

YAPAY SİNİR AĞLARI YÖNTEMİ: MEME KANSERİ TEŞHİSİ

Fatih Es

Bursa Teknik Üniversitesi

18360859021@ogrenci.btu.edu.tr

Özet – Meme kanseri, meme dokusunu oluşturan hücre gruplarından birinin değişime uğraması ve kontrolsüz olarak çoğalması nedeniyle oluşan tümör sonucu ortaya çıkan bir hastalıktır.[1] Günümüzde artarak devam eden meme kanseri sorunun teşhis edilmesinde teknoloji son yıllarda büyük bir rol oynamaktadır. Bu araştırma da bu alanda yapılan çalışmalara destek olmak amacıyla oluşturulmuştur. Amerika Birleşik Devletleri’nin, Wisconsin eyaletindeki meme kanseri hastalarına ait veriler baz alınmış ve yapay sinir ağları kullanılarak bir yapay zeka modeli geliştirilmiştir. Model sonuçları, daha önce aynı veri setini kullanarak oluşturulmuş bir akademik makale ile karşılaştırılmış ve karşılaştırma sonuçları makale içeriğinde yer edinmiştir. Meme kanserinin erken teşhisi konusunda her geçen yıl gelişen teknoloji ile birlikte ilerlemeler kaydedilmiştir. Yapay sinir ağları kullanılarak geliştirilen sınıflandırıcı modeli %96 doğruluk oranına sahiptir. Bu oranla birlikte modelin erken teşhiste etkin bir rol oynayabileceği öngörülmüştür.

Anahtar Kelimeler – Meme kanseri, Yapay Sinir ağları, Sınıflandırıcı, Veri Analizi, Wisconsin, USA

I. GİRİŞ

Türk Tıbbi Onkoloji Derneği Yönetim Kurulu Üyesi Doç. Dr. Özlem Sönmez, meme kanseri ile ilgili şu değerlendirmede bulundu:

“Meme kanseri tüm dünyada kadınlarda en sık görülen kanserdir, Erkeklerde de çok nadir olmak üzere meme kanseri görülür. Dünyada her yıl 2.1 milyon kadın, ülkemizde ise 20 bin kadın meme kanserinden etkilenmektedir. Yaşam boyu her 8 kadından biri meme kanseri riski, her 38 kadından birisi ise meme kanserinden ölme riski

ile karşı karşıyadır. Dünyanın her bölgesinde meme kanseri görülme sıklığı artmaktadır. Ancak, özellikle gelişmiş ülkelerde tarama programlarının artması ve erken tanı sayesinde son 30 yılda meme kanserinden ölüm oranı % 40 azalmış olup, hastalıktan kurtulma oranı neredeyse % 95’tir.” [2]

İnsan vücudu, her biri kendine özgü işlevi olan milyonlarca hücreden oluşur. Bu hücrelerden herhangi birinin düzensiz büyümesi olduğunda kanser olarak adlandırılır. Bunda hücreler kontrolsüz bir şekilde bölünür ve büyür, tümör adı verilen anormal bir doku kütlesi oluşturur. Tümör hücreleri büyür ve vücudun normal işleyişini bozan sindirim, sinir ve dolaşım sistemlerini istila eder. Her bir tümör kanserli olmasa da.[3]

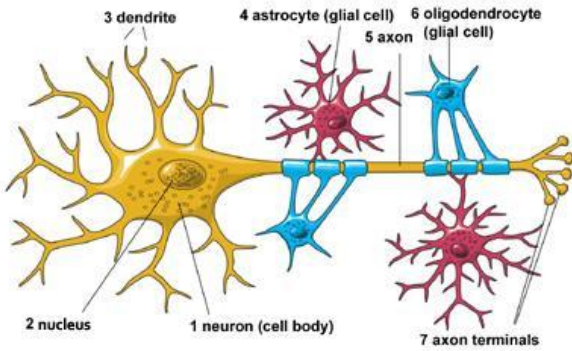
Kanser, etkilenen hücre tipine göre sınıflandırılır ve 200’den fazla kanser türü bilinmektedir. Bu makale meme kanserine odaklanmıştır. Meme kanseri, dünya genelinde kadınlar arasında en sık görülen kanser türüdür [4]

Veri toplama ve saklama tekniklerindeki son gelişmeler, çeşitli tıbbi şirketlerin ve hastanelerin, ilaç ve bir hastalığın semptomlarına ilişkin tıbbi kayıtlarıyla ilgili çok miktarda veri tutmasını mümkün kılmıştır. Resmi olarak, veri madenciliği, yararlı bilgileri çıkarmak için veriler üzerinde güçlü algoritmalar çalıştırma sürecidir. Bu metodolojilerin kullanımları ve potansiyelleri kapsamını tıbbi verilerde bulmuştur. Bir hastalığın sonucunu tahmin etmek zor bir iştir. Veri madenciliği teknikleri, tahmin segmentini basitleştirme eğilimindedir. Otomatik araçlar, tıbbi araştırma gruplarına sunulan büyük hacimli tıbbi verilerin toplanmasını mümkün kılmıştır. [5]

II. METODOLOJİ

A. NÖRON

Nöronlar bilgi habercileridir. Beynin farklı alanları arasında ve beyin ile sinir sisteminin geri kalanı arasında bilgi iletmek için elektriksel uyarılar ve kimyasal sinyaller kullanırlar. Düşündüğümüz, hissettiğimiz ve yaptığımız her şey, nöronlar ve onların destek hücreleri olan astrositler (4) ve oligodendrositler (6) adı verilen gliyal hücreler olmadan imkansız olurdu.



