



Tek Kanallı Akciğer Seslerinde Süzgeç Tipi Özellik Seçim Yöntemlerini Kullanarak Solunum Patolojisinin Teşhisi

M. Alptekin Engin^{1*}, Latif Akçay², Selim Aras³

^{1*} Bayburt Üniversitesi, Mühendislik Fakültesi, Elektrik Elektronik Mühendisliği, Bayburt, Türkiye, (ORCID: 0000-0003-3399-9343), maengin@bayburt.edu.tr

² Bayburt Üniversitesi, Mühendislik Fakültesi, Elektrik Elektronik Mühendisliği, Bayburt, Türkiye, (ORCID: 0000-0003-2580-2643), lakcay@bayburt.edu.tr

³ Ondokuz Mayıs Üniversitesi, Mühendislik Fakültesi, Elektrik Elektronik Mühendisliği, Samsun, Türkiye, (ORCID: 0000-0003-1231-5782), selim.aras@omu.edu.tr

(2nd International Conference on Applied Engineering and Natural Sciences ICAENS 2022, March 10-13, 2022)

(DOI: 10.31590/ejosat.1082560)

ATIF/REFERENCE: Engin, M. A., Akçay, L., Aras, S. (2022). Diagnosing Respiratory Pathology Using Filter-Type Feature Selection Methods in Single-Channel Lung Sounds. *European Journal of Science and Technology*, (34), 374-380.

Öz

Yapılan çalışmada, tek kanallı yaygın akciğer sesleri kullanılarak patolojik ve sağlıklı denekler üzerinde detaylı bir öznelik analizi gerçekleştirilmiştir. 94 kişiden elde edilen normal, ronküs, ince ral ve kaba ral seslerine ait 594 adet solunum döngüsünün otomatik tespiti ile elde edilen veri tabanı kullanılmıştır. Daha sonra, sıfır geçiş oranı, enerji, enerjinin entropisi, spektral merkezilik, bir çerçevenin yayılımı, spektral entropi, spektral akı, spektral devrilme, Mel-frekans cepstral katsayıları, harmonik oran, pencerenin temel frekansı ve renk vektörü öznelik çıkarma yöntemleri veri tabanına uygulanmıştır. Sonsuz gizli öznelik seçimi, sonsuz öznelik seçimi, özvektör merkeziliği, minimum artıklık, maksimum ilgililik, relief, karşılıklı bilgi, laplace skoru, çoklu küme, fisher, denetimsiz ayırmacı, yerel öğrenmeye dayalı kümeleme, korelasyona dayalı öznelik seçim yöntemleri eğitim aşamasında kullanılmıştır. Sınıflandırma için destek vektör makinesi, k en yakın komşu, naive bayes ve karar ağaçları algoritmaları kullanılmıştır. Sonuç olarak, öznelik sayısı sınırlı olmadığı durumda, k en yakın komşuluk sınıflandırıcısı ve çoklu küme öznelik seçim yöntemi kullanılarak %97,5 sınıflandırma doğruluğu elde edilmiştir. Öznelik sayısı 3 ile sınırlandırıldığında ise k en yakın komşu sınıflandırıcısı ve özvektör merkeziliği veya sonsuz öznelik seçimi yöntemleri kullanılarak %91,6 sınıflandırma doğruluğu elde edilmektedir.

Anahtar Kelimeler: Akciğer sesleri, Solunum Döngüsü, Sınıflandırma, Öznelik Çıkarımı, Öznelik Seçimi.

Diagnosing Respiratory Pathology Using Filter-Type Feature Selection Methods in Single-Channel Lung Sounds

Abstract

In this study, detailed feature analysis was performed on the pathological and healthy subjects via single-channel common lung sounds. The database obtained from the automatic detection of 594 respiratory cycles of normal, rhonchi, fine crackle, and coarse crackle sounds obtained from 94 people was used. Then, zero-crossing rate, energy, the entropy of energy, spectral centroid, the spread of a frame, spectral entropy, spectral flux, spectral roll-off, Mel-frequency cepstral coefficients, harmonic ratio, the fundamental frequency of a window, and chroma vector feature extraction methods have been applied to the database. Infinite latent feature selection, infinite feature selection, eigenvector centrality, minimum redundancy maximum relevance, relief, mutual information, laplacian score, multi-cluster, fisher, unsupervised discriminative, local learning-based clustering, correlation-based feature selection methods are used in the training phase. Support vector machine, k nearest neighbors, naive bayes, and decision trees algorithms were used for classification. As a result, 97.5% classification accuracy was achieved by using the k nearest neighbor classifier and the multi-cluster feature selection method when the number of features is not limited. When the number of features is limited to 3, 91.6% classification accuracy is achieved by using the k nearest neighbor classifier and the eigenvector centrality or infinite feature selection methods.

Keywords: Lung Sounds, Respiratory Cycle, Classification, Feature Extraction, Feature Selection.

* Corresponding Author: maengin@bayburt.edu.tr

1. Giriş

Teknolojik gelişmeler, hastalık teşhisinde gün geçtikçe kullanımı kolay ve yüksek doğrulukta yöntemlerin geliştirilmesine ön ayak olmaktadır. Özellikle sinyal işleme alanındaki gelişmeler ile birçok hastalık erken dönemde teşhis edilebilmekte bu sayede tedavi ve iyileşme süresi de kısalmaktadır. Akciğer seslerini dinleyerek akciğer rahatsızlıklarının tespiti de bu yöntemlerden biridir. Akciğer seslerinin vücudun belirli bölgelerinden dinlenmesi işleme oksültasyon adı verilir. Oksültasyon kadar kolay, hızlı ve girişimsel olmayan (noninvasiv) başka bir tanı kaynağı bulunmamaktadır (Emeksiz vd., 2018). Akciğer seslerinin dinlenilmesi ile hastalık teşhisi ilk uygarlıklardan bu yana hekimin kulağını hastanın göğsüne dayaması ile uygulanan bir yöntem olmasına rağmen 1816'da Fransız bir doktor olan Laennec'in stetoskopu keşfi ile boyut değiştirmiştir (Lehrer, 2018). Günümüze kadar çok az değişime uğrayan stetoskop doktorun damgası olarak isimlendirilmektedir (Bohadana vd., 2014). Fakat akciğer seslerinden hastalık teşhisinde doktorun iyi bir işitme yeteneğine sahip olması, iyi eğitilmiş olması ve yeterli tecrübeye sahip olması gerekmektedir. Tıp fakültesini yeni bitiren doktorların bu konuda yetersiz deneyime sahip olmalarından dolayı zorlandıkları bilinmektedir (Aras vd., 2018). Ancak elektronik stetoskopların kullanılması akciğer seslerinin sınıflandırılmasında dijital ortamdaki faydalanılmasına imkân sağlamıştır. Elektronik stetoskoplar, akciğer seslerinden hastalık teşhisinde standart bir karar verme mekanizmasına, depolayıp tekrar dinlemeye ve gürültü bileşenlerini süzmeye olanak sağlamaktadır.

