

---

Makale / Research Paper

---

## Yapay Sinir Ağları İle Göğüs Kanseri Tahmini

Zehra KARAPINAR ŞENTÜRK<sup>1</sup>, Arafat ŞENTÜRK<sup>1</sup>

<sup>1</sup>Mühendislik Fakültesi, Bilgisayar Müh. Bölümü, Düzce Üniversitesi, Düzce – Türkiye

**Geliş/Received:** 10.11.2015

**Düzeltilme/Revised:** 25.11.2015

**Kabul/Accepted:** 25.11.2015

**Özet:** Dünya Sağlık Örgütü'nün verilerine göre, hem gelişmiş hem de gelişmekte olan ülkelerde, göğüs kanseri kadınlar arasındaki en yaygın kanser türüdür. Kentleşme ve batılı hayat tarzına uyum sağlamanın gelişen dünyada bu kanser türünü tetiklediği düşünülmektedir. Çoğu kanser vakası hastalığın son evrelerinde teşhis edilebilmekte ve tedaviler bu evreden sonra çoğu zaman cevap verememekte ve hasta kayıpları yaşanmaktadır. Bu sebeple, diğer kanser türlerinde olduğu gibi, göğüs kanserinin de erken tanısı hayati önem taşımaktadır. Bu çalışmada, göğüs kanserinin erken tanısına katkıda bulunabilmek amaçlanmıştır. Hastalara belirtileri doğrultusunda göğüs kanseri olup olmadıklarına dair teşhis konulmuştur. Bu amaçla, veritabanında mevcut olan, önceden teşhisi konulmuş vakaların verileri kullanılarak, henüz teşhisi belli olmayan örneklerle tahmin yoluyla teşhis konulmuştur. Tahmin işi, tasarlanan bir yapay sinir ağına yaptırılmış ve doğruluk oranı %77 olarak bulunmuştur.

**Anahtar kelimeler:** Yapay sinir ağları, Meme kanseri, Erken tanı.

---

## Neural Networks with Breast Cancer Forecast

**Abstract:** According to the data provided by World Health Organization, breast cancer is the most common cancer type among women in both developed and developing countries. Urbanization and western life style are thought to trigger this cancer type. Most of the cancer events are diagnosed at the late phases of the illness and treatments mostly does not work after this time, and patient deaths faced. Therefore, as in other cancer types, the early diagnosis of breast cancer is vital. In this paper, to contribute the early diagnosis of breast cancer is aimed. Patients are diagnosed by the help of their symptoms. For this purpose, by the help of the data of diagnosed patients, the diagnosis of new samples are tried to be predicted. An artificial neural network is designed for prediction and its accuracy rate is 77%.

**Keywords:** Artificial neural networks, Breast cancer, Early diagnosis.

---

### 1. GİRİŞ

Geçmişten günümüze doğru gelindikçe sahip olunan verilerde oldukça büyük bir artış olmuştur ve buna bağlı olarak da verilerin kontrolü ve yönetimi hususunda da zorluklar ortaya çıkmıştır. Bilgisayar teknolojilerinin ucuzlamasıyla, çok miktarda veri sabit disklerde depolanabilmiş ve veritabanı sorguları ile nispeten yönetilebilmiş. İşlemler daha kısa sürede ve daha az hata oranıyla yapılabilmiş. Ancak zamanın gereksinimleri bunun da ötesine gitmeyi gerektirince ileriye yönelik tahmin yapabilen mekanizmalar doğmaya başlamıştır. Eldeki veriden daha çok bilgi edinebilme ihtiyacı her alanda hissedilmiş ve geliştirilen yöntemlerle, eldeki veri üzerinde yapılan analizler sayesinde geleceğe yönelik gerçekçi tahminler yapılmıştır. Aslında bu sürece “veri madenciliği” deniyor. Yani, ham veriden altın veriye ulaşma süreci. Veri madenciliği, veriyi kullanışlı bilgi elde

*Bu makaleye atf yapmak için*

Karapınar Ş.Z. ve Şentürk A, “Yapay Sinir Ağları ile Göğüs Kanseri Tahmini”, El-Cezerî Fen ve Mühendislik Dergisi, 2016, 3(2); 345-350.

*How to cite this article*

Karapınar Ş.Z. ve Şentürk A, “Neural Networks with Breast Cancer Forecast”, El-Cezerî Journal of Science and Engineering, 2016, 3(2); 345-350.

etmek üzere farklı açılardan analiz etme ve özetleme olarak tanımlanabilir. Buradaki bilgi, geliri artırma ya da harcamaları kısma amaçları ile kullanılabilir. Teknik olarak ise veri madenciliği, büyük ilişkisel veritabanlarındaki düzinelerce alan arasında belli ilişkiler veya modeller keşfetme sürecidir [1].

