



# Kahramanmaraş Sütçü İmam University

## Journal of Engineering Sciences



Geliş Tarihi : 30.06.2022  
Kabul Tarihi : 11.08.2022

Received Date : 30.06.2022  
Accepted Date : 11.08.2022

### GÖMÜLÜ VE SARICI ÖZNETELİK SEÇİM YÖNTEMLERİNİN KULLANILMASI İLE AKCİĞER RAHATSIZLIKLARININ TESPİTİ

#### DETECTION OF LUNG DISORDERS USING EMBEDDED AND WRAPPER FEATURE SELECTION METHODS

M. Alptekin ENGİN<sup>1\*</sup> (ORCID: 0000-0003-3399-9343)  
Selim ARAS<sup>2</sup> (ORCID: 0000-0003-1231-5782)

<sup>1</sup> Bayburt Üniversitesi, Elektrik Elektronik Mühendisliği Bölümü, Bayburt, Türkiye  
<sup>2</sup> Ondokuz Mayıs Üniversitesi, Elektrik Elektronik Mühendisliği Bölümü, Samsun, Türkiye

\*Sorumlu Yazar / Corresponding Author: M. Alptekin ENGİN, maengin@bayburt.edu.tr

#### ÖZET

Son yıllarda biyomedikal sinyal işleme alanındaki gelişmelere rağmen, akciğer rahatsızlıklarının tespiti üzerine hızlı ve yüksek doğrulukta çalışan teşhis sistemlerine duyulan ihtiyaç artmaktadır. Yapılan çalışmada fiziki muayene ile 94 farklı kişiden, solunum döngülerinin otomatik olarak tespit edilmesiyle elde edilen 150 adet normal ve 444 adet normal olmayan akciğer sesleri veri tabanı olarak kullanılmıştır. Sınıflandırmada öznetelik olarak frekans ve zaman bölgesinde 12 farklı yöntem uygulanmıştır. Elde edilen öznetelikler gömülü ve sarıcı öznetelik seçim yöntemleri kullanılarak değerlendirilmiştir. Bu yöntemler; özyinelemeli öznetelik eliminasyonu, uyarlanabilir yapı öğrenimi ile öznetelik seçimi, bağımlılık kılavuzlu denetimsiz öznetelik seçimi, sıralı yerellik ile denetimsiz öznetelik seçimi, içbükey küçültme yoluyla öznetelik seçimi, en küçük mutlak büzülme ve seçim operatörü öznetelik seçim yöntemleri olarak isimlendirilmektedir. İncelenen bu öznetelikler doğrusal destek vektör makineleri, k en yakın komşuluk, karar ağaçları ve naive Bayes yöntemleri ile sınıflandırılmıştır. Sonuç olarak öznetelik sayısının sınırlandırılmadığı durum için, özyinelemeli öznetelik eliminasyonu yönteminin k en yakın komşuluk sınıflandırma ile beraber kullanıldığı durum için %97,3 doğruluk değerindeki başarıma ulaşılmaktadır. Öznetelik sayısının üç ile sınırlandırıldığı durumda ise uyarlanabilir yapı öğrenimi ile öznetelik seçimi yönteminin karar ağaçları yöntemi ile beraber kullanılması ile %91,4 değerinde başarıma ulaşılmıştır. Sınıflandırmada solunum döngülerinin otomatik tespit edilmesi ve en uygun özneteliklerin seçilmesi aracılığı ile yüksek doğruluk değerine ulaşılmıştır.

**Anahtar Kelimeler:** Solunum sesleri, öznetelik çıkarımı, öznetelik seçimi, sınıflandırma

#### ABSTRACT

Despite the advances in biomedical signal processing in recent years, the need for fast and highly accurate diagnostic systems for the detection of lung disorders continues. In the study, 150 normal and 444 abnormal lung sounds obtained by automatic detection of respiratory cycles from 94 different people by physical examination were used as a database. Then, 12 different feature extraction methods were applied. Features were evaluated using embedded and wrapper selection methods. These methods are recursive feature elimination, adaptive structure learning, dependence-guided unsupervised feature selection, unsupervised feature selection with ordinal locality, feature selection via concave minimization, least absolute shrinkage, and selection operator feature selection methods. Features are classified by linear support vector machines, k nearest neighbor, decision trees, and naive Bayes methods. As a result, when the number of features is not limited, 97.3% accuracy is obtained when the recursive feature elimination is used together with the k nearest neighbor. In the case where the number of features is limited to three, the classification accuracy of 91.4% was achieved using the adaptive structure learning feature selection and the decision trees. High classification accuracy has been achieved through automatic detection of respiratory cycles and selection of the most appropriate features.

**Keywords:** Respiratory sounds, feature extraction, feature selection, classification

## GİRİŞ

Nefes alış veriş sırasında solunan havanın basıncının hızla değişmesi ve akciğer dokularına çarpması sonucunda oluşan akciğer seslerini kullanarak hastalık tespiti uzun yıllara dayanan bir yöntemdir. Fransız Doktor Laennec'in stetoskopu 1816 yılında keşfi, akciğer seslerinden hastalık teşhisi üzerine bir dönüm noktası teşkil etmektedir (Lehrer, 2008). Daha sonra bu aygıt doktorun simgesi olarak nitelendirilmiştir (Bohadana et al.,2014). Fakat akciğer seslerini sinyal işleme yöntemleriyle analiz etmeye olanak sağlayan elektronik stetoskoplar, günümüze kadar fazla bir değişikliğe uğramayan klasik stetoskopun yerini almaya başlamıştır. Tıp fakültesinden yeni mezun olan doktorlar, akciğer seslerinden hastalık teşhisinde yeterli tecrübeye sahip olmamalarından dolayı zorlanmaktadır (Aras et al., 2018). Bu bakımdan elektronik stetoskoplar sayesinde bu seslerin kaydedilip tekrar dinlenmesi ve değerlendirilmesi mümkündür.

