

Kütle Spektrometresi Verileri Kullanılarak Yumurtalık Kanserinin Yapay Sinir Ağlarıyla Sınıflandırılması

Demet YEŞİLBAŞ*¹, Ayşegül GÜVEN¹

¹Erciyes Üniversitesi, Mühendislik Fakültesi, Biyomedikal Mühendisliği Bölümü, Kayseri

Geliş tarihi: 05.03.2021

Kabul tarihi: 13.09.2021

Öz

Kanserin fark edilme aşaması, diğer kanser türlerinde olduğu gibi iyileşme oranını etkiler. Yaşı ilerlemiş kadınlar için ciddi bir hastalık olan yumurtalık kanseri başlangıç aşamasında fark edilmez, çoğu zaman diğer hastalıklarla karıştırılır. Yüzey Güçlendirmeli Lazer Desorpsiyon/İyonizasyon Uçuş Zamanlı Kütle Spektrometresi (SELDI-TOF-MS) kompleks numunelerin incelenmesine olanak sağlayarak yumurtalık ve diğer kanser türlerinin ayırt edilmesinde potansiyel belirteçtir. Bu çalışmada, FDA-NCI web sitesinde yer alan yumurtalıklara ait iki Düşük Çözünürlüklü SELDI-TOF-MS veri setini Yapay Sinir Ağları (YSA) ile sınıflandırarak, veri setlerini karşılaştırıyoruz. Ön işleme adımı olarak, Temel Bileşenler Analizi (PCA-Principal Component Analysis) kullandık. PCA uygulanmış verinin en yüksek varyans oranına sahip 20 bileşeni seçildi, 10 nörondan oluşan tek gizli katmanlı İleri Yönlü YSA ile sınıflandırma yapıldı ve 4-3-02 isimli veri seti için %95 doğruluk elde edilirken, 8-7-02 isimli veri seti için %100 doğruluk elde edilmiştir.

Anahtar Kelimeler: Yapay sinir ağları, Temel bileşenler analizi, Yumurtalık kanseri, SELDI-TOF-MS

Classification of Ovarian Cancer with Neural Networks Using Mass Spectrometry Data

Abstract

The stage of cancer diagnosis affects the rate of recovery, as in other types of cancer. Ovarian cancer is a serious disease for older women, is not noticed at the initial stage and is often confused with other diseases. Surface-Enhanced Laser Desorption/Ionization Time-of-Flight Mass Spectrometry (SELDI-TOF-MS) is a potential biomarker for distinguishing ovarian and other types of cancer by allowing the examination of complex samples. In this study, we classified two Low-Resolution SELDI-TOF-MS ovarian datasets from the FDA-NCI website with Artificial Neural Networks (ANN) and compared them. We used Principal Component Analysis (PCA) as a preprocessing step of classification. 20 components of maximum variance in the PCA-applied data are selected and classified with the feed-forward ANN consists of a single hidden layer with 10 neurons, 95% accuracy was achieved for the data set named 4-3-02 and 100% accuracy achieved for the data set named 8-7-02.

Keywords: Artificial neural networks, Principal components analysis, Ovarian cancer, SELDI-TOF-MS

*Sorumlu yazar (Corresponding author): Demet YEŞİLBAŞ, demetyesilbass@gmail.com

1. GİRİŞ

Yumurtalık kanseri, 2020 yılı istatistiklerine göre kadınlarda görülen en yaygın sekizinci kanser türüdür [1]. Global Kanser Gözlemevi (Global Cancer Observatory)'nin dünya çapında gelecekteki kanser insidansını ve ölüm oranını tahmin eden aracı Cancer Tomorrow'a göre, mevcut 313.825 yeni vaka sayısının 2040 yılında 427.959 yeni vaka sayısına; 207.155 mortalitenin 306.929'a yükseleceği öngörülmektedir [2]. Yumurtalık kanseri genellikle menopoz sonrası dönemde ortaya çıkar. İlk aşamada fark edildiğinde gerekli tedavi ve uygulamalar ile %90'a varan iyileşme oranı gözlemlenirken, kanserin son aşamalarında bu oran, %30'un altına düşmektedir [3,4]. Tümörün gizli olarak büyümesi semptomların geç görülmesine neden olur ve hastalık ileri aşamalarda fark edilir. Bu da ölüm oranının yüksek olmasına neden olur [5]. Bu süreçte teşhis sistemlerinden elde edilen bilginin hatalı yorumlanması problemin ciddiyetini artırır.

Kötü huylu tümör ve diğer birçok hastalık, protein düzeyindeki mutasyonlar neticesinde meydana gelir [6]. Proteomik, bir hücre veya organizmanın tüm protein bileşiminin yani proteomunun faaliyetlerini incelenmesini içerir. Proteomik teknolojisindeki ilerlemeler, birçok hastalığın moleküler seviyelerde çok daha iyi anlaşılmasına imkân tanır [7]. Kanserli dokuya ait protein imzaları teşhis, hastalığın seyri ve tedaviye yanıt hakkında faydalı olabilecek bilgiler sağlar [8].

