

Yapay Sinir Ağları ile Osteoporoz Tahmini

İçindekiler

Giriş	2
1. Proje Amacı.....	2
2. Kullanılan Veri Seti.....	2
3. Veri İncelemesi ve Görselleştirme	3
3.1. Histogramlar:.....	4
3.2. Kutu Grafiği:	4
3.3. Korelasyon Matrisi:	4
3.4. Scatter Plotlar:	4
4. Veri Ön İşleme ve Normalizasyonu	4
5. Yapay Sinir Ağı Modeli Oluşturma ve Eğitim	5
5. 1. Veri Seti Bölme	5
5. 2. Modelin Oluşturulması.....	6
6. Model Performansı	6
7. Sonuçlar.....	7
8. KAYNAKÇA	8

Giriş

Bu rapor, kemik erimesi tanısında yapay sinir ağlarının kullanılmasını incelemekte ve bir yapay sinir ağı modelinin oluşturulması ve eğitilmesi sürecini detaylandırmaktadır. Kemik erimesi, kemik dokusunun yoğunluğunun azalması ve kemik kırılmaları riskinin artmasıdır. Bu projede, kemik erimesi olup olmadığını belirlemek için çeşitli demografik ve biyolojik öznitelikler kullanılarak bir yapay sinir ağı modeli oluşturulmuştur.

1. Proje Amacı

Bu projenin amacı, kemik erimesi tanısında yapay sinir ağlarının etkinliğini değerlendirmektir. Kemik erimesi, kemik dokusunun yoğunluğunun azalması ve kemik kırılmaları riskinin artmasıyla karakterize edilen bir durumdur. Bu proje, demografik ve biyolojik özniteliklerin (yaş, cinsiyet, beslenme alışkanlıkları, yaşadığı yer vb.) kemik erimesi tanısındaki rolünü anlamak ve yapay sinir ağlarının bu tanıda kullanılabilirliğini araştırmak için tasarlanmıştır. Yapılan çalışma, yapay sinir ağlarının kemik erimesi tanısında kullanımının potansiyelini ortaya koymayı amaçlamaktadır.

2. Kullanılan Veri Seti

Veri setimiz, kemik erimesi tanısında kullanılan 22 adet öznitelik içermektedir. Bu öznitelikler şu şekildedir:

1	yaş	meslek	kilo	boy	cinsiyet	yasadigi_y	egitim_dur	sut_urunle	et	kahve	alkol	fiziksel_ak	ldl	hdl	trigliserit	crp	tsh	wbc	hg	plt	krea	alt	kemik_erin
2	35	1	72	168	1	3	1	0	1	0	0	2	126,26	50,8	156,2	14,12	1,41	6,74	12,6	300	0,73	17,5	0
3	38	2	76	174	2	3	3	0	1	0	0	2	192	50,6	117	7,29	1,81	8,52	14,6	199	0,85	16,3	0
4	69	1	66	155	1	3	1	1	2	1	0	1	109,62	37,6	176,9	1,92	0,84	11,95	17,4	206	1,29	46,9	1
5	53	3	65	160	1	3	3	1	2	1	0	3	85,86	56,1	81,2	2,44	2	6,99	13,2	232	0,86	7,7	0
6	53	4	85	178	2	3	4	2	2	1	0	3	65,3	37,5	81	2,44	1,47	9,05	12,4	291	0,6	22,1	0
7	24	5	55	168	1	3	4	1	2	1	0	2	187,46	56,9	102,2	1,49	1,99	6,76	13,6	331	0,71	13,9	0
8	23	5	83	190	2	3	4	0	2	1	0	3	127,24	55,9	107,3	0,81	3,07	5,21	13,1	212	0,75	10,4	0
9	41	1	67	165	1	3	1	1	3	0	0	4	129,62	41,2	101,9	3,57	0,99	10,7	13,2	327	0,55	20,9	0
10	50	6	75	169	2	3	1	1	1	0	0	4	86,5	69,5	48,5	0,34	2,19	6,05	11,6	297	0,81	12	0
11	48	6	85	170	1	3	1	1	1	0	0	0	94,8	39,2	169,5	1,26	1,49	6,81	12,7	284	0,65	15,8	0
12	23	7	43	169	1	3	4	1	2	1	0	0	62,2	30,8	73,5	2,21	5,83	7,57	12,9	82	0,67	42,1	1
13	41	8	86	165	1	3	4	1	2	1	0	1	117,74	38,9	121,8	1,3	1,43	6,72	14,7	280	0,83	13,4	0
14	38	9	90	174	2	3	2	0	3	0	0	0	118,94	36,4	255,8	1,59	4,63	8,49	14,7	282	0,84	48	0
15	33	9	60	158	1	3	3	1	3	0	0	0	77,46	51,5	56,7	1,84	0,61	7,43	14,8	447	0,74	73,1	0
16	57	6	85	171	2	3	4	2	1	1	0	2	177,44	62,7	85,3	6,33	0,93	7,11	14,7	336	0,83	12,7	0
17	54	6	69	160	1	3	3	1	1	1	0	1	80,92	33	182,9	1,27	1,58	5,76	12,3	313	0,84	37,2	1
18	26	10	55	164	1	3	2	1	2	0	0	0	119,2	38,6	76	0,6	1,92	7,41	15	279	0,95	34,6	0
19	32	10	53	167	1	3	1	1	1	1	0	0	110,5	49,7	174,5	4,82	3,7	6,43	11,7	322	0,56	12,3	0
20	25	10	57	160	1	3	2	1	1	0	0	0	143,36	37,9	207,2	2,48	0,81	5,01	14,4	191	0,71	21,7	0