Şekil 1: Nöron Hücresi Görünümü

Nöronların üç temel parçası vardır: bir hücre gövdesi ve akson (5) ve dendrit (3) adı verilen iki uzantı. Hücre gövdesi içinde hücrenin faaliyetlerini kontrol eden ve hücrenin genetik materyalini içeren bir çekirdek (2) bulunur. Akson uzun bir kuyruğa benzer ve hücreden gelen mesajları iletir. Dendritler bir ağacın dalları gibi görünürler ve hücre için mesajlar alırlar. Nöronlar, komşu nöronların aksonları ve dendritleri arasında, sinaps adı verilen küçük bir boşluk boyunca nörotransmitter adı verilen kimyasallar göndererek birbirleriyle iletişim kurarlar.

Üç sınıf nöron vardır:

1. Duyusal nöronlar, duyu organlarından (gözler ve kulaklar gibi) beyne bilgi taşır.
2. Motor nöronlar, konuşma gibi istemli kas aktivitelerini kontrol eder ve beyindeki sinir hücrelerinden gelen mesajları kaslara taşır.
3. Diğer tüm nöronlara internöron denir.

Bilim adamları, nöronların vücuttaki en çeşitli hücre türü olduğunu düşünüyor. Bu üç nöron sınıfı

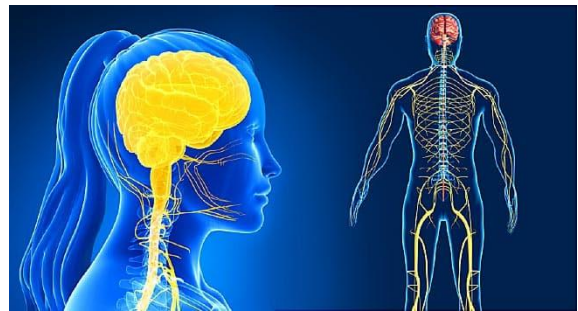
içinde, her biri belirli mesaj taşıma yeteneklerine sahip yüzlerce farklı tip vardır.

Bu nöronların bağlantı kurarak birbirleriyle nasıl iletişim kurdukları, her birimizi nasıl düşündüğümüz, hissettiğimiz ve hareket ettiğimiz konusunda benzersiz kılın şeydir. [6]

B. SİNİR SİSTEMİ

Benzer işlevleri yerine getiren devreler, daha geniş davranışsal amaçlara hizmet eden sinir sistemlerinde gruplandırılmıştır. En genel işlevsel tanım, nöral sistemleri, çevreden bilgi alan ve işleyen görme veya işitme gibi duyu sistemler ve organizmanın bu bilgilere hareketler üreterek yanıt vermesini sağlayan motor sistemler olarak ikiye ayırır. Bununla birlikte, nispeten iyi tanımlanmış bu giriş ve çıkış sistemleri arasında uzanan çok sayıda hücre ve devre vardır. Bunlar topluca çağrışım sistemleri olarak adlandırılır ve en karmaşık ve en az iyi karakterize edilmiş beyin fonksiyonlarını yürütürler.

Bu geniş fonksiyonel ayrımlara ek olarak, sinirbilimciler ve nörologlar geleneksel olarak omurgalı sinir sistemini anatomik olarak merkezi ve çevresel bileşenlere ayırmışlardır. Merkezi sinir sistemi beyin (beyin, beyincik ve beyin sapı) ve omurilikten oluşur. Periferik sinir sistemi, vücut yüzeyindeki duyu reseptörlerini ve ayrıca kulak gibi özel reseptör yapılarında merkezi sinir sistemindeki işlem devreleriyle birbirine bağlayan duyu nöronlarını içerir. Periferik sinir sisteminin motor kısmı iki bileşenden oluşur. Beyni ve omuriliği iskelet kaslarına bağlayan motor aksonlar, periferik sinir sisteminin somatik motor bölümünü oluşturur. Viseral veya otonomik motor bölümü, düz kasları, kalp kasını ve bezleri innerve eden hücreler ve aksonlardan oluşur. [7]