Kütle spektrometresi (Mass Spectrometry- MS), manyetik veya elektriksel alanda hareket eden yüklü parçacıkların kütle/yük (m/z) oranına bağlı olarak molekülün yapısı hakkında bilgi verilmesini sağlar. Yaygın olarak kullanılan MS sistemleri Matris ile Desteklenmiş Lazer Desorpsiyon/İyonizasyon Uçuş Zamanlı Kütle Spektrometresi (MALDI-TOF-MS) ve daha gelişmiş versiyonu olan Yüzey Güçlendirmeli Lazer Desorpsiyon/İyonizasyon Uçuş Zamanlı Kütle Spektrometresidir (Surface-Enhanced Laser Desorption/Ionization Time-of-Flight Mass Spectrometry-SELDI-TOF-MS) [9]. SELDI-TOF-MS, kromatografi ve kütle spektrometrisi tekniklerini birleştirir ve çeşitli biyolojik

numunelerde binlerce analitin kapsamlı bir şekilde değerlendirerek, kanserin erken tespit edilmesine imkân sağlayan bir yaklaşımdır [10]. SELDI-TOF-MS'nden bilgi nasıl elde edilir diye baktığımızda: Kromatografik ayırma işlemi uygulanarak protein seçici yüzeye tutunmuş numune, lazerle iyonize edilir ve bir vakum odasındaki ters yüklü elektrotlara doğru hareket etmesi sağlanır. Burada kütle/yük oranı (m/z) ile orantılı olan detektöre varma süresi (Time of Flight-TOF) ölçülür. Uygun yazılımlarla m/z değerine karşılık gelen veriler gösterilir [11]. Bu veriler üzerinden numunenin ayırt edilmesini sağlayan pikler üzerinden biyoinformatik araçlarla analizler yapılarak hastalıklar (yumurtalık kanseri, akciğer kanseri, prostat kanseri, meme kanseri, kolon kanseri ve karaciğer kanseri) hakkında bilgiler elde edilebilir [12].

SELDI-TOF-MS ile, x ekseninde kütle-yük oranlarının sayısını (m/z), y ekseninde göreceli yoğunluğu gösteren yüksek boyutlu bir sinyal elde edilir. Elde edilen veriler kullanılarak sınıflandırma çalışmaları yapılmaktadır. Bu yüksek boyutlu verinin tümünü kullanmak yerine, verileri büyük oranda temsil eden, yeni özellikler tanımlayabilen ve veri içerisindeki gürültüyü azaltan bir özellik çıkarma yöntemi uygulanarak daha hızlı bir sınıflandırma yapılabilir [13]. Bu konuda yapılmış çalışmalar, genel olarak, verinin belirgin istatistiksel özelliklerini kullanarak sınıflandırılması üzerinedir. Günümüzde sağlık alanında kullanımı artan yapay zekânın alt kümesi olan makine öğrenmesi teknikleri, bu veriler üzerinde sınıflandırma yapmak için kullanılmaktadır.

Bazı araştırma ekipleri, SELDI verisindeki piklerin protein kimliklerini potansiyel teşhis için önem teşkil ettiğini ortaya koymuşlardır. Yapılan bir çalışmada, SELDI verisindeki pikler üzerinden özellik çıkartıp Çok Katmanlı Perceptron kullanılarak %93 doğrulukla sınıflandırma yapılmıştır [14]. Hu ve arkadaşları pankreas kanserini teşhis etmek için CIPHERGEN ProteinChip yazılımı kullanarak pikleri tespit edip Yapay Sinir Ağları (YSA) ile sınıflandırmış ve 0,3 olarak belirlenen en uygun kesme değeri için doğruluğu %93,47 olarak bulmuşlardır [15]. CIPHERGEN

Protein Chip Yazılımı ile piklerin bulunmasına dayanan başka bir çalışmada Destek Vektör Makineleri (Support Vector Machines-SVM) ve Çapraz Doğrusal Ayırım Analizi (Diagonal Linear Discriminant Analysis-DLDA) ile yumurtalıklara ait veriler kanser ve iyi huylu tümör olarak sınıflandırılmıştır [16].

Yapılan çalışmaların bir kısmı, istatistiksel yöntemler ile özellik çıkarımı yapıp sınıflandırma algoritmasıyla hastalığın tespit edilmesine yöneliktir. Wu ve arkadaşları Olasılıksal Temel Bileşenler Analizi (Probabilistic Principal Component Analysis-PPCA) ile Yumurtalık kanseri için SELDI-TOF-MS verilerinin boyutunu azaltıp SVM ile sınıflandırarak PPCA'nin (doğruluk=%93,93), Temel Bileşenler Analizi (Principal Component Analysis-PCA) (doğruluk=%83,34)'den daha başarılı bir sonuç ortaya koyduğunu göstermişlerdir [18]. Meme kanseri için elde edilen düşük kütleli SELDI spektrumları dalgacık dönüşümü ile T²'li PCA istatistik metodlarıyla ön işlemeye tabi tutulduktan sonra yine SVM ile, kanser ve kontrol verisi %100 doğrulukta sınıflandırılmıştır [19].

