



## BİLGİSAYAR TABANLI SES ANALİZİNİN TIBBİ TANIDA KULLANILMASI

(COMPUTER BASED VOICE ANALYSIS ON MEDICAL DIAGNOSIS)

Erkan Zeki ENGİN\*, Mehmet ENGİN\*

### ÖZET/ABSTRACT

Sesin oluşmasını sağlayan organlarındaki patolojik durumlardan kaynaklanan ses hastalıklarının birçoğu sesin kalitesinde değişime sebep olur. Uzmanlar, sesteki hastalıklara tanı koymak için değişik yöntemler kullanmaktadır. Bu çalışmada; örselemesiz tabanlı analiz ile, doktorun tanı koymasına yardımcı olunmaktadır. Genlik değişim oranı, perde değişim oranı, sessizlik derecesi, Teager enerji ortalamalı dalgacık dönüşüm katsayıları ve yüksek dereceli istatistik parametreleri ile öznel vektörleri oluşturulmuştur. Sağlıklı veya farklı hastalık sınıflarına ait ses bölütleri, geriye yayılım temelli çok katmanlı algılayıcı ağlar ile sınıflandırılmıştır. Geriye yayılım temelli ağlar; esnek, ölçekli-eşlenik gradyan ve Brodyen-Fletcher-Goldfarb-Shanno (BFGS) öğrenme algoritmaları ile eğitilmiştir. Benzetim çalışmaları sonucunda, ölçekli-eşlenik gradyan algoritmasının en iyi sonucu verdiği bulunmuştur.

*The change in voice quality is affected by many of voice disorders that coming from pathological conditions of voice generation organs. The aim of this study is to help that the clinicians could be diagnosed about voice disorders with non-invasive based analysis. In our work, amplitude perturbation quotient, pitch period perturbation quotient, degree of unvoiceness, Teager Energy Operators averages of wavelet transform coefficients, and higher-order statistics parameters have formed the feature vectors. The voice segments belonging to different pathological or normal classes were classified by backpropagation based multilayer perceptron networks. In backpropagation based multilayer perceptron networks, resilient, scaled-conjugate gradient, and Brodyen-Fletcher-Goldfarb-Shanno learning algorithms were used in training. According to the results of the simulation studies, scaled-conjugate gradient algorithm gave the best results.*

### ANAHTAR KELİMELER/KEYWORDS

Ses analizi, Akustik parametreler, Dalgacık dönüşümü, Yüksek dereceli istatistikler, Sınıflandırma, Yapay sinir ağları

*Voice analysis; Acoustic parameters; Wavelet transform; Higher-order statistics; Classification; Artificial neural networks*

## 1. GİRİŞ

Ses tellerindeki patolojik durumlar, ses tellerinin normal titreşimlerinde önemli değişikliklere yol açar ve sesin kalitesini bozar (Hansen vd., 1998). Bu hastalıklara, erken tanı konulmalı ve tedavi edilmelidir. Ses üretimindeki sorunlar, işlevsel veya organik gırtlaksal patolojilerden dolayı oluşabilir ve farklı terapiler ve cerrahi operasyonlar ile tedavi edilebilir.

Kliniksel veya araştırma amacı ile gırtlak incelenmesi için çeşitli yardımcı laboratuvar yöntemleri bulunmaktadır. Bu yardımcı laboratuvar yöntemleri; esnek ve sert fiberoptik larengoskopi (fiberoptik alet ile inceleme), video-stroboskopi (gırtlak hareketlerinin incelenmesi için gırtlığa vuru ışığı gönderilmesi), elektromiyografidir (gırtlığın işlevsel durumunun dolaylı gözlemlenmesi) (Martinez ve Hugo, 2000). Akustik analiz, yardımcı laboratuvar yöntemlerine alternatif bir tanı koyma yöntemidir. Ses işaretinin sayısal işlenmesine dayanan akustik analiz örselensiz bir yöntem olup, aşağıdaki sebeplerden dolayı kabul görmektedir (Godino-Llorente vd., 2000):

- Tanıda nesnellik: Uzman olmayan doktorların veya glotisi (ses telleri arasındaki boşluk) görmek için uygun ekipmanı olmayan doktorların, öznel tanı koymalarını önler,
- Vokal ve ses hastalıklarına erken tanı konulabilir,
- Vokal işlevdeki bozulma nesnel olarak belirlenebilir,
- Cerrahi, farmakoloji tedavilerin ve rehabilitasyonun değerlendirilmesi: Tıbbi tedavi ve farmakoloji uygulanmadan önceki ses ile sonraki ses nesnel olarak karşılaştırılabilir,
- Otomatik analiz ile bazı basit hastalıkların tespiti uzman olmadan yapılabilir.

Jo ve Kim çalışmalarında, sesleri Yapay Sinir Ağları (YSA) ile sağlıklı ve hastalıklı (iyi ve kötü huylu) olmak üzere ikiye ayırmışlardır (Jo ve Kim, 1999). Bu amaçla; jitter, shimmer ve harmonik gürültü oranı parametrelerini kullanmışlardır. Godino-Llorente vd. ise öğreticili ağ yapısı ve Öz Düzenlemeli ağ Yapısı (ÖDY) ile normal ve hastalıklı sesleri sınıflandırmışlardır (Godino-Llorente vd., 2000). Ağların girişlerine MDVP programın hesapladığı akustik parametreleri uygulamışlardır.

