak varyans değeri belirli bir eşikten düşük olan özellikleri kaldırmaktadır. Sadece sayısal sütunlar için bu yöntem kullanılmaktadır (Filter ve Filter, 2014).
- Ki-kare (CS): Sınıflandırma modellerinin performansını artıran özellikleri seçmek için "Ki-kare" istatistiklerini kullanmaktadır. Yöntem özelliklerin hedef sınıfa bağımlılığı veya bağımsızlığı kavramına dayanmaktadır (Sikri vd., 2023).
- Temel Bileşen Analizi (PCA): PCA, potansiyel olarak bağlantılı değişkenlerin bir koleksiyonunu, temel bileşen olarak bilinen bir dizi doğrusal olarak ilişkisiz değişkene dönüştürmek için ortogonal bir dönüşüm kullanan bir boyut azaltma işlemidir. Karşılık gelen her öge, önceki bileşenin dik olması koşuluyla mümkün olan en yüksek varyasyona sahiptir, yani mümkün olduğu kadar fazla veri belirsizliği hesaba katılmaktadır (Mulla vd., 2021).

### 3.4. Makine Öğrenme Yöntemleri (Machine Learning Methods)

Eğitim ve test aşamasında makine öğrenme yöntemleri kullanılmaktadır. Kullanılan makine öğrenme yöntemleri:

- Naïve Bayes (NB): İstatistik tabanlı algoritmalarından biridir. Bu algortmada veri setinde daha önce sınıflandırılmış veriler kullanılarak mevcut sınıflardan hangisinin yeni veriye ait olduğu bulma imkanı bulunmaktadır. Bayes sınıflandırıcısı nitelik verilerinin olasılık değerlerini dikkate almaktadır (Çelik, 2022).
- Rastgele Orman (RO): Eğitim aşamasında çok sayıda rastgele oluşturulmuş karar ağacının çıktısını birleştirerek regresyon ve sınıflandırma problemlerini çözmek için kullanılmaktadır (Çavuş ve Sancaktar, 2022).
- k-En Yakın Komşu: Sınıfları belirli olan bir örnek kümedeki değerler aracılığıyla, kullanıcı tarafından tanımlanan kadar komşuya bakarak yeni bir verinin hangi sınıf kümesine atanacağını tespit etmektedir (Sağbaş vd., 2022).
- Lojistik Regresyon (LR): Derlemin her kategorisindeki örnekler, lojistik fonksiyona göre sınıflandırılmaktadır. Lojistik regresyon, verileri tanımlamak ve bir derlemdaki bağımlı ikili değişken ile bağımsız değişkenler arasındaki ilişkiyi açıklamak için kullanılmaktadır (Altinel, 2021).

- Ekstrem Gradient Artırma (XGB): Gradient Artırma algoritmasının optimize edilmiş versiyonudur. XGB, ağaç yapısını oluştururken maksimum derinlik değerinden yararlanmaktadır. Oluşturulan ağaç aşağı yönde aşırı ilerleme gösterirse, ağaç üzerinde budama gerçekleştirilmektedir. Böylece aşırı öğrenme önlenmektedir (Kelle ve Yüce, 2022).
- Destek Vektör Makineleri (DVM): Örüntü tanıma ve regresyon problemlerinin çözümü için temelleri istatistiksel öğrenme teorisine dayanarak önerilmiş bir makine öğrenme algoritmasıdır. Yöntem, verileri optimal olarak kategoriye ayırıp n-boyutlu bir hiperdüzlem oluşturmaya çalışmaktadır (Çevik ve Kayakuş, 2020).
- Derin Sinir Ağları (DNN): Bir DNN, nöronların önceki katmandan nöron aktivasyonlarını girdi olarak aldığı ve basit bir hesaplama gerçekleştirdiği birden çok katman dizisinde düzenlenmiş bir nöron koleksiyonudur. Ağın nöronları, girdiden çıktıya karmaşık doğrusal olmayan bir eşlemeyi (mapping) ortaklaşa uygulamaktadır. Bu eşleme, hata geri yayılımı yöntemi kullanılarak her bir nöronun ağırlıklarının uyarlanmasıyla verilerden öğrenilmektedir (Montavon vd., 2018).

### 3.5. Derin Öğrenme Tabanlı Modeller (Deep Learning Based Models)

CNN, RNN, LSTM, EfficientNet, ResNet, DenseNet, MobileNet modelleri görsel özellikleri ses analizi için eğitim-test aşamasında kullanan modellerdendir. Bu modeller evrimsel sinir ağlarından yararlanarak geliştirildiğinden yapısal olarak derin öğrenme tabanlıdır. Çalışmada kullanılan tüm modeller Keras'taki uygulamalardan alınmıştır. Modellere değinmek gerekirse:

- **CNN:** Farklı konumlardaki bilgileri başarılı doğrulukla algılayabilen bir yapay sinir ağı biçimidir. Bu model, görüntü işleme ve doğal dil işlemedeki çeşitli sorunları çözmüştür. Öğrenmeyi kolaylaştıran özel bir mimariyle karakterize edilmektedir. CNN çok katmanlı bir ağıdır, dolayısıyla bir katmanın çıktısı bir sonraki katmanın girdisi olmaktadır. Genellikle bir girdi, bir veya birkaç gizli katman ve bir çıktından oluşmaktadır (Rhanoui vd., 2019).
- **RNN:** Tekrarlayan sinir ağı (RNN), nöronların w ağırlığındaki yaylarla bağlandığı, birbirine bağlı ve etkileşimli bir nöron ağıdır. Bu tür bir ağ, değişken boyutlardaki girdilerin yanı sıra zaman serileri, yani otomatik konuşma tanıma ve otomatik kalıp tanıma gibi durumlarda çok kullanışlıdır. Bu tür yapay sinir ağlarında bilginin yayılma yönü çift yönlüdür; veri sırasını korur; ve dahili hafızası sayesinde bir döngüye dayandığı için uzun dizilerden oluşan bir giriş arasında bağlantı kurabilmektedir (Rhanoui vd., 2019).
- **LSTM:** Hafızası sayesinde gradient kaybolması sorununu çözebilen RNN'nin bir uzantısıdır. Üç kapı aracılığıyla verilerin okunmasını, yazılmasını ve silinmesini mümkün kılmaktadır. Giriş kapısı, güncellemelere izin verirken veya engellerken, unutmaya kapısı algoritma tarafından öğrenilen ağırlıklara göre önemli değilse bir nöronu devre dışı bırakmaktadır. Çıkış kapısında ise çıktındaki nöron durumu kontrol edilmektedir (Rhanoui vd., 2019).
- **EfficientNet:** EfficientNet Modelleri, basit ve oldukça etkili bileşik ölçeklendirme yöntemlerine dayalıdır. Bu yöntem, transfer öğrenme veri kümeleri için kullanılan model verimliliğini korurken temel bir CNN'i herhangi bir hedef kaynak kısıtlamasına ölçeklendirmeyi sağlamaktadır (Tan ve Lee, 2019). EfficientNet, bilgisayarla görme görevleri için yeni bir yaklaşım getirmektedir. EfficientNet, B0'dan B7'ye modelleri içerir ve her birinin 5.3 milyondan 66 milyona kadar farklı parametreleri vardır (Marques vd., 2020). Çalışmada B7 modeli kullanılmıştır. Ayrıca bu zamana kadar literatür incelendiğinde Türkçe ses kayıtlarından duygu tespitine ilişkin öğrenmeyi aktarmak için EfficientNet'i kullanan benzer bir çalışma yoktur.
- **ResNet:** ResNet, Artık (Residual) Ağ'ın kısa biçimidir. Derin evrimsel sinir ağları, yıllar içinde görüntü tanıma ve sınıflandırma alanında oldukça fazla kullanılmaktadır. Ancak, kaybolan gradyan ve bozulma problemleri nedeniyle daha derin sinir ağlarını eğitmek zor olmuştur. ResNet, bu iki sorunu da çözmeye odaklanmaktadır. Sinir ağlarında her katman, eldeki görev için eğitilirken düşük veya yüksek seviyeli özellikleri öğrenirken, ResNet modeli bazı artıkları öğrenmeye çalışmaktadır. ResNet modelinde Relu aktivasyonları kullanılmaktadır (Reddy ve Juliet, 2019). ResNet50, 50 katmanlı bir artık ağıdır ve çalışmada bu modelin iki versiyonu kullanılmıştır.
- **DenseNet:** Her katmanı diğer tüm katmanlara ileri beslemeli bir şekilde bağlayan Yoğun Konvolüsyon Ağı (DenseNet) olarak bilinmektedir. L katmanlı geleneksel evrimsel ağlar L bağlantılarına sahipken, her katman ve onu takip eden katman arasında  $L*(L+1)/2$  doğrudan bağlantıya sahiptir. Her katman için, önceki tüm katmanların özellik haritaları girdi olarak kullanılmaktadır. Ayrıca kendi özellik haritaları, sonraki tüm katmanlarda girdi olarak kullanılmaktadır. DenseNet'ler yok olan gradyan problemini hafifletirler, özellik yayılımını güçlendirmektedir. Özelliklerin yeniden kullanımını sağlarken, kullanılan parametre sayısını önemli ölçüde azaltmaktadır (Huang vd., 2017). Çalışmada DenseNet121 modeli kullanılmıştır.
- **MobileNet:** MobileNet'ler, hafif, derin sinir ağları oluşturmak için derinlemesine ayrılabilir evrimsel modeller kullanan modern bir mimariye dayanmaktadır. Gecikme ve doğruluk arasında verimli bir şekilde denge

kuran iki basit küresel hiper parametre sunmaktadır. Bu hiper parametreler, model oluşturucunun, problemin kısıtlamalarına dayalı olarak uygulamaları için doğru boyutlu modeli seçmesine olanak tanımaktadır (Howard vd., 2017). Çalışmada MobileNet'in iki versiyonundan yararlanılmıştır.

#### 4. Deneysel Sonuçlar (Experimental Results)

Veri setine ait çalışmalar, yöntemler ve modellere ait analizler ve duygu etiketlerine ait çıkarımlar bu bölümdeki alt başlıklar altında detaylı biçimde anlatılmaktadır. Sınıflandırma ölçütü olarak F1-skor, doğruluk, geri çağırma ve kesinlik değerleri kullanılmıştır. Sınıflandırma doğruluğu, doğru sınıflandırılmış örnek sayısının toplam örnek sayısına oranıdır. Gerçek Pozitifler (True Positive-TP) için gözlem olumludur ve olumlu olduğu tahmin edilirken, gerçek negatifler (True Negative-TN) için gözlem olumsuz olup olumsuz olarak tahmin edilmiştir. Yanlış Pozitifler (False Positive-FP)'de ise gözlem olumsuz olmasına rağmen olumlu tahmin edilirken, yanlış negatifler (False Negative-FN) için gözlem olumlu olup olumsuz tahmin yapılmaktadır (Karcioğlu ve Bulut, 2021). Geri çağırma, kesinlik ve F1-skor değerleri bu değerler aracılığı ile hesaplanmaktadır. Geri çağırma TP'in TP+FN değerine oranıyla elde edilirken; kesinlik ise TP'in TP+FP değerine oranıdır. F1-skor ise geri çağırma ve kesinlik değerinin harmonik ortalamasıdır (Karcioğlu ve Yaşa, 2020).