Yaş: Bireylerin yaşları

Kilo: Bireylerin kiloları

Boy: Bireylerin boyları

Cinsiyet: (1=kadın, 2=erkek)

Meslek: (1= Ev Hanımı, 2= Site görevlisi, 3= Kütüphaneci, 4= Öğretim üyesi, 5= Öğrenci, 6= Emekli, 7= Sekreter, 8= Öğretmen, 9= Kasap, 10= Kuaför, 11=Direksiyon Eğitimi, 12= Güvenlik, 13= Memur, 14= İşçi, 15= Spiker, 16= Grafiker, 17= Katip)

Et Tüketimi: (1= 1/hafta, 2= 2-4/hafta, 3= 5 günden fazla/hafta)

Yaşadığı Yer: (1=Köy, 2=İlçe, 3=Merkez)

Fiziksel Aktivite: 0= Hiç, 1= Nadiren, 2= Az düzeyde, 3= Düzenli, 4= Haftanın her günü)

LDL (Düşük Yoğunluklu Lipoprotein) Kolesterol: Bireylerin LDL kolesterol seviyeleri.

Trigliserit: Bireylerin trigliserit seviyeleri.

HDL (Yüksek Yoğunluklu Lipoprotein) Kolesterol: Bireylerin HDL kolesterol seviyeleri.

CRP (C-reaktif Protein): Bireylerin CRP seviyeleri.

WBC (White Blood Cell): Bireylerin beyaz kan hücresi seviyeleri.

HG (Hemoglobin): Bireylerin hemoglobin seviyeleri.

PLT (Platelet): Bireylerin trombosit seviyeleri.

Kreatinin: Bireylerin kreatinin seviyeleri.

Eğitim Durumu: (1=İlkokul, 2=Ortaokul, 3=Lise, 4=Üniversite)

Kahve İçmesi: (0, 1= 1-3 fincan/gün, 2= 5 fincandan fazla)

Alkol Tüketimi: ((0, 1= 1-3/hafta, 2= 4/hafta)

TSH (Tiroid Uyarıcı Hormon): Bireylerin TSH seviyeleri.

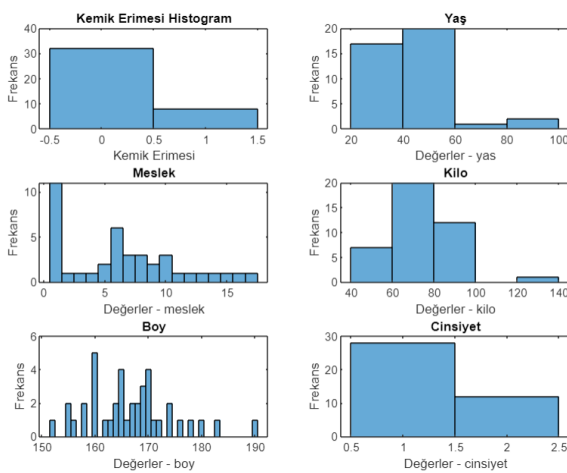
ALT (Alanin Aminotransferaz): Bireylerin ALT seviyeleri.