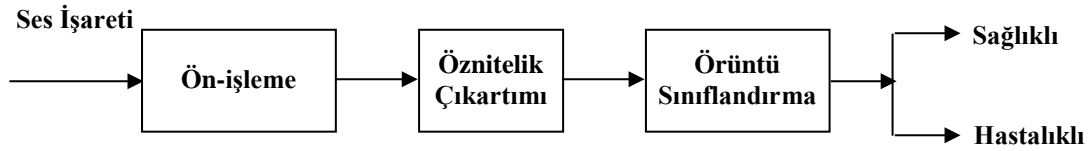
Martinez ve Hugo çalışmalarında, frekans ortamında akustik analizler yaparak otomatik karar verme üzerine çalışmışlardır (Martinez ve Hugo, 2000). Kepstrum, mel-kepstrum, delta-kepstrum, delta-mel-kepstrum ve hızlı Fourier dönüşümü gibi farklı işleme tekniklerini ses üzerinde uygulamışlardır. Elde edilen veriler YSA'ya uygulanarak sesleri sınıflandırmışlardır.

Zaman, izgesel ve kepstral tabanlı parametrelerin patolojik seslerin algılanmasında etkili olduğu gösterilmiştir (Boyanov ve Hadjitodorov, 1997). Dört farklı yöntem ile yapılan sınıflandırmalar birleştirilerek sınıflandırma başarısında bir miktar artış elde etmişlerdir. Hadjitodorov vd. patolojik seslerin akustik analizi için sistem tasarlamışlardır (Hadjitodorov vd., 2000). Sistemlerinde; jitter, shimmer, çeşitli harmonik-gürültü oranları ve diğer çok kullanılan akustik parametreleri kullanmışlardır. Ayrıca, turbülans gürültüsünün hesaplanması için turbülans gürültü göstergesi ve nefesli sesleri karakterize etmek için normalize edilmiş birinci harmonik enerjisi olan yeni akustik parametreler tanımlamışlardır. Bu parametrelerle yapılan sınıflandırma sonucunda, bir önceki sisteme göre başarı artmıştır (Boyanov ve Hadji, 1997).

Heris vd. çalışmalarında; 53 normal ve 54 tek-terafli ses teli felci, 20 polip, 20 nodül hastasından alınan /a/ kayıtlarını, destek vector makinesi ve k-en yakın komşuluğu sınıflandırıcıları ile dört gruba ayırmışlardır (Heris vd., 2009). Bu amaçla; dalgacık paket dönüşümü (dB10) ile 5. seviyeye ayrılan işaretlerden; yaklaşık entropi ve benzerlik parametreleri çıkartmışlardır. Ayrıca her izgesel alt-bantlardan, enerji ve Shannon entropi parametrelerini hesaplamışlardır. Bu parametrelerden etkin olanların tespiti sonucunda yapılan sınıflandırmada en yüksek başarı destek vector makinesi ile elde edilmiştir.

Wang vd. çalışmalarında, 53 normal ve 173 hastalıklı deneye ait kayıtları, Hamming penceresi ile 40 ms'lik bölütlere (20 ms'lik örtüşme) ayırmışlar ve her bölütden 18 mel-frekans kepsrum katsayıları çıkartarak sınıflandırma yapmışlardır (Wang vd., 2011). Diğer bir çalışmada ise 148 denekten alınan /a/ harfi ile sağlıklı ve hastalıklı olarak sınıflandırma yapabilmek için 10 farklı öznelik kümesine (akustik parametreler, frekans parametreleri, mel-frekans parametreleri, kepsral enerji parametreleri, öz-ilişki parametreleri, izgesel harmonik-gürültü oranı, kepsral harmonik-gürültü oranı, doğrusal kestirim katsayıları, doğrusal kestirim cosine dönüşüm katsayıları) ait parametreler çıkartılmıştır (Uloza vd., 2011). Bu öznelik kümesinden genetik algoritma ile seçilen uygun parametreler ile de sınıflandırma yapmışlardır.

Bu çalışmada, patolojik seslerin değerlendirilmesi ve algılanması için örselemesiz tanıya yardımcı olabilecek bir sistem tasarlanmıştır. Tasarlanan sistemin blok diyagramı Şekil 1'de görülmektedir. Çalışmanın ilk aşaması olan ön işlemede /a/ harfiden dc ve dudak etkileri elendikten sonra işaret bölütlere ayrılmıştır. Öznelik çıkartım aşamasında, sesli olduğu tespit edilen bölütlere; genlik değişim oranı, perde değişim oranı, sessizlik derecesi, Teager enerji ortalamalı dalgacık dönüşüm katsayıları ve yüksek dereceli istatistik parametreleri hesaplanmıştır. Son kısımda ise YSA sınıflandırıcı ile sesin sağlıklı veya hastalıklı olduğuna karar verilmektedir.