Literatürdeki akciğer sesleri üzerine yapılan çalışmalar incelendiğinde, ses sinyallerinin kayıt edilme şartlarına göre çok kanallı ve tek kanallı akciğer seslerinin kullanıldığı görülmektedir. Farklı web kaynaklarından elde edilen 70 adet wheeze ve 42 adet normal akciğer sesi kullanılan mevcut bir çalışmada kısa zamanlı fourier dönüşümü vasıtası ile sınıflandırmada %90,17 değerindeki doğruluğa ulaşılmıştır (Li & Hong, 2015). Başka bir çalışmada ise ral ve ronküs seslerinin sınıflandırılmasında 60 adet kayıt kullanılmıştır (İçer & Gengeç, 2014). Anlık frekans ortalaması, frekans oranı, değişim zamanı ve öz değerleri kullanılan bu çalışmada destek vektör makineleri (DVM) sınıflandırıcısı kullanarak %80 değerinin üzerinde bir doğruluk elde edilmiştir. Himeshima et al. tarafından gerçekleştirilen bir çalışmada ise 63 sağlıklı 63 hasta akciğer sesinin sınıflandırılması üzerine birini dışarda bırak çapraz doğrulama yöntemi kullanılarak %88.7 duyarlılık ve %91,5 özgüllük değerlerine ulaşılmıştır (Himeshima et al., 2012). Mondal et al. tarafından tek kanallı normal ve normal olmayan akciğer seslerinin sınıflandırılması üzerine yapılan çalışmada ise aşırı öğrenme makineleri ve çapraz doğrulama teknikleri beraber kullanılmış ve %86.3 duyarlılık ile %86.9 özgüllük başarımları elde edilmiştir (Mondal et al., 2014). Xie et al. öznelik olarak çok ölçekli temel bileşenler analizini deneysel sınıflandırma yöntemi ile beraber kullanarak normal ve normal olmayan akciğer seslerinin sınıflandırılması için % 98.3 değerinde doğruluk değeri elde etmiştir (Xie et al., 2012). 14 kanallı kayıt yöntemi ile elde edilmiş 21 adet normal olmayan ve 27 adet normal solunum sesleri içeren bir veri tabanı kullanan başka bir çalışmada ise solunum sesi verileri nefes alma ve verme eylemleri için ayrı ayrı incelenmiştir. Nefes alışıta %77.8, nefes verişte %68.9 sınıflandırma doğruluğu değerlerine ulaşılan bu çalışmada autoregressive model ve frekans spektrumu yöntemleri, k en yakın komşuluk sınıflandırıcısı ile kullanılmıştır (Sankur et al.,1994). Palaniappan et al. öznelik olarak autoregressive (AR) model parametreleri ve mel frekans katsayıları (MFCC) kullanarak destek vektör makineleri (DVM) sınıflandırıcısı ile sırası ile %88.7 ve % 89.7 değerindeki doğruluğa ulaşmıştır (Palaniappan et al., 2016). Ayrıca akciğer seslerinin sınıflandırılmasında dalgacık dönüşümü (Koeipensri et al., 2016), doğrusal öngörülü kodlama (Azmy, 2015) gibi yöntemlerin kullanıldığı çalışmalar da bulunmaktadır.

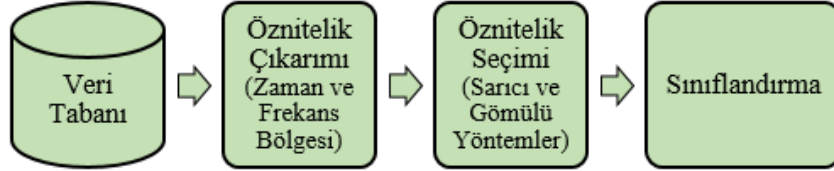
Akciğer sesleri üzerine yapılan çalışmalar incelendiğinde çok kanallı akciğer sesleri daha çok veri barındırmasına rağmen elde edilmiş hastanın kilolu ve kıllı olması durumunda çok sayıdaki mikrofonun göğüs duvarına yerleştirilmesi oldukça meşakkatli bir işlemdir. Bu bakımdan akciğer seslerinin kayıt edilmesinde tek kanallı kayıt yöntemi daha avantajlıdır. Fakat daha az veri barındıran tek kanallı akciğer seslerinden akciğer hastalıklarının otomatik olarak başarılı bir şekilde tespit edilmesi için eldeki verinin en uygun şekilde değerlendirilmesi gerekmektedir. Bu bakımdan literatürde akciğer seslerinin sınıflandırılmasında gömülü ve sarıcı öznelik seçimi üzerine detaylı bir çalışma bulunmamaktadır. Ayrıca literatürdeki mevcut çalışmalar incelendiğinde veri tabanı olarak az sayıda veri üzerine sınıflandırma işlemi yapıldığı görülmektedir. Veri tabanındaki veri sayısını artırabilmek için ise elde edilmiş yöntemleri farklı olan veri tabanlarının birleştirildiği gözlenmektedir. Ayrıca geçmiş çalışmalarda solunum döngüleri genellikle elle belirlenmekte ya da belirli bir zaman tespit edilerek kayıtlar bu zamana göre bölünmektedir. Bu bakımdan aynı kişiye ait solunum döngülerin bile farklı zamanlarda gerçekleşebildiği göz önüne alındığında bu durum olumsuz bir yapıya tekabül etmektedir.

Yapılan çalışmada, yeterli miktarda veri bulunan ve aynı kayıt yöntemiyle elde edilmiş akciğer ses sinyallerinden oluşan bir veri tabanı toplanmış ve kullanılmıştır. Birden fazla solunum döngüsü içeren bu kayıtlar solunum döngülerinin otomatik olarak tespit edilmesinin ardından uzman hekimler tarafından etiketlenmiştir. Elde edilen tüm akciğer solunum döngüsü seslerine 12 adet farklı öznelik çıkarım yöntemi uygulanmıştır. 35 birim uzunluğundaki öznelik vektörüne sarıcı ve gömülü öznelik seçim yöntemlerinden 6 tanesi uygulanmış ve elde

edilen öznelikler 4 farklı sınıflandırma yöntemi kullanılarak hasta ve sağlıklı akciğer seslerinin sınıflandırılması gerçekleştirilmiştir.

## MATERYAL VE YÖNTEM

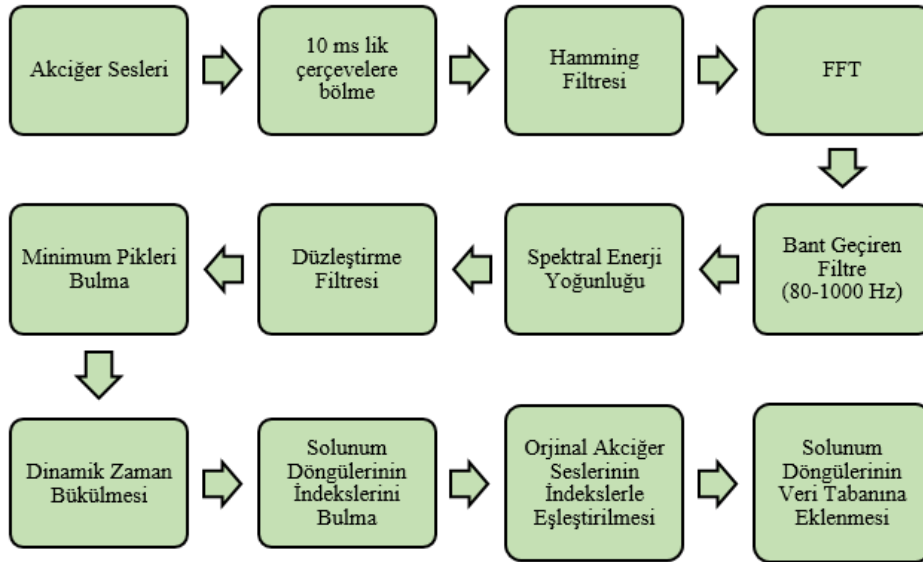
Kullanılan sisteme ait blok diyagram Şekil 1’de gösterilmektedir. Blok diyagram içeriğindeki tüm aşamalar bu bölüm içerisinde ilgili alt başlıklar altında detaylı olarak anlatılmaktadır.