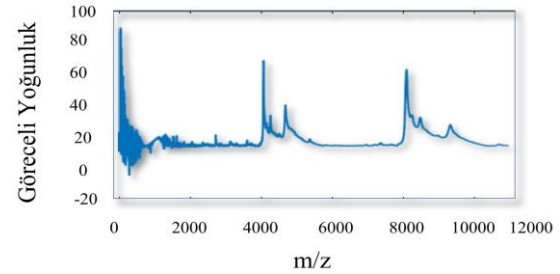
Verilerden öğrenme yoluna dayalı YSA kullanılarak yapılmış çalışmalar mevcuttur. Thakur ve arkadaşları, SELDI-TOF-MS verileri üzerinde student t-testi kullanarak özellik seçimi ve İleri Beslemeli Sinir Ağları kullanarak sınıflandırma yapmışlardır. Kontrol verisi için %98,50, kanser verisi için %99,16 test doğruluğu elde etmişlerdir [20]. SELDI-TOF-MS verileri kullanılarak yumurtalık kanserini teşhis etmek için YSA ile %98 doğruluk elde edilen bir çalışma mevcuttur [21]. Pei ve arkadaşları Yüksek Çözünürlüklü SELDI-TOF setleri için yumurtalık kanseri ve yüksek risk kontrol verisini sınıflandırmak için anahtar özelliklerin seçilmesi (rank features), PCA ve sinir ağları nöron sayısının optimizasyonu ile %100 doğruluk elde etmişlerdir [22]. Rahman ve arkadaşları, Toguchi yöntemini kullanarak optimum nöron sayısını belirlemişler ve 15 gizli nörondan oluşan bu YSA modeli %98,7 doğrulukla diğer sınıflandırma algoritmalarından (Fine Tree, SVM, K-En Yakın Komşuluk, Ensemble) daha yüksek bir performans göstermiştir [23].

Önerilen çalışmada, literatürde yapılan çalışmalarda YSA ile elde edilen başarının yüksek olması sebebiyle YSA kullanılmıştır. FDA-NCI web sitesindeki SELDI-TOF-MS tabanlı protein görüntüleme tekniği ile elde edilmiş veri setleri PCA ile 20 temel bileşene indirgenip sınıflandırılarak veri setleri arasındaki başarı kıyaslanmıştır.

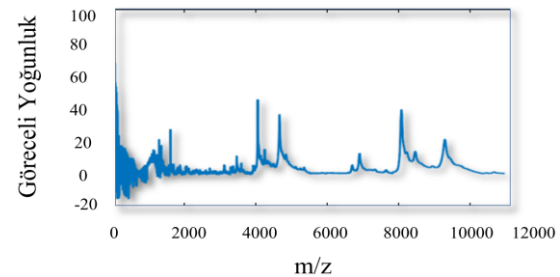
3. MATERYAL VE METOT

3.1. SELDI-TOF-MS Verileri

Çalışmada FDA-NCI web sitesinde bulunan Düşük Çözünürlüklü SELDI-TOF Veri Kümeleri kullanılmıştır. Bu düşük çözünürlüklü veri setleri elle işlenmiş 4-3-02 yumurtalık çalışması seti ve robotik şekilde işlemenin spektral kalite üzerindeki etkisini araştırmak amacıyla oluşturulmuş 8-7-02 çalışma setleridir. Her iki veri seti 15154 boyutludur. Yapılan çalışmada 4-3-02 verisinin 100 kontrol 100 kanser örneği kullanılmıştır. 8-7-02 çalışma seti ise 91 normal, 162 kanserden oluşmaktadır [24]. 4-3-02 veri setine ait kanserli ve kontrol verisi örnek grafikleri Şekil-1,2'de görülmektedir. Grafiğin y eksenini göreceli yoğunluğu, x eksenini m/z değerini ifade etmektedir.



Şekil 1. 4-3-02 veri seti için kontrol verisi örneği



Şekil 2. 4-3-02 veri seti için kanser verisi örneği

3.2. Temel Bileşenler Analizi

Büyük boyutlu veriler üzerinde örüntü tanıma veya sınıflandırma algoritmalarını uygularken bilgisayarın işlem yükünü azaltmak için öncelikli olarak yapılması gereken boyut azaltma tekniklerinin kullanılmasıdır.

PCA, büyük boyutlu verileri temel özelliklerini yakalayarak daha az boyutlu temsil edilmesini sağlayan istatistiksel bir metottur [25].

PCA uygulanırken yapılan işlemler beş adımda gösterilebilir [26]:

1. Ortalamaları Hesaplama: Verinin ortalamaları alınır (Eşitlik 1).

$$\bar{x} = \frac{1}{n} \sum_{j=1}^n x_j \quad (1)$$

2. Ortalamadan Çıkarma: Veri setlerinin her bir boyutu kendi ortalamalarından çıkarılır (Eşitlik 2). Çalışmamızda elde edilen yeni B verileri 15154x200 ve 15154x253 boyutundadır.

$$B = x - \bar{x} \quad (2)$$

3. Kovaryans Matris: İki değişkenin birlikte ne kadar değiştiklerinin ölçüsü olan 200x200 ve 253x253 boyutunda C kovaryans matrisleri hesaplanır (Eşitlik 3).

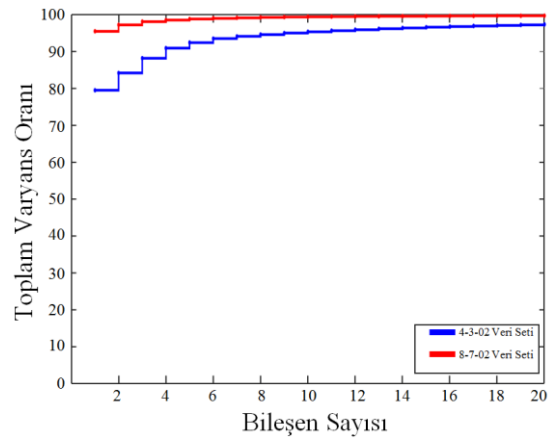
$$C = \frac{1}{n-1} B^* B \quad (3)$$

4. Öz değerler ve Özvektörler: C matrisi ile çarpıldığında yönü değişmeden kalan vektörlere denir ve Eşitlik 4'teki gibi gösterilir. Öz vektörleri hesaplayabilmek için matris formunun karesel olması gerekir bu sebeple kovaryans matris hale getirilerek işlem gerçekleştirilir.

