All Sciences Proceedings

5th International Conference on Applied Engineering and Natural Sciences

July 10-12, 2023 : Konya, Turkey

All Sciences Proceedings
http://as-proceeding.com/



https://www.icaens.com/

# Geleneksel Makine Öğrenmesi ile Rakam Seslerinin Sınıflandırılması

M. Alptekin Engin[1*], Latif Akçay [2]

[1]*Elektrik Elektronik Mühendisliği, Bayburt Üniversitesi, Türkiye*
[2] *Elektrik Elektronik Mühendisliği, Bayburt Üniversitesi, Türkiye*

[*]*(maengin@bayburt.edu.tr)*

***Özet*** **–** Belirli seslerin konuşmacıdan bağımsız olarak yüksek doğrulukta sınıflandırılması problemi günümüzde popülerliğini korumaktadır. Bu sesler içerisinde olan rakam seslerinin makineler tarafından algılanabilmesi ve başarılı bir şekilde sınıflandırılması ise günlük hayatımızda önemli bir yere sahiptir. Yapılan çalışmada toplam 3000 adet farklı ses verisi içeren ve açık bir veri tabanı olan Free Spoken Digit Dataset (FSDD) kullanılmıştır. Bu veri tabanı, ana dilleri İngilizce ve farklı birçok dil olan kişiler tarafından 0'dan 9'a kadar rakamları İngilizce olarak seslendirmeleri ile oluşturulmuştur. Sınıflandırma işleminde her bir rakama ait ses kayıtları bir sınıf olarak kullanılmıştır. Daha sonra ilgili veri tabanına öznitelik çıkarım işlemi tatbik edilmiştir. Tüm seslere ait öznitelikleri elde etmek için, ses işaretlerinden öznitelik çıkarımı üzerine yapılan mevcut çalışmalarda yaygın olarak kullanılan toplam 12 adet farklı öznitelik çıkarım yöntemi uygulanmıştır. Bahsedilen yöntemler kullanılarak hesaplanan tüm özniteliklerin, sınıflandırma işleminin gerçekleştirilmesi için %90'ı eğitim %10'u ise test aşamasında kullanılmak üzere ikiye bölünmüştür. Bu bölünme işlemi rastgele olarak gerçekleştirilmiştir. Sınıflandırma aşamasında ise farklı popüler makine öğrenmesine dayalı sınıflandırıcıların başarımları karşılaştırılmıştır. Bu karşılaştırmada eğitim ve test verilerinin rastgele seçilmesi neticesinde tüm işlemler defaatle tekrar edilmiş ve sınıflandırma doğruluklarının ortalama değerleri hesaplanmıştır. Sonuç olarak kullanılan sınıflandırma yöntemlerinin içinde en başarılı yöntem olan destek vektör makinesi yöntemine ait sınıflandırma doğruluk değeri %96.7 olarak tespit edilmiştir.

*Anahtar Kelimeler – Sinyal İşleme, Sınıflandırma, Makine Öğrenmesi, Öznitelik Çıkarımı, Rakam Seslerinin Sınıflandırılması*

## I. GİRİŞ

İnsanlık tarihi boyunca konuşmacının ağzından çıkan ses işaretlerinin dinleyicinin kulağı vasıtasıyla dinlenilmesi ile gerçekleşen kişiler arası iletişim, son yıllarda makinelerin de hayatımızdaki vazgeçilmez yeri doğrultusunda, insanlar ile makineler arasında olan bir yapıya kavuşmaktadır. Günümüzde makineler tarafından gerçekleştirilen konuşma tanımlama, doğal dil işleme (NLP) alanında bir teknoloji olarak karşımıza çıkmaktadır [1]. Konuşma seslerinin sınıflandırılması, ses işaretlerinin işlenmesi üzerine olan çalışmalar için önemli bir başlıktır. Bu tanımlanan seslerin içerisinde rakamların yeri ise oldukça mühimdir. Vatandaşlık işlemlerinden bankacılık sektörüne, telefon haberleşmesinden sanal asistanlara ve sesli yanıt sistemlerine kadar [2] hayatımızın birçok aşamasında işlemlerin doğru gerçekleştirilebilmesi için rakamları tek tek sıralamak zorunda kalmaktayız. Makine öğrenmesi aracılığı ile rakam seslerinin sınıflandırılması üzerine geliştirilen sistemlerde, okunuşları birbirlerine benzeyen rakamların var olması ve rakam seslendirme sürelerinin kısa olması gibi üstesinden gelinmesi gereken birçok zorlukla karşılaşılmaktadır [3]. Bu bakımdan rakam seslerinin sınıflandırılmasında yüksek doğruluk ile çalışacak sistemlere ihtiyaç duyulmaktadır. Bununla beraber rakam seslerinin sınıflandırılması üzeri sınırlı sayıda çalışma bulunmaktadır. Geçmiş bir çalışmada 20 ile 50 yaş aralığında olan 33 kişiden rakamların Portekizce ve İngilizce rastgele sırada okunuşlarına ait ses örnekleri toplanmıştır. Toplanan bu örnekler genlik tabanlı bir detektör ile parçalara ayrılmış ve konuşmacıdan bağımsız bir sınıflandırma için veri