Süt Tüketimi: Bireylerin süt tüketim alışkanlıkları. (0, 1= 1-3 porsiyon, 2= 3 porsiyondan fazla)

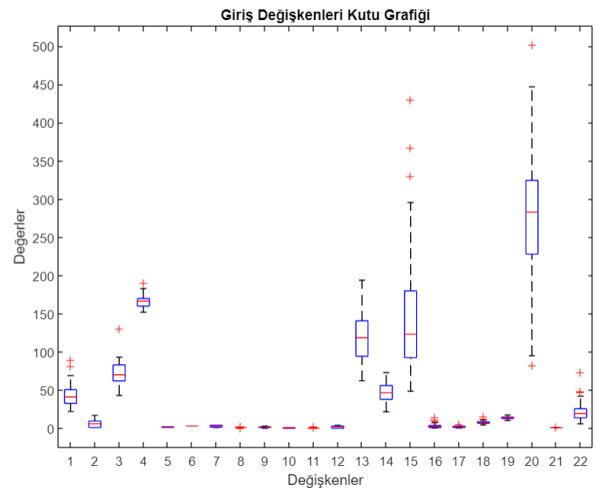
Kemik Erimesi: Bireylerin kemik erimesine sahip olup olmadığını belirten hedef değişken. (1= Kemik erimesine sahip, 0= Kemik erimesi yok)

3. Veri İncelemesi ve Görselleştirme

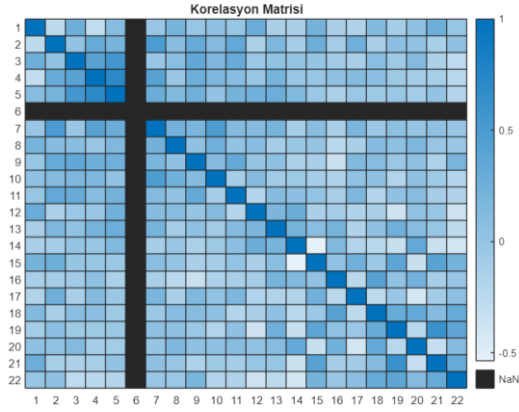
Proje kapsamında kullanılan veri setinin yapısını ve özelliklerini anlamak için veri görselleştirme yapılmıştır. Bu aşamada, veri setindeki özniteliklerin dağılımı ve birbiriyle olan ilişkileri incelenmiştir. Aşağıda, kullanılan görselleştirme yöntemleri ve elde edilen sonuçlar detaylı bir şekilde açıklanmıştır:



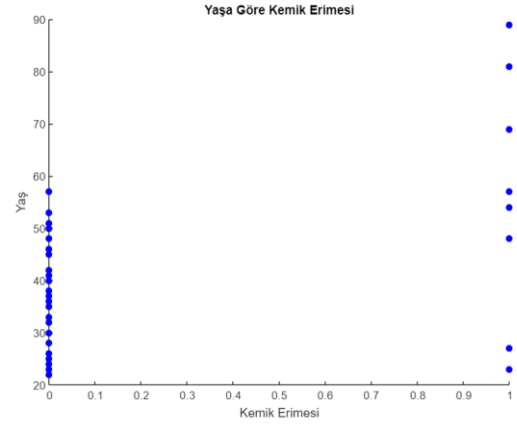
Şekil 1. Histogram Grafiği



Şekil 2. Kutu Grafiği



Şekil 3. Korelasyon Matrisi



Şekil 4. Scatter Plot

3.1. Histogramlar:

Şekil 1'deki histogramlar, veri setindeki sayısal özniteliklerin dağılımını görsel olarak göstermek için kullanılmıştır. Örneğin, yaş, kilo, boy gibi özniteliklerin dağılımları histogramlarla analiz edilmiştir. Histogramlar, özniteliklerin merkezi eğilimini, dağılım şeklini ve olası aykırı değerleri görmek için faydalıdır.