##### 4.1. Veri Seti Analizi (Analysis of Dataset)

Çalışmada TurEV-DB veri seti için birçok analiz uygulanmıştır. İlk aşamada veri setindeki ses kayıtları librosa kütüphanesi aracılığıyla okunarak özellik çıkarım işlemi gerçekleştirilmiştir. Toplam 175 tane çıkarılan özelliklere ait örnek Tablo 3'te gösterilmiştir. Tabloda görüldüğü üzere veriler sayısal değer içermektedir ve her biri sesin karakteristiklerini yansıtmaktadır.

**Tablo 3.** Librosa kütüphanesi ile çıkarılan bazı özellikler (Some features extracted with the Librosa library)

id	centroid_max	centroid_min	centroid_mean	centroid_std	centroid_kurtosis	centroid_skew	...	tempo
0	4469.312	685.198	2129.420	1007.268	-0.380	0.538	...	215.332
1	3612.809	546.315	1505.730	742.842	0.870	1.117	...	234.907
2	4544.681	1155.373	2848.935	1092.199	-1.436	-0.128	...	151.999
3	5716.467	707.610	2153.877	1366.089	0.395	1.143	...	234.908
4	5372.452	1204.042	3379.245	1208.404	-0.988	-0.202	...	184.570

Veri setine ait tüm özelliklere ait çıkarılan istatistikler ise Tablo 4'te verilmiştir. Her özelliğe ait miktar (count), ortalama (mean), standart sapma (std), minimum (min), maksimum (max) ve yüzdeler (25%, 50% ve 75%) istatistikler elde edilmiştir.

**Tablo 4.** Çıkarılan bazı özelliklerin istatistikleri (Statistics of some extracted features)

	centroid_max	centroid_min	centroid_mean	centroid_std	...	mfcc_d1_9_kurtosis	mfcc_d1_9_skew	tempo
<b>Miktar (count)</b>	1735	1735	1735	1735	...	1735	1735	1735
<b>Ortalama (mean)</b>	5142.721	570.066	2267.097	1318.238	...	-0.480	-0.060	186.601
<b>Standart sapma (std)</b>	1489.205	350.368	719.092	434.259	...	0.717	0.576	45.645
<b>Minimum (min)</b>	1173.855	0	603.804	165.848	...	-1.739	-1.903	73.828
<b>25%</b>	4410.100	373.520	1730.448	1039.286	...	-1.001	-0.446	151.999
<b>50%</b>	5051.705	552.518	2245.288	1315.450	...	-0.608	-0.061	184.570
<b>75%</b>	5838.0165	768.125	2774.925	1585.062	...	-0.098	0.330	215.332
<b>Maksimum (max)</b>	10449.296	2244.712	4459.542	2847.500	...	3.496	1.743	287.109

##### 4.2. Yöntemler ve Modellerin Analizi (Analysis of Methods and Models)

Librosa kütüphanesi aracılığıyla çıkarılan sayısal özellikler ile Türkçe ses kayıtlarına ait duygu analizi gerçekleştirimi için makine öğrenme yöntemleri ve derin sinir ağları eğitilmiştir. Eğitim aşamasında verilerin %80'i, test aşamasında ise %20 kullanılmıştır. Eğitilen bu yöntemlere ait sistemin başarısı Tablo 5'te verilmiştir. Tablo incelendiğinde en başarılı XGB yöntemi olurken, en başarısız DVM yöntemi olmuştur. XGB'nin en başarılı



olma nedeni olarak optimize edilerek geliştirilmiş, ağaç budama yöntemiyle aşırı öğrenmeyi engelleyen bir algoritma olması belirtilebilir. Yöntemler arasında K-NN için k değeri sınıf sayısı olan 4 olarak belirlenirken, DNN yönteminde, 4 tane Dense ve 4 tane aşırı öğrenmeyi engelleyen Dropout katmanı ve softmax aktivasyon fonksiyonu kullanılmıştır.

**Tablo 5.** Çıkarılan özellikler ile yöntemlerin başarısı (%) (Success of methods with extracted features)

Yöntem	Geri Çağırma	F1-skor	Doğruluk
NB	60.23	60.22	60.23
K-NN	49.86	49.57	49.86
RO	83.00	82.97	83.00
DVM	39.48	34.35	39.48
LR	58.50	57.85	58.50
XGB	<b>87.03</b>	<b>87.02</b>	<b>87.03</b>
DNN	82.51	82.62	82.42

Tablo 5 incelendiğinde, Librosa ile çıkarılan tüm sayısal özellikler eğitilen yöntemlerde girdi olarak verilmiştir. Ses analizinde her özelliğin sisteme olumlu etkisi olmayabileceği için BFE, LVF, CS VE PCA özellik azaltım teknikleri kullanılarak, yeni bir özellik verileri elde edilmiştir. Yeni özellikler ile beraber tüm sistem, makine öğrenme yöntemleri ve DNN ile yeniden eğitilerek test edilmiştir. Bu yöntemlere ait elde edilen sonuçlar Tablo 6'da gösterilmiştir. Tablo incelendiğinde LVF özellik çıkarımı en iyi doğruluk ve F1-skor değerini vermiştir, ancak Tablo 5'e göre başarı artış göstermemiştir. Bunun nedeni olarak azaltılan özelliklerin eğitim aşamasında faydalı olduğu, silinmesinin eğitimi olumsuz etkilemesi belirtilebilir. BFE ve CS için özellik sayısı girdi olarak verildiğinden faydalı olan özellikler de silinmesi negatif etki etmiştir. LVF ve PCA için ise eşik değerine göre özellik sayısı belirlenmektedir. PCA'de ise varolan özelliklerden yeni özelliklerin çıkarım işlemi de başarıya olumsuz yansımıştır.