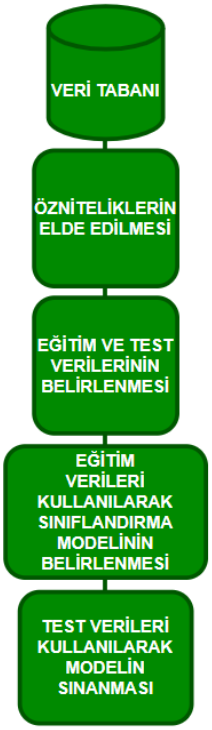
tabanı oluşturulmuştur. En başarılı yöntem olarak çizgisel frekans spektrumu öznitelikleri ile 3. dereceden polinom çekirdeği kullanan destek vektör makinesi sınıflandırıcısı ile İngilizce için %87,27 ve Portekizce için %89,09 doğruluk değeri elde edilmiştir [2]. Diğer bir çalışmada 1230 adet Bengalce rakam seslerinden oluşan bir veri tabanına evrişimsel sinir ağlarına (CNN) dayalı sınıflandırma uygulanmış ve %98,37 doğruluk değeri tespit edilmiştir [4]. Peştuca rakam seslerinin sınıflandırılması üzerine yapılan bir çalışmada ise Mel-frekans kepstrum katsayıları kullanılarak CNN yöntemi ile %84,17 değerinde bir sınıflandırma doğruluk değeri elde edilmiştir [5]. Rakamların İngilizce seslendirilmesi üzerine 2400 ses kaydı içeren bir diğer çalışmada ise rastgele orman sınıflandırıcısı kullanılarak %97,5 değerinde sınıflandırma doğruluğu elde edilmiştir [6]. Yapılan bir diğer çalışmada ise farklı yaş, cinsiyet ve aksana sahip 60 kişinin rakamları İngilizce seslendirmeleri ile elde edilen Audio MNIST veri tabanı [7] kullanılarak yapay sinir ağları aracılığı ile bir sınıflandırma modeli geliştirilmiştir [8]. Modelin test sonuçları için %99,56 sınıflandırma doğruluğu elde edilmiştir. Ayrıca bu çalışmada geliştirilen sınıflandırma modeli, Free Spoken Digit Dataset (FSDD) veri tabanı [9] ile de sınanmıştır. Bu sınama sonucunda yaklaşık %80,60 sınıflandırma doğruluğu elde edilmiştir.

Rakam seslerinin sınıflandırılması üzerine yapılan önceki çalışmalar incelendiğinde, İngilizce söylenen rakam seslerinin sınıflandırılması üzerine fazla çalışmanın bulunmadığı tespit edilmiştir. Ayrıca çalışmalarda, veri setinin eğitim ve test aşaması için bölünmesinde belirli kümelerin kullanıldığı görülmektedir. Kör ve adil bir sınıflandırmanın gerçekleştirilmesi için eğitim ve test kümelerinin bölünmesi işleminin rastgele olarak defaatle gerçekleştirilmesi gerekmektedir. Model performans ölçümü için her seferinde alınan sonuçların ortalamasının verilmesi daha genel geçer sonuçların elde edilmesine vesile olacaktır.