Literatürde akciğer sesleri üzerine yapılan çalışmalar, ses sinyallerinin kaydedilmesinde tek kanallı veya çok kanallı kayıt yöntemlerinin uygulanmasına dayanmaktadır. Çok kanallı kayıt yöntemleri akciğer sesleri hakkında daha fazla veri barındırmasına rağmen çoklu mikrofonların hastanın vücuduna yerleştirilmesi oldukça karmaşık ve zor bir eylemdir. Hastaların kilolu ve kılı olmaları durumunda bu işlem oldukça fazla zaman almaktadır. Bu yüzden yapılan çalışmada, tek kanallı kayıt yöntemi sonucunda elde edilen akciğer sesleri kullanılmıştır. Çok kanallı akciğer sesleri üzerine yapılan mevcut bir çalışmada 21 adet patolojik ve 27 adet sağlıklı deneklerin solunum sesleri frekans spektrumu ve autoregressive (AR) model parametreleri aracılığıyla analiz edilmiştir. Bu çalışmada 14 kanallı solunum sesi verileri inspirasyon ve ekspirasyon için ayrı ayrı oluşturulmuştur. K en yakın komşuluk (k-NN) sınıflandırma yöntemini kullanan bu çalışmada sırası ile %77.8 ve %68.9 sınıflandırma başarımına ulaşılmıştır (Yılmaz vd., 2006). Diğer bir çalışma ise 5 farklı akciğer sesi üzerine 120 kayıttan oluşan bir veri tabanı üzerinde gerçekleştirilmiştir (Palaniappan vd., 2016). Bu çalışmada da AR Katsayıları ve Mel Frekans Cepstral Katsayıları (MFCC) öznitelikleri ayırt edilmiş solunum döngülerinden elde edilerek kullanılmıştır. Destek vektör makineleri (DVM) sınıflandırıcısı kullanılan bu çalışmada AR katsayılarının kullanılması ile %88.72, MFCC katsayılarının kullanılması ile %89.68 sınıflandırma başarımına ulaşılmıştır. Akciğer seslerinden öznitelik çıkarımı üzerine hızlı fourier dönüşümü (Göğüş vd., 2016), dalgacık dönüşümü (Koeipensri vd., 2016), Doğrusal Öngörülül Kodlama (Sankur vd., 1994) gibi farklı yöntemleri kullanan çalışmalar da mevcuttur.

Literatürdeki akciğer sesleri üzerine yapılan mevcut birçok çalışmanın içerisinde bu çalışmada tamamen otomatik bir sınıflandırma yöntemi amaçlanmıştır. Ayrıca literatürdeki var olan mevcut çalışmalar incelendiğinde detaylı bir öznitelik

araştırmasının yapıldığı bir çalışma bulunmamaktadır. Tek kanallı kayıt yöntemi ile elde edilen akciğer seslerinin, çok kanallı kayıt yöntemi ile elde edilen akciğer seslerinden daha az veriye sahip olması nedeniyle yüksek sınıflandırma başarımı değerlerine ulaşabilmek oldukça zordur. Yapılan çalışmada bu zorluğun üstesinden gelebilmek için detaylı bir öznitelik araştırılmasına gidilmiştir. Ayrıca yapılan çalışmada kaydedilen akciğer seslerine ait solunum döngülerinin otomatik olarak tespitini sağlayan yöntem ile veri tabanı oluşturulmuştur. Bu sayede seslerin elde edilmesinden sınıflandırılmasına yani hastalığın teşhisine kadar tam otomatik bir sistemin tasarımı mümkün olmaktadır.

Çalışmanın geri kalan bölümleri: 2. Bölümde materyal ve metod, 3. Bölümde araştırma sonuçları, 4. Bölümde sonuçlar ve tartışmalar şeklinde organize edilmiştir. Materyal ve metod bölümü, hem çalışmanın çerçevesi hem de veri seti hakkında bilgi sağlamaktadır. Bulgular bölümünde, çalışma veri seti kullanılarak değerlendirilmiştir. Sonuç ve tartışma bölümünde, çalışmanın ana hatları doğrultusunda diğer çalışmalara göre farklılıkları ve performansı yorumlanmıştır.

2. Materyal ve Metod

Yapılan çalışmanın blok şeması Şekil 1'de gösterilmektedir. Veri tabanı oluşturmak için 94 farklı kişiden elde edilen tek kanallı akciğer seslerine ait solunum döngüleri otomatik olarak tespit edilmiştir. Elde edilen 150 adet sağlıklı, 444 adet patolojik solunum döngüsü veri tabanı olarak kullanılmıştır. Daha sonra zaman ve frekans bölgesinde yaygın olarak kullanılan 12 farklı öznitelik çıkarım yöntemi uygulanmasıyla 35 birim uzunluğunda öznitelik vektörü elde edilmiştir. Elde edilen bu özniteliklere 12 adet filtre tipi öznitelik seçim yöntemi kullanılarak 4 farklı sınıflandırma yöntemi uygulanmıştır.