Donald Hebb (1949) bugünün sinir ağı teorisinin temelini oluşturduğu bilinmektedir. Nörolog olan Hebb, beynin nasıl öğrendiği ile ilgili çalışmalar yapmıştır. Çalışmalarına beynin en temel birimi olan sinir hücrelerini ele alarak başlamıştır. İki sinir hücresinin birbirleriyle nasıl bir korelasyon sergilediklerini incelemiş ve sinir ağı teorisini bu temel üzerine oturtmuştur. Bu temel kuşkusuz tek gerçek değildir. Çünkü beynin nasıl bir çalışma sergilediği şu an dahi teoriler yardımıyla açıklanmaktadır. Ancak Hebb'in yardımıyla bu fikir ile yola çıkılmış ve günümüzdeki yüzlerce ayrı teoriyle geniş bir yelpazeye hitap eder hale gelmiştir. Şu an gerçek yaşamda kullanılan ve başarı oranı %99'lar ile ifade edilen bir sürü yapay sinir ağı (YSA) modeli mevcuttur. Tüm geliştirilen modeller bilgisayar dünyasında "çözumsuz" veya "nı karmaşık" olarak nitelendirilen problemlerin çözümünü hedeflemekte ve hatta bir kısmını başarıyla çözmektedir [2].

Paralellik, hata toleransı, gerçekleştirme kolaylığı, öğrenilebilirlik, adaptasyon, doğrusal olmayan yapı ve gerçek zamanlı kullanım yapay sinir ağlarının temel özellikleridir. YSA, insan beyninin yapısından esinlenerek geliştirilmiş, öğrenilebilir ve analiz edilebilir bir algılayıcıdır. Literatürde YSA ile makine öğrenmesi, kümeleme, sınıflandırma, tahmin, karakter tanıma, ses işleme [3], görüntü işleme [4, 5], en uygun şekle sokma ve veri filtreleme gibi çok sayıda uygulamaya rastlamak mümkündür. Öğrenme olayı biyolojik sınırlarda sinapslarda gerçekleşirken, yapay sinir ağlarında öğrenme ağırlıklarda gerçekleşmektedir. Kullanılan algoritmadan bağımsız olarak her tür veriye uygulanabilmesi, hatayı minimize etmesi ve öğrenilebilmesi ve buna bağlı olarak gerçekçi tahminler yapabilmesi yapay sinir ağlarının her alanda tercih edilmesinin başlıca sebeplerindedir.

Göğüs kanseri özellikle kadınların ölüm sebeplerinden biridir [6, 7]. Bu kanser türü, her sekiz kadından birinde yaşamları boyunca etkili olmakta, erkeklerde ise nadiren görülmektedir. Göğüs kanseri kötü huylu bir tümördür ve göğüs hücrelerinde gelişir. Bilim adamları, göğüs kanserinin gelişmesine yardımcı olan yaşlanma, genetik risk faktörü, aile geçmişi, çocuk sahibi olamama, obezite ve benzeri birçok etkeni bilmelerine rağmen henüz bu etkenlerin etkilerini tamamen ortadan kaldıracak çözümlere ulaşamamışlardır [9, 10]. Erken dönemde tespiti ve tümörün lokalizasyonu, ölüm oranını azaltmak için tek yoldur. X-ışını mamografi göğüs kanseri taraması için en çok kullanılan tanı aracıdır. Ancak, göğüs kanseri tespiti için X-ışınının kullanımı önemli derecede sınırlamaları vardır [11].

Klinik raporlarda, YSA'ların çok farklı özellikleri kullanarak göğüs kanseri tanısında uygulanmıştır [12, 13]. Literatürde kanser tanısının YSA'lar ile tahmin edilmeye çalışıldığı pek çok çalışma mevcuttur. Farklı kanser türleri üzerinde denemeler yapılmıştır. YSA'lar, diğer sezgisel yöntemlerle birleştirilerek etkinliği artırılmaya çalışılmıştır. Sadece bu alanda değil çok geniş yelpazede çalışılan bir konu durumundadır. Fogel ve ark. [14] sinir ağları ile radyoaktif özellikleri ve hastanın yaşını kullanarak göğüs kanseri tespiti için verileri eğitmişlerdir. Revett ve ark. [15] ve Gorunescu ve ark. [16] kaba setleri ve olasılıklı sinir ağları içeren melez bir modele dayalı göğüs kanseri için tıbbi bir karar destek sistemi tasarlamışlardır. Hsiao ve ark. [17] göğüs tümörlerinin iyi huylu ya da kötü huylu olup olmadığının belirlenmesi için damarlanma indeksleri (harmonik ve harmonik olmayan 3D güç Doppler görüntüleme) kullanarak bir MLP sınıflandırıcı eğitimi gerçekleştirmişlerdir.

