

EĞİTİCİLİ VE EĞİTİCİSİZ NÖRAL ALGORİTMALAR KULLANARAK GÖĞÜS KANSERİ TEŞHİSİ

Tüba KIYAN¹

Tülay YILDIRIM²

Yıldız Teknik Üniversitesi
Elektronik ve Haberleşme Müh. Böl
34349 Beşiktaş-İstanbul

¹e-posta: tkiyan@yildiz.edu.tr

²e-posta: tkiyan@yildiz.edu.tr

Anahtar Sözcükler: Göğüs kanseri, çok katmanlı algılayıcı, radyal temelli fonksiyon ağı, kendi kendine organize olan haritalar, vektör kuantalamalı öğrenme

ÖZET

Göğüs kanseri, kadınlarda en yaygın ikinci kanser çeşididir. Bu kanser türünde teşhis, tedaviden elde edilecek verimi artırdığı gibi tedavi yönteminin seçiminde de kullanılabilir. Bu bildiride, göğüs kanserinin teşhisine yardım amacıyla eğitici ve eğitici olmayan nöral algoritmalar kullanılmıştır. Literatürdeki Wisconsin göğüs kanseri teşhisi (WBCD) veritabanı, çok katmanlı algılayıcı, radyal temelli fonksiyon ağı, kendi kendine organize olan özellik haritaları ve vektör kuantalamalı öğrenme yöntemleri kullanılarak sınıflandırılmış ve önceki çalışmalara göre başarılı sonuçlar elde edilmiştir.

1.GİRİŞ

Kanser, bir hücrenin kontrolsüz olarak bölünmesiyle başlar ve sonuçta tümör olarak bilinen gözle görülür bir kütle oluşur. Tümör, iyi huylu ya da kötü huylu olabilir. Kötü huylu tümörler, hızla gelişir ve çevresindeki dokuları istila ederek bu dokulara ve organlara zarar verir. Göğüs kanseri, göğüste büyümeye başlayan kötü huylu bir dokudur. Kadınlarda görülen en yaygın ikinci kanserdir. Göğüs kanserinin belirtileri göğüste görülen yumru, göğsün şekil ve boyutunda değişiklik, derinin renginde görülen değişiklikler, göğüs ağrısı vb. [1]

Nöron ağı literatüründe Wisconsin göğüs kanseri teşhisi (WBCD) verisi ile göğüs kanserinin medikal teşhisi üzerine yapılmış bir çok araştırma vardır. [2]'de logaritmik benzetimli tavlama ile Algılayıcı algoritmasını birleştiren bir algoritma kullanılmış ve %98.8 doğruluk elde edilmiştir. [3]'de bütün WBCD verisi üzerinde doğruluk derecesi %98.98'e ulaşan bulanık modelleme ve işbirlikçi eşdeğişim'in kullanıldığı bir sınıflama tekniği uygulanmıştır. [4]'de ise İleri Beslemeli Sinir Ağı Kural Çıkartma Algoritmasına dayalı bir sınıflama yapılmış ve %98.24 doğruluk elde edilmiştir. Kronolojik olarak ilk 367 örneğin alındığı [5]'de doğruluk %93.7'dir.

Bu makale, WBCD veritabanı üzerinde eğitici ve eğitici olmayan nöral algoritmaların uygulanması ile elde edilen sonuçları vermektedir. Çok katmanlı algılayıcı (Multilayer Perceptron-MLP), radyal temelli fonksiyon ağı (Radial Basis Function-RBF), kendi kendine organize olan özellik haritaları (Self-Organizing Maps-SOM) ve vektör kuantalamalı öğrenme (Learning Vector Quantization-LVQ) WBCD verisini sınıflamada kullanılmıştır ve bu sonuçlar daha öncekilerle karşılaştırılmıştır.

2. UYGULANAN NÖRAL YAPILAR

2.1.ÇOK KATMANLI ALGILAYICI (MLP):

Çok katmanlı algılayıcı bazı zor ve farklı problemleri başarılı bir şekilde çözmek için uygulanır. Oldukça popüler olan hatanın geriye yayılması mantığına dayalıdır.