Şekil 1. Yöntemin Blok Diyagramı

### Veri Tabanı

Sınıflandırma sonucunu etkileyen en önemli bileşenlerden biri veri tabanıdır. Akciğer sesleri üzerine yapılan çalışmalarda standart özelliğe sahip bir veri tabanının bulunmayışı ticari bir yazılım ve aygıt geliştirmedeki en önemli eksiklik olarak belirtilmektedir (Gurung et al., 2011). Bu problemin üstesinden gelebilmek için ülkemizde geniş kapsamlı bir çalışma gerçekleştirilmiştir (Aras, 2018). Bu önceki çalışmamızdaki akciğer sesleri üzerine bir veri tabanının oluşturulmasında ilgili etik kurul izin belgesi alınmıştır. Bu veri tabanındaki sesler, Karadeniz Teknik Üniversitesi Tıp Fakültesi Göğüs Hastalıkları Bölümü’nde görev yapan iki uzman hekim tarafından uygun diyafram aralığında (20-2000 Hz.) tek kanallı elektronik stetoskop (Thinklabs ds32a+) kullanılarak kaydedilmiş ve etiket değerleri isimlendirilmiştir. Kayıtlarda birden fazla solunum döngüsü bulunduğu için her bir solunum döngüsünü otomatik olarak ayırmak için (Aras et al., 2018)’de belirtilen yöntem kullanılmıştır. Kullanılan yöntemin akış şeması Şekil 2’de gösterilmektedir.



Şekil 2. Akciğer Seslerindeki Solunum Döngülerinin Otomatik Olarak Bulunması

Şekil 2’de aşamaları belirtilen yöntemin kullanılması ile solunum döngülerinin sınırları yani başlangıç ve bitiş süreleri tespit edilmektedir. Tespit edilen bu sınırlar orijinal akciğer seslerine uygulanarak solunum döngüleri otomatik olarak ayırt edilmiştir. Bu şekilde standart bir veri tabanı oluşturulması amaçlanmaktadır. Tüm sınıflandırma işlemlerinde solunum döngüleri otomatik olarak ayrılan orijinal akciğer seslerine ait veri tabanı kullanılmıştır. Bu yöntem ile 94 kişiden elde edilen akciğer sesleri ile oluşturulan veri tabanı Tablo 1’de detaylı olarak gösterilmektedir.

**Tablo 1.** Veri Setinin Akciğer Ses Tipleri İle Oluşturulması

Kişi Sayısı	Solunum Döngüsü Tipi	Solunum Döngüsü Sayısı
23	Ronküs	150
20	İnce Ral	148
21	Kaba Ral	146
30	Normal	150

Tablo 1’de belirtilen akciğer sesleri incelendiğinde, sağlıklı ve sağlıklı olmayan kişiler üzerine olan bu çalışmada solunum döngüleri normal (150 adet) ve normal olmayan (444 adet) sesler olarak iki kısma ayrılmıştır.

### Özniteliklerin Hesaplanması

Yapılan çalışmada, veri tabanında bulunan ses sinyalleri kullanılarak öznitelik elde etme konusunda literatürdeki mevcut çalışmalarda kullanılan 12 farklı zaman ve frekans bölgesi öznitelik çıkarım yöntemi kullanılmıştır. Bu yöntemler; sıfır geçiş oranı ( $f_1$ ) Denklem 1’de,

$$f_1 = \frac{1}{2N} \sum_{n=1}^N |sgn[x_i(n)] - sgn[x_i(n-1)]| \quad (1)$$

Enerji ( $f_2$ ) Denklem 2’de,

$$f_2 = E(i) = \sum_{n=1}^N [x_i(n)]^2 \quad (2)$$

Enerji entropisi ( $f_3$ ) Denklem 3’de,

$$f_3 = H_i = - \sum_{j=1}^M e_j \cdot \log_2(e_j) \quad (3)$$

Spektral merkezilik ( $f_4$ ) Denklem 4’de,

$$f_4 = C_i = \frac{\sum_{k=1}^{N/2} k X_i(k)}{\sum_{k=1}^{N/2} X_i(k)} \quad (4)$$

Bir çerçevenin yayılımı ( $f_5$ ) Denklem 5’de,

$$f_5 = S_i = \sqrt{\frac{\sum_{k=1}^{N/2} (k - C_i)^2 X_i(k)}{\sum_{k=1}^{N/2} X_i(k)}} \quad (5)$$

Spektral entropi ( $f_6$ ) Denklem 6’da,

$$f_6 = H = - \sum_{f=0}^{L-1} n_f \cdot \log_2(n_f) \quad (6)$$

Spektral akı ( $f_7$ ) Denklem 7’de,

$$f_7 = SF_{(i,i-1)} = \sum_{k=1}^{\frac{N}{2}} (X_i(k) - X_{i-1}(k))^2 \quad (7)$$

Spektral devrilme ( $f_8$ ) Denklem 8’de belirtilmektedir.

$$f_8 = \sum_{k=1}^m X_i(k) = 0.85 \sum_{k=1}^{N/2} X_i(k) \quad (8)$$

Ayrıca yapılan çalışmada MFCC (Volkman et al. 1937) yöntemi kullanılarak elde edilen katsayılar da öznitelik olarak kullanılmıştır ( $f_9 \dots f_{21}$ ). Tüm bu yöntemlere ek olarak, harmonik oran ( $f_{22}$ ) ve pencerenin temel frekansı

( $f_{23}$ ) (Kim et al., 2007) ve renk vektörü (Bartsch & Wakefield, 2005) ( $f_{24} \dots f_{35}$ ) yöntemleri de kullanılarak toplam 35 birim uzunluğunda bir öznitelik vektörü elde edilmiştir.

### Özniteliklerin Seçilmesi

Öznitelik seçimi, makine öğrenmesi, örüntü tanıma ve veri madenciliği konularında verimli ve aktif bir araştırma alanı olarak karşımıza çıkmaktadır (Mitra & Murthy, 2002). Sınıflandırma problemlerinde öznitelik seçiminin temel avantajı, fazlalık içeren gereksiz verileri indirgeyerek vektör boyutunu düşürmek yani depolama alanından tasarruf etmek ve sınıflandırma verimliliğini artırmaktır (Ladha & Deepa, 2011). Bu bakımdan öznitelik seçimi, ilgili sınıfı en iyi tanımlayabilecek ve birbirleri ile ilişkisi bulunmayan en uygun özniteliklerin elde edilmesidir (Forman, 2003). Öznitelik seçimi, süzgeç, sarıcı ve gömülü yöntemler olmak üzere üç temel başlıkta ele alınmaktadır (Guyon, 2006). Süzgeç tipi yöntemlerde sınıflandırıcıdan bağımsız olarak özniteliklerin muhtevassından çıkarılan bilgiler ile analiz gerçekleştirilmektedir. Sarıcı yöntemlerde öznitelik vektöründen elde edilen alt kümelerin sınıflandırıcı kullanarak başarımlarının ölçümü ile en uygun özniteliklerin elde edilmesi sağlanır. Gömülü yöntemlerde ise en başarılı özniteliklerin seçim işlemi ile sınıflandırıcının öğrenme süreci aynı anda gerçekleşmektedir. Ayrıca öznitelik seçim yöntemlerinin incelenmesinde bir diğer husus ise sınıflandırmadaki etiket değerlerini kullanan denetimli ve etiket değerlerini kullanmayan denetimsiz yöntemlerdir.