Şekil 1. Tasarlanan sistemin blok diyagramı

## 2. MATERYAL VE YÖNTEM

### 2.1. Materyal

123 deneye ait olan sesler, Ege Üniversitesi Kulak-Burun-Boğaz ABD'nde kaydedilmiştir. Ses kayıtları, 70 sağlıklı (47 bayan ve 23 erkek) ve 53 hastalıklı (10 bayan ve 43 erkek) denekten alınmıştır. Hastalıklı seslerin; 33'ü kanserli, 12'si nodüllü, 3'ü kronik larenjitli, 2'si ödemli ve 3'ü Reinke ödemlidir. Tanılar, rutin kliniksel yöntemler ile gerçekleştirilmiştir. Ses işaretleri, bilgisayar hafızasına 22050 Hz örnekleme hızı ve 16-bitlik ADC ile örneklenecek kaydedilmiş olup, ağız-mikrofon uzaklığı 20 cm olarak seçilmiştir.

Deneklerden kayıt olarak /a/ ünlü harfi alınmıştır. Bu harfin seçilmesindeki neden, ses kanalının akustik etkilerinin yeterli derecede uzun aralıkta oldukça kararlı olduğundan /a/ harfinin transfer işlevinin küçük hareketlerden nispeten etkilenmemesidir. Buna ek olarak, bu ünlü harf için glotisdeki bir devirdeki hava akışının tepe değerinin kolayca belirlenebilmesidir (Yumoto ve Gould, 1982).

### 2.2. Yöntem

#### 2.2.1. Ön İşleme

Öznelik vektörünü oluşturacak parametreleri hesaplanmadan önce bazı ön işlemler gerçekleştirilmiştir. Ses işaretinden ilk olarak dc bileşen elenir ve daha sonra işaret transfer

işlevi;  $1/(1-\mu z^{-1})$ ,  $\mu \leq 1$  olan bir ön işleme süzgecinden geçirilerek dudakların oluşturduğu yayının yükü etkisi azaltılır (Manfredi vd., 2000).

Ses işareti, durağan olmayan yapıdadır ve konuşma sırasında ses kanalının şekli ve glotisin titreşim hızı değişir. Bu yüzden ses analizi, bölütlere ayrılmış işaret üzerinde yapılmalıdır. Bölütlerin boyutu, işaretin karakteristiğine göre değişir ve temel olarak kestirilmek istenen perde değerine (temel periyoda) göre değişmektedir. Doğru sonuçlar elde edebilmek için peş peşe iki perde dönemi içeren bölütlerde çalışmak gerekmektedir (Manfredi vd., 2000). Bu durumun kesin olarak sağlanması için Hamming pencere kullanılarak bölüt boyutu 30 ms ve örtüşme değeri 20 ms yapılmıştır (Boyanov ve Hadjitodorov, 1997; Boersma, 1993).

Temel olarak, sesli ve sessiz olmak üzere iki tip uyarım vardır. Diğer uyarım tipleri ise sesli, sessiz ve nefesin birleşimidir. Dolayısıyla, her bölütte uyarım tipi belirlenmeli ve parametrelerin hesaplanmasında sadece sesli bölütler kullanılmalıdır.

Ses işaretini bölütlere ayırmadan önce tepe değeri ve daha sonra ise her bir bölütün tepe değeri bulunur. Her bölütün tepe değeri, ses işaretinin tepe değerinin %5'i ile karşılaştırılır ve bu değerden küçük olan bölütler nefes olarak kabul edilir (Veprek ve Scordilis, 2002).

Bölütün öz ilişki işlevinin en büyük değeri, bölütün enerjisinin %40'ından daha küçükse veya temel frekansı 40-500 Hz'in dışında ise bölüt sessiz olarak kabul edilir. Bölütün, sessiz olup olmadığına karar vermede başarıyı artırıcı bazı algoritmalar bulunmaktadır (Veprek ve Scordilis, 2002).

- Sessiz olduğuna karar verilen bölütten önceki iki bölüt de sesli ise daha düşük bir eşik değeri ile karşılaştırılır. Bölütün enerjisi, bu yeni eşik değerinden büyükse ve temel frekansı 40-500 Hz arasında ise bu bölüt sesli olarak kabul edilir. Bu şartlara uymuyorsa, sessiz olarak kalır.
- Kısa sessiz bölütler, bazen gürültünün katkısıyla (elektronik veya akustik) sesli olarak alınabilir. Böyle bölütler genellikle yanlarındaki sesli bölütlerden net olarak ayrılmıştır. , Bu bölütler normal seste oluşmadığından sessiz olarak alınmalıdır. Böyle ardışık sesli bölütlerin sayısı dörtten küçük ve yanlarındaki sessiz bölütlerinin sayısı da ikiden büyük ise bu sesli bölütler sessiz olarak kabul edilir.
- Tek bir sessiz bölüt, iki sesli bölüt arasında ise sesli olarak kabul edilir ve perde dönemi yanlarındaki sesli bölütlerin perde dönemlerinin ortalaması olarak alınır.

### 2.2.2. Zaman ve Genlik Tabanlı Parametreler

Önemli parametrelerden biri Sessizlik Derecesi (SD) olup, aşağıdaki eşitlikle hesaplanır (Boyanov ve Hadjitodorov, 1997).

$$SD = N_{sessiz} / N_{top} \quad (1)$$

$N_{top}$  : toplam bölüt sayısı ve  $N_{sessiz}$  : sessiz bölüt sayısıdır.