$$CV = VD \quad (4)$$

5. Bileşenlerin Seçilmesi ve Özellik Vektörünün Oluşturulması: Bu aşamada veri setini en iyi temsil edecek bileşenler seçilir. Neticede elde

edilecek veri seti orijinal veri setinden daha az boyuta indirgenmiş olur. Şekil 3'te PCA uygulanmış veri setlerinin ilk 20 bileşenin varyans oranı görülmektedir. Mavi renkle gösterilen 4-3-02 veri seti için en yüksek varyansa sahip ilk 20 temel bileşenin tüm bileşenlere oranı %97,7 iken aynı şekil üzerinde kırmızı çizgi ile ifade edilen 8-7-02 veri seti için 99,8'dir.



Şekil 3. Veri setleri için ilk 20 bileşenin varyans oranı

3.3. Yapay Sinir Ağları

YSA, insan nöronlarının öğrenme yeteneğinin matematiksel olarak modellenmesiyle oluşturulmuş bir algoritmadır. YSA mimarisi, bünyesinde bulundurduğu yapay sinir hücrelerinin katmanlar halinde paralel şekilde bir araya gelerek bağlanmasıyla oluşturulur ve problemin türüne göre farklı katmanlardan oluşan yapı kullanılır. YSA'da hücreler ağırlık değerleri ile birbirlerine bağlıdır. Giriş katmanına verilen değerler bağlantı ağırlıkları ile çarpılıp gizli katmana iletilir (Şekil 4). Gizli katman sayısı ve bu katmanlardaki nöron sayısı problemin türüne göre deneme yanılma yoluyla belirlenir. Gizli katmanındaki bilgi işlenerek çıktı katmanına gönderilir [27].

Deneyler için MATLAB R2020b versiyonu kullanılmıştır. Oluşturulan YSA modelindeki ara katmandaki nöron sayısını saptamak için veriyi en çok temsil eden 20 girişli yapı kullanılmıştır.

Nöron sayısı 1'den 25'e kadar denenmiş ve 10 nöron olarak saptanmıştır.

Girdilere uygun çıktıları elde etmek için ağırlık değerlerinin iyi belirlenmesi gerekir. Ağın öğrenme fonksiyonunu gerçekleştirmesi doğru ağırlıkların bulunması işlemi ile mümkündür. T_i olması istenen çıkış, O_i gerçek çıktı, α öğrenme oranı ve x_j ise j. ağ girdisi olmak üzere ağırlık değişiminin hesaplanması Eşitlik 5'te görülmektedir [26,28].

$$\Delta w_{ij}^k = \alpha (T_i^k - O_i^k) x_j^k \quad (5)$$

Herhangi bir k. katmandaki örnek çifti için toplam hata ve ortalama hata fonksiyonu (Eşitlik 6,7) verilmektedir.

$$E = \frac{1}{2} \sum (T_k - O_k)^2 \quad (6)$$

$$E = \frac{1}{2p} \sum_p \sum_j (T_j - O_j)^2 \quad (7)$$

Önerdiğimiz çalışmada, giriş katmanından çıkış katmanına tek yönlü bilgi iletimi yapılmasına dayanan İleri Yönlü Yapay Sinir Ağı Modeli kullanılmıştır.

Geliştirilen YSA modelinde en başarılı öğrenme metodunu Scaled Conjugate Gradient (SCG) geri yayılım algoritması olduğu belirlenmiştir. SCG, Levenberg-Marquardt (LM) algoritmasından bilinen model güven bölgesi yaklaşımını Conjugate Gradient (CG) ile kombine ederek oluşturulur. Bu yöntemde adım boyutunu tahmin etmede satır arama tekniği yerine tasarım parametreleri, her yineleme kullanıcısından bağımsız olarak güncellenir. Eşitlik 8'de görüldüğü üzere s Hesse matrisi yaklaşımı, E toplam hata fonksiyonudur. E' , E nin gradyanıdır. λ_k ve σ_k ölçekleme faktörleridir.

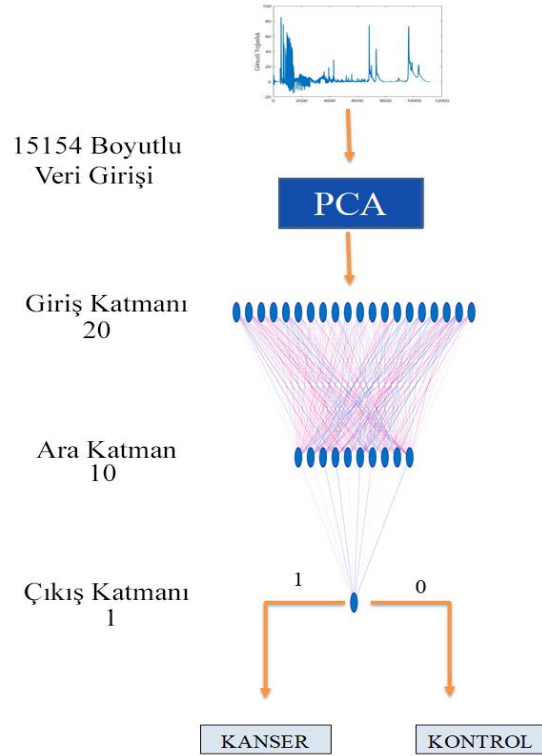
$$s_k = \frac{E'(w_k + \sigma_k p_k) - E'(w_k)}{\sigma_k} + \lambda_k p_k \quad (8)$$

SCG Algoritması dışında diğer öğrenme algoritmaları (BP, LM, CG gibi) ile de deneyler

yapılmıştır. Çalışmada kullanılan algoritma, başarı oranı daha yüksek olduğu için SCG kullanılmıştır.