Yapılan çalışmada sarıcı ve gömülü öznitelik seçim yöntemleri kullanılmıştır. Sınıflandırıcıdan bağımsız olarak hesaplanmaları, daha hızlı çalışmaları gibi özelliklere sahip olmalarına rağmen başarımları sınırlı olan süzgeç yöntemleri bu çalışmada göz ardı edilmiştir. Kullanılan öznitelik seçim yöntemleri, tipi ve sınıfı Tablo 2'de belirtilmektedir. Ayrıca Tablo 2'de verilen yöntemlerin hesaplama karmaşıklığı da belirtilmektedir. Burada T, örnek sayısı, n, seçilen ilk öznitelik sayısı, i, yinelemeli algoritmalarda kullanılan iterasyon sayısı ve C sınıf sayısına karşılık gelmektedir.

**Tablo 2.** Öznitelik Seçimi Üzerine Kullanılan Yöntemler

Yöntem	Tipi	Sınıfı	Hesaplama Karmaşıklığı
RFE	Sarıcı	Denetimli	$O(T^2 n \log_2 n)$
FSASL	Sarıcı	Denetimsiz	$O(n^3 + Tn^2)$
DGUFs	Sarıcı	Denetimsiz	$O(n^3 + Tn)$
UFSOL	Sarıcı	Denetimsiz	$O(iTCn^3)$
FSV	Gömülü	Denetimli	$O(T^2 + n^2)$
LASSO	Gömülü	Denetimli	$O(T^2 + n^2)$

Özyinelemeli Öznitelik Eliminasyonu (Recursive Feature Elimination, RFE) (Guyon et al., 2002) yönteminde öznitelikler ilk olarak belirli bir ölçüt kullanılarak sıralanır. Daha sonra her bir adımda ağırlık vektörüne bağlı olarak sıralı bir şekilde geriye doğru eleme şeklinde en küçük sıralama puanına sahip olan öznitelik elenir. Uyarlanabilir Yapı Öğrenimi ile Öznitelik Seçimi (Feature Selection with Adaptive Structure Learning, FSASL) yöntemi temel olarak doğrusal regresyona dayanmaktadır. Bu yöntemde yapı öğrenme ve öznitelik seçimi aynı anda gerçekleşmektedir (Du & Shen, 2015). Güncel yöntemlerden biri olan Bağımlılık Kılavuzlu Denetimsiz Öznitelik Seçimi (Dependence Guided Unsupervised Feature Selection, DGUFs) yönteminde orijinal veriler, küme etiketleri ve seçilen öznitelikler arasındaki karşılıklı bağımlılık dikkate alınır (Guo & Zhu, 2018). Sıralı Yerellik ile Denetimsiz Öznitelik Seçimi (Unsupervised Feature Selection with Ordinal Locality, UFSOL) görelî komşulukların korunması ve mesafeye dayanan kümeleme tabanlı bir yaklaşımdır (Guo et al., 2017). İçbükey küçültme yoluyla öznitelik seçimi (Feature Selection via Concave Minimization, FSV) yönteminde özniteliklerin seçimi lineer programlama tekniği ile ağırlık minimizasyonuna dayanmaktadır (Bradley & Mangasarian, 1998). En Küçük Mutlak Büzülme ve Seçim Operatörü (Least Absolute Shrinkage and Selection Operator, LASSO) (Tibshirani, 1996) yöntemi, çıkış değeri ile model parametrelerinin arasındaki karesel hatalar toplamının yanında regresyon katsayılarının mutlak değerlerinin toplamını da belirli bir değerden az tutarken, tahmin hatasını minimuma indirmeye çalışır.

### Sınıflandırma Yöntemleri

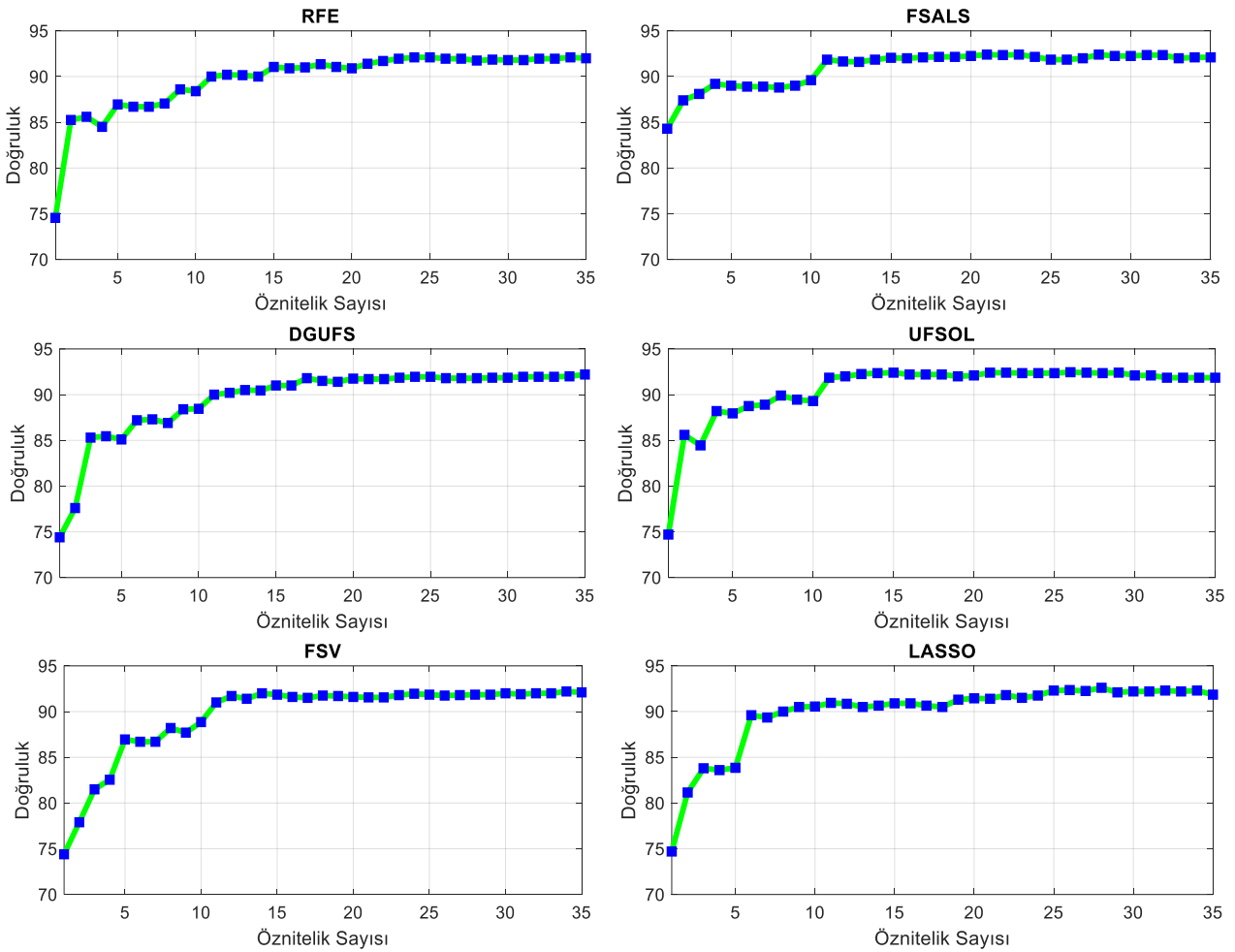
Yapılan çalışmada, öznitelik seçim aşamasında doğrusal destek vektör makineleri (DVM) sınıflandırıcısı kullanılmıştır. Veri setinin doğrusal vektörlerle bölünmesi üzerine olan DVM yöntemi, Cortes ve Vapnik tarafından örüntü tanıma ve sınıflandırma problemlerinin çözümü için geliştirilmiştir (Cortes & Vapnik, 1995). Ayrıca yapılan çalışmada öznitelik seçim aşamasındaki en başarılı sonucun elde edilmesinde kullanılan özniteliklerin, farklı sınıflandırıcılardaki performansları da incelenmiştir. Bu sınıflandırıcılar; k en yakın komşuluk sınıflandırıcısı