İkinci parametre olarak Genlik Değişim Oranı (GDO) kullanılmıştır. Bu parametre, 11 dönemlik düzeltme faktörü ile bölütler arasındaki tepe genliklerindeki değişimin ölçüsü olup, aşağıdaki eşitlikle hesaplanır (Kılıç ve Okur, 2001).

$$GDO = \frac{1}{n-10} \left[ \sum_{i=6}^{n-5} \left| \frac{A_{i-5} + A_{i-4} + \dots + A_{i+4} + A_{i+5} - A_i}{11} \right| \right] \times 100 \quad (2)$$

$$\frac{1}{n} \sum_{i=1}^n A_i$$

$A_i$  : ilgili bölütün tepe değerinin genliği ve  $n$  : sesli bölüt sayısıdır.

Ses kanalını uyaran hava akışı kısmen dönemseldir. Ses tellerinin titreşim hızı harfin temel frekansı (perde) olup, ses tellerinin tam açılımları arasındaki süre de temel dönemdir (Manfredi vd., 2000). Temel frekans, konuşmacının ses tellerinin boyutuna ve gerilmesine bağlıdır ve erkeklerin ses tellerinin ortalama boyutları bayanlarınkinden büyüktür. Bunun sonucunda, bayanların ortalama temel frekansı genellikle erkeklerden düşüktür.

Perdenin hesaplanmasında birçok algoritma bulunmakta olup, bu çalışmada öz ilişki işlevi kullanılmıştır. Durağan işaretlerde, öz ilişki işlevi aşağıdaki eşitlikle hesaplanır (Boersma, 1993).

$$r(\tau) = \sum_{t=1}^N x(t)x(t-\tau) \quad (3)$$

Her bölüt için öz ilişki işlevinin tepe değerinin yeri bulunur. Bölütlerdeki perde dönemi; her bölütteki yerel tepe değerinin bulunduğu noktanın, bölütün başlangıç noktası ile arasındaki süreye eşittir.

Öz ilişki işlevinde, formantlardan ve harmoniklerden dolayı çok sayıda sahte tepe noktası bulunmaktadır. Bu tepeleri elemek için işaret önce kesme frekansı 900 Hz olan alçak geçiren süzgeçten geçirilir ve daha sonra merkez sınırlama yöntemi uygulanır (Sonhdi, 1968).

Zaman tabanlı parametre olarak Perde Değişim Oranı (PDO) kullanılmıştır. 5 dönemlik düzeltme faktörü ile perde değişimi aşağıdaki eşitlikle hesaplanır (Kılıç ve Okur, 2001).

$$PDO = \frac{1}{n-4} \left[ \sum_{i=3}^{n-2} \left| \frac{P_{i-2} + P_{i-1} + P_i + P_{i+1} + P_{i+2} - P_i}{5} \right| \right] \times 100 \quad (4)$$

$$\frac{1}{n} \sum_{i=1}^n P_i$$

$P_i$  : ilgili bölütün perde dönemi ve  $n$  : sesli bölüt sayısı.

GDO ve PDO parametreleri bulunduktan sonra, GDO ve PDO'nun büyük olan değerine göre normalizasyon işlemi yapılır. SD için buna gerek yoktur; çünkü, bu değer her zaman birden küçük çıkmaktadır. Bu değerler bütün ses üzerinde hesaplanan ortalama değerlerdir ve bu yüzden, her sese ait bölütlerin öznelilik vektörlerine aynı değerler girilir.

### 2.2.3. Dalgacık Dönüşümü Tabanlı Parametreler

Dalgacık analizinde işaret, zaman-frekans yerine zaman-ölçek ortamında incelenir ve en önemli üstünlüklerinden biri; işareti belirli bir bölgesinin analizinin mümkün olmasıdır. Dalgacık analizi, işareti ana dalgacığın kaydırılmış ve ölçeklendirilmiş biçimlerine açarak incelemektedir.

Dalgacık dönüşümü ile işaret, sabit katsayılı ve değişken genişlikli frekans alt bantlarına ayrılır (Tewfik vd., 1992). Bu analiz için çeşitli işlev tipleri kullanılmaktadır. Teorik olarak sıfır ortalamalı ve sonlu enerjiye sahip herhangi bir işlev ana dalgacık sayılabilir. Ancak, ana dalgacığı seçmek için birçok kriter vardır. Ana dalgacığın, zaman ve frekans ortamındaki sönümlenmesi önemlidir. Zaman ve frekans ortamında iyi lokalize olabilmek için dalgacık, zaman ve frekans ortamında hızlı sönümlenmelidir (Tüfekci ve Gowdy, 2000).

Sürekli dalgacık dönüşümü; bütün bir işaretin, dalgacık işlevinin bütün ölçek ve kaymaları ile çarpımlarının toplamıdır:

$$C(\text{ölçek}, \text{konum}) = \int_{-\infty}^{\infty} f(t) \Psi(\text{ölçek}, \text{konum}) dt \quad (5)$$

$f(t)$ : analiz edilen işaret ve  $\Psi$ : ana dalgacık işlevidir. Eşitlik 5'deki C'ler ölçek ve konumuna bağlı bulunan dalgacık dönüşümü katsayılarıdır. Düşük ölçeklerde dalgacık daralır ve ani değişimler (yüksek frekanslar) yakalanır. Yüksek ölçeklerde ise dalgacık genişleyerek düşük frekanslar yakalanır. Sürekli dalgacık dönüşümü, bilgisayarda hesaplanırken ayrık olarak işlenir. Ancak, ayrık dalgacık dönüşümünden farkı ölçeğin ve kaymanın sürekli olmasıdır.