3.2. Kutu Grafiği:

Şekil 2'deki kutu grafiği, veri setindeki sayısal özniteliklerin dağılımını ve merkezi eğilimini kutu ve uçlarındaki çizgiler aracılığıyla gösterir. Bu grafiği kullanarak, özniteliklerin medyan değerlerini, çeyrekler arası aralıklarını ve potansiyel aykırı değerleri görsel olarak inceleyebiliriz.

3.3. Korelasyon Matrisi:

Şekil 3'deki korelasyon matrisi, veri setindeki öznitelikler arasındaki ilişkileri göstermek için kullanılmıştır. Korelasyon matrisi, her bir öznitelik çifti arasındaki Pearson korelasyon katsayılarını içerir. Bu sayede, öznitelikler arasındaki pozitif, negatif veya zayıf ilişkileri anlayabiliriz.

3.4. Scatter Plotlar:

Şekil 4'teki scatter plotlar, iki sayısal öznitelik arasındaki ilişkiyi görsel olarak göstermek için kullanılmıştır. Örneğin, kemik erimesi ile yaş ve trigliserit seviyeleri arasındaki ilişkiyi anlamak için scatter plotlar çizilmiştir. Bu sayede, öznitelikler arasındaki ilişkiler ve olası desenler görsel olarak incelenmiştir.

Veri görselleştirme aşaması, projede kullanılan veri setinin yapısını anlamak ve yapay sinir ağı modelinin eğitimine rehberlik etmek için kritik bir öneme sahiptir. Elde edilen görselleştirmeler, veri setindeki önemli özellikleri ve ilişkileri anlamamıza yardımcı olmuştur.

4. Veri Ön İşleme ve Normalizasyonu

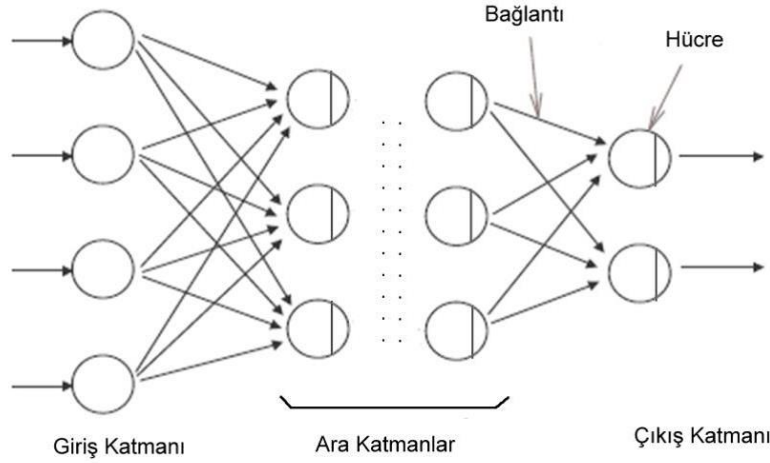
Veri setindeki özniteliklerin min-max normalizasyonu uygulanarak [0, 1] aralığına ölçeklendirilmiştir. Bu işlem, veri setindeki özniteliklerin farklı aralıklarda olmasından kaynaklanan ölçek farklılıklarını gidermeyi amaçlamaktadır. Yapay sinir ağı modellerinin daha iyi performans göstermesine yardımcı olan bu normalizasyon işlemi, özellikle gradient tabanlı optimizasyon algoritmaları kullanıldığında önemlidir.

Normalizasyon, modelin daha hızlı ve daha istikrarlı bir şekilde eğitilmesine yardımcı olur. Ölçek farklılıklarının giderilmesi, modelin optimizasyon sürecini iyileştirir. Aynı zamanda, aşırı öğrenmeyi azaltabilir. Ölçek farklılıklarının büyük olması durumunda, modelin bazı öznitelikleri diğerlerine göre daha fazla vurgulaması ve bu da aşırı öğrenmeye yol açabilir. Normalizasyon, bu tür durumları önlemeye yardımcı olur. Ayrıca, modelin

genel performansını artırır. Ölçek farklılıklarının giderilmesi, modelin daha doğru ve tutarlı tahminler yapmasını sağlar.