(KNN), karar ağaçları (KA) ve naive bayes (NB) sınıflandırıcılarıdır. Uygulanan KNN (Fix & Hodges, 1951) sınıflandırma yönteminde birinci dereceden en yakın komşuluklar Öklid mesafesi kullanılarak ele alınmıştır. Uygulanan KA yönteminde ise CART algoritması kullanılmıştır (Sezer et al., 2010). Uygulanan NB yönteminde ise Gauss tipi çekirdek kullanılmıştır (Berrar, 2018).

## BULGULAR

Yapılan çalışmada, solunum döngülerinin otomatik olarak algılandığı akciğer seslerinden oluşan veri tabanı kullanılarak 12 farklı yöntem ile 35 birim uzunluğundaki öznitelik vektörü toplam 594 adet ses verisi için elde edilmiştir. Veri seti %80 eğitim ve %20 test olmak üzere iki kısma ayrılmıştır. Daha sonra eğitim aşamasında, elde edilen bu özniteliklere 4 adet sarıcı 2 adet gömülü öznitelik seçim algoritması tatbik edilmiştir. Seçim yöntemlerinde doğrusal destek vektör makineleri sınıflandırıcısı kullanılmıştır.



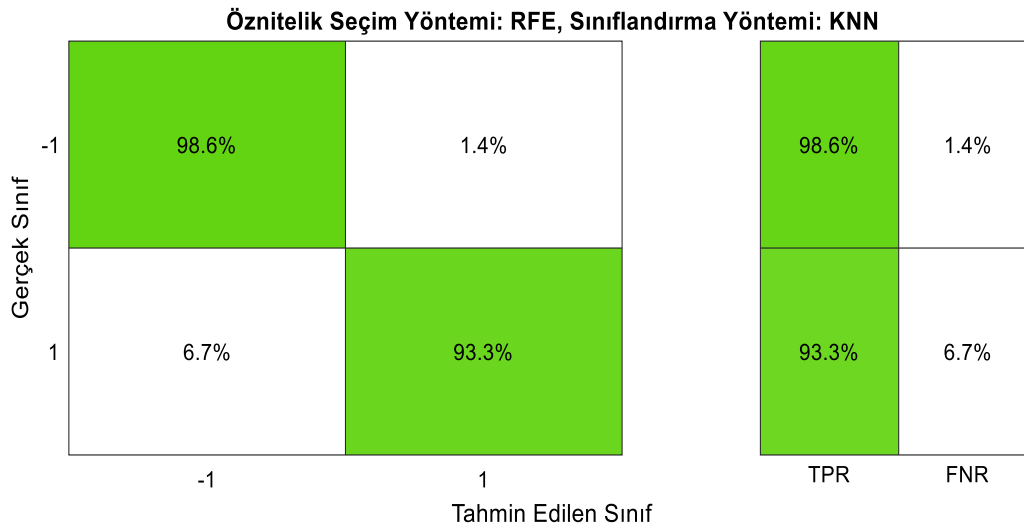
Şekil 3. Kullanılan Yöntemlerin Öznitelik Sayısına Göre Sınıflandırma Doğruluklarının Karşılaştırılması

Ayrıca eğitim aşamasında öznitelik alt kümelerinin öznitelik seçim yöntemleri vasıtasıyla elde edilmesinde 10 kat çapraz doğrulama yöntemi kullanılmıştır. Bu yöntemde eğitim seti 10 eşit parçaya bölünmekte ve bir bölme doğrulama kümesi olmak üzere 10 kez farklı doğrulama kümesi seçilecek şekilde tekrar eden eğitim kümelerine karşılık doğrulama kümesinin başarımının ortalaması ölçülmektedir. Verilen sonuçlar ise eğitim modelinin oluşturulması akabinde eğitimde kullanılmayan test kümesine aittir. Ayrıca test ve eğitim kümelerinin rastgele seçilmelerinden dolayı tüm bu işlemler 10 defa tekrarlanmış ve test işlemi akabindeki doğruluk değerlerinin ortalamaları hesaplanmıştır. Şekil 3'de kullanılan öznitelik seçim yöntemlerinin öznitelik vektörü boyutuna göre sınıflandırma doğruluklarının karşılaştırılması verilmektedir. Tablo 3'de ise, Şekil 3 de verilen yöntemlerin en başarılı sonuçlara ulaştıkları durumun farklı sınıflandırıcılarda incelenmesine ait sınıflandırma doğrulukları gösterilmektedir. Tablo 3'de belirtilen sınıflandırma yöntemlerinde de 10 kat çapraz doğrulama kullanılmıştır. Rastgele seçimin üstesinden gelebilmek için sınıflandırma işlemi de 10 defa tekrarlanmış ve test işlemi sonucu doğruluk değerlerinin ortalamaları verilmiştir. Tablo 3'de verilen en yüksek başarımdaki sonucun hata matrisi Şekil

4'de verilmiştir. Burada normal akciğer sesleri 1 ile etiketlenirken, normal olmayan akciğer sesleri -1 ile etiketlenmiştir. Yeşil renk doğru pozitif oranı (TPR) gösterirken, renksiz kısım ise yanlış pozitif oranını (FNR) göstermektedir. Sunulan yöntem doğrultusunda hata matrisi yorumlandığında normal olmayan akciğer sesleri %98,6 başarıyla tespit edilirken, normal akciğer sesleri %93,3 başarıyla tespit edildiği gözlemlenmektedir.