Bu çalışmanın amacı, göğüs kanseri hastalığının teşhisi konusunda bir analiz yapmaktır. Böylelikle, hayati derecede önemli olan teşhis sürecindeki zaman kaybı ortadan kalkacak ve kalan süre hastalığın etkili tedavisinde kullanılacaktır.

## 2. Materyal ve Yöntem

### 2.1 Kullanılan Veritabanı

Uygulamada kullanılan veritabanı UCI Machine Learning Repository'den alınan Breast Cancer veritabanıdır. Madison'daki Wisconsin Üniversitesi Hastanesi'nden Dr. William H. Wolberg tarafından oluşturulmuştur. Veritabanında 699 örneğe ait veri ve bunların ait oldukları 2 sınıf mevcuttur. Sınıflar benign (iyi huylu) ve malignant (kötü huylu) şeklindedir. Örneklerin 458'i benign sınıfına aitken, 241'i malignant sınıfına aittir. Veri setinde toplam 11 öznitelik var. Bunlar ve değer aralıkları Tablo1'de gösterilmektedir.

**Tablo 1.** Veri Setleri ve değer aralıkları

Veri Seti	Değer Aralığı
1	Sample code number
2	ID Numarası
3	Clump Thickness
4	1 - 10
5	Uniformity of Cell Size
6	1 - 10
7	Uniformity of Cell Shape
8	1 - 10
9	Marginal Adhesion
10	1 - 10
11	Single Epithelial Cell Size
12	1 - 10
13	Bare Nuclei
14	1 - 10
15	Bland Chromatin
16	1 - 10
17	Normal Nucleoli
18	1 - 10
19	Mitoses
20	1 - 10
21	Class
22	2 - 4

Tahminde kullanılacak öznitelikler 2-10 arasındaki özniteliklerdir. 1. öznitelik şahsi kodu, 11. öznitelik ise sınıf bilgisini içermektedir.

### 2.2 Yapılan İş

Veritabanında hastalığın teşhisinde kullanılmak üzere 9 adet öznitelik bulunmaktadır. Dolayısıyla, giriş katmanında 9 nöron kullanılmıştır. Sonuç olarak da kullanıcılar ya benign ya da malignant sınıflarına dâhil edileceklerinden, çıkış katmanında 2 adet nöron bulunmaktadır. Ara katmanda ise 10 nöron vardır. Başlangıçta rastgele ağırlık ve bias değerleri oluşturulmuş ve eğitim boyunca bu ağırlıklar güncellenerek son değerler elde edilmiştir. Çıkış nöronlarında oluşan hatanın geri yayılımı sayesinde tahmindeki hata oranı minimuma indirilmiştir.

Aktivasyon fonksiyonu olarak kolay türevlenebilir olması sebebiyle sigmoid fonksiyonu tercih edilmiştir.

$$\text{sigmoid}(x) = \frac{1}{1+e^{-x}} \quad (1)$$

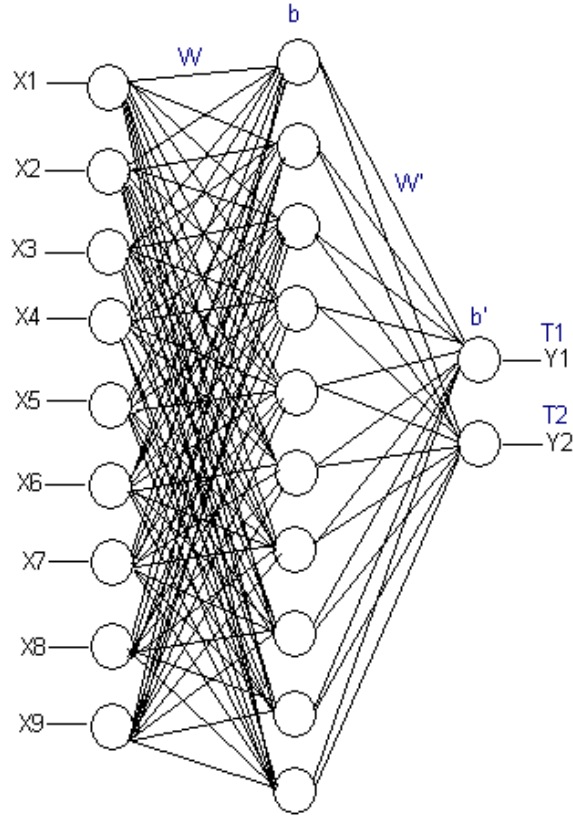
Hesaplamalar, çıkarılan aşağıdaki formüllerle yapılmıştır.