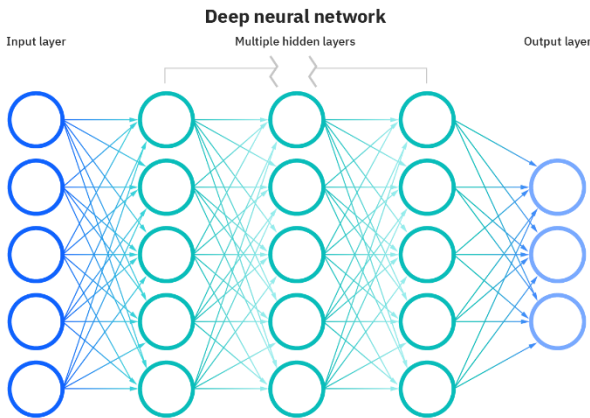
Şekil 2: İnsan Sinir Sistemi

C. YAPAY SİNİR AĞLARI

Sinir ağı, bilgisayarlara verileri insan beyninden esinlenerek işlemeyi öğreten bir yapay zeka yöntemidir. Bu, insan beynine benzeyen katmanlı bir yapıda birbirine bağlı düğümleri veya nöronları kullanan ve derin öğrenme adı verilen bir tür makine öğrenimi sürecidir. Bilgisayarların hatalarından ders çıkarmak ve sürekli olarak gelişmek için kullandığı uyarlanabilir bir sistem oluşturur. Böylece, yapay sinir ağları, belgelerin özetini çıkarma ya da yüzleri tanıma gibi karmaşık sorunları daha fazla doğrulukla çözmeye çalışır. [8]

Yapay sinir ağları (YSA) veya simüle edilmiş sinir ağları (SNN'ler) olarak da bilinen sinir ağları, makine öğreniminin bir alt kümesidir ve derin öğrenme algoritmalarının kalbinde yer alır. Adları ve yapıları, biyolojik nöronların birbirine sinyal gönderme şeklini taklit ederek insan beyninden esinlenmiştir.

Yapay sinir ağları (YSA), bir girdi katmanı, bir veya daha fazla gizli katman ve bir çıktı katmanı içeren bir düğüm katmanından oluşur. Her düğüm veya yapay nöron diğerine bağlanır ve ilişkili bir ağırlık ve eşik sahiptir. Herhangi bir düğümün çıktısı belirtilen eşik değerinin üzerindeyse, o düğüm etkinleştirilir ve ağırlık bir sonraki katmanına veri gönderilir. Aksi takdirde, ağırlık bir sonraki katmanına hiçbir veri iletilmez.



Şekil 3: Yapay Sinir Ağları Modeli

Sinir ağları, zaman içinde doğruluklarını öğrenmek ve geliştirmek için eğitim verilerine güvenir. Bununla birlikte, bu öğrenme

algoritmaları doğruluk için ince ayar yapıldığında, bilgisayar bilimi ve yapay zekada güçlü araçlardır ve verileri yüksek bir hızda sınıflandırmamıza ve kümelememize olanak tanır. Konuşma tanıma veya görüntü tanımadaki görevler, insan uzmanlar tarafından yapılan manuel tanımlamayla karşılaştırıldığında dakikalar yerine saatler sürebilir. En iyi bilinen sinir ağlarından biri Google'ın arama algoritmasıdır.

C.1. Sinir Ağları Nasıl Çalışır?

Her bir düğümü, girdi verileri, ağırlıklar, bir sapma (veya eşik) ve bir çıktıdan oluşan kendi doğrusal regresyon modeli olarak düşünün. Formül şöyle bir şeye benzeyecektir:

$$\sum_{i=1}^m w_i \cdot x_i + bias = w_1 \cdot x_1 + w_2 \cdot x_2 + w_3 \cdot x_3 + bias$$

$$output = f(x) = \begin{cases} 1 & \text{if } \sum w_i \cdot x_i + b \geq 0 \\ 0 & \text{if } \sum w_i \cdot x_i + b < 0 \end{cases}$$

Bir girdi katmanı belirlendikten sonra ağırlıklar atanır. Bu ağırlıklar, herhangi bir değişkenin önemini belirlemeye yardımcı olur ve daha büyük olanlar diğer girdilere kıyasla çıktıya daha fazla katkıda bulunur. Tüm girdiler daha sonra ilgili ağırlıklarıyla çarpılır ve toplanır. Daha sonra çıktı, çıktıyı belirleyen bir aktivasyon fonksiyonundan geçirilir. Bu çıktı belirli bir eşik aşarsa, düğümü "ateşler" (veya etkinleştirir), verileri ağıdaki bir sonraki katmana iletir. Bu, bir düğümün çıktısının bir sonraki düğümün girdisi olmasına neden olur. Verileri bir katmandan sonraki katmana geçirme işlemi, bu sinir ağını ileri beslemeli bir ağ olarak tanımlar.