**Tablo 6.** Özellik azaltım teknikleri sonrası yöntemlerin başarısı - sırasıyla F1-skor, doğruluk (%) (Success of methods after feature reduction techniques - respectively F1-score, accuracy (%))

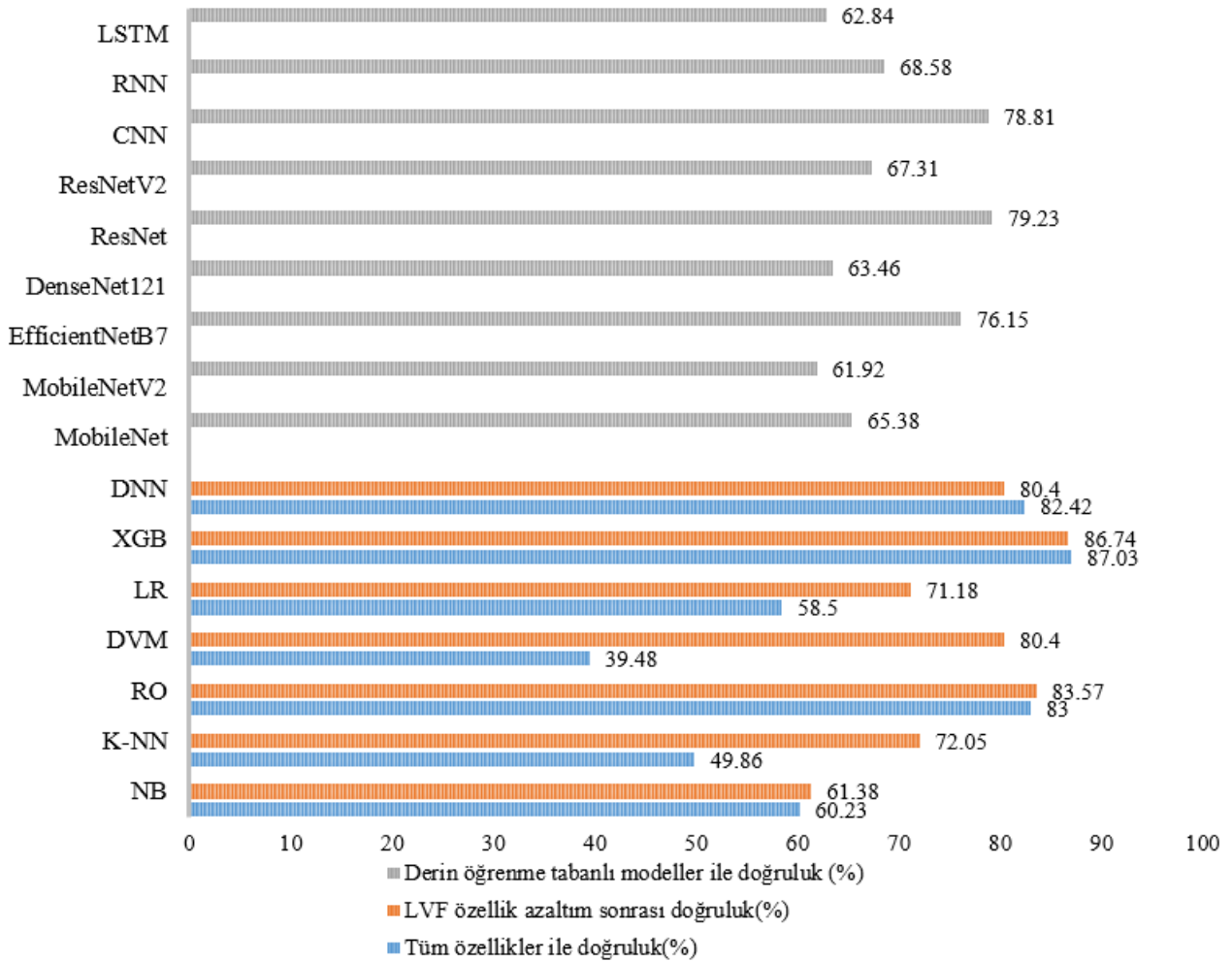
Yöntem	NB	K-NN	RO	DVM	LR	XGB	DNN
BFE (n=50)	62.25,	72.98,	82.66,	79.86,	73.20,	84.49,	81.95,
	61.38	73.20	82.71	79.83	73.20	84.44	81.84
BFE (n=100)	61.16,	69.44,	80.65,	79.60,	77.61,	85.61,	81.38,
	61.80	69.74	80.69	79.54	77.52	85.59	81.56
CS (n=50)	61.64,	71.94,	82.32,	78.22,	71.38,	81.04,	82.11,
	60.81	72.05	82.42	78.10	71.47	80.98	81.84
CS (n=100)	62.05,	76.59,	84.73,	79.83,	74.46,	84.96,	80.88,
	61.96	76.66	84.73	79.83	74.35	85.01	80.98
LVF (n=52)	55.46,	71.20,	85.24,	77.12,	64.16,	86.19,	83.15,
	55.04	71.47	85.30	77.23	64.55	86.17	83.29
LVF (n=133)	61.52,	71.99,	83.55,	80.50,	71.46,	<b>86.78,</b>	80.82,
	61.38	72.05	83.57	80.40	71.18	<b>86.74</b>	80.40
PCA (n=70)	64.72,	70.78,	71.32,	83.93,	70.75,	73.85,	75.64,
	64.27	70.89	71.76	83.86	70.61	73.78	75.50
PCA (n=107)	66.10,	69.29,	72.35,	83.93,	74.14,	71.71,	74.19,
	65.71	69.45	72.91	83.86	74.06	71.76	74.06