Neticede oluşturulan YSA mimarisi;

- * 20 boyutlu elemanın verildiği giriş katmanı,
- * Tansig (Hyperbolic tangent sigmoid transfer function) transfer fonksiyonu kullanılarak oluşturulan 10 nöronlu gizli katman,
- * Kontrol (0) ve kanser (1) olmak üzere bir ve sıfırdan oluşan tek nöronlu çıktı katmanı şeklindedir.



Şekil 4. Önerilen model

4. SONUÇLAR

YSA oluşturulması 10 tekrarlı çapraz doğrulama ile 200 örnekten oluşan 4-3-02 verisi, 180'i eğitim ve 20'si test olarak ayrılmıştır. Bu şartlar altında, giriş katmanı için 4 bileşenden 20 bileşene kadar

sınıflandırma yapılmıştır ve 20. bileşende en yüksek sınıflandırma yüzdesine ulaşılmıştır. Elde edilen sonuçlar, bileşen sayısı ve doğruluk yüzdeleri gösteren Çizelge-1'de yer almaktadır. 20 bileşen ile ağ eğitildiğinde elde edilen doğruluk eğitim için %99,4'e, test için %95'e eşittir.

Çizelge 1. 4-3-02 veri seti için eğitim ve test doğruluğunun seçilen bileşen sayısı ile değişimi

Bileşen sayısı	Eğitim doğruluğu (%)	Test doğruluğu (%)
4	70,0	80,0
5	78,9	80,0
6	92,2	80,0
8	90,6	90,0
10	91,7	90,0
12	96,7	70,0
14	93,3	90,0
16	97,2	90,0
18	98,9	85,0
20	99,4	95,0

Aynı deneyler 91 kontrol ve 162 kanser verisinden oluşan 8-7-02 veri seti için yapılmıştır ve 10 tekrarlı çapraz doğrulama ile 228'i eğitim 25 test olarak şekilde ayrılmıştır ve sonuçlar Çizelge 2'de görülmektedir. 20 bileşen ile ağ eğitildiğinde elde edilen doğruluk eğitim için %96,4'e, test için %100'e eşittir.

Çizelge 2. 8-7-02 veri seti için eğitim ve test doğruluğunun seçilen bileşen sayısı ile değişimi

Bileşen Sayısı	Eğitim doğruluğu (%)	Test doğruluğu (%)
4	91,1	95,0
5	93,3	85,0
6	99,4	85,0
8	96,1	95,0
10	94,4	95,0
12	98,3	85,0
14	95,6	85,0
16	99,4	90,0
18	95,0	100
20	96,4	100

Bu veri setinin gruplarındaki örnek sayısı eşit olmadığı için 91 normal verinin tamamı ve rastgele olarak 162 kanser verisinden 91 veriyi alarak yeniden sınıflandırma yaptığımızda 20 bileşen için yine %100 doğruluk elde edilmiştir.

PCA uygulanmadan 15154 boyutlu veri setleri sınıflandırıldığında elde edilen eğitim ve test doğruluğu Çizelge 3'te görülmektedir.

Çizelge 3. Verinin tamamı kullanılarak elde edilen eğitim ve test doğruluğu

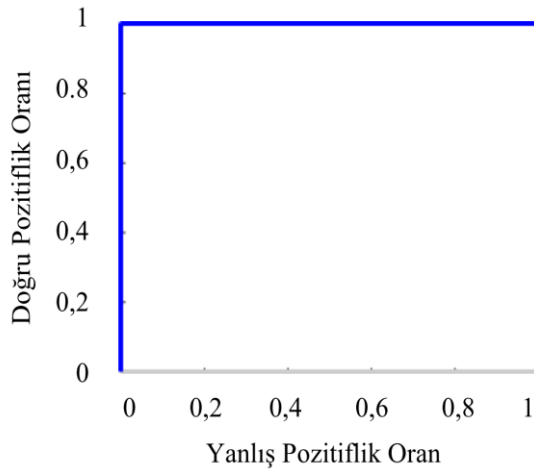
Veri Seti	Eğitim doğruluğu (%)	Test doğruluğu (%)
4-3-02	98,9	95,0
8-7-02	100	100

Oluşturulan YSA modelinin performans değerlendirmesi yapılmasına olanak sağlayan yüzde hata ve sınıflandırılmayan örnek oranını gösteren hata (confusion) matrisleri aşağıda verilmiştir. Matrislerdeki yeşil kareler doğru yüzdeyi, kırmızı kareler doğru olmayan yüzdeyi, koyu gri alan tüm doğruluğu ifade eder. '1' olarak gösterilen grup kanser verisini '0' olarak gösterilen grup kontrol grubunu göstermektedir. 20 temel bileşen ile ifade edilen, 4-3-02 veri seti için eğitim hata matrisi Çizelge 4'te görülmektedir.