Şekil 1. Sistemin blok şeması

2.1. Veri Tabanı

Veri tabanı sınıflandırma işleminde en önemli materyallerden biridir. Gurung, akciğer seslerinin sınıflandırılmasında standart bir veri tabanının bulunmamasının ticari bir aygıt geliştirmekteki en önemli eksiklik olduğunu vurgulamıştır (Gurung vd., 2011). Yapılan çalışmada kullanılan akciğer sesleri, ülkemizde gerçekleştirilen bir çalışma ile elde edilmiştir. Bu veri tabanının oluşturulmasında standartlar önceden belirlenmiş ve Karadeniz Teknik Üniversitesi (KTÜ) Etik Kurulu'ndan çalışma için izin belgesi alınmıştır. Veri toplama aşamasında, Karadeniz Teknik Üniversitesi Tıp Fakültesi Göğüs Hastalıkları bölümünde bulunan iki uzman hekim vasıtasıyla tüm ses kayıtları alınmış ve elde edilen kayıtların etiketleri belirlenmiştir. Kayıtları elde edilmesinde tek kanallı, analog çıkışlı ve akciğer ses frekans aralığına (20-2000 Hz.) uygun olan diyafram seçeneğinde bir elektronik stetoskop (Thinklabs ds32a +) kullanılmıştır. Kayıtlar birden fazla solunum döngüsü içerdiği için nefes alış ve veriş evresini içeren bir solunum döngüsünün otomatik olarak tespiti için (Aras vd., 2018)'de belirtilen yöntem kullanılmıştır. Bu şekilde 94 farklı denekten elde edilen 150 adet sağlıklı, 444 adet

patolojik solunum döngüsü veri tabanı olarak kullanılmıştır. Kullanılan veri tabanın detayları Tablo 1’de belirtilmiştir.

Tablo 1. Kaydedilen akciğer seslerinden otomatik olarak belirlenen solunum döngüleri

Ses Tipi	Denek Sayısı	Solunum Döngü Sayısı
Normal	30	150
Ronküs	23	150
İnce Ral	20	148
Kaba Ral	21	146
Toplam	94	594

2.2. Özniteliklerin Elde Edilmesi

Yapılan çalışmada, sıfır geçiş oranı (zero crossing rate-ZCR) Denklem 1’de, enerji (energy) Denklem 2’de, enerji entropisi (entropy of energy) Denklem 3’de, spektral merkezi (spectral centroid) Denklem 4’de, bir çerçevenin yayılımı (spread of a frame) Denklem 5’de, spektral entropi (spectral entropy) Denklem 6’da, spektral akı (spectral flux) Denklem 7’de, spektral devrilme (spectral rolloff) Denklem 8’de belirtilmek üzere öznitelik çıkarma yöntemleri kullanılmıştır. Özniteliklerin elde edilmesinde kullanılan formüller Tablo 2’de gösterilmiştir. Ayrıca bu özniteliklere ek olarak, MFCC, harmonik oran (harmonic ratio), pencerenin temel frekansı (Fundamental frequency of a window) ve renk vektörü (chroma vector) öznitelikleri de kullanılmıştır.

Tablo 2. Kullanılan öznitelik çıkarım yöntemleri ve formülleri

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

MFCC insan işitme sistemini modelleyerek kullanan bir yöntemdir. Yapılan çalışmada MFCC yöntemi, solunum

döngülerindeki dengeli akustik özelliklere sahip olduğu düşünülen kısa çerçeveler üzerinde kullanılmıştır. Bu yöntemde Önce FFT uygulanarak frekans bölgesine geçilir daha sonra güç spektrumu bulunur. Bu güç spektrumu, 20 üçgen filtreden oluşan bir Mel filtre bankasından geçirilir. Daha sonra duyarlılığı azaltmak için logaritma hesaplanır ve son aşamada ayırık kosinüs dönüşümü (DCT) uygulanarak MFCC katsayıları hesaplanır. Yapılan çalışmada MFCC yöntemi kullanılarak 13 adet katsayı elde edilmiş ve bu katsayılar öznitelik olarak kullanılmıştır ($f_9 \dots f_{21}$). Harmonik oran ve pencerenin temel frekans değerleri (Kim vd., 2007)’de belirtilen yöntem ile hesaplanmıştır (f_{22}, f_{23}). Armoni ile yakından ilişkili olan renk tabanlı ses özellikleri, müzik verilerinin işlenmesi ve analiz edilmesinde yaygın kullanım alanına sahip bir öznitelik çıkarım yöntemidir. Yapılan çalışmada 12 adet öznitelik (Bartsch vd., 2005)’de belirtilen renk vektörü yöntemi ile elde edilmiştir ($f_{24} \dots f_{35}$). Sonuç olarak hesaplanan tüm öznitelikler Denklem 9 da gösterilmektedir.

$$F = (f_1, f_2, f_3, \dots, f_{35}) \quad (9)$$

2.3. Özniteliklerin Seçimi

Elde edilen özniteliklerin ilgili sınıfın ayırt edici özelliklerini tanımlayabilmesi ve özgün yani birbirleri cinsinden ifade edilmemeleri, sınıflandırma sürecinin en önemli gereksinimlerindedir. Bu bakımdan öznitelik seçimi, veri setinden elde edilen en uygun özniteliklerin belirlenmesi işlemidir (Forman, 2003).

Öznitelik seçiminin avantajları (Ladha vd., 2011) şu şekilde belirtilmektedir;

- Öznitelik vektör boyutunu düşürerek depolama alanını azaltır.
- Gereksiz ve gürültülü verileri kaldırarak veri kalitesini yükselir.
- Veri toplama ve kullanımında kaynak tasarrufu sağlar,
- Veri kümesinin daha basit şekilde görselleştirilmesini sağlayarak anlaşılabilirliği artırır.
- Veri analizinde öğrenme algoritmalarının çalışma süresini kısaltır.
- Sonuç modelinde doğruluk değerini yükseltir.