Analizin son aşamasında derin öğrenme tabanlı CNN, RNN, LSTM geleneksel modeller ile EfficientNet, ResNet, DenseNet, MobileNet yenilikçi modellerinin spektrogram görselleri ile eğitilmesi ve test edilmesi gerçekleştirilmiştir. Eğitim aşamasında verilerin %80'i, test aşamasında ise %20 kullanılmıştır. Modeller üzerinde Türkçe veri seti ile ince ayar (fine-tuning) aşaması uygulanarak, belirlenen öğrenme oranı (keras kütüphanesi ReduceLRonPlateau metodu ile factor = 0.5, patience = 10, mode = 'min') ve erken durdurma (keras kütüphanesi EarlyStopping metodu ile patience = 20) kriterleriyle eğitilmesi sağlanmıştır. Eğitilen modellere ait başarı Tablo 7'de verilmiştir. Tablodaki verilere göre ResNet modeli Türkçe veri seti ile diğer modellere göre daha iyi öğrenmiş ve daha başarılı bir sonuç vermiştir. DenseNet, MobileNetV2 modellerinin ise eğitimdeki kayıp (loss) değerleri fazla olduğundan ve modeller iyi öğrenme gerçekleştirmediğinden dolayı başarı düşmüştür. Burada modellerin başarısının düşük olma sebebi olarak modeller milyonlarca parametre ile eğitildiğinden, ince ayar aşamasındaki veri setinin yetersiz gelmesi söylenebilir.

**Tablo 7.** Derin öğrenme tabanlı modellerin başarısı (The success of deep learning-based models)

Yöntem	Kayıp	Geri Çağırma (%)	F1-skor (%)	Doğruluk (%)
MobileNet	0.93	65.40	65.37	65.38
MobileNetV2	0.98	61.96	61.38	61.92
EfficientNetB7	0.61	76.27	76.25	76.15
DenseNet121	0.96	63.23	61.00	63.46
ResNet	<b>0.55</b>	<b>79.31</b>	<b>79.29</b>	<b>79.23</b>
ResNetV2	0.76	67.48	65.88	67.31
CNN	0.24	79.00	78.20	78.81
RNN	0.74	69.00	68.40	68.58
LSTM	0.86	63.00	62.30	62.84

Her aşama için en iyi sonuçları veren yöntemlerin ve tüm derin öğrenme tabanlı modellerin doğruluk değerleri Şekil 3'te verilmiştir. Sonuçlar incelendiğinde Librosa kütüphanesi ile çıkarılan tüm özelliklerin kullanımı eğitim-test aşamasında en başarılı doğruluk değerini vermiştir. En başarısız ise derin öğrenme tabanlı MobileNetV2 modeli olmuştur.

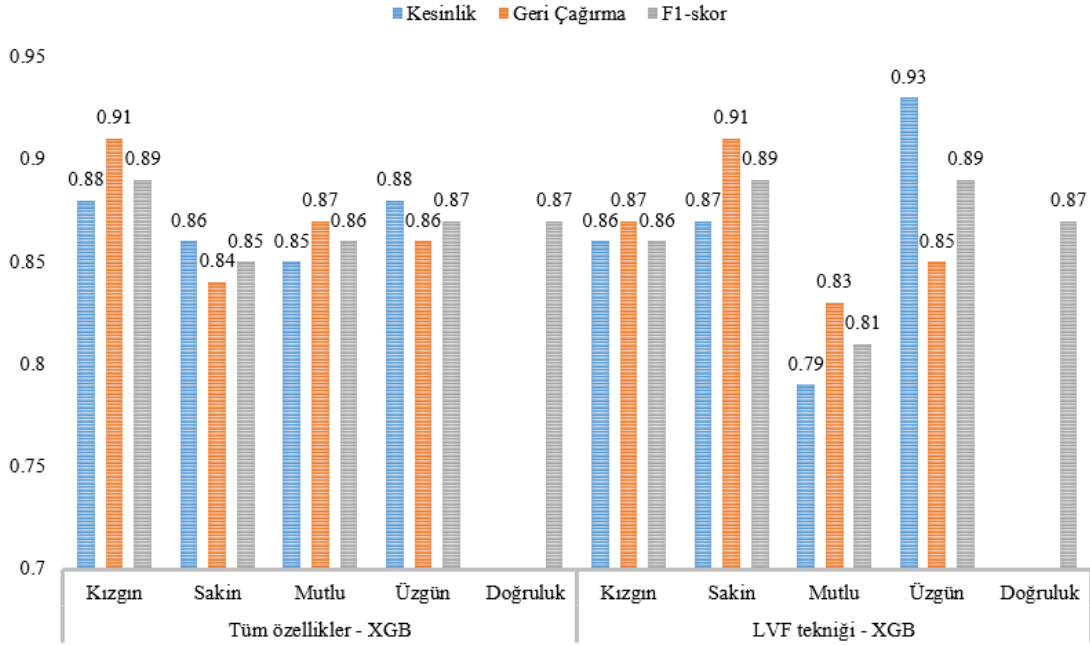


**Şekil 3.** En başarılı yöntemler ve derin öğrenme tabanlı modellerin doğruluk değerleri (Accuracy values of deep learning-based models and the most successful methods)

