

FIRAT ÜNİVERSİTESİ

Teknoloji Fakültesi  
Yazılım Mühendisliği Bölümü

**YMT410 – YAPAY ZEKA VE UZMAN SİSTEMLER**

**PROJE RAPORU**

**Projenin Adı**

**YAPAY ZEKA İLE HASTALIK TEŞHİSİ**

**Proje Yöneticisi**

**Doç. Dr. FATİH ÖZKAYNAK**

**Proje Sorumlusu**

**İsim: Abdulkadir KARACA**

**Numara: 12542531**

# 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 veri tabanı 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.

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 hücresini 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 “çözümsüz” veya “np karmaşık” olarak nitelendirilen problemlerin çözümünü hedeflemekte ve hatta bir kısmını başarıyla çözmektedir.

## Göğüs Kanseri

Göğüs kanseri özellikle kadınların ölüm sebeplerinden biridir. 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 kaldırabilecek çözümlere ulaşamamışlardır. 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.

Klinik raporlarda, YSA’ların çok farklı özellikleri kullanarak göğüs kanseri tanısında uygulanmıştır.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.kaba setleri ve olasılıklı sinir ağları içeren melez bir modele dayalı göğüs kanseri için tıbbi bir karar destek sistemi tasarlamışlardır. Hsiao ve ark. 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 kullanılabilecektir.

## Kullanılan Veri Seti

Uygulamada kullanılan veri tabanı UCI Machine Learning Repository’den alınan Breast Cancer veri tabanıdır. Madison’daki Wisconsin Üniversitesi Hastanesi’nden Dr. William H. Wolberg tarafından oluşturulmuştur. Veri tabanında 699 örneğe ait veri ve bunların ait oldukları 2 sınıf mevcuttur. Sınıflar benign yani 0’a yakın ise (iyi huylu) ve malignant 1’e yakın ise (kötü huylu) şeklindedir. Örneklerin 458’i 0 sınıfına aitken, 241’i 1sınıfına aittir. Veri setinde toplam 10 öznitelik var. Bunlar ve değer aralıkları Tablo1’de gösterilmektedir.

|  |  |  |
| --- | --- | --- |
| **VERİ SETİ** | | **DEĞER ARALIĞI** |
| **1** | **Kaba Kalınlığı** | **1-10** |
| **2** | **Hücre Boyutu Düzgünlüğü** | **1-10** |
| **3** | **Hücre şekli tekdüzelik** | **1-10** |
| **4** | **Marjinal adhezyon** | **1-10** |
| **5** | **Tek epi\_Cell\_Size** | **1-10** |
| **6** | **Çekirdek** | **1-10** |
| **7** | **Chromatin** | **1-10** |
| **8** | **Normal Nükleol** | **1-10** |
| **9** | **Mitoz** | **1-10** |
| **10** | **Sınıf** | **0-1** |

*Tablo 1*

## Yapılan İş

Veri tabanında 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.

## Oluşturulan Yapay Sinir Ağı



Şekil 1: Oluşturulan yapay sinir ağı

## Sonuçlar

Bu çalışmada, edinilen veri tabanı kullanılarak yapay sinir ağları ile göğüs kanserinin teşhisine yönelik bir tahmin mekanizması oluşturulmuştur. Veri tabanındaki her bir örneğe ait öznitelik 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(0’a yakın), 197’si malignant(1’e yakın) ve testte kullanılan örneklerin 155’i benign(0’a yakın), 44’ü malignant(1’e yakın) 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.