Bu çalışmada, dalgacık dönüşümü tabanlı parametreleri bulurken dalgacık ayrışım ağacı ile bölütler altıncı ölçeğe kadar açılmıştır. Bu ayrışım için 'Daubechies-5' dalgacığı kullanılmıştır. Daha sonra, her ayrışımın sonundaki ayrıntı katsayıları alınmıştır. Her ölçekteki ayrıntı katsayılarına, Teager Enerji İşleci (TEİ) uygulanmıştır (Jabloun vd., 1999).

$$\Psi_d [s(n)] = s(n)^2 - s(n+1)s(n-1) \quad (6)$$

$s(n)$ : ayrık-zaman işaretidir. Her ölçek için Teager enerjisi  $e_1$ :

$$e_1 = \frac{1}{N_l} \sum_{n=1}^{N_l} |\Psi[s(n)]| \quad l = 1, \dots, L \quad (7)$$

şeklinde hesaplanır (Jabloun vd., 1999). Eşitlikteki  $N_l$ :  $l$ . ölçekteki katsayısı sayısıdır. Her ölçek için bulunan parametrelerde kendi içlerinde normalize edilerek ağa girecek şekilde hazırlanmış olmaktadır.

#### 2.2.4. Yüksek Dereceli İstatistik Tabanlı Parametreler

Yüksek Dereceli İstatistikler (YDİ), öz ilişki işlevi ve güç izgeseli gibi ikinci derece ölçümlerin daha üst mertebelere genişletilmesidir. Gaussian olasılık yoğunluğu işlevine sahip işaretler, ikinci derece istatistikler tarafından temsil edilebilir. Ancak; pratikte birçok işaret ve biyolojik işaretler, Gaussian olmayan karakterde olup, düşük dereceli istatistikler tarafından temsil edilemezler. Bu yüzden, yüksek dereceli istatistikler (kümülanlar) tanımlanmıştır.

Durağan bir süreçte; birinci kümülan, sürecin ortalamasıdır:  $C_{1,x} = E(x(t))$ . Yüksek mertebeli kümülanlar ortalamadaki kaymalardan bağımsızdır. Bu yüzden, ortalamayı sıfır kabul ederek tanımlamak uygundur. Ortalama sıfırdan farklı ise ortalama çıkarılır ve çıkan sonucun kümülanları bulunur. Sıfır ortalamalı durağan bir sürecin kümülanları aşağıdaki eşitlikde tanımlanmaktadır.

$$\begin{aligned}
C_{2x}(k) &= E\{x^*(n)x(n+k)\} \\
C_{3x}(k,l) &= E\{x^*(n)x(n+k)x(n+l)\} \\
C_{4x}(k,l,m) &= E\{x^*(n)x(n+k)x(n+l)x^*(n+m)\} \\
&\quad - C_{2x}(k)C_{2x}(l-m) - C_{2x}(l)C_{2x}(k-m) - M_{2x}^*(m)M_{2x}(k-l)
\end{aligned} \tag{8}$$

$M_{2x}(m) = E\{x(n)x(n+m)\}$  gerçek değerli bir süreçte  $C_{2x}(m)$ 'e eşittir. Eşitlikteki,  $E$ : beklenti işleci ve  $k$ ,  $l$ , and  $m$ 'lerde zaman gecikmeleridir.

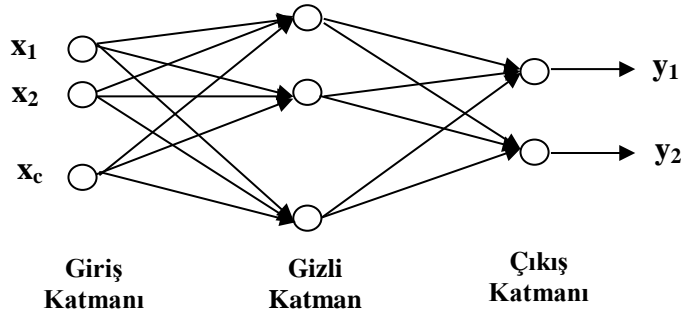
Gaussian gürültüyü bastırması ve fazı önlemesinden dolayı sesli/sessiz sınıflandırılması, perde döneminin bulunması ve geçişleri bulmak gibi ses uygulamalarında YDİ kullanılmaktadır (Nemer vd., 2001).

Normalizasyon işlemi her kümülantın kendi içinde yapılmıştır. Normalizasyon işleminden sonra her kümülantdan 5 (eşit aralıkta) değer seçilmiştir (Osowski ve Linh, 2001). Böylece öznitelik vektöründe, her bölüt için farklı 15 kümülant parametresi kullanılmıştır.

### 2.2.5. Örüntü Sınıflandırma

1960'lı yılların ortasından itibaren sayısal bilgisayarların yaygın kullanımı ile birlikte modern örüntü analiz teknikleri uygulanmaktadır (Ciaccio vd., 1993). Tüm örüntü analiz kuramı ve teknikleri ağırlıklı olarak matematik tabanlı olup; istatistiksel, sentaktik ve yapay zeka olmak üzere üç gruba ayrılır.

Çalışmamızda kullanılan YSA, yapay zeka grubuna girmektedir. YSA (Şekil 2); giriş, gizli ve çıkış katmanlarından oluşmaktadır. Giriş katmanı, öğrenilecek veya sınıflandırılacak bilgiyi alan nöronların olduğu katmandır. Çıkış katmanındaki nöronlar ise girişin hangi sembole daha yakın olduğunu vermektedir. Gizli katmanda, giriş ile çıkış katmanlarını birbirine bağlamaktadır.