**Tablo 3.** Öznitelik Seçim Yöntemlerinin Farklı Sınıflandırıcılardaki Başarımları

Yöntem	Öznitelik Boyutu	DVM	KNN	NB	KA
<b>RFE</b>	24	92,1	<b>97,3</b>	88,7	91,8
FSASL	21	92,4	95,3	79,8	90,4
DGUFSS	35	92,2	96,4	85,4	90,4
UFSOL	26	92,5	96,2	84,3	89,9
FSV	34	92,2	96,2	86,2	90,8
LASSO	28	92,6	96,5	84,0	90,9



**Şekil 4.** RFE ve KNN Yöntemlerinin Beraber Kullanılması İle Elde Edilen Hata Matrisi

En başarılı yöntemin sınıflandırmada %97,3 değerinde bir doğruluğa tekabül etmesine karşılık 24 adet özniteliği kullanması pratikte çok ideal bir durum değildir. Bu sebeple kullanılan öznitelik sayısının üç ile sınırlandırılması durumundaki sınıflandırıcı başarımının karşılaştırılması Tablo'4 de verilmiştir. Burada uygulanan sınıflandırma yöntemlerinin zaman yönünden performansları ise Tablo 5'de verilmiştir. Tablo 4 ve Tablo 5'deki verilen değerler aynı işlemin 10 defa tekrarlanması sonucunda elde edilen ortalama değerlerdir.

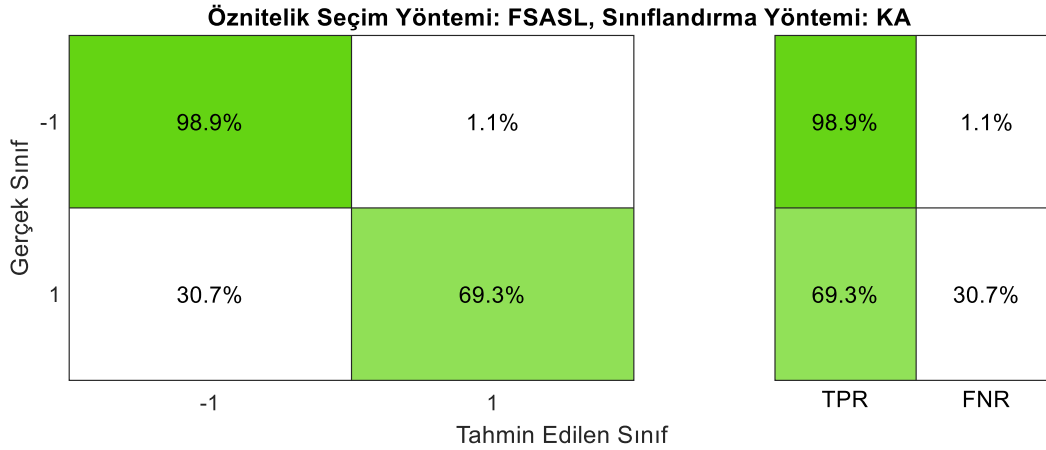
**Tablo 4.** Belirli Sayıda Öznitelik İle Elde Edilen Sınıflandırma Performansı

Yöntem	DVM	KNN	NB	KA
RFE	85,8	89,4	86,9	87,7
<b>FSASL</b>	88,9	88,2	89,4	<b>91,4</b>
DGUFSS	85,5	88,4	87,2	88,2
UFSOL	85,7	90,9	86,9	87,7
FSV	81,5	87,0	80,0	83,5
LASSO	83,8	85,2	86,0	87,8

Ayrıca Şekil 5'de, Tablo 4'de belirtilen en iyi sonuç için hata matrisi verilmektedir. Burada da normal akciğer sesleri 1 ile etiketlenirken, normal olmayan akciğer sesleri -1 ile etiketlenmiştir. Yeşil renk doğru pozitif oranı (TPR) gösterirken, renksiz kısım ise yanlış pozitif oranını (FNR) göstermektedir. Üç adet özniteliğin belirtilen yöntem ile kullanılması durumundaki hata matrisi yorumlandığında normal olmayan akciğer seslerinin %98,9 başarıyla tespit edildiği görülmektedir.

**Tablo 5.** Belirli Sayıda Öznitelik İle Elde Edilen Sınıflandırmadaki Eğitim Süreleri

Yöntem	DVM (sn.)	KNN (sn.)	NB (sn.)	KA (sn.)
RFE	1,123	1,056	1,084	1,179
FSASL	1,171	1,162	1,192	1,189
DGUFSS	1,025	1,016	1,125	1,012
<b>UFSOL</b>	1,006	0,989	<b>0,979</b>	0,981
FSV	1,038	1,064	1,032	1,043
LASSO	1,138	1,089	1,083	1,090

**Şekil 5.** FSASL ve KA Yöntemlerinin Beraber Kullanılması İle Elde Edilen Hata Matrisi

Tablo 6'da en başarılı sonuçlara ait farklı performans metrikleri kullanılarak başarımın ölçülmesine yer verilmiştir.

**Tablo 6.** En Başarılı Yöntemlerin Farklı Metriklerle Sınıflandırma Performansının Ölçülmesi

Yöntem	Duyarlılık	Özgüllük	Kesinlik	F1 Skoru
RFE ve KNN	0,93	0,99	0,99	0,96
FSASL ve KA	0,76	0,98	0,99	0,86

## TARTIŞMA VE SONUÇ

Yapılan çalışmada tek kanal akciğer sesleri kullanılarak normal ve normal olmayan akciğer seslerinin sınıflandırılması için kapsamlı bir öznitelik araştırması gerçekleştirilmiştir. Çok kanallı kayıt yöntemi ile oluşturulan akciğer sesleri daha çok bilgi içermesine rağmen çoklu mikrofonların hastaya bağlanması zor bir işlem olduğundan yapılan çalışmada, tek kanal akciğer sesleri tercih edilmiştir. Buna karşılık tek kanallı kayıt yöntemi ile elde edilen akciğer sesleri muhtevassında daha az veri barındırmaktadır. Bu bakımdan yapılan çalışmada daha az veri ile daha yüksek sınıflandırma başarımına erişebilmek için özniteliklerin verimli bir şekilde kullanılabilmesi amacıyla farklı öznitelik seçim yöntemleri kullanılmıştır.

Tek kanallı akciğer sesleri kayıt işlemleri sırasında birden fazla solunum döngüsüne sahip olabilmektedirler. Bu bakımdan yapılan çalışmada bu akciğer seslerinin solunum döngülerinin otomatik tespiti ve ayrıştırılması da gerçekleştirilmiştir.