5. Yapay Sinir Ağı Modeli Oluşturma ve Eğitime

İleri beslemeli yapay sinir ağı (YSA), en temel yapay sinir ağı türlerinden biridir ve bilgi akışının bir yönde, girişten çıkışa doğru (ileri), geriye dönüş olmadan ilerlediği bir modeldir. Yani, veri girişi ağıya verilir, ardından gizli katman(lar) boyunca işlenir ve sonunda çıkış katmanından bir çıkış üretilir.



Şekil 5. İleri Beslemeli Sinir Ağı

Yapay sinir ağı modeli, feedforwardnet fonksiyonu kullanılarak oluşturulmuştur. Bu fonksiyon, MATLAB'un Neural Network Toolbox'undaki bir fonksiyondur. Bu fonksiyon, giriş katmanı, gizli katman(lar) ve çıkış katmanı olmak üzere üç tip katmandan oluşan bir YSA oluşturur.

Veri setinizde 22 öznitelik (giriş değişkeni) ve bir hedef değişken (çıkış değişkeni) bulunuyor:

Giriş Katmanı (Input Layer): Giriş katmanı, modelin dış dünyadan aldığı verileri alır ve ağına içine aktarır. Veri setimizde 22 öznitelik bulunduğu için, giriş katmanının boyutu 22 olacaktır.

Gizli Katmanlar (Hidden Layers): Gizli katmanlar, giriş verilerini alır ve içsel karmaşıklıkları temsil eden nöronlar içerir. Bu örnekte, gizli katman boyutlarını belirledik, sırasıyla 15, 20 ve 15 nöron içeren üç gizli katmanımız var.

Çıkış Katmanı (Output Layer): Çıkış katmanı, modelin tahmin etmeye çalıştığı değerleri üretir. Veri setindeki çıkış değişkeni kemik erimesi olup, bu değişkenin 0 veya 1'dir. Dolayısıyla, çıkış katmanının boyutu 1 olacaktır.

Yapay sinir ağı modelinin giriş, gizli ve çıkış katmanlarının boyutları şu şekildedir:

Giriş Katmanı: 22 nöron

Gizli Katmanlar: 15, 20 ve 15 nöron içeren üç gizli katman

Çıkış Katmanı: 1 nöron

5. 1. Veri Seti Bölme

Veri seti, modelin eğitim ve test verileri üzerinde doğru bir şekilde değerlendirilmesini sağlamak amacıyla %70 eğitim ve %30 test olacak şekilde rastgele ikiye ayrılmıştır. Bu bölme işlemi, veri setinin eğitimde kullanılarak modelin parametrelerinin öğrenilmesi ve daha sonra test setindeki performansının değerlendirilmesi için önemlidir. Kodlarımız kod parçası, veri setini Hold-Out yöntemi kullanarak ikiye bölmektedir. cvpartition

fonksiyonu, veri setini belirtilen oranlarda rastgele ikiye ayırmak için kullanılır. Bu şekilde, X_train ve y_train değişkenleri eğitim verilerini, X_test ve y_test değişkenleri ise test verilerini temsil eder.

5. 2. Modelin Oluşturulması

Yapay sinir ağı modeli, kemik erimesi tanısında kullanılmak üzere belirlenen özneliklerden öğrenme yapmak için oluşturulmuştur. Model, [15 20 15] gizli katman boyutlarına sahip bir yapı kullanılarak inşa edilmiştir. Ayrıca, modelin eğitilmesi için trainlm geri yayılım algoritması tercih edilmiştir. Kodumuz, yapay sinir ağı modelinin oluşturulmasını ve eğitimi gerçekleştirir. feedforwardnet fonksiyonu, giriş ve çıkış katmanları arasındaki bağlantıları ve gizli katmanlarının boyutunu belirler. trainlm geri yayılım algoritması, ağın ağırlıklarını eğitmek için kullandık.

```
hiddenLayerSizes = [15 20 15]; % Gizli katman boyutları
```

```
net = feedforwardnet(hiddenLayerSizes, 'trainlm');
```

Burada hiddenLayerSizes, gizli katmanların boyutlarını belirten bir vektördür. Bu vektördeki her bir sayı, bir gizli katmandaki nöron sayısını temsil eder.