İkili değerler kullanarak tek bir düğümün nasıl görünebileceğini inceleyelim. Bu kavramı daha somut bir örneğe uygulayabiliriz, örneğin sörf yapıp yapmamanız gibi (Evet: 1, Hayır: 0). Gitme ya da gitmeme kararı bizim tahmin edilen sonucumuzdur ya da y-hat. Karar vermenizi etkileyen üç faktör olduğunu varsayalım:

1. Dalgalar iyi mi? (Evet: 1, Hayır: 0)
2. Sıralama boş mu? (Evet: 1, Hayır: 0)
3. Yakın zamanda köpekbalığı saldırısı oldu mu? (Evet: 0, Hayır: 1)

Ardından, bize aşağıdaki girdileri vererek aşağıdakileri varsayalım:

- $X_1 = 1$, çünkü dalgalar pompalanıyor.
- $X_2 = 0$, çünkü kalabalıklar dışarıda.
- $X_3 = 1$, çünkü yakın zamanda bir köpekbalığı saldırısı olmadı.

Şimdi, önemi belirlemek için bazı ağırlıklar atamamız gerekiyor. Daha büyük ağırlıklar, belirli değişkenlerin karar veya sonuç için daha büyük önem taşıdığını gösterir.

- $W_1 = 5$, çünkü büyük dalgalar sık sık ortaya çıkmaz.
- $W_2 = 2$, çünkü kalabalığa alışsınız.
- $W_3 = 4$, çünkü köpekbalıklarından korkuyorsunuz.

Son olarak, -3'lük bir sapma değerine dönüşecek olan 3'lük bir eşik değeri de varsayacağız. Tüm çeşitli girdilerle, istenen çıktıyı elde etmek için değerleri formüle eklemeye başlayabiliriz.

$$Y = (1 \times 5) + (0 \times 2) + (1 \times 4) - 3 = 6$$

Bu bölümün başından itibaren aktivasyon fonksiyonunu kullanırsak, 6'nın 0'dan büyük olması nedeniyle bu düğümün çıktısının 1 olacağını belirleyebiliriz. Bu durumda sörf yapacaksınız; ancak ağırlıkları veya eşiği ayarlarsak modelden farklı sonuçlar elde edebiliriz. Yukarıdaki örnekte olduğu gibi bir kararı gözlemlediğimizde, bir sinir ağının önceki kararların veya katmanların çıktılarına bağlı olarak nasıl giderek daha karmaşık kararlar alabileceğini görebiliriz.

Yukarıdaki örnekte, burada kullanılan matematiğin bir kısmını göstermek için algılayıcıları kullandık, ancak sinir ağları 0 ile 1 arasında değerlere sahip olmakla ayırt edilen

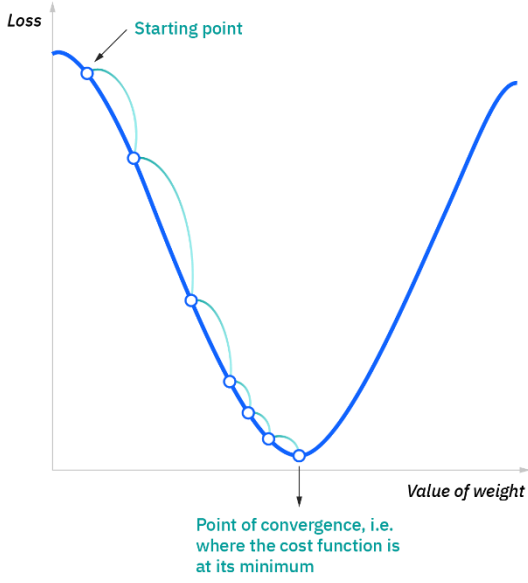
sigmoid nöronlardan yararlanır. diğerine göre, 0 ile 1 arasında x değerlerine sahip olmak, tek bir değişkendeki herhangi bir belirli değişikliğin herhangi bir belirli düğümün çıktısı ve ardından sinir ağının çıktısı üzerindeki etkisini azaltacaktır.

Görüntü tanıma veya sınıflandırma gibi sinir ağları için daha pratik kullanım durumları hakkında düşünmeye başladığımızda, algoritmayı eğitmek için denetimli öğrenmeden veya etiketli veri kümelerinden yararlanacağız. Modeli eğitirken, bir maliyet (veya kayıp) işlevi kullanarak doğruluğunu değerlendirmek isteyeceğiz. Bu aynı zamanda genel olarak ortalama kare hatası (MSE) olarak da adlandırılır. Aşağıdaki denklemde,

- i örneğin indeksini temsil eder,
- tahmin edilen sonuçtur,
- y' gerçek değerdir,
- m örnek sayısıdır.

$$Cost Function = MSE = \frac{1}{2m} \sum_{i=1}^m (y' - y)^2$$

Nihai olarak amaç, herhangi bir gözlem için uygunluğun doğruluğunu sağlamak için maliyet fonksiyonumuzu en aza indirmektir. Model ağırlıklarını ve önyargısını ayarlarken, yakınsama noktasına veya yerel minimuma ulaşmak için maliyet fonksiyonunu ve pekiştirmeli öğrenmeyi kullanır. Algoritmanın ağırlıklarını ayarladığı süreç, gradyan iniş yoluyla ve modelin hataları azaltmak (veya maliyet fonksiyonunu en aza indirmek) için alacağı yönü belirlemesine izin verir. Her eğitim örneğinde, modelin parametreleri minimumda kademeli olarak yakınsamak için ayarlanır.