Yapılan bu çalışmada FSDD veri tabanı kullanılmıştır. Sunulan yöntemde ses işaretlerinden öznitelik elde etmede 12 farklı yöntem uygulanmıştır. Sınıflandırma modelinin eğitiminde bu öznitelikler rastgele olacak şekilde eğitim ve test olarak bölünmüştür. Daha sonra hesaplanan bu öznitelikler ile makine öğrenmesine dayalı sınıflandırıcılar vasıtasıyla model geliştirilmiş ve sonuçlar elde edilmiştir. Tüm bu süreç 10 kez gerçekleştirilmiş ve hesaplanan ortalama başarım değerleri literatürdeki diğer çalışmalar ile karşılaştırılmıştır.

## II. MATERYAL VE YÖNTEM

Yapılan çalışmaya ait uygulanan yöntemlerin blok şeması Şekil 1'de gösterilmektedir.




Şekil 1. Kullanılan yöntemin blok şeması

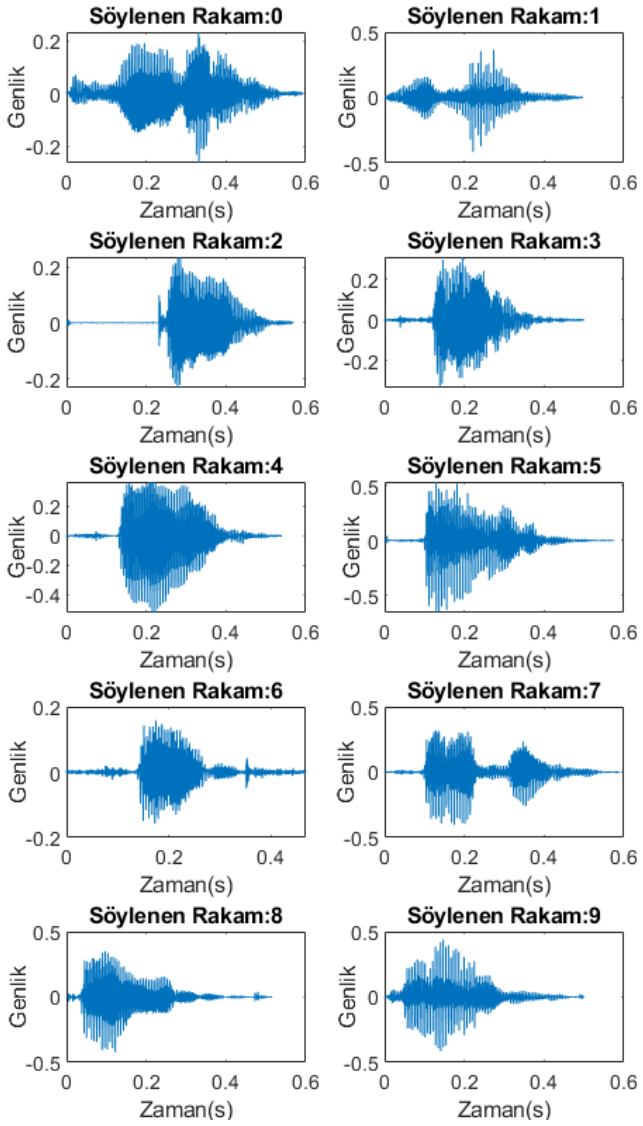
### A. *Veri Tabanı*

Yapılan çalışmada Free Spoken Digit Dataset (FSDD) kullanılmıştır [1]. Bu veri tabanı farklı kişiler tarafından konuşulan rakam ses kayıtlarından oluşan açık ve büyüyen bir veri tabanıdır. İlgili kayıtlar gürültüden en az etkilenecek şekilde alınmıştır. Veri tabanı oluşturulmasında, toplam 6 farklı konuşmacı, 0'dan 9'a kadar olan her rakamı 50 defa söylemektedir. Toplam 3000 adet ses kaydı bulunan bu veri tabanındaki konuşmacılara ait detaylı bilgiler Tablo 1'de verilmiştir.

Tablo 1. Konuşmacılara ait veriler

| İsim | Cinsiyet | Uyruk/Ana Dil | Kayıt Dili |
|---|---|---|---|
| Jackson | Erkek | Amerika/İngilizce | İngilizce |
| Nicolas | Erkek | Belçika/Fransızca | İngilizce |
| Theo | Erkek | Amerika/İngilizce | İngilizce |
| Yweweler | Erkek | Almanya/Almanca | İngilizce |
| George | Erkek | Yunanistan/Yunanca | İngilizce |
| Lucas | Erkek | Almanya/Almanca | İngilizce |

Kullanılan veri tabanındaki rastgele bir konuşmacıya ait ilk kez seslendirdiği rakam ses işaretlerinin zaman ekseninde gösterimi Şekil'2 de verilmektedir.