'trainlm' ise eğitim fonksiyonunu belirtir. Burada 'trainlm', Levenberg-Marquardt geri yayılım algoritması kullanılarak ağı eğitmek için seçilmiştir.

Yapay sinir ağı modelinin eğitim parametreleri aşağıdaki gibi ayarlanmıştır:

```
net.trainParam.show = 20; Her 20 iterasyonda bir eğitim ilerlemesinin ekrana yazdırılması sağlanmıştır.
```

```
net.trainParam.epochs = 20; Eğitimin maksimum iterasyon sayısı 20 olarak belirlenmiştir.
```

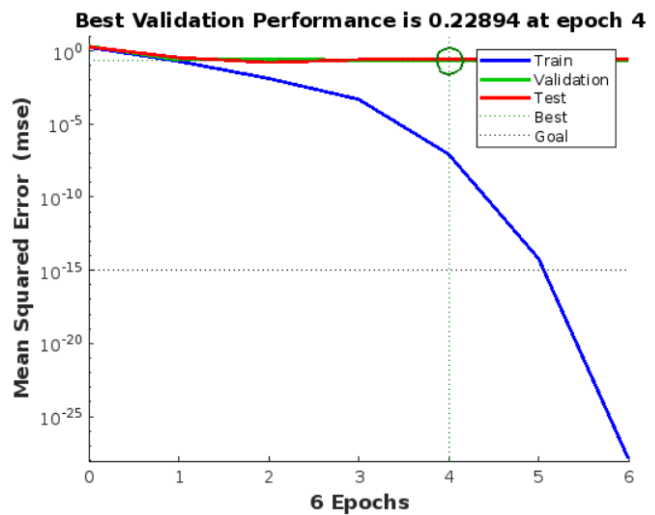
```
net.trainParam.goal = 1e-25; Eğitim hatasının minimum değeri 1e-25 olarak belirlenmiştir.
```

Yapay Sinir Ağı'nın Eğitilmesinde ise oluşturulan YSA modeli, veri seti üzerinde eğitilmiştir. Eğitim işlemi, train fonksiyonu kullanılarak gerçekleştirilmiştir.

```
[net,tr] = train(net, X_train', y_train'); Bu satır ile modelin eğitilmesi sağlanmıştır.
```

6. Model Performansı

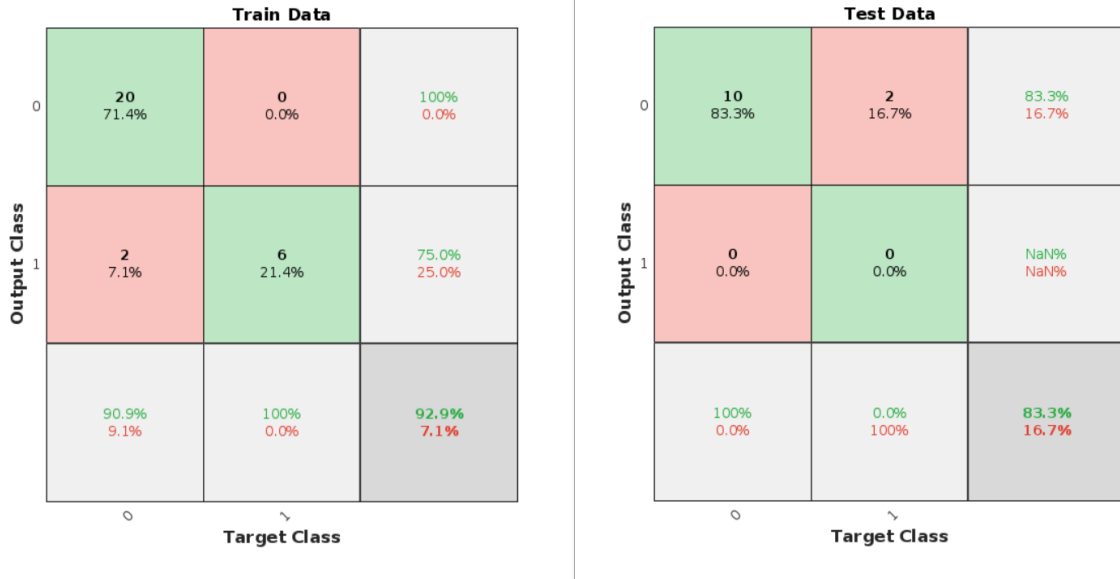
Eğitim verileri üzerinde yapılan tahminler ile gerçek çıkış değerleri arasındaki hatalar hesaplanır. Bu, modelin eğitim setinde ne kadar iyi veya kötü performans gösterdiğini anlamak için önemlidir. Hata hesaplama adımı, tahmin edilen çıkışlar ve gerçek çıkışlar arasındaki farkların karesinin ortalamasını alarak MSE (Mean Squared Error - Ortalama Kare Hatası) değerini elde eder.