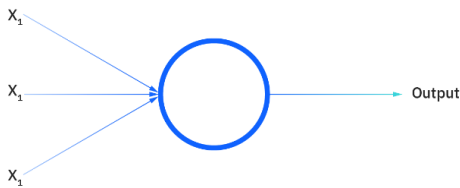
Şekil 4: Yapay Sinir Ağlarında Ağırlık-Kayıp İlişkisi

Derin sinir ağlarının çoğu ileri beslemelidir, yani girdiden çıktıya yalnızca bir yönde akarlar. Ancak, modelinizi geri yayılım yoluyla da eğitebilirsiniz; yani çıkıştan girişe ters yönde hareket eder. Geri yayılım, her bir nöronla ilişkili hatayı hesaplamamıza ve ilişkilendirmemize izin vererek, modelin parametrelerini uygun şekilde ayarlamamıza ve uydurmamıza izin verir.

C.2. Sinir Ağlarının Türleri

Sinir ağları, farklı amaçlar için kullanılan farklı türlerde sınıflandırılabilir. Bu kapsamlı bir tür listesi olmasa da, aşağıdakiler, yaygın kullanım durumları için karşılaşacağınız en yaygın sinir ağı türlerini temsil etmektedir:

Algılayıcı (perceptron), 1958'de Frank Rosenblatt tarafından oluşturulan en eski sinir ağıdır. Tek bir nörona sahiptir ve bir sinir ağının en basit şeklidir:



Şekil 5: Tek Nörona Sahip Sinir Ağı Modeli

İleri beslemeli sinir ağları veya çok katmanlı algılayıcılar (MLP'ler), bu makalede öncelikli olarak odaklandığımız şeydir. Bir girdi katmanı, bir gizli katman veya katmanlar ve bir çıktı katmanından oluşurlar. Bu sinir ağlarına yaygın olarak MLP'ler de denilse de, gerçek dünya problemlerinin çoğu doğrusal olmadığı için algılayıcılardan değil, aslında sigmoid nöronlardan oluştuklarına dikkat etmek önemlidir. Veriler genellikle onları eğitmek için bu modellere beslenir ve bunlar bilgisayarla görme, doğal dil işleme ve diğer sinir ağlarının temelidir.

Evrişimli sinir ağları (CNN'ler), ileri beslemeli ağlara benzer, ancak genellikle görüntü tanıma, görüntü tanıma ve/veya bilgisayarla görme için kullanılırlar. Bu ağlar, bir görüntü içindeki kalıpları tanımlamak için doğrusal cebirden, özellikle matris çarpımından ilkeler kullanır.

Tekrarlayan sinir ağları (RNN'ler) geri besleme döngüleriyle tanımlanır. Bu öğrenme algoritmaları, borsa tahminleri veya satış tahminleri gibi gelecekteki sonuçlar hakkında tahminler yapmak için zaman serisi verileri kullanıldığında öncelikle yararlanılır.

C.3. Sinir Ağları ve Derin Öğrenme

Derin Öğrenme ve sinir ağları, kafa karıştırıcı olabilen konuşmalarda birbirinin yerine kullanılma eğilimindedir. Sonuç olarak, derin öğrenmedeki "derin" in sadece bir sinir ağındaki katmanların derinliğine atıfta bulunduğunu belirtmekte fayda var. Girdileri ve çıktıyı içeren üçten fazla katmandan oluşan bir sinir ağı, derin öğrenme algoritması olarak kabul edilebilir. Yalnızca iki veya üç katmanı olan bir sinir ağı, yalnızca temel bir sinir ağıdır.

C.4. Sinir Ağlarının Tarihi

Sinir ağlarının tarihi çoğu insanın düşündüğünden daha uzundur. "Düşünen bir makine" fikrinin izini Antik Yunanlılara kadar götürebilsek de, bizler,

yıllar içinde popülaritesi azalan ve akan sinir ağları etrafında düşünmenin evrimine yol açan önemli olaylara odaklanacağız:

1943: Warren S. McCulloch ve Walter Pitts, “Sinirsel aktiviteye içkin fikirlerin mantıksal bir hesabı” yayınladı. Bu araştırma, insan beyninin birbirine bağlı beyin hücreleri veya nöronlar aracılığıyla nasıl karmaşık modeller üretebileceğini anlamaya çalıştı. Bu çalışmadan çıkan ana fikirlerden biri, ikili eşiği olan nöronların Boole mantığıyla (yani, 0/1 veya doğru/yanlış ifadeleri) karşılaştırılmasıydı.