Çizelge 4. 4-3-02 veri seti için eğitim hata matrisi

Gerçek Sınıf	1	86 47,8%	1 0,6%	98,9% 1,1%
	0	0 0,0%	93 51,7%	100% 0,0%
		100% 0,0%	98,9% 1,1%	99,4% 0,6%
		0	1	Hedef Sınıf

Alıcı Çalışma Karakteristikleri (Receiver Operating Characteristic-ROC) sınıflandırma modelinin performansı hakkında bilgi verir. ROC eğrisinin altında kalan alanın (Area Under the Curve- AUC) mümkün olduğunda yüksek olması istenir. Şekil 5'te oluşturulan YSA için elde edilen ROC eğrilerinden ağır başarılı bir eğitim gerçekleştirildiği görülmektedir.



Şekil 5. 4-3-02 veri seti için ROC eğrisi

4-3-02 veri setinin 20 temel bileşeni için elde edilen test hata matrisi Çizelge 5'te görülmektedir. Test sonucunda %95 doğruluk elde edilmiştir.

Çizelge 5. 4-3-02 veri seti için test hata matrisi

Gerçek Sınıf	1	13 65,0%	0 0,0%	100% 0,0%
	0	1 5,0%	6 30 %	85,7% 14,3%
		92,9% 7,1%	100% 0,0%	95% 5,0%
		0	1	
		Hedef Sınıf		

5. TARTIŞMA

SELDI-TOF-MS yumurtalık kanserinin erken tanısı için etkili bir metottur. Elde edilen verinin yüksek boyutlu olması dezavantaj olsa da PCA gibi ön işleme adımlarıyla bu sorunun üstesinden gelinebilir. SELDI-TOF-MS üzerinde çalışmaya devam edilen bir yöntemdir ve kanser için gelecekte potansiyel teşhis aracı olması beklenmektedir [11]. Bu teknikte elde edilen verilerin hatasız yorumlanması kritik önem taşır.

Sinir ağları ve diğer sınıflandırma algoritmaları kullanılarak çeşitli kanser türlerini sınıflandırmaya yönelik çalışmalar mevcuttur ve yüksek doğruluk elde edilmiştir. Bu çalışmalar mevcut sınıflandırma algoritmalarının SELDI-TOF-MS verilerini gruplandırmadaki başarısını göz önüne koymaktadır. Literatürde SELDI-TOF-MS verileri kullanılarak YSA ile kanser tespit edilmesine yönelik çalışmalar Çizelge 6'da görülmektedir. Yapılan bir çalışmada, YSA ile yumurtalık kanserini tespit etmek için ön işleme adımı olarak Rank Özellikleri ve PCA kullanılmış, verinin boyutu 15000'den 66'ya indirgenmiş ve optimize edilerek 7 nöron ile yüksek çözünürlüklü yumurtalık kanseri verisi %100 doğrulukta sınıflandırılmıştır [22]. Yumurtalık kanseri verilerinin tespit ederek ve student t-test uygulayarak sınıflandıran çalışmalar da mevcuttur. Araştırmacılar bu çalışmalarla yüksek boyutlu SELDI-TOF-MS verisinin boyutunun küçülterek ağır işlem yükünü azaltılmasına katkı sağlamışlardır [20].

Ön işlemeye tabi tutulmadan yumurtalık kanserinin sınıflandırıldığı çalışmalar için de yüksek doğruluk elde edilmiştir [21,23].

Literatürdeki YSA ile yapılmış çalışmalar Yüksek Çözünürlüklü SELDI-TOF-MS verilerine odaklanmışlardır. Düşük çözünürlüklü SELDI-TOF-MS kullanılarak yapılan çalışmada verinin tamamı kullanılarak 15154 giriş katmanı, 15 nörondan oluşan gizli katman ve 2 çıkışlı bir ağ yapısı oluşturmuşlardır [23]. Bu çalışmada kullanılan örnek sayısından yola çıkarak 8-7-02 veri seti olduğu anlaşılmaktadır bu veri seti ile elde ettiğimiz doğruluk oranı PCA kullanılarak veya kullanılmayarak daha yüksektir.

Çizelge 6. Yumurtalık kanseri için SELDI-TOF-MS verilerinin YSA ile Sınıflandırılması üzerine yapılmış çalışmalar

Veri	Yöntem	Doğruluk
[20] FDA-NCI Yüksek çözünürlüklü 95 kontrol 121 kanser	1. Student t-testi 2. İleri beslemeli sinir ağları	Kanser: %98,50 Kontrol: %99,16
[21] FDA-NCI Bilgi verilmemiştir	1. İleri Beslemeli Sinir Ağları	%98
[22] FDA-NCI Yüksek çözünürlüklü 95 normal 121 kanser	1. Rank özellikleri 2. PCA 3. Nöron optimizasyonu 4. İleri beslemeli sinir ağları	%100
[23] NCI PBSII Data Düşük çözünürlüklü 91 normal 162 kanser	1. Optimum nöron sayısı için Taguchi 2. Yapay sinir ağları	%98,7

Çalışmamızda, mevcut çalışmalardan farklı olarak, yumurtalık kanserinin sınıflandırılma başarısına, SELDI-TOF-MS verilerinin elde edilmiş biçiminin etkisini inceledik. Çizelge 7’de iki veri seti için 15154 boyutlu verinin tamamı ve 20 temel bileşen kullanılarak ağın eğitilmesi sonucu elde edilen doğruluk yüzdeleri görülmektedir.