Şekil 6. MSE

Şekil 6'da görüldüğü üzere eğitim sonucunda elde edilen performans, ortalama karesel hata (Mean Squared Error - MSE) değeri kullanılarak değerlendirilmiştir. Test Seti İçin Mean Squared Error: 0.16128, Eğitim Seti İçin Mean Squared Error: 0.0707860.01 olarak hesaplanmıştır ve modelin gerçek değerler ile tahmin ettiği değerler arasındaki ortalama karesel farkın düşük olduğunu göstermektedir.

predictedTrainOutput, modelin eğitim seti üzerindeki çıkışlarını temsil ederken y_train ise gerçek çıkış değerlerini içerir. plotconfusion fonksiyonu, confusion matrixi çizmek için kullanılır. Bu şekilde, modelin eğitim seti üzerindeki performansını daha iyi anlamak için hem sayısal hem de görsel bir değerlendirme yapılır.



Şekil 7. Confusion Matrisi

Confusion matrisi, sınıflandırma modellerinin performansını değerlendirmek için kullanılan bir araçtır. Her bir sınıf için gerçek ve tahmin edilen etiketlerin sayısını içeren bir tablodur. Bu matris, doğru ve yanlış sınıflandırmaları görselleştirerek modelin sınıflandırma yeteneğini değerlendirir. Şekil 7’ de görüldüğü gibi test datası için accuracy değeri %83.3, train datası için accuracy değeri ise %92.9 çıkmaktadır.

7. Sonuçlar

Sonuç olarak bu proje, kemik erimesi tanısında yapay sinir ağlarının etkinliğini göstermek amacıyla yapılmıştır. Yapılan analizler, kullanılan özniteliklerin (yaş, meslek, kilo, boy, cinsiyet vb.) kemik erimesi tahmininde önemli bir rol oynadığını ortaya koymaktadır. Özellikle, kemik erimesi ile ilişkilendirilebilecek özniteliklerin belirlenmesi ve bunların yapay sinir ağı modeline entegre edilmesi, modelin tahmin performansını artırmıştır.

MSE (Mean Squared Error - Ortalama Kare Hatası) değerinin düşük olması ve doğruluk değerinin yüksek olması, modelin kemik erimesi tanısında başarılı olduğunu göstermektedir. Bu bulgular, yapay sinir ağlarının kemik erimesi tanısında potansiyel bir araç olarak kullanılabileceğini ve klinik uygulamalarda yardımcı olabileceğini öne sürmektedir.

8. KAYNAKÇA

1. Web Sitesi:

Derin Öğrenme (2017, Mart 4) Yapay Sinir Ağları [Web Site] <https://www.derinogrenme.com/2017/03/04/yapay-sinir-aglari/>

2. Doküman:

Dergi Park (2019, Haziran 24). Osteoporoz ve Tedavisi [PDF]. <https://dergipark.org.tr/tr/download/article-file/760328>

Dicle Tıp Dergisi (2008, Ocak 10). Osteoporoz Riskinin Yapay Sinir Ağları Yöntemi ile Saptanması [PDF]. <https://dergipark.org.tr/tr/download/article-file/760328>