#### 4.3. Duygu Etiketlerinin Analizi (Analysis of Emotion Labels)

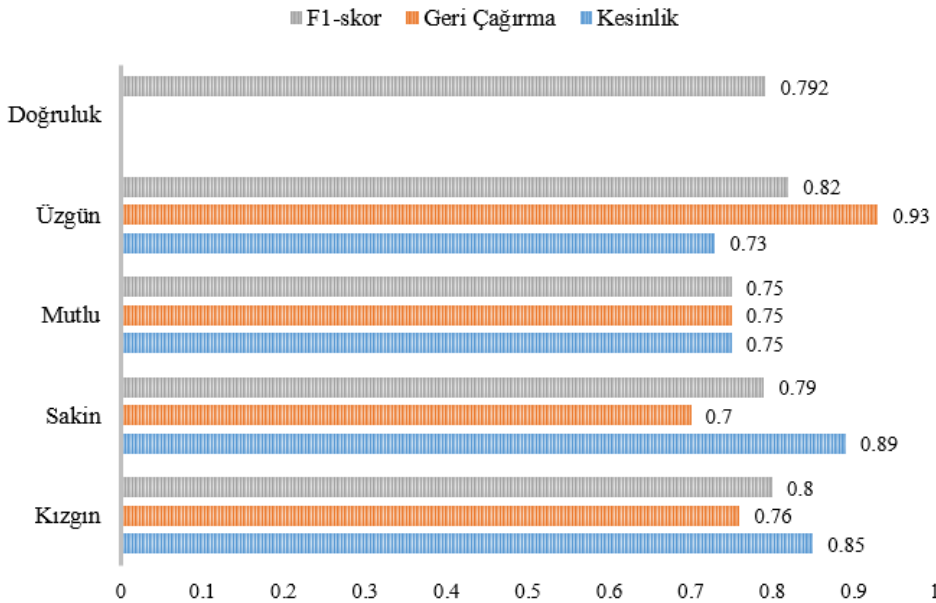
Yöntem ve modellerin başarısına göre hangi duygu etiketlerinin daha başarılı veya başarısız sınıflandırıldığı bu başlık altında analiz edilmiştir. Bu analiz için her aşamada en başarılı yöntemlerin sınıflandırma raporu elde edilmiştir. Tüm çıkarılan özellikler ile en başarılı olan XGB yöntemi ve LVF özellik azaltım teknikleri ile en başarılı yöntem XGB'nin sınıflandırma raporuna ait grafik Şekil 4'te gösterilmiştir. Şekil incelendiğinde, tüm özelliklerin

kullanımı ile XGB yöntemi duygu etiketlerini dengeli bir şekilde tespit etmiştir. Ancak LVF ile özellik azaltım uygulanmasıyla XGB yönteminin duygu etiketleri analiz edildiğinde en zor tespit edilen mutlu etiketine sahip ses kayıtları olmuştur.



Şekil 4. Çıkarılan tüm özelliklerde ve özellik azaltım uygulanmasında en başarılı XGB yöntemi için sınıflandırma raporu (Classification report for the most successful XGB method in all extracted features and feature reduction implementation)

Spektrogramlar ile uygulanan derin öğrenme tabanlı modellerdeki eğitim sonucunda en başarılı olan ResNet modeline ait duygu etiketlerinin sınıflandırma raporu ise Şekil 5'te verilmiştir. Bu modelde ise mutlu duygu etiketine sahip ses kayıtları test aşamasında daha zor tespit edilmiştir.



Şekil 5. ResNet modelinin duygu etiketleri için sınıflandırma raporu (Classification report for emotion labels of the ResNet model)

## 5. Sonuç ve Tartışma (Result and Discussion)

Çalışmada Türkçe dili açısından literatüre katkı sağlamak amacıyla, Türkçe ses kayıtları üzerinde duygunun tespit edilme işlemi gerçekleştirilmiştir. Bunun için makine öğrenme yöntemlerini eğitme aşamasında, Librosa kütüphanesi ile elde edilen sayısal özellikler, derin öğrenme tabanlı modelleri eğitmek için ise spektrogram görsel özelliklerinden yararlanılmıştır. Canpolat vd. (2020) uyguladığı çalışmada sadece spektrogram özellikleri

eğitimde CNN modeli kullanılırken, bu çalışmada ise sayısal karakteristik ve görsel özellikler olmak üzere iki tür özelliklerin kullanılması ile literatüre yenilik kazandırılmıştır. Ayrıca Türkçe için bu özellikler geçmişte çok az sayıda çalışmada karşılaştırılmıştır.

İlk aşamada Librosa ile elde edilen sayısal karakteristik özellikler makine öğrenme yöntemleri ve derin sinir ağları ile eğitilerek yöntemlerin başarısı ölçülmüştür. Bu analiz sonucunda XGB yöntemi ile yaklaşık %87.03'lük bir doğruluk değeri elde edilmiştir. Eğitim aşamasında çok sayıda özellik kullanılmasından dolayı, sonraki aşamada BFE, LVF, CS ve PCA özellik azaltım yöntemleri aracılığıyla azaltılan yeni özelliklerin yöntemlerdeki başarıya etkisi analiz edilmiştir. Analizler sonucunda LVF yöntemi ile özelliklerin azalması ile XGB yöntemi %86.74 doğruluk değeri elde ederken, başarıda yaklaşık %0.3'lük bir düşüş olmuştur. Bu azalışın nedeni olarak, özellik azaltım sonrası başarıya olumlu yönde etki eden özelliklerin silinmesi gösterilebilir.