Şekil 2. Standart bir ileri yönlü YSA yapısı

Bu çalışmada, geriye yayılım temelli çok katmanlı algılayıcı ağlar kullanılmıştır. Geriye yayılım temelli ağlarda; esnek, Ölçekli Eşlenik Gradyan (ÖEG) ve BFGS öğrenme algoritmaları kullanılmıştır. Bu ağlardaki en önemli faktör, gizli katmandaki nöron sayısının seçimidir. Nöron sayısının az seçilmesi durumunda hata yeterli derecede küçük elde edilemez ve çok büyük seçildiği durumda ise ağın genelleme kabiliyeti zarar görür (Osowski ve Linh, 2001). Farklı sayılarda gizli nöron sayısı ile sistem test edilmiş ve hatayı kabul edilebilir bir seviyeye çeken en küçük nöron sayısı seçilmiştir.

Öznitelik vektöründeki parametreler normalize edildiği için giriş vektöründeki katsayılar sıfır ile bir değerleri arasındadır. Bu çalışmada, yirmi dört giriş nöronu (GDO, PDO, DSD,

dalgacık dönüşümünden altı parametre ve kümülanlıdan on beş parametre), on gizli nöron ve iki çıkış nöronu (normal ve patolojik) kullanılmıştır.

Ağın öğrenme aşamasında katsayılar, gradyan yöntemi ile güncellenmiştir. Gradyan öğrenme yönteminde, hata işlevinin gradyanı ile katsayılar güncellenmektedir (Osowski ve Linh, 2001).

$$w(k+1) = w(k) + \eta p(k) \quad (9)$$

$\eta$ :  $k$ . yaklaşımdaki öğrenme katsayısı ve  $p(k)$ :  $k$ . yaklaşımdaki hatanın gradyanıdır. Katsayıların ayarlanmasında kullanılan öğrenme katsayısı aşağıdaki algoritmalar ile optimize edilmiştir.

Esnek Öğrenme Algoritması : Çok katmanlı ağlarda, genellikle gizli katmanlarda sigmoit transfer işlevini kullanılır. Bu işlevler, sonsuz aralıktaki girişleri sonlu aralıktaki değerlere sıkıştırırlar. Sigmoid işlevlerin eğimi, giriş büyüdükçe sıfıra yaklaşır. Dolayısıyla, gradyan değeri çok küçük değerler alır. Ağırlık ve eşik değerinde küçük değişimler olması sonucu en uygun değerlerden oldukça uzaktır (Demuth ve Beale, 2002).

Esnek Geriye Yayılım (EGY) yönteminin kullanılmasının amacı, kısmi türevin genliğindeki bozucu etkinin giderilmesidir. Türevin işareti, güncellenecek ağırlıkların yönüne karar vermek için kullanılır (türevin genliğinin, ağırlıkların güncellenmesinde bir etkisi yoktur). Ağırlıkların ne kadar değişeceği, ayrı bir güncelleme değeri ile belirlenir (Demuth ve Beale, 2002).

ÖEG Öğrenme Algoritması : Azalan gradyan yönteminde öğrenme katsayısı sabittir ve bu her zaman en uygun değerlere ulaşmak için en kısa yol değildir. Eşlenik gradyan yöntemlerin çoğunda değişimlerin değerleri her adımda ayarlanır. Değişim değeri, hatayı yol boyunca en aza indirmek için eşlenik gradyan yönünde araştırma yapılır (Principe vd., 2000).

ÖEG tam otomatik olup kullanıcı tarafından parametre girilmesine gerek duymamaktadır. ÖEG yöntemi, zaman gecikmesine sebep olan yol arama algoritması kullanmamaktadır (Moller, 1993).

BFGS Öğrenme Algoritması : Newton yöntemi, en uygun noktayı hızlı bulmasından dolayı eşlenik gradyan yöntemine alternatif bir yöntemdir. Newton yönteminin temeli Eşitlik 9'da verilmektedir.

$$w(k+1) = w(k) - H_k^{-1} p(k) \quad (10)$$

$H(k)$ : Hessian matrisidir (ikinci türev). Newton yöntemi genellikle eşlenik gradyan yönteminden daha hızlıdır. Ancak ağlar için Hessian matrisini hesaplama yöntemi yavaşlatmaktadır. Newton yöntemine dayanan bazı yöntemler ile ikinci türevleri hesaplamak gerekmemektedir. Her yaklaşımda yaklaşık Hessian matrisi güncellenir (güncelleme gradyanın işlevidir). Bu yaklaşık Newton yöntemleri içinde en başarılı olanı BFGS güncellemesidir (Demuth ve Beale, 2002).

### 3. SONUÇLAR

Zaman-frekans tabanlı parametreler (GDO, PDO, SD), dalgacık dönüşümü ve kümülanlı tabanlı parametreler ile oluşturulan öznelik vektörü ile farklı ağ tipleri eğitilmiş ve test edilmiştir. Ağın sesleri sağlıklı ve hastalıklı olmak üzere iki sınıfa ayırmasını gerçekleştirmek için eğitim aşamasında sağlıklı seslerin 35'i ve hastalıklı seslerin 28'i kullanılırken, test için



35 sağlıklı ve 25 hastalıklı ses kullanılmıştır. Daha sonra bu sesler, YSA ile sağlıklı ve hastalıklı olmak üzere iki sınıfa ayrılmıştır.