$$O_n = f\left(\sum_{i=1}^9 x_i \cdot w_{in} + b_n\right), n = 1, \dots, 10. \quad (2)$$

Burada,  $O_n$  ara katman çıkışlarını,  $w_{in}$  ve  $b_n$  ise ara katman ağırlık ve biaslarını ifade etmektedir.

$$Y_n = f\left(\sum_{i=1}^{10} O_i \cdot w'_{in} + b'_n\right), n = 1, 2. \quad (3)$$

Burada,  $Y_n$  çıkış katmanı çıkışlarını,  $W'_{in}$  ve  $b'_n$  ise çıkış katmanı ağırlık ve biaslarını ifade etmektedir. Öğrenme katsayısı 0,5 alınmıştır.



Şekil 1. Oluşturulan Yapay Sinir Ağı

Güncellemeler ise şu formüllerle gerçekleştirilmiştir:

$$\delta_i = (T_j - Y_j) \cdot Y_j \cdot (1 - Y_j)$$

$$\Delta W'_{ij}(t+1) = 0,5 \cdot \delta_j \cdot O_i, j = 1,2 \text{ ve } i = 1, \dots, 10.$$

$$W'^{yeni}_{ij} = W'^{eski}_{ij} + \Delta W'_{ij}(t+1)$$

$$\Delta W_{ij}(t+1) = 0,5 \cdot O_j \cdot (1 - O_j) \cdot \delta_{1,2} \cdot W_{ij}, X_i, j = 1, \dots, 10 \text{ ve } i = 1, \dots, 9.$$

$$W^{yeni}_{ij} = W^{eski}_{ij} + \Delta W_{ij}(t+1)$$

(4)

$$\Delta b'_{ij}(t+1) = 0,5 \cdot \delta_j, i = 1,2.$$

$$b'^{yeni}_{ij} = b'^{eski}_{ij} + \Delta b'_{ij}(t+1)$$

$$\Delta b_{ij}(t+1) = 0,5 \cdot O_i \cdot (1 - O_i) \cdot \delta_{1,2} \cdot W_{ji}, i = 1, \dots, 10 \text{ ve } j = 1, \dots, 9.$$

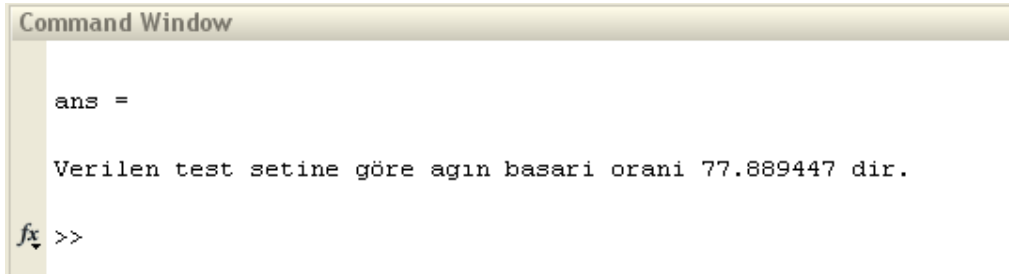
$$b^{yeni}_{ij} = b^{eski}_{ij} + \Delta b_{ij}(t+1)$$

$$hata = \frac{1}{2} \sum_{i=1}^2 (t_n - y_n)^2$$

$t_{1,2}$ , hedeflenen çıkış değerlerini ifade etmektedir.

Kodlamalar Matlab'da yapılmıştır ve Matlab'ın hazır fonksiyonları ve YSA aracı kullanılmamıştır.