Bir çok katmanlı ağ üç önemli özelliğe sahiptir:

(1) Ağdaki her nöron modeli çıkışta non-lineerlik içerir. Burada önemli bir nokta, non-lineerliğin Rosenblatt'ın perseptronunda kullanılan keskin-geçişli fonksiyona göre yumuşak geçişli olmasıdır.

(2) Ağ, çıkış ya da girişe ait olmayan bir ya da daha fazla saklı nörona sahiptir. Bu saklı nöronlar ağı karmaşık işleri öğrenmesini sağlar.

(3) Ağda, her nöron birbiriyle bağlıdır. Bağlantılardaki bir değişiklik, sinaptik bağlantılarda ve ağırlıklarda değişikliğe neden olur.

Geriye yayılım algoritmasının gelişimi, nöron ağlarında bir dönüm noktasıdır [6].

2.2.RADYAL TEMELLİ FONKSİYON AĞLARI (RBF):

RBF, çok boyutlu uzayda eğri uyurma problemini ortaya koyan farklı bir yaklaşımdır. Bu açıdan, öğrenme, istatistiksel olarak birden fazla boyutlu uzayda eğitme verisi en uygun yüzeyi bulmaya eşdeğerdir. RBF ağı, üç tane birbirinden tamamen farklı katmandan oluşur. Giriş katmanı, kaynak düğümlerinden oluşmaktadır. İkinci katman, çok

katmanlı ağa göre farklı bir yapıdadır. Çıkış katmanı ise, giriş katmanına uygulanan aktivasyon paternlerinin cevabını verir. Giriş uzayından saklı katman uzayına transformasyon nonlineerken saklı katman uzayından çıkış uzayına lineerdir [6].

2.3. KENDİ KENDİNE ORGANİZE OLAN ÖZELLİK HARİTALARI (SOM):

SOM, giriş uzayında birbirine yakın olan benzer verileri gruplar ve iki boyutlu haritada yakın birimlere atar [7]-[8].

Böyle bir algoritmanın özellikleri şunlardır [6]:

- (1) Bir ya da iki boyutlu kafes yapısındaki nöronlar, girişten alınan bilgilerin basit diskriminant fonksiyonlarını hesaplar.
- (2) Bu diskriminant fonksiyonları karşılaştırılır ve en büyük değerli diskriminant fonksiyonuna sahip olan nöron seçilir.
- (3) Seçilen nöronlar ve komşuları aktive edilir.
- (4) Aktive olan nöronlar giriş işaretine göre diskriminant fonksiyonlarını artırırlar [6].

2.4. VEKTÖR KUANTALAMALI ÖĞRENME (LVQ):

LVQ, eğitici bir öğrenmedir. Sınıf yoğunluklarını modellemek yerine, vektörlerle belirlenmiş diskriminant fonksiyonlarını modeller ve bu vektörler ile veri arasındaki en yakın komşuyu bulur. Sınıflamada, bir veri noktası x_i en yakın vektörün sınıfına dahil edilir. Eğitim algoritması, kazanan birimin iterative gradient güncellemesini içerir. Kazanan birim m^c şu şekilde ifade edilir:

$$c = \arg \min_k \|x_i - m^k\| \quad (1)$$

Eğer bir veri doğru şekilde sınıflandırılmışsa, model vektör o veriye yaklaşır. Eğer yanlış sınıflandırılmışsa uzaklaşır. Kazanan birim m^c için en yakın komşu kuralına göre güncelleme denklemi şöyledir:

$$m^c(t+1) := m^c(t) \pm \alpha(t)[x(t) - m^c(t)] \quad (2)$$

Denklemdaki işaret verinin doğru sınıflanmasına göre değişir. Eğer doğru sınıflanmışsa pozitif, yanlış sınıflanmışsa negatiftir. Öğrenme oranı zamanla monoton olarak azalmalıdır. Bu işlemler yakınsama oluncaya kadar devam ettirilir [9].

3. WBCD VERİTABANI

Bu çalışmada kullanılan göğüs kanseri veritabanı University of Wisconsin Hospitals, Madison Dr. William H. Wolberg 'den elde edilmiştir [10]. 699 örnekten her biri dokuz özellik ve iyi huylu ya da kötü huylu olmak üzere bir sınıf bilgisi içerir. 16 örnek eksik özellik içerdiğinden simülasyonlar 683 örnek

üzerinde yürütülmüştür. 444 tane örnek iyi huylu sınıfına, 239 örnek ise kötü huylu sınıfına aittir.