Şekil 2. Bir konuşmacıya ait örnek veriler

## B. *Özniteliklerin Elde Edilmesi*

Yapılan çalışmada veri tabanında bulunan tüm ses işaretleri için öznitelik çıkarımı üzerine toplam 12 farklı yöntem uygulanmıştır. Bu 12 yöntem sonucunda 35 adet eleman barındıran bir öznitelik vektörü elde edilmiştir. Bu yöntemlerden sıfır geçiş oranı $(f_1)$, enerji $(f_2)$, enerji entropisi $(f_3)$, spektral merkezilik $(f_4)$, bir çerçevenin yayılımı $(f_5)$, spektral entropi $(f_6)$, spektral akı $(f_7)$, spektral devrilme $(f_8)$, yöntemlerine ait matematiksel eşitlikler Tablo 2'de verilmiştir. Bu özniteliklere ek olarak, Mel-frekans kepstrum katsayıları $(f_9 \cdots f_{21})$, harmonik oran ve pencerenin temel frekansı $(f_{22}, f_{23})$ renk vektörü $(f_{24} \cdots f_{35})$ yöntemleri de kullanılmıştır.

Tablo 2. Kullanılan yöntemlere ait eşitlikler

| Yöntem | Öznitelik | Formül | |
|---|---|---|---|
| Sıfır geçiş oranı | $f_1$ | $\frac{1}{2N}\sum_{n=1}^{N}\lvert sgn[x_i(n)] - sgn[x_i(n-1)]\rvert$ | (1) |
| Enerji | $f_2$ | $\sum_{n=1}^{N}[x_i(n)]^2$ | (2) |
| Enerji entropisi | $f_3$ | $-\sum_{j=1}^{M} e_j . log_2(e_j)$ | (3) |
| Spektral merkezi | $f_4$ | $\frac{\sum_{k=1}^{N/2} kX_i(k)}{\sum_{k=1}^{N/2} X_i(k)}$ | (4) |
| Bir çerçevenin yayılımı | $f_5$ | $\sqrt{\frac{\sum_{k=1}^{N/2}(k-C_i)^2 X_i(k)}{\sum_{k=1}^{N/2} X_i(k)}}$ | (5) |
| Spektral entropi | $f_6$ | $-\sum_{f=0}^{L-1} n_f . log_2(n_f)$ | (6) |
| Spektral akı | $f_7$ | $\sum_{k=1}^{\frac{N}{2}}\left(X_i(k) - X_{i-1}(k)\right)^2$ | (7) |
| Spektral devrilme | $f_8$ | $\sum_{k=1}^{m} X_i(k) = 0.85\sum_{k=1}^{N/2} X_i(k)$ | (8) |

## C. *Verilerin Ayrılması*

Yapılan çalışmada veri tabanında bulunan tüm veriler, sınıflandırma modelinin eğitilmesinde ve eğitilen modelin sınanmasında kullanılmak üzere eğitim ve test olarak ikiye bölünmüştür. Bu ayırma işlemi rastgele olacak şekilde tüm verilerin %90'ı eğitim, %10'u ise test aşamasında kullanılması şeklinde gerçekleştirilmiştir. Kör ve adil bir sınıflandırmanın gerçekleştirilebilmesi için eğitim aşamasında kullanılan veriler, modelin test edilmesinde kullanılmamıştır.

## D. *Sınıflandırma Modelinin Belirlenmesi*

Yapılan çalışmada sınıflandırma modelinin belirlenmesinde geleneksel makine öğrenmesi tabanlı literatürde yaygın olarak kullanılan destek vektör makineleri (DVM) [10], k en yakın komşuluk (KNN) [11], karar ağaçları (KA) [12] ve naive bayes (NB) [13] sınıflandırıcıları kullanılmıştır.

Sınıflandırma modeline ait kullanılan yöntemler belirlendikten sonra eğitim verileri aracılığı ile sınıflandırma modeli oluşturulmuştur. Ayrıca

model eğitilirken başarımı artırmak 5 katlı çapraz doğrulama yöntemi de kullanılmıştır.