Verilerin alındığı örneklerin farklı olduğunu belirterek, robot kolla hazırlanmış 8-7-02 çalışma seti manuel oluşturulmuş 4-3-02 çalışma setine göre daha yüksek başarıyla sınıflandırılmıştır. Bu iki veri setinin deney süreçlerindeki değişikliklerin sınıflandırma başarısına etkisini göstermektedir.

Çizelge 7. 4-3-02 ve 8-7-02 veri setleri için ağın doğruluk yüzdesi

YSA giriş katmanı		Eğitim (%)	Test (%)
4-3-02 Veri seti	PCA uygulanmış	99,4	95,0
	Verinin tamamı	98,9	95,0
8-7-02 Veri seti	PCA uygulanmış	99,6	100
	Verinin tamamı	100	100

Ağın işlem yükünü azaltmak için veri setleri PCA ile ön işlemeye tabi tutulmuştur. Düşük çözünürlüklü veriler için %70’ten yüksek sınıflandırma başarısı elde etmek için ilk 4 temel bileşen ile sınıflandırma yapılmasının yeterli olduğu görülmektedir (Çizelge 1 ve Çizelge 2).

PCA işlem yükünü azaltarak ağın hızlanmasını sağlamıştır Aynı şartlar altında 15154 boyutlu veri için eğitim süresi 21 saniye sürerken 20 temel bileşen ile 2,7 saniyeye düşürülerek daha hızlı bir sistem oluşturulmuştur (Çizelge 8).

Ek olarak, 4-3-02 veri setinin tamamı kullanılarak eğitilen ağın başarısı %95’e eşittir. Yani PCA kullanarak oluşturulan veri ile yapılan sınıflandırmanın başarısı verinin tamamı kullanılarak yapılan sınıflandırmanın başarısına eşittir. Aynı durum 8-7-02 veri seti için de geçerlidir 15 saniye süren işlem PCA ile 2,8 saniyeye düşürülüp her iki durum için %100 başarı elde edilmiştir. Bu da kullanılan PCA bileşenlerinin, daha az boyutla veriyi temsil ederek başarılı bir sonuç elde ettiğini göstermektedir.

Çizelge 8. PCA uygulandığında ve uygulanmadığında geçen süre

	PCA uygulanmış	PCA uygulanmamış
4-3-02	2,74 s	21,76 s
8-7-02	2,83 s	15,09 s

Ayrıca PCA uygulanmadan sınıflandırma yapıldığında 8-7-02 veri seti içerdiği örnek sayısı 53 tane daha fazla olmasına rağmen sınıflandırılması için gerekli olan sürenin daha kısa olduğu görülmektedir. Böyle bir sonuç elde edilmesi verinin robotik şekilde elde edilmesinin sınıflandırma başarısını artırmasına ek olarak bilgisayarın işlem yükünü de azalttığını göstermektedir.

Bu çalışmayla, kullanımının yaygınlaşmasını öngördüğümüz SELDI-TOF-MS verileri üzerinden hızlı ve doğru tanıya yardımcı olacak sistemler geliştirerek, tıpta hatalı yumurtalık kanseri teşhisini en aza indirmeyi hedefleyen araştırmalara katkı sağlamayı amaçlamaktayız.

6. KAYNAKLAR

1. Global Cancer Observatory: Cancer Today: International Agency for Research on Cancer. <https://gco.iarc.fr/today>, Erişim Tarihi: 24.02.2021, 2020, Lyon, Fransa.
2. Global Cancer Observatory: Cancer Tomorrow. Lyon, Fransa: International Agency for Research on Cancer. <https://gco.iarc.fr/tomorrow>, Erişim Tarihi 24.02.2021, 2020, Lyon, Fransa.
3. Hamidou, Z., Causeret, S., Dabakuyo, T.S., Gentil, J., Arnould, L., Roignot, P., Altwegg, T., Poillot, M.L., Bonnetain, F., Arveux, P., 2010. Population-based Study of Ovarian Cancer in Côte d'Or: Prognostic Factors and Trends in Relative Survival Rates Over the Last 20 Years. *BMC Cancer*, 10, 622.
4. Torre, L.A., Trabert, B., DeSantis, C.E., Miller, K.D., Samimi, G., Runowicz, C.D., Gaudet, M.M., Jemal, A., Siegel, R.L., 2018. Ovarian Cancer Statistics. *CA: A Cancer Journal for Clinicians*, 68, 284-296.
5. Jacobs, I.J., Menon, U., 2004. Progress and Challenges in Screening for Early Detection of Ovarian Cancer. *Molecular & Cellular Proteomics*, 3(4), 355-366.
6. Zhang, D.Y., Ye, F., Gao, L., Liu, X., Zhao, X., Che, Y., Wang, H., Wang, L., Wu, J., Song, D., Liu, W., Xu, H., Jiang, B., Zhang, W., Wang, J., Lee, P., 2009. Proteomics, Pathway Array and Signaling Network-based Medicine in Cancer. *Cell Div.*, 4, 20.
7. Hood, L.E., Omenn, G.S., Moritz, R.L., Aebersold, R., Yamamoto, K.R., Amos, M., Hunter-Cevera, J., Locascio, L., 2012. New and Improved Proteomics Technologies for Understanding Complex Biological Systems: Addressing a Grand Challenge in the life Sciences. *Proteomics*, 12, 2773-2783.
8. Shruthi, B., Vinodhkumar, P., Selvamani, M., 2016. Proteomics: A New Perspective for Cancer. *Adv. Biomed. Res.* 5, 67.
9. Taşkın, V., Doğan, B., Ölmez, T., 2012. Yumurtalık Kanserinin Kütle Spektrometresi Verilerinden Kısmi En Küçük Kareler Yöntemi ile Teşhisi. *Biyomedikal Mühendisliği Ulusal Toplantısı (BİYOMUT)*, 151-154, İstanbul.
10. Issaq, H.J., Veenstra, T.D., Conrads, T.P., Felschow, D., 2002. The SELDI-TOF MS Approach to Proteomics: Protein Profiling and Biomarker Identification. *Biochemical and Biophysical Research Communications*, 292(3), 587-592.
11. Poon, T.C., 2007. Opportunities and Limitations of SELDI-TOF-MS in Biomedical Research: Practical Advices. *Expert Review of Proteomics*, 4(1), 51-65.
12. Muthu, M., Vimala, A., Mendoza, O.H., Gopal, J., 2016. Tracing the Voyage of SELDI-TOF MS in Cancer Biomarker Discovery and its Current Depreciation Trend-need for Resurrection? *TrAC Trends in Analytical Chemistry*, 76, 95-101.
13. Tang, K.L., Li, T.H., Xiong, W.W., Chen, K., 2010. Ovarian Cancer Classification Based on Dimensionality Reduction for SELDI-TOF Data. *BMC Biyoinformatik* 11, 109.
14. Bougioukos, P., Cavouras, D., Daskalakis, A., Kossida, S., Nikiforidia, G., Bezezerianos, A., 2006. Feature Extraction and Analysis of Prostate Cancer Proteomic Mass Spectra for