Özellikler şunlardır:

1. Kütle kalınlığı
2. Hücre büyüklüğünün uniformluğu
3. Hücre şeklinin uniformluğu
4. Marjinal yapışma
5. Tek epithelial hücre büyüklüğü
6. Yalın çekirdek
7. Hafif kromatin
8. Normal Nükleoli
9. Mitozlar

Bu özellikler dokuz farklı skalada dış görünüm ve kromozom değişikliklerini ölçer. Bütün veriler 1 ile 10 arasında değişen değerlere sahiptir. Sınıf bilgisi olarak 4 kötü huylu verileri, 2 ise iyi huylu verileri gösterir. Tablo 1 veri dağılımlarını göstermektedir.

TABLO 1.
SINIFLAR VE VERİ DAĞILIMLARI

Sınıf	Toplam Veri Sayısı	Eğitme Veri Sayısı	Test Veri Sayısı
2	444	222	222
4	239	120	119

4.SİMÜLASYON SONUÇLARI

Simülasyonlar MATLAB 6.0 Neural Network Toolbox kullanılarak gerçekleştirilmiştir. Dört farklı nöron ağı yapısı; çok katmanlı algılayıcı, radyal temelli fonksiyon ağı, kendi kendine organize olan özellik haritaları ve vektör kuantalamalı öğrenme, WBCD veritabanı üzerine eğitici ve eğitici olmayan algoritmaları karşılaştırmak amacıyla uygulanmıştır. RBF ağına genişlik ölçekleme parametre değeri 4.4, SOM ve MLP'de öğrenme oranları sırasıyla 0.9 ve 0.6 olarak alınmıştır. Ağların genelleştirme yeteneklerini test etmek amacıyla farklı eğitim ve test setleri uygulanmıştır.

4.1. EĞİTME VERİSİ SİMÜLASYONU:

Eğitim için veri kümesinin yarısı kullanılmıştır. Eğitim setinde 222 örnek iyi huylu sınıfına, 120 örnek ise kötü huylu sınıfına dahildir. Eğitim setinin sınıflama sonuçları MLP, RBF, SOM, ve LVQ için sırasıyla Tablo 2, 3, 4 ve 5'te verilmiştir. RBF ağı, 342 verinin tamamını doğru sınıflandırarak en iyi sınıflama oranını sağlarken; SOM 331 doğru sınıflama ile en düşük sınıflama oranını vermiştir. LVQ, 333 örneği doğru sınıflarken; MLP, 335 doğru sınıflama ile ikinci en iyi ağıdır. Görüldüğü gibi eğitici algoritmalar eğitim için daha iyidir. Tablo 6, ağların sınıflama performanslarını göstermektedir.

TABLO 2.
MLP İLE EĞİTİM DATASININ
SINIFLANDIRILMASI (ORTALAMA)

Sınıf	İyi Huylu	Kötü Huylu
Doğru	217	118
Yanlış	5	2

TABLO 3.
RBF İLE EĞİTİM DATASININ
SINIFLANDIRILMASI

Sınıf	İyi Huylu	Kötü Huylu
Doğru	222	120
Yanlış	0	0

TABLO 4.
SOM İLE EĞİTİM DATASININ
SINIFLANDIRILMASI

Sınıf	İyi Huylu	Kötü Huylu
Doğru	213	118
Yanlış	9	2

TABLO 5.
LVQ İLE EĞİTİM DATASININ
SINIFLANDIRILMASI

Sınıf	İyi Huylu	Kötü Huylu
Doğru	215	118
Yanlış	7	2

TABLO 6.
EĞİTİM VERİSİ İÇİN SINIFLANDIRMA
PERFORMANSI

Nöron ağı tipi	Performans
MLP	%98.04
RBF	%100
SOM	%96.7
LVQ	%97.3

4.2. TEST VERİSİ SİMÜLASYONU:

Test verisi olarak toplam 341 örnek yani bütün veri setinin %50'si seçilmiştir. Bunların 222 tanesi iyi huylu sınıfına 119 tanesi ise kötü huylu sınıfına dahildir. MLP, RBF, SOM, ve LVQ için sonuçlar Tablo 7, 8, 9, ve 10'da verilmiştir.