Öznitelik seçim yöntemleri süzgeç, sarıcı ve gömülü yöntemler olmak üzere üç farklı sınıfta incelenmektedir (Guyon vd., 2006). Gömülü yöntemlerde öznitelik seçim süreci ile sınıflandırıcının öğrenme yöntemi eş zamanlı olarak gerçekleşmektedir. Sarıcı yöntemlerde ise belirli öznitelik alt kümelerinin sınıflandırmadaki doğrulukları hesaplanarak en uygun alt kümenin tespiti gerçekleştirilmektedir. Süzgeç yöntemlerinde ise sınıflandırıcı göz ardı edilerek özniteliklerin iç özellikleri analiz edilir (Şekil 2). Süzgeç tipi yöntemler daha az karmaşık ve hızlı çalışmalarından dolayı büyük veri setlerinin analizinde daha çok tercih edilmektedir (Guyon vd., 2003). Ayrıca öznitelik seçim yöntemleri denetimli ve denetimsiz olarak iki grupta da incelenebilmektedir. Denetimli öznitelik seçim yöntemlerinde ilgili sınıfa ait etiket değerleri kullanılmaktadır. Denetimli öznitelik seçim yöntemlerindeki temel hipotez, iyi bir öznitelik alt kümesinin, ilgili sınıfla yüksek oranda ilişkili (öngörülebilir) ancak birbirleri ile ilişkisiz (öngörülemez) özellikleri barındıran bir küme olmasıdır (Gennari vd.,1989) Denetimsiz öznitelik seçim yöntemlerinde ilgili sınıflara ait etiket

değerleri verilmez. Denetimsiz öznitelik seçim yöntemleri, ilgili özniteliklerin de aynı zamanda tespiti gerektiğinden oldukça zor bir problemdir.

Yapılan çalışmada filtre tipi öznitelik seçim algoritmaları kullanılmıştır. Kullanılan teknikler Tablo 3’de gösterilmektedir. Bu tekniklerden Relief-F (Liu vd., 2007), özniteliklerin kalitesini, değerleri birbirine yakın olan veri örneklerini ne kadar iyi ayırt ettiğine göre tahmin eden iteratif, rastgele ve denetimli bir yaklaşımdır. Bu yöntem gereksiz özellikler arasında ayırım yapmaz ve az sayıda veri kullanımı durumunda düşük performans gösterir. Etkili ve hızlı bir süzgeç yöntemi olan Fisher (Gu vd., 2012) yönteminde, bir özelliğin puanı sınıflar arası ayırım ve sınıf içi varyansın oranı olarak hesaplanır. Burada öznitelikler bağımsız olarak sıralama ile değerlendirilir.



Şekil 2. Süzgeç tipi öznitelik seçim yöntemlerinin genel blok şeması

Tablo 3. Kullanılan öznitelik seçim yöntemleri

Sıra No	Yöntem	Tipi
1	ILFS	Denetimli
2	Inf-FS	Denetimsiz
3	EC-FS	Denetimli
4	mRMR	Denetimli
5	Relief-F	Denetimli
6	MI	Denetimli
7	LS	Denetimsiz
8	MCFS	Denetimsiz
9	Fisher	Denetimli
10	UDFS	Denetimsiz
11	LLCFS	Denetimsiz
12	CFS	Denetimsiz

Yaygın kullanım alanına sahip karşılıklı bilgiye dayanan MI (Mutual Information) (Zaffalon vd., 2002) yönteminde, belirli bir öznitelik değerinin dağılımı ile belirli bir sınıfa üyelik arasındaki karşılıklı bilgi bir seçim kriteri olarak kabul edilir. Özniteliklerin seçimi bu kriter ile sıralanmaları ile gerçekleştirilir. Sonsuz Gizli Özellik Seçimi (Infinite Latent Feature Selection, ILFS) (Roffo vd., 2016), birleşimsel problemi atlayarak tüm olası öznitelik alt kümelerini göz önünde bulundurarak sıralama adımını gerçekleştiren, olasılık tabanlı bir öznitelik seçim yaklaşımıdır. Bu yöntemin en karakteristik özelliği, kaynağını olasılıksal gizli semantik analizden (Probabilistic Latent Semantic Analysis, PLSA) alan bir süreç kullanarak ilgili öznitelikleri modelleme hedefidir. Özvektör Merkeziliği (Eigenvector Centrality EC-FS) yoluyla öznitelik seçimi (Roffo vd., 2016), özniteliklerin düğümler olarak gösterildiği bir yakınlık grafiği üzerinde haritalayan bir yöntemdir. Bu yöntemde özniteliklerin sırası, öz vektör merkeziliği göstergeleri aracılığıyla düğümlerin önemi değerlendirilerek verilir. Denetimsiz öğrenme yöntemlerinden biri olan Sonsuz Öznitelik Seçiminde (Infinite Feature Selection, Inf-FS) (Roffo vd., 2015), her bir öznitelik grafik üzerinde bir düğüm olarak temsil edilmektedir. Bu yöntemde olası tüm öznitelik alt kümeleri grafikte yollar olarak hesaba katılarak her özniteliğe bir