1958: Frank Rosenblatt, "The Perceptron: A Probabilistic Model for Information Storage and Organisation in the Brain" adlı araştırmasında belgelenen algılayıcının gelişimiyle tanınır. Denklem ağırlıklar ekleyerek McCulloch ve Pitt'in çalışmasını bir adım daha ileri götürüyor. Bir IBM 704'ten yararlanan Rosenblatt, bir bilgisayara solda işaretli kartların sağda işaretli kartlarla nasıl ayırt edileceğini öğrenmesini sağladı.

1974: Çok sayıda araştırmacı geri yayılım fikrine katkıda bulunurken, Paul Werbos ABD'de onun sinir ağları içindeki uygulamasını doktora tezinde not eden ilk kişiydi.

1989: Yann LeCun, geri yayılımdaki kısıtlamaların kullanımının ve bunun sinir ağı mimarisine entegrasyonunun algoritmaları eğitmek için nasıl kullanılabileceğini gösteren bir makale yayınladı. Bu araştırma, ABD Posta Servisi tarafından sağlanan elle yazılmış posta kodu rakamlarını tanımak için bir sinir ağından başarıyla yararlandı. [9]

D. MEME KANSERİ VERİ SETİ

Bu projede ABD'nin Wisconsin eyaletinde teşhis edilen meme kanseri hastalarının UCI (University of California, Irvine) üzerinde yayınlanan veri seti kullanılmıştır.

Ayrıca veri seti, California Üniversitesi-Irvine'de bulunan Makine Öğrenmesi Deposunda (Machine Learning Repository)

çevrimiçi olarak mevcuttur. Yarıçap, doku, çevre uzunluğu, alan, pürüzsüzlük, kompaktlık, konkavlık, konkav noktalar, simetri ve fraktal boyut olmak üzere 10 gerçek değerli nitelik vardır.[10]

Özniteliklerin isimler ve anlamları aşağıdaki çizelgede listelenmiştir.