TABLO 7.
MLP İLE TEST DATASININ SINIFLANDIRILMASI
(ORTALAMA)

Sınıf	İyi Huylu	Kötü Huylu
Doğru	212	114
Yanlış	10	5

TABLO 8.
RBF İLE TEST DATASININ SINIFLANDIRILMASI

Sınıf	İyi Huylu	Kötü Huylu
Doğru	215	113
Yanlış	7	6

TABLO 9.
SOM İLE TEST DATASININ SINIFLANDIRILMASI

Sınıf	İyi Huylu	Kötü Huylu
Doğru	220	116
Yanlış	2	3

TABLO 10.
LVQ İLE TEST DATASININ SINIFLANDIRILMASI

Sınıf	İyi Huylu	Kötü Huylu
Doğru	220	110
Yanlış	2	9

Test seti için SOM, 336 doğruyla en yüksek sınıflama oranını verirken MLP 326 doğru sınıflama ile en düşük sınıflama oranına sahiptir. RBF 328 tane örneği doğru sınıflarken LVQ 330 doğruyla ikinci en iyi ağıdır. Sonuçlardan görüldüğü gibi eğitici algoritmalar test için daha iyidir. Tablo 11 bütün ağların test seti üzerindeki sınıflama oranlarını vermektedir.

TABLO 11.
TEST VERİSİ İÇİN SINIFLANDIRMA
PERFORMANSI

Nöron ağı tipi	Performans
MLP	%95.74
RBF	%96.18
SOM	%98.5
LVQ	%96.7

5. SONUÇ

Bu çalışmada eğitici ve eğitici olmayan nöral algoritmalar göğüs kanseri teşhisi amacıyla karşılaştırılmıştır. RBF eğitme setindeki en iyi sınıflayıcı olmasına rağmen en önemli sonuç test verisi üzerinde düşünülmelidir. Çünkü test verisi ağın genelleştirme yeteneğini ortaya koyar. Test kümesi düşünüldüğünde SOM en iyi sınıflama oranını vermektedir. Genel olarak sonuçlara bakıldığında WBCD verisinin sınıflandırılmasında en uygun nöron ağı modeli RBF ve SOM olmuştur. Ayrıca sonuçlar, eğitici ve eğitici olmayan nöral algoritmaların göğüs kanseri teşhisinde büyük başarı elde ettiğini göstermiştir.

KAYNAKLAR

- [1] <http://medlineplus.gov/>, 2003.
- [2] A. A. Albrecht, G. Lappas, S. A. Vinterbo, C.K. Wong, L. Ohno-Machado, "Two applications of the LSA machine", ICONIP 2002.

- [3] C. A. Pena-Reyes, M. Sipper, "Fuzzy CoCo: A cooperative coevolutionary approach to fuzzy modeling", IEEE Transactions on Fuzzy Systems, 9(5): 727-737, 2001.
- [4] R. Setiono, "Generating concise and accurate classification rules for breast cancer diagnosis", Artificial Intelligence in Medicine, 18(3): 205-217, 2000.
- [5] J. Zhang, "Selecting typical instances in instance-based learning", Proc. Ninth International Machine Learning Workshop, sf. 470-479, Morgan-Kaufmann, 1992.
- [6] S. Haykin, "Neural Networks: A Comprehensive Foundation", Macmillan College Publishing Company, 1994.
- [7] Kohonen: "Self-Organizing Maps", Springer (1997).
- [8] T. Welstead, "Neural network and fuzzy logic applications in C/C++" John Wiley & Sons, Inc., 1994.
- [9] J. Hollmen, V. Tresp, O. Simula, "A learning vector quantization algorithm for probabilistic models", EUSIPCO 2000, Vol. II, sf.721-724.
- [10] [Http: // ftp.ics.uci.edu/pub/ml-repos/machine-learning-databases/](http://ftp.ics.uci.edu/pub/ml-repos/machine-learning-databases/)