### *E. Sınıflandırma Modelinin Sınanması*

Sınıflandırma modelinin sınanmasında hata matrisi ve sınıflandırma doğruluğu değerleri hesaplanmıştır. Hata matrisi bir sınıflandırma modelinde gerçek değerler ile sınıflandırıcının tahmin ettiği değerlerin karşılaştırılmasını sağlayan bir tablodur.

Tablo 3. Hata matrisi

| **Hata Matrisi** | | **Gerçek** | |
|---|---|---|---|
| | | **Pozitif** | **Negatif** |
| **Tahmin** | **Pozitif** | Doğru Pozitif (DP) | Yanlış Pozitif (YP) |
| | **Negatif** | Yanlış Negatif (YN) | Doğru Negatif (DN) |

Tablo 3'de gösterilen hata matrisi değerleri kullanılarak bir sınıflandırıcıya ait doğruluk değeri Denklem 9'da verilen eşitlikle hesaplanmaktadır.

$$Doğruluk\ (Accuracy) = \frac{DP + DN}{DP + DN + YP + YN} \qquad (9)$$

Sınıflandırma doğruluğu, modelin doğru olarak tahmin ettiği veri sayısının toplam tahmin sayısına oranı olarak açıklanmaktadır. Bu yöntem, dengeli yani eşit ya da benzer sayıda veri içeren sınıfların oluşturduğu bir veri tabanında isabetli karşılaştırma yapılabilmesi için uygun bir yöntemdir.

Yapılan çalışmada eğitim ve test kümeleri rastgele seçilmiş ve bu seçim 10 defa tekrarlanmıştır. Her seçilen eğitim kümesi ile sınıflandırma modeli geliştirilmiş ve test kümesi ile bu model sınanmıştır. Son olarak modelin sınanması sonucu elde edilen tüm doğruluk değerlerinin ortalaması hesaplanmıştır.

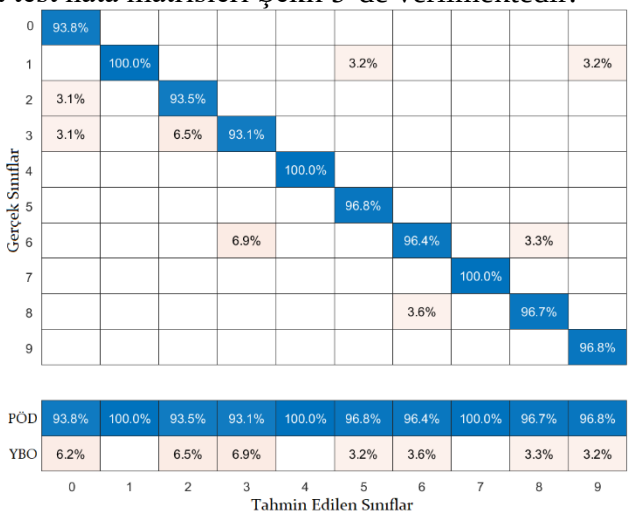
## III. BULGULAR

Yapılan çalışmada Matlab2021b yazılım ortamı kullanılmıştır. Test verileri için kullanılan sınıflandırıcılara ait hesaplanan doğruluk değerleri Tablo 4'de verilmiştir.

Tablo 4. Kullanılan sınıflandırıcılara ait doğruluk sonuçları

| **Yöntem** | **Sınıflandırma Doğruluğu** |
|---|---|
| DVM | %96,7 |
| KNN | %92,3 |
| KA | %73,0 |
| NB | %77,7 |

Tablo 4'de belirtilen yöntemler içerisindeki en başarılı sonuçların elde edildiği DVM yöntemine ait test hata matrisleri Şekil 3'de verilmektedir.