- Biomarkers Discovery. General Secretariat for Research and Technology, 1.
15. Hu, Q.Y., Wang, K.Z., Ding, Y.H., Zheng, L.F., Liang, S.H., Lei, Z.M., Fu, W.G., Yan, L., 2010. Application of SELDI-TOF-MS Coupled with an Artificial Neural Network Model to the Diagnosis of Pancreatic Cancer. *Laboratory Medicine*, 41(11), 676–681.
 16. Wegdam, W., Moerland, P.D., Meijer, D., Jong, S.M de., Hoefsloot H.C.J., Kenter, G.G., Buist, M.R., Aerts, J.M.F.G., 2012. A Critical Assessment of SELDI-TOF-MS for Biomarker Discovery in Serum and Tissue of Patients with an Ovarian Mass. *Proteome Sci.*, 10, 45.
 17. Simsek, C., Sonmez, O., Yurdakul, A.S., Ozmen, F., Zengin, N., Keyf, A.I., Ozturk, C., 2013. Importance of Serum SELDI-TOF-MS Analysis in the Diagnosis of Early Lung Cancer. *Asian Pacific Journal of Cancer Prevention*, 14(3), 2037-2042.
 18. Wu, J., Ji, Y., Zhao, L., Ji, M., Ye, Z., Li, S., 2016. A Mass Spectrometric Analysis Method Based on PPCA and SVM for Early Detection of Ovarian Cancer. *Computational and Mathematical Methods in Medicine*, 2016, 6169249.
 19. Cohen, A., Messaoudi, C., Badir, H., 2018. A New Wavelet-based Approach for Mass Spectrometry Data Classification, in *New Frontiers of Biostatistics and Bioinformatics*. Springer, Cham, 175-189.
 20. Thakur, A., Mishra, V., Jain, S.K., 2011. Feed Forward Artificial Neural Network: Tool for Early Detection of Ovarian Cancer. *Scientia pharmaceutica*, 79(3), 493-506.
 21. Sharma, A., Singh, S., 2016. Neural Network for Diagnosis of Ovarian Cancer Based on Proteomic Patterns in Serum. *Journal of Scientific and Technical Advancements*, 2(2), 25-27.
 22. Pei, S., Tong, L., Li, X., Jiang, J., Huang, J., 2017. Feed-forward Network for Cancer Detection, in *2017 13th International Conference on Natural Computation. Fuzzy Systems and Knowledge Discovery (ICNC-FSKD)*, (697-701). IEEE.
 23. Rahman, M.A., Muniyandi, R.C., Islam, K.T., Rahman, M.M., 2019. Ovarian Cancer Classification Accuracy Analysis Using 15-Neuron Artificial Neural Networks Model, in *2019 IEEE Student Conference on Research and Development (SCOREd)*, IEEE, 33-38.
 24. FDA-NCI Klinik Proteomik Program Veri Bankası: <https://home.ccr.cancer.gov/ncifdaproteomics/ppatterns.asp>.
 25. Akkücü, U., 2009. Birçok Boyutlu Ölçekleme Tekniği Olarak Torgersen Ölçekleme Yöntemi ve Temel Bileşenler Analizi ile Karşılaştırması. *Dumlupınar Üniversitesi, Sosyal Bilimler Dergisi*, (25), 311-322.
 26. Brunton, S.L., Kutz, J.N., 2019. *Data-driven Science and Engineering: Machine Learning, Dynamical Systems and Control*. Cambridge University Press.
 27. Gümüş, V., Soydan, N., Simsek, O., Aköz, M., Kırkgöz, M., 2016. Yağış-Akış İlişkisinin Belirlenmesinde Farklı Yapay Sinir Ağı Yöntemlerinin Karşılaştırılması. *Çukurova Üniversitesi Mühendislik-Mimarlık Fakültesi Dergisi*, 28(1), 37-50.
 28. Ataseven, B., 2013. Yapay Sinir Ağları ile Öngörü Modellemesi. *Öneri Dergisi*, 10(39), 101-115.