Yapılan çalışmada, zaman ve frekans bölgesine ait ve ses sinyalleri üzerine yapılan geçmiş çalışmalarda çokça tercih edilen, 12 farklı öznitelik elde etme yöntemi kullanılmıştır. Bu 12 farklı yöntem ile elde edilen 35 birim uzunluğundaki öznitelik vektörü, gömülü ve sarıcı öznitelik seçim yöntemlerinden 6 tanesi uygulanarak en başarılı alt kümeler tespit edilmiştir. Kullanılan gömülü ve sarıcı öznitelik seçim yöntemlerinin hesaplama karmaşıklığı göz önünde bulundurulduğunda işlem yükü ağır olan yöntemlerdir. Fakat en başarılı sonuçlara ulaşılması ve sınıflandırma modelinin oluşturulması için tek sefer için gerçekleştirilecek bu işlemlerin yoğunlukları kabul edilebilir düzeydedir. Yapılan çalışmada öznitelik sayısı sınırlandırılmadığı durum için RFE öznitelik seçim yönteminin KNN sınıflandırma yöntemi ile beraber kullanıldığı durum için 24 adet öznitelik kullanılması ile %97,3 değerinde sınıflandırma başarımına ulaşılmaktadır. Fakat 24 adet öznitelik hem elde edilmiş hem de sınıflandırma işlem süresi açısından oldukça büyük bir sayıya tekabül etmektedir. Dolayısıyla öznitelik sayısının üç ile sınırlandırıldığı durum için de başarımlar ölçülmüştür.



Öznitelik sayısının 3 ile sınırlandırıldığı durum için FSASL öznitelik seçim yönteminin KA sınıflandırma yöntemi ile beraber kullanılması ile %91,4 değerinde sınıflandırma başarımına ulaşılmıştır. Ayrıca araştırılan bir diğer konu ise sınıflandırma işlemlerinde sınıflandırıcıların eğitim süresinin hesaplanmasıdır. Öznitelik sayısının 3 ile sınırlı tutulduğu durum için kullanılan 4 farklı sınıflandırma yönteminde de yaklaşık aynı eğitim süresine ulaşıldığı gözlemlenmiştir. Bu eğitim süresi ise yaklaşık 1 sn. civarındadır. Ayrıca yapılan tüm çalışmalarda Intel i5-11400H işlemciye sahip bir bilgisayar ve Matlab2021b ortamı kullanılmıştır.

Akciğer sesleri üzerine yapılan çalışmalarının dezavantajlarından biri, bu çalışmalarda kullanılan yaklaşımların standardizasyonunun olmamasıdır. Ayrıca bu alanda yapılan çalışmaları karşılaştırmak için ortak bir uluslararası standart veri tabanı bulunmamaktadır. Bu nedenle, yapılan çalışmanın sonuçlarını diğer çalışmaların sonuçlarıyla birebir adil bir karşılaştırma mümkün değildir. Buna rağmen sunulan yöntem vasıtası ile elde edilen sonuçların farklı kayıt yöntemleri ve farklı akciğer sesleri üzerine yapılan diğer çalışmalar ile karşılaştırılmasına ait sonuçlar Tablo 7’de gösterilmektedir.

**Tablo 7. Sunulan Yöntemin Karşılaştırılması**

Yöntem	Veri	Sınıflandırma	Doğruluk
(Yılmaz & Kahya, 2006)	27 Normal-21 Normal olmayan Çok Kanal Akciğer Sesleri	AR Model ve KNN	Nefes Veriş %77,8 Nefes Alış %68,9
(Şen et al., 2015)	20 Normal-20 Normal olmayan Çok Kanal Akciğer Sesleri	AR Model & Gauss Karışım Modeli ve DVM	%90,0
(Kim et al., 2021)	1222 Normal-696 Normal olmayan Tek Kanal Akciğer Sesleri	Derin Öğrenme	%85,7
Sunulan	150 Normal-444 Normal olmayan Tek Kanal Akciğer Sesleri	RFE ve KNN	%97,3
Sunulan	150 Normal-444 Normal olmayan Tek Kanal Akciğer Sesleri	FSASL ve KA	%91,4

Tablo 7 incelendiğinde Yılmaz ve Kahya tarafından gerçekleştirilen çalışmada solunum döngülerine ait nefes alış ve veriş aşamaları el ile ayrılmıştır. Çok kanal akciğer sesleri için bu işlemin elle gerçekleştirilmesi oldukça zordur. Solunum döngülerinin bu şekilde ayrılması yüzünden sınırlı sayıda veri ile gerçekleştirilen bu çalışmada karşılaştırılan diğer yöntemlere nazaran daha düşük değerlerde sınıflandırma doğruluğuna ulaşılmaktadır (Yılmaz & Kahya, 2006). Şen et al. tarafından gerçekleştirilen çalışmada ise yine çok kanallı akciğer sesleri kullanılmış ve Gauss karışım modeli vasıtasıyla yine küçük çapta bir veri tabanı kullanılarak daha yüksek bir başarıma ulaşılmaktadır (Şen et al., 2015). Çok kanallı akciğer seslerinin sağlıklı ve patolojik ayrımının yapıldığı ve sınırlı sayıda veri kullanılan bu iki çalışma daha çok özniteliklerin çıkarımı üzerinedir. Kim et al., tarafından gerçekleştirilen çalışmada ise yeterli miktarda tek kanal akciğer sesleri kullanılmaktadır. Bu çalışmada veri tabanının oluşturulmasında belirli süre boyunca elde edilen birden fazla solunum döngüsü el ile ayrılarak kullanılmıştır. Yöntem olarak ses sinyallerinin spektrum görüntülerinin evrimsel sinir ağı yapıyla beraber kullanılması sonucunda yüksek başarıma ulaşılabileceği gösterilmiştir. Tarafımızdan sunulan yöntemlerde ise farklı özniteliklerin başarımlarının öznitelik seçimi ile farklı sınıflandırıcılardaki tek kanallı akciğer sesleri üzerine etkisi yeterli veri kullanılarak incelenmektedir. Sonuçlar incelendiğinde, tek kanallı akciğer seslerinin belirtilen öznitelik seçim yöntemleri ile beraber kullanılması durumunda daha yüksek başarımlara ulaşılabileceği görülmektedir. Sunulan yöntem, kullanılan veri tabanındaki solunum döngülerinin otomatik olarak tespit edilmesi yönüyle de diğer çalışmalardan ayrılmaktadır. Hem öznitelik sayısının sınırlandırılmadığı RFE ve KNN yöntemlerinin beraber kullanıldığı durum için, hem de öznitelik sayısının sınırlandırıldığı FSASL ve KA yöntemlerinin beraber kullanıldığı durum için, doğruluk değerleri karşılaştırılan diğer çalışmalardan daha yüksektir.

Akciğer seslerinin solunum döngülerinin otomatik olarak elde edilmesi ile sınıflandırılmasının gerçekleştirildiği bu çalışmadan elde edilen sonuçlar, ilgili alanda tasarlanacak tamamen otomatik olarak çalışan ve sadece doktorların onayı ile hastalık teşhisinde kullanılabilecek bir cihazın tasarımı için temel teşkil etmektedir.