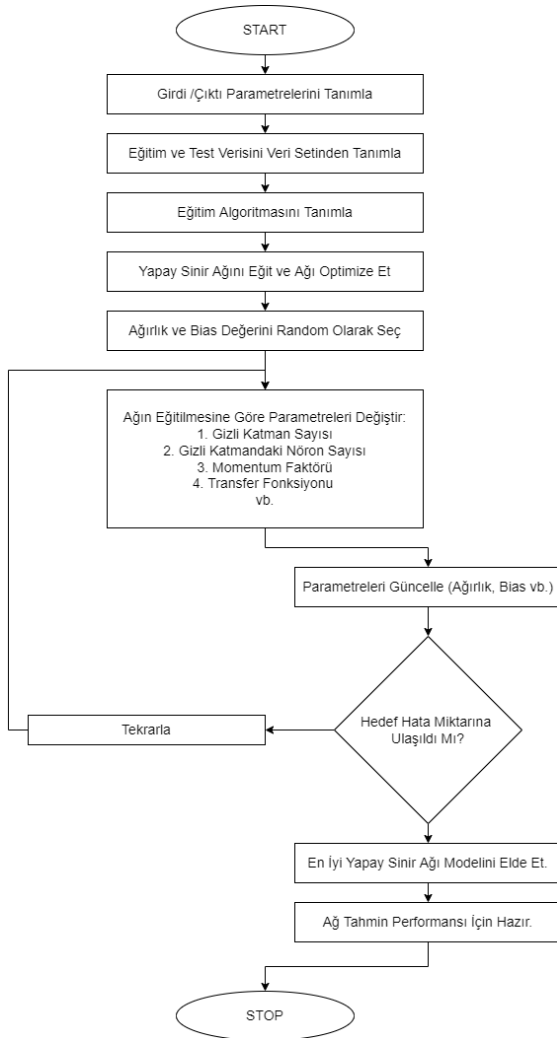
Tablo 1: Öznitelikler - Anlamları

Öznitelik Adı	Öznitelik Anlamı
id	Benzersiz ayraç
diagnosis	Tümör teşhisi (B= Kötü huylu, M= İyi huylu)
radius_mean	Merkezden çevre üzerindeki noktalara olan mesafelerin ortalaması
texture_mean	Gri-skala değerinin standart sapması
perimeter_mean	Çekirdek tümörün ortalama boyutu
area_mean	Tümörün yayıldığı alanın ortalaması
smoothness_mean	Yarıçap uzunluklarındaki yerel değişimin ortalaması
compactness_mean	Perimeter ve area değerlerine bağlı yoğunluk oranlarının ortalaması
concavity_mean	Konturun içbükey kısımlarının ortalama şiddeti
concave_points_mean	Konturun içbükey kısımlarının sayısının ortalaması
symmetry_mean	## Bilinmiyor
fractal_dimension_mean	“Kıyı şeridi yaklaşımı için ortalama” - 1
radius_se	Merkezden çevre üzerindeki noktalara olan mesafelerinin standart hatası
texture_se	Gri-skala değerinin standart sapmasının standart hatası
perimeter_se	## Bilinmiyor
area_se	## Bilinmiyor
smoothness_se	Yarıçap uzunluklarındaki yerel standart hatası
compactness_se	Perimeter ve area değerlerine bağlı yoğunluk oranlarının standart hatası
concavity_se	Konturun içbükey kısımlarının ciddiyeti için standart hata
concave_points_se	Konturun içbükey kısımlarının sayısı için standart hata
symmetry_se	## Bilinmiyor
fractal_dimension_se	“Kıyı şeridi yaklaşımı için standart hata” - 1
radius_worst	Merkezden çevre üzerindeki noktalara olan mesafeleri için “en kötü” veya en büyük ortalama değer
texture_worst	Gris kala değerinin standart sapması için “en kötü” veya en büyük ortalama değer
perimeter_worst	## Bilinmiyor
area_worst	## Bilinmiyor
smoothness_worst	Yarıçap uzunluklarındaki yerel değişim için “en kötü” veya en büyük ortalama değer
compactness_worst	Perimeter ve area değerlerine bağlı “en kötü” veya en büyük ortalama değer
concavity_worst	Konturun içbükey kısımlarının ciddiyeti için “en kötü” veya en büyük ortalama değer
concave_points_worst	Konturun içbükey kısımlarının sayısı için “en kötü” veya en büyük ortalama değer
symmetry_worst	## Bilinmiyor
fractal_dimension_worst	“Kıyı şeridi yaklaşımı için en kötü veya en büyük ortalama değer” - 1

Veri seti bilgileri:

- Veri seti karakteristiği çok değişkenlidir.
- Öznitelikler gerçek sayılardan oluşmuştur.
- 569 kayıt bulunmaktadır (357 kayıt iyi huylu, 212 kayıt kötü huyludur.)
- 32 adet öznitelik bulunmaktadır.
- Kayıtlarda eksik veriler bulunmamaktadır.

Verileri sınıflandırılırken kullanılan algoritma yapay sinir ağları modellenerek tasarlanmıştır. Algoritma aşağıdaki şekilde gösterilmiştir.



Şekil 6: Model Geliştirme Algoritması Akış Şeması

Veri seti üzerinde yapılan tüm işlemler 16 GB DDR4 RAM, Intel Core i7-9750H CPU işlemcili ve Windows 10 Home Single

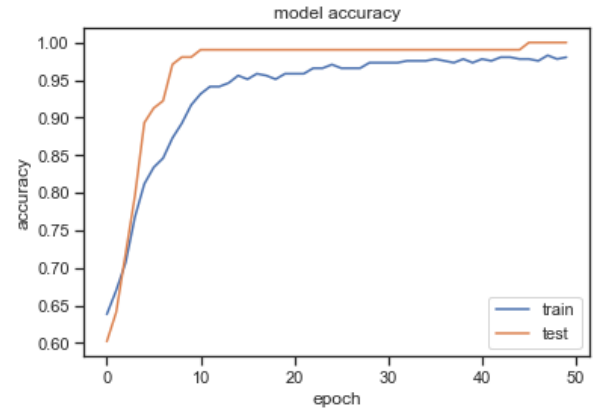
Language işletim sistemine sahip makinede gerçekleştirilmiştir.

III. SONUÇ

Yapay sinir ağları kullanılarak geliştirilmiş olan model, kendisine verilen veriler ile bir hücrenin iyi huylu (B) veya kötü huylu (M) tümör şeklinde sınıflandırmaktadır. Veri set içerisinde bulunan öznitelikler arasında bağlantılı olanlar bulunmaktadır. Bu ilişkiler göz ardı edilerek sınıflandırıcı eğitilmiştir. Tüm verilerin %70'lik bir kısmı eğitim seti olarak, %30'luk bir kısmı ise test seti olarak ayrılmıştır.

Model eğitim seti ile eğitilmiştir ve eğitim setinde gösterdiği başarı ölçüldüğünde doğruluk değeri %98.04 olarak belirlenmiştir.

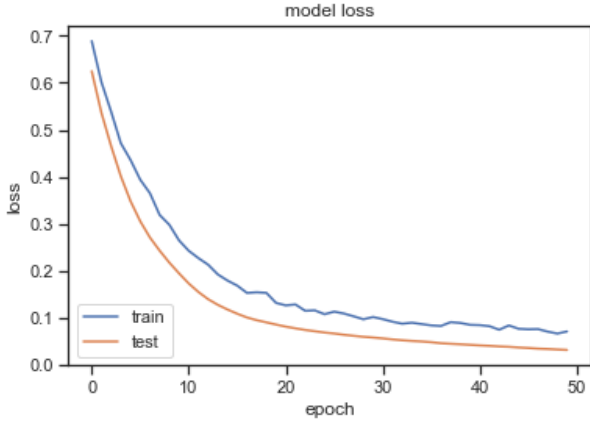
Model toplamda 40 epoch işleminden geçirilmiştir. Bir epoch, bir döngü için tüm eğitim verileriyle sinir ağının eğitilmesi anlamına gelir. Bir epoch, tüm verileri tam olarak bir kez kullanırız. Modelin eğitim veri seti ve test veri seti üzerindeki doğruluk değerleri aşağıdaki grafikte belirtilmiştir.