önem puanı verilir. En büyük veya en farklı değerdeki merkezilik puanı, en önemli özniteliği belirtir. Bir başka denetimsiz yöntem ise bir öznitelik önemi yerellik koruma gücüyle değerlendiren Laplacian Skorudur (Laplacian Score, LS) (He vd., 2005). Bu yöntemde yerel bir geometrik yapıyı modellemek için en yakın komşuluk grafiği oluşturulur. Öznitelik sıralaması bu grafiğin yapısına göre gerçekleştirilir. Korelasyon Tabanlı Özellik Seçimi (Correlation based Feature Selection, CFS) (Hall, 1999) yönteminde, özniteliklerin arasındaki iç korelasyon değerleri dikkate alınmaktadır. Bu yaklaşım, en iyi öznitelik alt kümesinin ilgili sınıf ile yüksek oranda ve bu alt kümeyi oluşturan özniteliklerin birbirleri ile düşük oranda korelasyona sahip olması hipotezine dayanmaktadır. Minimum Artıklık Maksimum İlgililik (Min Redundancy Max Relevance, mRMR) (Ding vd., 1989), yönteminde ilgili sınıfla en ilgili öznitelikler seçilmeye çalışılırken bu öznitelikler arasındaki fazlalığı en aza indirmeye çalışılmaktadır. Denetimsiz Ayırıcı Öznitelik Seçimi (Unsupervised Discriminative Feature Selection, UDFS) (Yang vd., 2011) yönteminde, diğer denetimsiz öznitelik seçim algoritmalarından farklı olarak, toplu olarak tüm öznitelik kümesinden en ayırt edici öznitelik alt kümesi seçilmektedir. Bu yöntemde dağılım matrislerinde bulunan ayırt edici bilgileri ve özellik korelasyonlarını aynı anda kullanarak seçim işlemi gerçekleştirilir. Çoklu Küme Öznitelik Seçimi (Multi-Cluster Feature Selection, MCFS) (Cai vd., 2010) yönteminde, ilk olarak farklı öznitelikler arasındaki korelasyonu tespit edebilmek için spektral analiz uygulanır. Daha sonra Laplacian grafiğinin en üst öz vektörleri kullanılarak özniteliklerin önemini tespit edebilmek için L1-norm düzenlemesine sahip bir regresyon modeli kullanılır ve bir öznitelik puanı hesaplanır. Yöntem bu şekilde en küçük kareler probleminin çözümü ile elde edilen katsayıların en yüksek mutlak değerlerine dayalı olarak öznitelikleri seçer. Yerel Öğrenmeye Dayalı Kümeleme (Local Learning Based Clustering Feature Selection, LLCFS) (Zeng vd., 2011) yönteminde, her bir öznitelik bir ağırlık ile ilişkilendirilir. Daha sonra her bir öznitelik uygunluğunu dikkate almak için yerel öğrenme tabanlı kümeleme algoritmasının yerleşik düzenlemesi kullanılır. Buna göre, kümeleme sürecindeki ağırlıklar iteratif olarak tahmin edilir. Bu yakınsamadan sonra çıktı, ayrıklaştırılmış bir küme gösterge matrisi ile birlikte elde edilir.

2.4. Kullanılan Sınıflandırma Yöntemleri

Yapılan çalışmada sınıflandırma işlemlerinde Karar Ağaçları (KA), Destek Vektör Makineleri (DVM), k En Yakın Komşuluk (k-NN) ve Naive Bayes (NB) sınıflandırma yöntemleri kullanılmıştır. KA yöntemi sınıflandırma sürecinde oldukça yaygın olarak kullanılan güvenilirliği yüksek, düşük maliyetli, yorumu ve tatbiki kolay bir yöntemdir (Ozkaya vd., 2020). Bu yöntemde veri ilişkileri ağaç diyagramları şeklinde gösterilmektedir. Eğitim sürecinde önceden bilinen veriler analiz edilir ve sınıflandırma kuralları belirlenerek model oluşturulur. Test sürecinde ise kuralların doğrulukları tespit edilir. Doğruluğun kabul edilebilir düzeyde olması ile kurallar yeni verilerin sınıflandırılmasında kullanılır. Yapılan çalışmada Gini indeksinden yararlanan CART karar ağacı algoritması kullanılmıştır (Sezer vd., 2010). Bir diğer sınıflandırma yöntemi olan lineer DVM sınıflandırıcısında, ikili veri kümesinin doğrusal vektörlerle ayrılması söz konusudur. Ayrıca bu sınıflandırma yönteminde çekirdek fonksiyonları kullanarak sınıflandırma başarımının artırılması da mümkündür (Meltek vd., 2020). Yapılan çalışmada doğrusal destek vektör makineleri sınıflandırıcısı kullanılmıştır. K-NN sınıflandırma yöntemi ise, bir veriye en yakın komşuların sayısı belirlenerek verinin komşulara

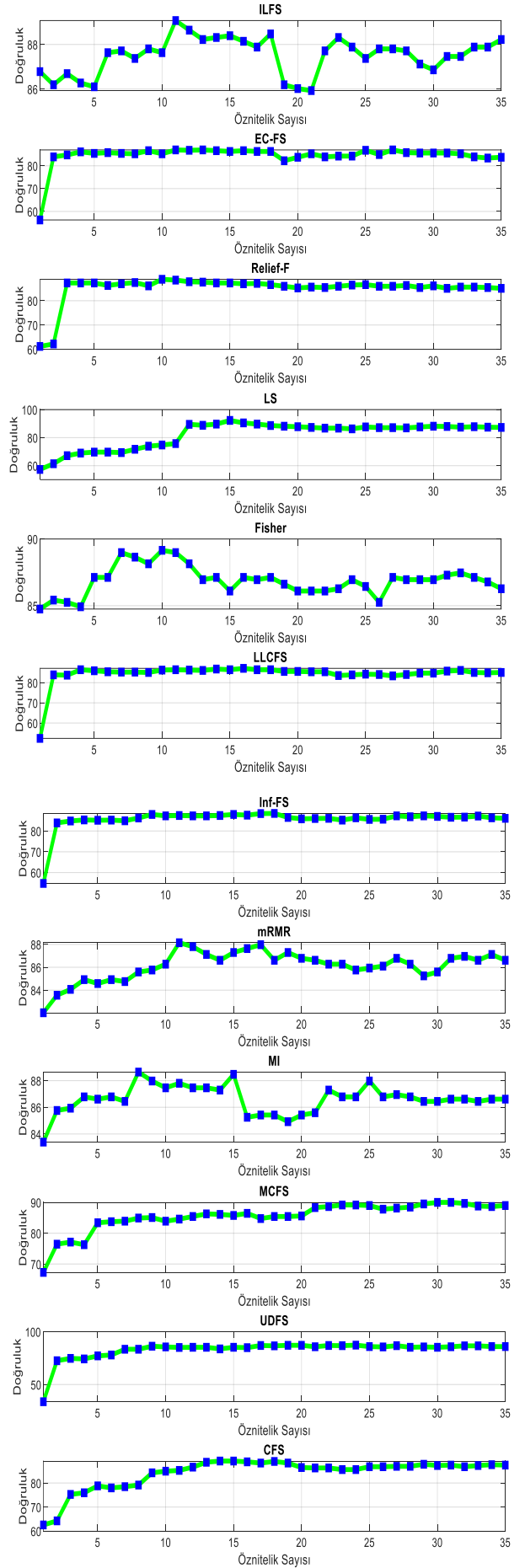
olan uzaklığı ile yeni verinin sınıfının belirlenmesine dayanır (Fix vd., 1951). Yapılan çalışmada komşuluk mesafeleri Öklid mesafesi kullanılarak hesaplanmış ve birinci dereceden en yakın komşuluklar incelenmiştir. Naive Bayes sınıflandırma yönteminde ise, aynı kıstasların sınıflandırma sonucuna olan etkilerinin olasılıksal olarak tespiti kullanılmaktadır (Berrar vd., 2018).