Ses kayıtlarının frekans spektrumunun görsel olarak gösterildiği spektrogram özellikleri ise derin öğrenme tabanlı modellerin eğitiminde kullanılmıştır. Geleneksel derin öğrenme modelleri incelendiğinde, CNN modeli %78.58 doğruluk değeri ile RNN ve LSTM'den daha iyi sonuç elde etmiştir. Modellerin başarısı makine öğrenme yöntemleri ile karşılaştırıldığında daha düşük doğruluk değerine ulaşmıştır. Bunun nedeni olarak spektrogramların sınıflar için daha belirgin bir örüntü göstermemesi olabilir.

Normalize edilmiş spektrogramların eğitimde kullanılması sonucunda ise yenilikçi modellerden ResNet %79.23'lük en iyi doğruluk değerine ulaşmıştır. Burada modelin makine öğrenme yöntemlerine göre düşük başarı gösterme nedeni, milyonlarca parametre içeren modele, ince ayar yapılan veri setinin yetersiz kalması gösterilebilir. Veri setindeki veri sayısının az olması, modellerin eğitim sürecinde iyi öğrenememesine neden olmuş ve başarısının makine öğrenme yöntemlerine göre daha düşük olmasını sağlamıştır. Modeller arasında en başarısız olan ise MobileNetV2 olmuştur. Makine öğrenme yöntemlerinde ise veri setine bağımlı olarak eğitim gerçekleştirildiğinden, ince-ayar uygulanan derin öğrenme modellerine göre doğruluk oranı daha yüksek elde edilmiştir.

Çalışmanın sonucunda Türkçe ses kayıtları üzerinde duygu tespitinin yapılması hem makine öğrenme yöntemleri hem de derin öğrenme tabanlı modeller ile analiz edilerek karşılaştırılması sağlanmış ve literatüre katkı sağlamıştır. Gelecek çalışmalarda Türkçe müzik veri seti, şive, konuşma veri seti elde edilerek ses analizi üzerinden sınıflandırma yapılması amaçlanmaktadır. Böylece Türkçe dilinin literatürdeki çalışma sayısı artırılacaktır.

### Çıkar Çatışması (Conflict of Interest)

Yazarlar tarafından herhangi bir çıkar çatışması beyan edilmemiştir. No conflict of interest was declared by the authors.

### Kaynaklar (References)

- Ali, L., Zhu, C., Zhou, M., Liu, Y. 2019. Early diagnosis of Parkinson's disease from multiple voice recordings by simultaneous sample and feature selection. *Expert Systems with Applications*, 137, 22-28.
- Altınel, A. B. 2021. Cluds: Combining Labeled and Unlabeled Data With Logistic Regression for Social Media Analysis. *Mühendislik Bilimleri ve Tasarım Dergisi*, 9(4), 1048-1061.
- Alu, D. A. S. C., Zoltan, E., & Stoica, I. C. (2017). Voice based emotion recognition with convolutional neural networks for companion robots. *Science and Technology*, 20(3), 222-240.
- Anand, S., Patra, S. R. 2022. Voice and Text Based Sentiment Analysis Using Natural Language Processing. In *Cognitive Informatics and Soft Computing: Proceeding of CISC 2021*, pp. 517-529. Singapore: Springer Nature Singapore.
- Canpolat, S. F., Ormanoğlu, Z., Zeyrek, D. 2020. Turkish Emotion Voice Database (TurEV-DB). In *Proceedings of the 1st Joint Workshop on Spoken Language Technologies for Under-Resourced Languages (SLTU) and Collaboration and Computing for Under-Resourced Languages (CCURL)*, pp. 368-375.
- Çavuş, E., Sancaktar, İ. 2022. Batarya sağlık durumunun makine öğrenmesi ile kestirimi. *Niğde Ömer Halisdemir Üniversitesi Mühendislik Bilimleri Dergisi*, 11(3), 601-610.
- Çelik, A. 2022. Predicting Diagnosis of Covid-19 Disease With Adaboost and Naive Bayes Machine Learning Algorithms. *Mühendislik Bilimleri ve Tasarım Dergisi*, 10(4), 1212-1221.
- Çevik, K. K., Kayakuş, M. 2020. Bilişim Teknolojileri Departmanında Kullanıcıların Taleplerine Cevap Verme Süresinin Makine Öğrenmesi ile Tahmin Edilmesi. *Mühendislik Bilimleri ve Tasarım Dergisi*, 8(3), 728-739.
- Elbir, A., Aydın, N. 2020. Music genre classification and music recommendation by using deep learning. *Electronics Letters*, 56(12), 627-629.
- Ergenç, İ., Bekar Uzun, İ. P. 2017. *Türkçenin Ses Dizgesi* (1st ed.). Ankara: Seçkin Yayıncılık.
- Filter, L. V., Filter, P. 2014. Seven techniques for dimensionality reduction. Technical report
- Howard, A. G., Zhu, M., Chen, B., Kalenichenko, D., Wang, W., Weyand, T., Andrietto, M., Adam, H. 2017. Mobilenets: Efficient convolutional neural networks for mobile vision applications. arXiv preprint arXiv:1704.04861.