Şekil 7: Model Eğitim-Test Doğruluk Değerleri

Grafik incelendiğinde 10-20 arasındaki epoch değerinden sonra doğruluk değeri belirgin bir şekilde artmamıştır. Model test verisiyle de eğitim verisine yakın bir doğruluk değerine sahip olmuştur. Böylelikle modelin aşırı beslenme durumunda kalmadığı gözlemlenir.

Bu tespiti destekleyen diğer bir veri ise eğitim süresinde her iki veri seti için de loss yani kayıp değerinin düzenli olarak düştüğüdür. Aşağıdaki grafikte eğitim veri seti ve test veri seti için doğruluklar listelenmiştir.

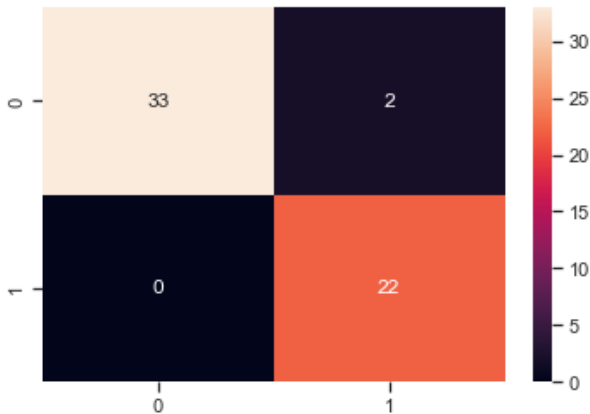


Şekil 8: Model Eğitim-Test Kayıp Değerleri

Modelin test verileri üzerinde hesaplanan karışıklık (confusion) matrisi:

- True Positive(TP): 33
- False Negative(FN): 0
- False Positive(FP): 2
- True Negative(TN): 22

şeklinde hesaplanmıştır. Bu değerler aşağıdaki şekil üzerinde de belirtilmiştir.



Şekil 9: Test Verisine Uygulanan Tahmine Dayalı Karmaşıklık Matrisi

Modelin aynı veri seti ve aynı modelleme yöntemi (yapay sinir ağları) kullanılarak 2022 yılında yapılan bir çalışma [11] ile sonuçlarının kıyaslanması Çizelge 3.1 ile sunulmuştur .

Tablo 2: Verileri Karşılaştırma (Kalın olarak belirtilen veriler daha kabul edilebilir olduğu anlamına gelir.)

Özellik	Geçerli Makale(%)	Diğer Akademik Makale (%)
Accuracy	96.49	86.67
Precision	96.00	87.4
Recall	97.00	86.67
F1- Score	96.50	94.00
Sensitivity	100.00	-
Specifity	91.66	-

IV. REFERANSLAR

- [1] <https://www.acibadem.com.tr/ilgi-alani/meme-kanseri/>
- [2] <https://www.ntv.com.tr/saglik/meme-kanseri-dunyada-her-yil-2-1-milyon-kadini-etkiliyor,hpmGtGjK6o0VEcL6YSEg>
- [3] Ronak Sumbaly, N. Vishnusri, S. Jeyalatha “Diagnosis of Breast Cancer using Decision Tree Data Mining Technique”, 2014
- [4] Ronak Sumbaly, N. Vishnusri, S. Jeyalatha “Diagnosis of Breast Cancer using Decision Tree Data Mining Technique”, 2014
- [5] Ying Tan, Yuhui Shi, and Milan Tuba, “Method Based on Data Mining Techniques for Breast Cancer Recurrence Analysis”, 2020
- [6] <https://www.ninds.nih.gov/health-information/patient-caregiver-education/brain-basics-life-and-death-neuron#:~:text=Neurons%20are%20information%20messengers.,rest%20of%20the%20nervous%20system.>
- [7] <https://www.ncbi.nlm.nih.gov/books/NBK11061/>
- [8] <https://aws.amazon.com/tr/what-is/neural-network/>
- [9] <https://www.ibm.com/cloud/learn/neural-networks>
- [10] https://www.researchgate.net/publication/335444489_Meme_Kanseri_Teshisi_Icin_Yeni_Bir_Skor_Fuzyon_Yaklasimi
- [11] Hager Saleh , Sara F. Abd-el ghany , Hashem Alyami, and Wael Alosaimi -Predicting Breast Cancer Based on Optimized Deep Learning Approach (<https://www.hindawi.com/journals/cin/2022/1820777/>)