Ekran çıktısı Şekil 2’de gösterilmiştir.



```
Command Window

ans =

Verilen test setine göre ağın başarı oranı 77.889447 dir.

fx >>
```

Şekil 2. Çalışmanın ekran çıktısı

### 3. Sonuçlar

Bu çalışmada, edinilen veritabanı kullanılarak yapay sinir ağları ile göğüs kanserinin teşhisine yönelik bir tahmin mekanizması oluşturulmuştur. Veritabanındaki her bir örneğe ait öznelik değerleri oluşturulan yapay sinir ağının girişlerine verilmiş ve hata oranı belli bir seviyenin altına inene kadar geri besleme yapılarak ağ eğitilmiştir. Eğitim sırasında 699 örneğin 500’ü kullanılmıştır. Kalan 199 örnek ise test aşamasında kullanılmıştır. Eğitim de kullanılan örneklerin 303’ü benign, 197’si malignant ve testte kullanılan örneklerin 155’i benign, 44’ü malignant sınıfına aittir. Ağın başarısı çoğu zaman (ağ birden fazla kere yeniden ağırlıklandırılarak çalıştırılmış ve doğruluğu denemiştir) %77’nin üzerinde çıkmaktadır.

### Kaynaklar

- [1] Hand D.J., Mannila H. ve Smyth P., “Principles of data mining”, 1st Edition, MIT Press Cambridge, (2001).
- [2] Fausett L., “Fundamentals of neural networks: architectures, algorithms and applications”, 1st Edition, Prentice-Hall International. (1994).
- [3] Uncini A., “Audio signal processing by neural Networks”, Neurocomputing, (2003), 55 (3-4), sayfa. 593 – 625.
- [4] Shi Z. ve He L., “Application of Neural Networks in Medical Image Processing”, Proceedings of the Second International Symposium on Networking and Network Security (ISNNS ’10) , China, (2010), sayfa. 2-4.
- [5] Ramirez-Quintana J.A., Chacon-Murguia M. I. ve Chacon-Hinojos J. F., “Artificial Neural Image Processing Applications: A Survey”, Engineering Letters, (2012), 20 (1), sayfa 68-81.
- [6] Klemm M., Craddock I., Leendertz J., Preece A. ve Benjamin R., “Radar-based breast cancer detection using a hemi-spherical antenna array experimental results”, IEEE Trans. Antennas Propag., (2009), 57 (6), sayfa 1692–1704.
- [7] Abbosh A., “Early breast cancer detection using Doppler frequency shift”, Asia-Pacific Microwave Conference Proceedings (APMC), Yokohama, (2010), sayfa 275-278.
- [8] Chou S. M., Lee T. S., Shao Y. E., Chen I. F., “Mining the breast cancer pattern using artificial neural networks and multivariate adaptive regression splines”, Expert Systems with Applications, (2004), 27 (1), sayfa 133-142.
- [9] Karabatak M., “Özellik seçimi, sınıflama ve öngörü Uygulamalarına yönelik birliktelik kuralı Çıkarımı ve yazılım geliştirilmesi”, Ph.D. Thesis, Fırat University Institute of Science, Elazığ, (2008).
- [10] Nass S., Henderson I. ve Lashof J., “Mammography and Beyond: Developing Technologies for the Early Detection of Breast Cancer”, National Cancer Policy Board, National Academy Press, USA, (2001).
- [11] Arbacha L., Stolpenb A. H., Berbaum K. S., Fajardo L. L., Reinhardt J. M., “Breast MRI lesion classification: Improved performance of human readers with a backpropagation neural network

- computer-aided diagnosis (CAD) system”, *J Magnetic Resonance Imaging*, (2007), 25 (1), sayfa 89-95.
- [12] Abdolmaleki P., Buadu L.D., ve Naderimanesh H., “Feature extraction and classification of breast cancer on dynamic magnetic resonance imaging using artificial neural network”, Elsevier, *Cancer Letters*, (2001), 171 (2), sayfa 183-191.
- [13] Fogel D. B., Wasson E.C., Boughton E.M. ve Porto V.W., “A step toward computer-assisted mammography using evolutionary programming and neural Networks”, *Cancer Letters*, (1997), 119 (1), sayfa 93-97.
- [14] Revett K., Gorunescu F., Gorunescu M., El-Darzi E. ve Ene M., “A breast cancer diagnosis system: a combined approach using rough sets and probabilistic neural Networks”, *Computer as a tool Eurocon 2005, Belgrade*, (2005), pp. 1124- 1127.
- [15] Gorunescu M., Gorunescu F., ve Revett K., “Investigating a Breast Cancer Dataset Using a Combined Approach: Probabilistic Neural Networks and Rough Sets”, *Proc. 3rd ACM International Conference on Intelligent Computing and Information Systems -ICICIS07, Cairo, Egypt*, (2007), pp. 246-249.
- [16] Hsiao Y.H., Huang Y.L., Liang W.M., Kuo S.J. and Chen D.R., “Characterization of benign and malignant solid breast masses: harmonic versus nonharmonic 3D power Doppler imaging”, *Ultrasound Medicine & Biology*, (2009), 35 (3), pp. 353-359.