Sınıflandırma yöntemlerinin başarılarını değerlendirmek için kullanılan değişkenler aşağıda tanımlanmaktadır (Hadjitodorov vd., 2000):

Doğru pozitif (DP): Hastalıklı sesin hastalıklı olarak sınıflandırılması.

Doğru negatif (DN): Sağlıklı sesin sağlıklı olarak sınıflandırılması.

Yanlış pozitif (YP): Sağlıklı sesin hastalıklı olarak sınıflandırılması.

Yanlış negatif (YN): Hastalıklı sesin sağlıklı olarak sınıflandırılması.

Hastalıklı seslerdeki başarı - HSB (sesin hastalıklı olduğunun tespit edilme olasılığı), sağlıklı seslerdeki başarı - SSB (sesin sağlıklı olduğunun tespit etme olasılığı) ve genel başarı - GB (sesin durumunun doğru tespit edilme olasılığı) hesaplamaları aşağıda verilmektedir.

$$HSB = \frac{DP}{DP + YN} \times 100$$

$$SSB = \frac{DN}{DN + YP} \times 100 \quad (11)$$

$$GB = \frac{DP + DN}{DP + DN + YN + YP}$$

Çizelge 1’de bölüt tabanındaki farklı ağ yapıları için başarı ölçütleri verilmektedir. Her deneğin ses kaydının, 0.5-2 saniyelik kısımları bölütlere ayrılmıştır. Dolayısıyla, her deneğin sesi için elimizde 50-200 arasında öznitelik vektörü bulunmaktadır. Ses tabanındaki farklı ağ yapıları için başarı ölçütleri Çizelge 2’de görülmektedir. Çizelge 2’deki sonuçlar, ÖEG öğrenme algoritmasının en iyi sonucu verdiğini göstermektedir. Bölüt tabanlı sınıflandırmada, eğitim aşamasında % 99,3 ve test aşamasında % 77,65 başarıya ulaşılmıştır. Ses tabanındaki sınıflandırmada ise eğitim aşamasında % 100 ve test aşamasında %88,3 başarıya ulaşılmıştır.

Çizelge 1. Farklı ağ algoritmaları için bölüt tabanlı başarı ölçütleri: (a) HSB, (b) SSB, (c) GB

	Eğitim			Test		
	DP	YN	Yüzde	DP	YN	Yüzde
ÖEG	4250	28	% 99.34	3315	930	% 78
EGY	4249	29	% 99.3	3160	1085	% 74.44
BFGS	4208	70	% 98.36	2870	1375	% 67.6

(a)

	Eğitim			Test		
	DN	YP	Yüzde	DN	YP	Yüzde
ÖEG	6706	46	% 99.31	5229	1549	% 77.38
EGY	6701	51	% 99.24	4849	1999	% 70.8
BFGS	6699	53	% 99.21	5223	1625	% 76.77

(b)

	Eğitim			Test		
	DP+DN	YN+ YP	Yüzde	DP+DN	YN+YP	Yüzde
ÖEG	10956	74	% 99.3	8614	2479	% 77.6
EGY	10950	80	% 99.3	8009	3084	% 72.1
BFGS	10907	123	% 98.3	8093	3000	% 72.9

(c)

Çizelge 2. Farklı ağ algoritmaları için ses tabanlı başarı ölçütleri: (a) HSB, (b) SSB, (c) GB

	Eğitim			Test		
	DP	YN	Yüzde	DP	YN	Yüzde
ÖEG	28	0	% 100	21	4	%84
EGY	28	0	% 100	21	4	%84
BFGS	28	0	% 100	21	4	%84

(a)

	Eğitim			Test		
	DN	YP	Yüzde	DN	YP	Yüzde
ÖEG	35	0	% 100	29	6	% 82.85
EGY	35	0	% 100	27	8	% 77.14
BFGS	35	0	% 100	28	7	% 80

(b)

	Eğitim			Test		
	DP+DN	YN+ YP	Yüzde	DP+DN	YN+YP	Yüzde
ÖEG	63	0	% 100	50	10	% 83.3
EGY	63	0	% 100	48	12	% 80
BFGS	63	0	% 100	49	11	% 81.6

(c)

#### 4. TARTIŞMA

Ses parametreleri; yaşa, cinsiyete ve patoloji tipine ve evresine göre değişmektedir. Dolayısıyla, bütün yaşlar ve patolojilere (farklı evreler) ait ses kayıtlarının çoğaltılması ile sistemin başarısıda arttırılabilecektir. Ayrıca daha geniş veri tabanı olduğu durumda, patolojilerin girtlağa olan etkisine göre birkaç sınıfa ayrılabilirler.

Çalışmamızda, geriye yayılım temelli çok katmanlı algılayıcı ağ yapıları denenmiştir. Diğer ağ yapıları olan ÖDY ve yinelenen ağ yapıları veya genetik ve bulanık gibi sınıflandırıcılar oluşturularak başarılar karşılaştırılabilir.

Öznelik vektörünü oluştururken zaman-frekans parametrelerinden GDO, PDO ve SD ile dalgacık dönüşümü ve yüksek dereceli model tabanlı parametreler kullanılmıştır. Diğer zaman-frekans ve model tabanlı parametrelerinde sınıflandırma başarıları test edilebilir.