## KAYNAKLAR

- Aras, S. (2018). Tek Kanallı Yaygın Akciğer Seslerinden Solunum Döngülerinin Otomatik Algılanması ve Sınıflandırılması, Doktora Tezi, Karadeniz Teknik Üniversitesi. Fen Bilimleri Enstitüsü. Elektrik-Elektronik Mühendisliği Ana Bilim Dalı, Trabzon 119s.
- Aras, S., Öztürk, M., & Gangal, A. (2018). Automatic detection of the respiratory cycle from recorded, single-channel sounds from lungs. *The Turkish Journal of Electrical Engineering & Computer Sciences*, 26, 11–22.
- Azmy, M. M. (2015). Classification of lung sounds based on linear prediction cepstral coefficients and support vector machine. *IEEE Jordan Conference on Applied Electrical Engineering and Computing Technologies (AEECT)*, 1–5.
- Bartsch, M. A., & Wakefield, G. H. (2005). Audio thumbnailing of popular music using chroma-based representations. *IEEE Transactions on Multimedia*, 7, 96–104.
- Berrar, D. (2018). Bayes' theorem and naive Bayes classifier. *Encyclopedia of Bioinformatics and Computational Biology: ABC of Bioinformatics*. Elsevier Science Publisher.
- Bohadana, A., Izbicki, G., & Kraman, S. S. (2014). Fundamentals of lung auscultation. *The New England Journal of Medicine*, 370, 744-751.
- Bradley, P. S., & Mangasarian, O. L. (1998). Feature selection via concave minimization and support vector machines. In *Machine Learning Proceedings of the Fifteenth International Conference* (pp. 82–90).
- Cortes, C., & Vapnik, V. (1995). Support-vector networks. *Machine Learning*, 20, 273–297.
- Du, L., & Shen, Y. D. (2015). Unsupervised feature selection with adaptive structure learning. In *Proceedings of the 21th ACM SIGKDD international conference on knowledge discovery and data mining* (pp. 209–218).
- Fix, E., & Hodges, J. L. (1951). Discriminatory Analysis, Nonparametric Discrimination: Consistency Properties USAF School of Aviation Medicine, Randolph Field (pp. 1-21). Texas, Tech. Report 4.
- Forman, G. (2003). An extensive empirical study of feature selection metrics for text classification. *Journal of Machine Learning Research*, 3, 1289–1305.
- Guo, J., Quo, Y., Kong, X., & He, R. (2017). Unsupervised feature selection with ordinal locality. In *IEEE International Conference on Multimedia and Expo (ICME)* (pp. 1213–1218).
- Guo, J., & Zhu, W. (2018, April). Dependence guided unsupervised feature selection. In *Proceedings of the AAAI Conference on Artificial Intelligence* (Vol. 32, No. 1).
- Gurung, A., Scrafford, C. G., Tielsch, J. M., Levine, O., & Checkley, W. (2011). Computerized lung sound analysis as diagnostic aid for the detection of abnormal lung sounds: a systematic review and meta-analysis. *Respiratory Medicine*, 105(9), 1396–1403.
- Guyon, I. (2006). *Feature extraction: foundations and applications*. Springer Science & Business Media.
- Guyon, I., Weston, J., & Barnhill, S. (2002). Gene selection for cancer classification using support vector machines. *Machine Learning*, 46, 389–422.
- Himeshima, M., Yamashita, M., Matsunaga, S., & Miyahara, S. (2012, August). Detection of abnormal lung sounds taking into account duration distribution for adventitious sounds. In *2012 Proceedings of the 20th European Signal Processing Conference (EUSIPCO)* (pp. 1821-1825). IEEE.
- İçer, S., & Gençgeç, Ş. (2014). Classification and analysis of non-stationary characteristics of crackle and rhonchus lung adventitious sounds. *Digital Signal Processing*, 28, 18–27.
- Kim, H. G., Moreau, N., & Sikora, T. (2007). *MPEG-7 audio and beyond: Audio content indexing and retrieval*. Nashville, TN: John Wiley & Sons.
- Kim, Y., Hyon, Y., Jung, S. S., Lee, S., Yoo, G., Chung, C., & Ha, T. (2021). Respiratory sound classification for crackles, wheezes, and rhonchi in the clinical field using deep learning. *Scientific Reports*, 11(1), 1-11.
- Koeipensri, T., Boonchoo, P., & Sueaseanak, D. (2016). The development of biosignal processing system (BPS-SWU V1. 0) for learning and research in biomedical engineering. In *9th Biomedical Engineering International Conference (BMEiCON)* (pp. 1–4). Laos.

- Ladha, L., & Deepa, T. (2011). Feature selection methods and algorithms. *International Journal on Computer Science and Engineering*, 3(5), 1787–1797.
- Lehrer, S. (2008). *Understanding lung sounds with audio CD* (3rd ed.). London, England: W B Saunders.
- Li, J., & Hong, Y. (2015, December). Wheeze detection algorithm based on spectrogram analysis. In 2015 8th International Symposium on Computational Intelligence and Design (ISCID) (Vol. 1, pp. 318-322). IEEE.
- Mitra, P., & Murthy, C. A. (2002). Unsupervised feature selection using feature similarity. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 24(3), 301–312.
- Mondal, A., Bhattacharya, P., & Saha, G. (2014). Detection of lungs status using morphological complexities of respiratory sounds. *ScientificWorld Journal*, <https://doi.org/10.1155/2014/182938>
- Palaniappan, R., Sundaraj, K., & Lam, C. K. (2016). Reliable system for respiratory pathology classification from breath sound signals. In IEEE International Conference on System Reliability and Science (ICSRS).
- Sankur, B., Kahya, Y. P., Çağatay, G. E., & Engin, T. (1994). Comparison of AR-based algorithms for respiratory sounds classification. *Computers in Biology and Medicine*, 24(1), 67–76.
- Şen, I., Saraclar, M., & Kahya, Y. P. (2015). A Comparison of DVM and GMM-Based Classifier Configurations for Diagnostic Classification of Pulmonary Sounds. *IEEE Transactions on Biomedical Engineering*, 62(7), 1768–1776.
- Sezer, E. A., Bozkır, A. S., Yağız, S., & Gökçeoğlu, C. (2010). Karar Ağacı Derinliğinin CART Algoritmasında Kestirim Kapasitesine Etkisi: Bir Tünel Açma Makinesinin İlerleme Hızı Üzerinde Uygulama. *Akıllı Sistemlerde Yenilikler ve Uygulamaları Sempozyumu*.
- Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society Series B (Methodological)*, 58, 267–288.
- Volkman, J., Stevens, S. S., & Newman, E. B. (1937). A scale for the measurement of the psychological magnitude pitch. *The Journal of the Acoustical Society of America*, 8, 185–190.
- Xie, S., Jin, F., Krishnan, S., & Sattar, F. (2012). Signal feature extraction by multi-scale PCA and its application to respiratory sound classification. *Medical & biological engineering & computing*, 50(7), 759-768.
- Yilmaz, C. A., & Kahya, Y. P. (2006). Multi-channel classification of respiratory sounds. In 28th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC) (pp. 2864–2867). New York, USA.