- Huang, G., Liu, Z., Van Der Maaten, L., Weinberger, K. Q. 2017. Densely connected convolutional networks. In Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 4700-4708.
- Karcioğlu, A. A., Bulut, H. 2021. Performance Evaluation of Classification Algorithms Using Hyperparameter Optimization. In 2021 6th International Conference on Computer Science and Engineering (UBMK), pp. 354-358.
- Karcioğlu, A. A., Yaşa, A. C. 2020. Automatic summary extraction in texts using genetic algorithms. In 2020 28th Signal Processing and Communications Applications Conference (SIU), pp. 1-4.
- Kelle, A. C., Yüce, H. 2022. MQTT Trafikinde DoS Saldırılarının Makine Öğrenmesi ile Sınıflandırılması ve Modelin SHAP ile Yorumlanması. *Journal of Materials and Mechatronics: A*, 3(1), 50-62.
- Koren, L., Stipancic, T. 2021. Multimodal emotion analysis based on acoustic and linguistic features of the voice. In International Conference on Human-Computer Interaction, pp. 301-311. Cham: Springer International Publishing.
- Marques, G., Agarwal, D., De la Torre Díez, I. 2020. Automated medical diagnosis of COVID-19 through EfficientNet convolutional neural network. *Applied soft computing*, 96, 106691.
- McFee, B., Raffel, C., Liang, D. P., McVicar, M., Battenberg, E., Nieto, O. 2015. librosa: Audio and music signal analysis in python. In Proceedings of the 14th python in science conference, 8, pp. 18-25.
- Montavon, G., Samek, W., Müller, K. R. 2018. Methods for interpreting and understanding deep neural networks. *Digital signal processing*, 73, 1-15.
- Mulla, G. A., Demir, Y., Hassan, M. 2021. Combination of PCA with SMOTE oversampling for classification of high-dimensional imbalanced data. *Bitlis Eren Üniversitesi Fen Bilimleri Dergisi*, 10(3), 858-869.
- Murwati, A. S., Aldianto, L. 2022. Exploring Voice of Customers to Chatbot for Customer Service with Sentiment Analysis. *The Asian Journal of Technology Management*, 15(2), 141-153.
- Najafian, M., Russell, M. 2020. Automatic accent identification as an analytical tool for accent robust automatic speech recognition. *Speech Communication*, 122, 44-55.
- Oflazoglu, C., Yildirim, S. Recognizing emotion from Turkish speech using acoustic features. *J AUDIO SPEECH MUSIC PROC.* 2013, 26 (2013). <https://doi.org/10.1186/1687-4722-2013-26>
- Özsönmez, D. B., Acarman, T., Parlak, İ. B. 2021. Optimal Classifier Selection in Turkish Speech Emotion Detection. 2021 29th Signal Processing and Communications Applications Conference (SIU), pp. 1-4.
- Pelchat, N., Gelowitz, C. M. 2020. Neural network music genre classification. *Canadian Journal of Electrical and Computer Engineering*, 43(3), 170-173.
- Reddy, A. S. B., Juliet, D. S. 2019. Transfer learning with ResNet-50 for malaria cell-image classification. In 2019 International Conference on Communication and Signal Processing (ICCSP), pp. 0945-0949. IEEE.
- Ren, Z., Jia, J., Guo, Q., Zhang, K., Cai, L. 2014. Acoustics, content and geo-information based sentiment prediction from large-scale networked voice data. In 2014 IEEE International Conference on Multimedia and Expo (ICME), pp. 1-4.
- Rhanoui, M., Mikram, M., Yousfi, S., Barzali, S. 2019. A CNN-BiLSTM model for document-level sentiment analysis. *Machine Learning and Knowledge Extraction*, 1(3), 832-847.
- Sağbaşı, E. A., Korukoğlu, S., BALLI, S. 2022. Mahalanobis uzaklığı tabanlı aykırı değer bulma ve ReliefF öznelik seçimine dayalı bir makine öğrenmesi yaklaşımı ile akıllı telefon verileri üzerinden stres tespiti. *Pamukkale Üniversitesi Mühendislik Bilimleri Dergisi*, 28(2), 336-345.
- Satyanarayana, G., Bhuvana, J., Balamurugan, M. 2020. Sentimental Analysis on voice using AWS Comprehend. In 2020 International Conference on Computer Communication and Informatics (ICCCI), pp. 1-4.
- Sikri, A., Singh, N. P., Dalal, S. 2023. 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.
- Singh, A. K. 2021. Prediction of Voice Sentiment using Machine Learning Technique. In 2021 10th International Conference on System Modeling & Advancement in Research Trends (SMART), pp. 162-166.
- Tan, M., Le, Q. 2019. Efficientnet: Rethinking model scaling for convolutional neural networks. In International conference on machine learning (pp. 6105-6114). PMLR.
- Tracy, J. M., Özkanca, Y., Atkins, D. C., Ghomi, R. H. 2020. Investigating voice as a biomarker: deep phenotyping methods for early detection of Parkinson's disease. *Journal of biomedical informatics*, 104, 103362.
- Wu, Y., Li, S., Li, H. 2019. Automatic pitch accent detection using long short-term memory neural networks. In Proceedings of the 2019 International Symposium on Signal Processing Systems, pp. 41-45.
- Yılmaz, Ü., Kuvat, Ö. Investigating the Effect of Feature Selection Methods on the Success of Overall Equipment Effectiveness Prediction. *Uludağ Üniversitesi Mühendislik Fakültesi Dergisi*, 28(2), 437-452.