3. Araştırma Sonuçları

Yapılan çalışmada akciğer seslerinden otomatik olarak solumun döngülerinin tespit edilmesiyle oluşturulmuş veri tabanı kullanılarak 12 farklı öznelik çıkarım yöntemi ile öznelikler elde edilmiştir. Daha sonra elde edilen veri tabanı %80 eğitim ve %20 test aşamasında kullanılmak üzere rastgele olacak şekilde ikiye bölünmüştür. Eğitim aşamasında, elde edilen öznelik vektörüne 12 adet süzgeç tipi öznelik seçim yöntemi uygulanmıştır. Bu öznelik seçim yöntemlerinin uygulanması ile öznelikler önem sırasına göre yeniden sıralanmıştır. Daha sonra bulunan bu öznelikler öznelik sayısına göre test aşamasında doğrusal destek vektör makineleri sınıflandırıcısı ile sınıflandırılarak doğruluk değerleri bulunmuştur (Şekil 3). Sınıflandırma işleminde normal akciğer sesleri 1, patojenik akciğer sesleri is -1 olarak etiketlenmiştir. Ayrıca eğitim aşamasında sınıflandırıcı doğruluğu 5 katlı çapraz doğrulama (cross validation) yöntemi ile ele alınmıştır. Bu yöntemde veri seti rastgele olacak şekilde beş adet kümeye bölünmektedir. Bu sayede eğitim verilerine karşı doğrulama verisinin performansı her yeni doğrulama kümesine karşı ölçülmektedir. Eğitim ve test veri kümeleri rastgele oluşturulduğundan her bir sınıflandırma işleminde farklı kümelerin seçilmesi için tüm bu işlemler on defa tekrarlanmış ve sonuçların doğruluk değerlerinin ortalaması hesaplanmıştır. Daha sonra doğruluk değerleri en yüksek olan öznelik sayıları her yöntem için tespit edilmiştir (Tablo 4). Ayrıca Tablo 4’de belirtilen öznelik seçim yöntemlerinin işlem sürelerinin yanında DVM sınıflandırıcısındaki en yüksek sınıflandırma başarımına sahip oldukları öznelikler farklı sınıflandırıcılarda da incelenmiştir. Bu sınıflandırma işleminde sınıflandırıcı doğruluğu yine 5 katlı çapraz doğrulama yöntemi ile ele alınmıştır. Eğitim ve test veri kümeleri rastgele oluşturulduğundan diğer sınıflandırıcıların kullanıldığı işlemler de on defa tekrarlanmış ve her bir sınıflandırıcı için elde edilen doğruluk değerlerinin ortalaması verilmiştir.

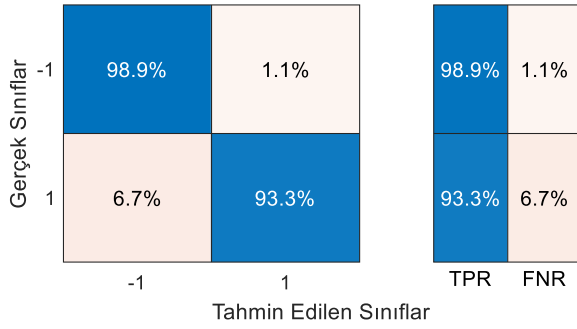
Tablo 4. Kullanılan öznelik seçim yöntemlerinin farklı sınıflandırıcılardaki başarımı

Yöntem	Süre (ms)	Öznelik Sayısı	DVM	KNN	NB	KA
ILFS	55,61	11	89,1	94,6	84,5	89,9
Inf-FS	15,25	18	88,5	92,4	83,0	89,4
EC-FS	101,67	11	87,0	94,6	84,0	89,9
mRMR	30,43	11	88,1	93,9	85,5	92,1
Relief-F	195,51	10	88,8	96,0	88,4	90,2
MI	33,79	8	88,6	93,4	85,0	89,6
LS	44,26	16	90,5	93,6	79,1	91,4
MCFS	113,56	30	90,0	97,5	86,0	89,2
Fisher	75,48	10	89,2	93,7	83,0	90,6
UDFS	63,68	24	87,3	97,0	85,9	90,4
LLCFS	850,18	16	87,1	93,6	78,7	90,8
CFS	11,16	14	89,2	95,6	87,4	90,6



Şekil 3. Öznelik seçim yöntemlerine göre kullanılan öznelik sayısının sınıflandırma başarımına etkisi

Tablo 4’de belirtilen başarıyı en yüksek sonuca ait hata matrisi (confusion matrix) Şekil 4’de gösterilmektedir. Hata matrisindeki normal akciğer sesleri 1 ile belirtilirken, patojenik akciğer sesleri -1 olarak gösterilmektedir.



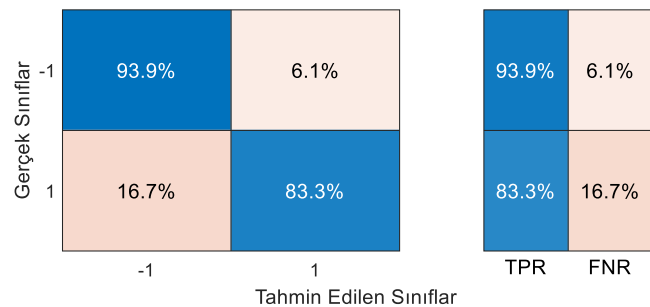
Şekil 4. öznitelik sayısının sınırlanmadığı durumdaki en başarılı sonuca ait hata matrisi

Yapılan çalışmada incelenen bir diğer husus ise, her bir öznitelik seçim yöntemi tarafından tespit edilen ilk üç öznitelige göre sınıflandırma başarılarının, farklı sınıflandırıcılarda incelenmesidir. (Tablo 5).