Şekil 3. En başarılı yönteme ait hata matrisi

## IV. TARTIŞMA VE SONUÇLAR

Yapılan çalışmada İngilizce olarak telaffuz edilen rakam seslerinin makine öğrenmesi ile sınıflandırılmasında, kuadratik çekirdek fonksiyon kullanan destek vektör makineleri sınıflandırıcısı ile tüm test verilerinde ortalama %96,7 değerinde bir doğruluk değeri tespit edilmiştir. Şekil 3'de verilen hata matrisine ait pozitif öngörme değeri (PÖD) ve yanlış bulgu oranı (YBO) verileri incelendiğinde yöntemin başarımının %93,5 değeri ile en düşük olduğu sesin üç rakamına ait seslerin bulunduğu sınıfta olduğu görülmektedir. Bununla beraber bir, dört ve yedi rakamlarına ait seslerin sınıflandırılmasının hatasız gerçekleştiği de tespit edilmiştir.

Yapılan çalışmanın sonuçlarının önceki çalışmalar ile sınıflandırma doğruluk değeri üzerinden karşılaştırılması Tablo 5'de verilmektedir.

Tablo 4. Kullanılan sınıflandırıcılara ait doğruluk sonuçları

| **Kaynak** | **Kullanılan Dil** | **Başarım** |
|---|---|---|
| [8] | İngilizce | %80,60 |
| [2] | İngilizce | %87,27 |
| Sunulan | İngilizce | **%96,7** |
| [6] | İngilizce | %97,5 |
| [2] | Portekizce | %89,09 |
| [4] | Bengalce | %98,37 |
| [5] | Peştuca | %84,17 |

Tablo 5 incelendiğinde sadece [8] ile sunulan yöntem aynı veri tabanını kullanmaktadır.

Geleneksel makine öğrenmesi tabanlı sunulan yöntem İngilizce rakam sesleri üzerine yapılan çalışmalar içerisinde yaygın birçok öznitelik elde etme yönteminin bir arada kullanılması sayesinde başarımı yüksek kabul edilen bir konumdadır.

## KAYNAKLAR


[1] D. Jurafsky and J. H. Martin, *Speech and Language Processing*, Pearson Education, Bengaluru, Karnataka, 2020.

[2] D. F. Silva, V. M. A. de Souza, and G. E. A. P. A. Batista, "A comparative study between MFCC and LSF coefficients in automatic recognition of isolated digits pronounced in Portuguese and English - doi: 10.4025/actascitechnol.v35i4.19825," *Acta Sci. Technol.*, vol. 35, no. 4, 2013.

[3] Oruh, J., & Viriri, S. (2022). Deep learning-based classification of spoken English digits. *Computational Intelligence and Neuroscience*, vol. 2022, p. 3364141, 2022,

[4] R. Sharmin, S. K. Rahut, and M. R. Huq, "Bengali spoken digit classification: A deep learning approach using convolutional neural network," *Procedia Comput. Sci.*, vol. 171, pp. 1381–1388, 2020.

[5] B. Zada and R. Ullah, "Pashto isolated digits recognition using deep convolutional neural network," *Heliyon*, vol. 6, no. 2, p. e03372, 2020.

[6] K. M. Srinivas and G. L. P. Ashok, "Spoken English digit classification using supervised learning," *International Journal of Research in Signal Processing*, vol. 5, pp. 49–53, 2019.

[7] S. Becker, M. Ackermann, S. Lapuschkin, K.-R. Müller, and W. Samek, "Interpreting and explaining deep neural networks for classification of audio signals," arXiv [cs.SD], 2018.

[8] A. J. M. Adoptante et al., "Spoken-digit classification using Artificial Neural Network," *ASEAN Engineering Journal*, vol. 13, no. 1, pp. 93–99, 2023.

[9] Z. Jackson, C. Souza, J. Flaks, Y. Pan, H. Nicolas, and A. Thite. Jakobovski/free-spoken-digit-dataset: v1.0.8.Zenodo. https://doi.org/10.5281/zenodo.1342401 2018.

[10] C. Cortes and V. Vapnik,"Support-vector networks". *Mach Learn*, vol. 20, pp. 273–297, 1995.

[11] E. Fix and J. L. Hodges, "Discriminatory Analysis, Nonparametric Discrimination: Consistency Properties". *USAF School of Aviation Medicine*, 41(128). 1951.

[12] J. R. Quinlan, "Induction of decision trees," *Mach. Learn.*, vol. 1, no. 1, pp. 81–106, 1986.

[13] G. I. Webb, E. Keogh, R. Miikkulainen, R. Miikkulainen, and M. Sebag, "Naïve Bayes," in *Encyclopedia of Machine Learning*, Boston, MA: Springer US, 2011, pp. 713–714.