## KAYNAKLAR

- P. Boersma (1993): "Accurate Short-term Analysis of the Fundamental Frequency and the Harmonics-to-Noise Ratio of a Sampled Sound", Proceedings of the Institute of Phonetic Sciences, Cilt 17, s. 97-110.
- B. Boyanov, S. Hadjitodorov (1997): "Acoustic Analysis of Pathological Voices", IEEE Engineering in Medicine and Biology, Cilt 16, s. 74-81.
- E. J. Ciaccio, S. M. Dunn, M. Akay (1993): "Biosignal Pattern Recognition and Interpretation Systems", IEEE Engineering in Medicine and Biology Magazine, Cilt 12, s. 89-97.
- H. Demuth, M. Beale (2002): "Neural Network Toolbox for Use with MATLAB".
- J. L. Godino-Llorente, S. Aguilero-Navarro, P. Gomez-Vilda (2000): "Non-supervised Neural Net Applied to the Detection of Voice Impairment", IEEE International Conference on Acoustics, Speech, and Signal Processing, s. 3594-3597.
- S. Hadjitodorov, B. Boyanov, B. Teston (2000): "Laryngeal Pathology Detection by Means of Class-specific Neural Maps", IEEE Transactions on Information Technology in Biomedicine, Cilt 4, s. 68-73.
- J. H. L. Hansen, L. Gavidia-Ceballos, J. F. Kaiser (1998): "A Nonlinear Operator-based Speech Feature Analysis Method with Application to Vocal Fold Pathology Assessment", IEEE Transactions on Biomedical Engineering, Cilt 45, s. 300-313.
- H. K. Heris, B. S. Aghazadeh, M. Nikkhah-Bahrami (2009): "Optimal Feature Selection for the Assessment of Vocal Fold Disorders", Computers in Biology and Medicine, Cilt 39, s. 860-868.
- F. Jabloun, A. E. Çetin, E. Erzin (1999): "Teager Energy Based Feature Parameters for Speech Recognition in Car Noise", Cilt 6, s. 259-261.
- C. W. Jo, H. Kim (1999): "Classification of Pathological Voice into Normal/Benign/Malign State", Proceedings of Eurospeech, s. 571-574.
- M. A. Kılıç, E. Okur (2001): "CSL ve Dr. Spech ile Ölçülen Temel Frekans ve Pertürbasyon Değerlerinin Karşılaştırılması" KBB İhtisas Dergisi, Cilt 8, s. 152-157.
- C. Manfredi, M. D'aniello, P. Brusagliani, A. Ismaelli (2000): "A Comparative Analysis of Fundamental Frequency Estimation Methods with Application to Pathological Voices", Medical Engineering & Physics, Cilt 22, s. 135-147.
- C. E. Martinez, R. L. Hugo (2000): "Acoustic Analysis of Speech for Detection of Laryngeal Pathologies", Proceedings of the 22<sup>nd</sup> Annual EMBS International Conference, s. 2369-2372.
- M. F. Moller (1993): "A Scaled Conjugate Gradient Algorithm for Fast supervised Learning", Neural Networks, Cilt 6, s. 525-533.
- E. Nemer, R. Goubran, S. Mahmoud (2001): "Robust Voice Activity Detection Using Higher-order Statistics in the LPC Residual Domain", IEEE Transactions on Speech and Audio Processing, Cilt 9, s. 217-231.
- S. Osowski, T. H. Linh (2001): "ECG Beat Recognition Using Fuzzy Hybrid Neural Network", IEEE Transactions on Biomedical Engineering, Cilt 48, s. 1265-1271.

- J. C. Principe, N. R. Euliano, W. C. Lefebvre (2000): "Neural and Adaptive Systems: Fundamentals through Simulations", John Wiley and Sons Inc.
- M. M. Sondhi (1968): "New Methods of Pitch Extraction", IEEE Transactions on Audio and Electroacoustics, Cilt 16, s. 262-266.
- A. H. Tewfik, D. Sinha, P. Jorgensen (1992): "On the Optimal Choice of a Wavelet for Signal Representation", IEEE Transactions on Information Theory, Cilt 38, s. 747-765.
- Z. Tüfekci, J. N. Gowdy (2000): "Feature Extraction Using Discrete Wavelet Transform for Speech Recognition", Proceedings of the IEEE Southeastcon, s. 116-123.
- V. Uloza, A. Verikas, M. Bacauskiene, A. Gelzinis, R. Pribuisiene, M. Kasetta, V. Saferis (2011): "Categorizing Normal and Pathological Voices: Automated and Perceptual Categorization", Journal of Voice, Cilt 25, s. 700-708.
- P. Veprek, M. S. Scordilis (2002): "Analysis, Enhancement and Evaluation of Five Pitch Determination Techniques", Speech Communication, Cilt 37, s. 249-270.
- X. Wang, J. Zhang, Y. Yan (2011): "Discrimination Between Pathological and Normal Voices Using GMM-SVM Approach", Journal of Voice, Cilt 25, s. 38-43.
- E. Yumoto, W. J. Gould (1982): "Harmonics-to-Noise Ratio as An Index of the Degree of Hoarseness", Journal Acoustical Society of America, Cilt 71, s. 1544-1550.