Tablo 5. Sınırlı öznitelik kullanılması ile elde edilen sınıflandırma başarıları

Yöntem	Öznitelik Sayısı	DVM	K-NN	NB	KA
ILFS	3	86,7	90,1	88,6	89,4
Inf-FS	3	87,8	91,6	89,9	91,2
EC-FS	3	87,8	91,6	89,9	91,2
mRMR	3	84,5	90,6	87,6	89,4
Relief-F	3	87,3	88,6	89,2	90,6
MI	3	85,9	90,0	90,3	91,4
LS	3	71,1	72,2	73,8	76,0
MCFS	3	78,1	85,2	83,0	88,5
Fisher	3	85,3	90,4	88,6	91,2
UDFS	3	84,8	87,4	79,6	84,2
LLCFS	3	85,8	86,8	86,3	89,6
CFS	3	78,3	79,0	81,0	84,7

Tablo 5’de elde edilen sonuçlarda da çapraz doğrulama yöntemi kullanılmıştır. Eğitim ve test veri kümeleri rastgele oluşturulduğundan tüm işlemler on defa tekrarlanmış ve ortalama değerler kullanılmıştır. Tablo 5’de belirtilen en başarılı iki yöntemde de aynı öznitelikler tespit edilmiştir. Bu yöntemlere ait hata matrisi ise Şekil 5 de gösterilmektedir.



Şekil 5. öznitelik sayısının sınırlanıldığı durumdaki en başarılı sonuca ait hata matrisi

4. Sonuçlar ve Tartışma

Geniş ölçekte öznitelik incelemesi gerçekleştirilen bu çalışmada, tek kanal akciğer sesleri kullanılmıştır. Bu bakımdan zaman ve frekans bölgesinde olmak üzere 12 farklı öznitelik çıkarım yöntemi ile elde edilen 35 tane öznitelige 12 adet farklı süzgeç tipi öznitelik seçim yöntemi uygulanmıştır. Sonuçlar süre bakımından incelendiğinde en hızlı çalışan seçim yönteminin 11,16 ms ile CFS yöntemi olduğu tespit edilmiştir. Bu yöntem, 14 adet öznitelik kullanılması durumunda k-NN sınıflandırıcısında %95,6 başarıya ulaşılmasını sağlamıştır. Buna karşılık en yavaş çalışan yöntem ise 850,18ms ile LLCFS yöntemidir. Bu yöntemde ise en yüksek başarıyı, 16 adet öznitelik ile yine k-NN sınıflandırıcısı kullanılarak %93,6 değerindedir. Süreye ve öznitelik sayısına bakılmaksızın en yüksek başarıyı, MCFS yöntemi ile sıralanmış 30 adet öznitelik k-NN sınıflandırıcısı kullanılması ile %97,5 değerinde elde edilmiştir. Başarımın yüksek olmasına rağmen 30 adet öznitelik kullanılması ve her öznitelik farklı bir yöntem ile elde edilmesi akciğer seslerinden hastalık tespit edilmesi üzerine tasarlanacak bir cihaz için çok uygun değildir. Bu bakımdan yapılan çalışmada öznitelik sayısının 3 ile sınırlanması ile en yüksek başarıya sahip yöntemlerin araştırılmasına gidilmiştir. Bu amaçla yapılan çalışmada EC-FS ve Inf-FS yöntemlerinden elde edilen ilk üç öznitelik aynı olduğu ve bu özniteliklerin k-NN sınıflandırıcısı kullanılması ile %91,6 değerindeki başarıya ulaşılmasını sağladığı tespit edilmiştir. Fakat bu iki öznitelik seçim yönteminin çalışma süreleri bakımından incelendiğinde EC-FS yönteminin 101,67 ms içerisinde, Inf-FS yönteminin ise, 15,25ms içerisinde öznitelikleri seçtiği tespit edilmiştir. Bu yüzden Inf-FS yönteminin öznitelik sayısının sınırlanıldığı durum için daha uygun olduğu tespit edilmiştir. Ayrıca genel olarak sınıflandırıcılar incelendiğinde k-NN yönteminin sonuçların çoğunda daha yüksek başarı gösterdiği tespit edilmiştir. Yapılan çalışmanın literatürdeki aynı konudaki diğer çalışmalar ile karşılaştırılması ise Tablo 6’da gösterilmektedir.

Tablo 6. Sunulan yöntemin diğer çalışmalar ile karşılaştırılması

Çalışma	Veri Seti	Sınıflandırma Yöntemi	Doğruluk
(Yılmaz vd., 2006)	27 Normal 21 patolojik Çok Kanallı Akciğer Sesleri	Autoregressive (AR) Model KNN	Nefes Verme %77,8 Nefes Alma %68,9
(Şen vd., 2015)	20 Normal 20 patolojik Çok Kanallı Akciğer Sesleri	Autoregressive (AR) Model Gauss Karışım Modeli (GKM) DVM	%90,0
(Kim vd., 2021)	1222 Normal 696 patolojik Tek Kanal Akciğer Sesleri	Derin Öğrenmeye Dayalı Algoritma	%85,7
Sunulan	150 Normal 444 patolojik Tek Kanal Akciğer Sesleri	MCFS KNN	%97,5
Sunulan	150 Normal 444 patolojik Tek Kanal Akciğer Sesleri	Inf-FS KNN	%91,6

Tablo 6’da gösterildiği üzere tek kanal kayıt yöntemi kullanılarak elde edilen akciğer seslerinin sınıflandırılmasında uygun öznitelik seçim yöntemlerinin kullanılması ile çok kanallı

kayıt yöntemleri ile elde edilen akciğer seslerinin sınıflandırılmasında elde edilen sınıflandırma başarımlarının değerlerine ulaşılabileceği görülmektedir. Ayrıca sunulan yöntemlerin, karşılaştırılan çalışmalara göre daha yüksek başarımlara sahip olduğu görülmektedir. Sunulan yöntemin tamamen otomatik olması ve akciğer seslerine ait solunum döngülerinin teker teker tespit edilmesi gibi avantajlarının bulunması ilgili alanda tasarlanacak bir cihaz için umut vadetmektedir.

Kaynakça

Aras, S., Öztürk, M., & Gangal, A. (2018). Automatic detection of the respiratory cycle from recorded, single-channel sounds from lungs, " *Turk. Turk. J. Of Electr. Eng. Comput. SCI*, 26, 11–22.

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*. Amsterdam, The Netherlands: Elsevier Science Publisher.

Bohadana, A., Izbicki, G., & Kraman, S. S. (2014). Fundamentals of lung auscultation, " *N. N. Engl. J. Med.*, 370(21).

Cai, D., Zhang, C., & He, X. (2010). Unsupervised feature selection for multi-cluster data. *Proceedings of the 16th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining - KDD '10*. New York, USA: ACM Press.

Ding, C., & Peng, H. C. (1989). Minimum Redundancy Feature Selection from Microarray Gene Expression Data. In *Proc. Second IEEE Computational Systems Bioinformatics Conf* (pp. 523–528). Boston: Kluwer Academic Publishers.

Emeksiz, Z., & Bostancı, İ. (2018). Güncel Pediatri, c. 16, s. 3, ss. *Akciğer: Sesimi Duyan Var Mı?*, 79–84.

Fix, E., & Hodges, J. L. (1951). Discriminatory analysis, nonparametric discrimination: consistency properties. *USAF School of Aviation Medicine*, 4.

Forman, G. (2003). An Extensive Empirical Study of Feature Selection Metrics for Text Classification. *Journal of Machine Learning Research*, 3, 1289–1305.

Gennari, J. H., Langley, P., & Fisher, D. (1989). Models of incremental concept formation, " *Artif. Artif. Intell.*, 40(1–3), 11–61.

Göğüş, F. Z., Karlık, B., & Harman, G. (2016). Identification of pulmonary disorders by using different spectral analysis methods. *International Journal of Computational Intelligence Systems*, 9(4), 595. doi:10.1080/18756891.2016.1204110

Gu, Q., Li, Z., & Han, J. (2012). Generalized Fisher score for feature selection. Retrieved from <http://arxiv.org/abs/1202.3725>

Gurung, A., Scraftford, C. G., Tielsch, J. M., Levine, O. S., & 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. doi:10.1016/j.rmed.2011.05.007

Guyon, I., & Elisseeff, A. (2003). An Introduction to Variable and Feature Selection. *JMLR*, 3, 1157–1182.

Guyon, I., Gunn, S., Nikravesh, M., & Zadeh, L. A. (2006). *Feature Extraction: Foundations and Applications*. Berlin, Germany: Springer.

Hall, M. A. (1999). *Correlation-based Feature Selection for Machine Learning*. Hamilton.

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. (2021). Respiratory sound classification for crackles, wheezes, and rhonchi in the clinical field using deep learning". *Sci Rep*, 11.

Koeipensri, T., Boonchoo, P., & Sueaseenak, 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), Laung Prabang (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. (2018). *Understanding lung sounds: Third edition*. North Charleston, SC: Createspace Independent Publishing Platform.

Liu, H., & Motoda, H. (2007). *Computational methods of feature selection*. London, England: CRC Press.

Metlek, S., & Kayaalp, K. (2020). *Makine Öğrenmesinde Teoriden Örnek Matlab Uygulamalarına Kadar Destek Vektör Makineleri*. Ankara, Türkiye: İktisad Yayınları.

Özkaya, U., Öztürk, Ş., & Barstugan, M. (2020). *Coronavirus (COVID-19) classification using deep features fusion and ranking technique*. In *Big Data Analytics and Artificial Intelligence Against COVID-19: Innovation Vision and Approach* (pp. 281–295). Springer, Cham.

Palaniappan, R., Sundaraj, K., & Lam, C. K. (2016). Reliable system for respiratory pathology classification from breath sound signals. *2016 International Conference on System Reliability and Science (ICSRS)*. IEEE.

Roffo, G., & Melzi, S. (2016). Features selection via eigenvector centrality. In *Proceedings of New Frontiers in Mining Complex Patterns*.

Roffo, Giorgio, Melzi, S., Castellani, U., & Vinciarelli, A. (2017). Infinite latent feature selection: A probabilistic latent graph-based ranking approach. *2017 IEEE International Conference on Computer Vision (ICCV)*. IEEE.

Roffo, Giorgio, Melzi, S., Castellani, U., Vinciarelli, A., & Cristani, M. (2021). Infinite Feature Selection: A graph-based feature filtering approach. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 43(12), 4396–4410. doi:10.1109/TPAMI.2020.3002843

Sankur, B., Kahya, Y. P., Çağatay Güler, E., & Engin, T. (1994). Comparison of AR-based algorithms for respiratory sounds classification. *Computers in Biology and Medicine*, 24(1), 67–76. doi:10.1016/0010-4825(94)90038-8

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 ilerleme hızı üzerinde uygulama. *Akıllı Sistemlerde Yenilikler ve Uygulamaları Sempozyumu*.

Yang, X.-K., He, L., Qu, D., Zhang, W.-Q., & Johnson, M. T. (2016). Semi-supervised feature selection for audio classification based on constraint compensated Laplacian score. *EURASIP Journal on Audio, Speech, and Music Processing*, 2016(1). doi:10.1186/s13636-016-0086-9

Yang, Y., Shen, H. T., Ma, Z., & Et, A. (2011). L2,1-norm regularized discriminative feature selection for unsupervised learning. In *Conf. International Joint Conference on Artificial Intelligence* (pp. 1589–1594).

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.

Zaffalon, M., & Hutter, M. (2002). Robust feature selection using distributions of mutual information. In *Proceedings of the 18th International Conference on Uncertainty in Artificial Intelligence (UAI-2002)* (pp. 577–584). San Francisco, CA.

Zeng, H., & Cheung, Y.-M. (2011). Feature selection and kernel learning for Local Learning-Based Clustering. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 33(8), 1532–1547. doi:10.1109/TPAMI.2010.215