



Makale / Research Paper

Veri İndirgeme ve Sınıflandırma Teknikleri ile Glokom Hastalığı Teşhisi

Esra ŞATIR¹, Ferdi AZBOY², Aziz AYDIN³, Hatice ARSLAN⁴, Şerife HACİEFENDİOĞLU⁵

^{1,2,3} Düzce Üniversitesi, Mühendislik Fakültesi, Bilgisayar Mühendisliği Bölümü, 81620, Düzce, TÜRKİYE

⁴ Düzce Üniversitesi, Mühendislik Fakültesi, Elektrik Elektronik ve Bilgisayar Mühendisliği Anabilim Dalı,
81620, Düzce, TÜRKİYE

⁵ Selçuk Üniversitesi, Mühendislik Fakültesi, Elektronik ve Bilgisayar Sistemleri Anabilim Dalı, 42000, Konya,
TÜRKİYE

esrasatir@duzce.edu.tr, ferdi46280@ogr.duzce.edu.tr, aziz45024@ogr.duzce.edu.tr,
hatice45048@ogr.duzce.edu.tr, manane@msn.com

Özet: Glokom göz içi basıncının sıklıkla yükselmesi nedeniyle görme sinirinin zarar görmesidir. Kendini hastalığın en son aşamalarında fark ettiren sinsi bir hastalık olan glokom, geç tanı konulduğunda görme sinirinde onarılması mümkün olmayan ciddi tahribatlar oluşturabilmektedir. Bu çalışmadaki amaç az nitelikte glokom hastalığının teşhis edilmesidir. Çalışmada Haciefendioğlu'nun yüksek lisans tezinden alınan hasta bilgileri veri kümesi olarak kullanılmıştır. Bu veri kümesindeki nitelik sayısı, kaba kümeler yöntemleri ile 8'den 5'e düşürülmüştür. İndirgenen veriler Karar Ağaçları ve Yapay Sinir Ağları kullanılarak sınıflandırılmış ve ardından da Çapraz Doğrulama ile performans değerlendirilmesi gerçekleştirilmiştir.

Anahtar Kelimeler: Glokom; kaba kümeler; karar ağaçları; yapay sinir ağları.

Diagnosis of Glaucoma Disease via Data Reduction and Classification Techniques

Abstract: Glaucoma is the damage on optic nerve due to the frequent increase of intraocular pressure. Accordingly, an individual's field of vision becomes narrower, slowly. Glaucoma, a sly disease that makes itself noticeable in the last stage, causes serious damages which are impossible to recover, if diagnosed late. The purpose here is the diagnosis of glaucoma with fewer attributes. In this study, the patient information which belongs to the graduate thesis of Haciefendioğlu, has been used. Number of the attributes in the dataset has been decreased from 8 to 5 by employing Rough Sets theory. The reduced data has been classified by using Decision Trees and Artificial Neural Networks and then performance evaluation has been performed via Cross Validation method.

Key Words: Glaucoma; rough sets; decision tree; artificial neural networks.

1. Giriş

Glokom, optik sinir adı verilen ana görme siniri ile ilgili bir hastalıktır. Optik sinir, retinadan gelen ışık üreten sinir sinyallerini alır ve bu elektrik sinyallerini görme olarak algıladığımız yere, beyne iletir. Glokom, genellikle hafif bir yan görüş kaybı ile başlayan, sinir hücrelerine ilerleyen hasarın

Bu makaleye atıf yapmak için

Şatır, E., Azboy, F., Aydın, A., Arslan, H., Haciefendioğlu, Ş., "Veri İndirgeme ve Sınıflandırma Teknikleri ile Glokom Hastalığı Teşhisi" El-Cezerî Fen ve Mühendislik Dergisi 2016, 3(3); 485-497.

How to cite this article

Şatır, E., Azboy, F., Aydın, A., Arslan, H., Haciefendioğlu, Ş., "Diagnosis of Glaucoma Disease via Data Reduction and Classification Techniques" El-Cezerî Journal of Science and Engineering, 2016, 3(3); 485-497.

yeterince erken bir safhada teşhis edilirse tedavi edilebilir [3]. Aşağıdaki paragraflarda, glokom teşhisi için gerçekleştirilen bazı çalışmalara yer verilmektedir.

Huang ve ark. tarafından 2005 yılında konuyla ilgili çalışmalarında glokomlu, glokomsuz kişilerin sınıflandırılmaları için ANFIS (Adaptive Neuro-Fuzzy Inference System) kullanmıştır. 135 hasta 206 sağlıklı bireyin verilerinin kullanıldığı çalışmada Stratus OCT cihazının verileri kullanılmıştır. Sistemde en küçük kareler metodu ile geri yayılım azalan gradyant yöntemi birleştirilmiştir. Çalışmada alt kadran kalınlığının glokomlu bireylerin sınıflandırılmasında karakteristik bir önemi olduğu belirtilmiştir [2].

Bock ve ark. tarafından 2007 yılında renkli fundus resimleri ile Almanya’da yapılmış bir diğer çalışmada ise obje ve yüz tanımadaki gibi doğrudan fundus resimleri ile sınıflandırma yöntemine gidilmiş, sınıflandırma için resimlerle ön çalışma yapıldıktan sonra Naive Bayes, k-Nearest Neighborhood, ve SVM (Support Vector Machine) kullanılarak karşılaştırılmıştır. 50 hasta ve 50 sağlıklı bireye ait fundus resimlerinde yapılan sınıflandırmada en başarılı performans SVM yöntemi ile elde edilmiştir [4].

Matsopoulos ve ark. tarafından 2008 yılında göz damarlarının şekillerinde glokom sebebiyle oluşan değişimleri belirlemek için yapılan çalışmada, önce fundus resimlerinde SOM (Self Organization Map) ile damarlar belirlenmiş daha sonra YSA (Yapay Sinir Ağları) ile sınıflandırma yöntemine başvurulmuştur. %87,5 sınıflandırma oranı elde edilmiştir [2].

Bizios ve ark. tarafından 2010 yılında yapılan çalışmada, glokom teşhisinde Optik Koherens tomografisi (OCT) ile Retina Sinir Lifi Tabakası Kalınlığı (RNFLT) ölçümlerine dayanan giriş ile Destek Vektör Makineleri (SVMs), Yapay Sinir Ağları (YSA) ve iki Makine Öğrenme Sınıflayıcısı (MLCs) performansı karşılaştırılmış ve farklı giriş parametrelerinin etkileri değerlendirilmiştir. 90 sağlıklı kişiden ve 62 glokom hastasından elde edilen Straous OCT verisi analiz edilmiştir [5].

Bizios ve ark. tarafından 2011 yılında gerçekleştirilen bir başka çalışmada ise hem SAP (Standard Automated Perimetry) hem de OCT (Optik Kohorens Tomografi) ölçümlerinden elde edilen veriler teker teker ve kombine olarak glokom teşhisi için YSA’ya giriş olarak verilmiştir. Her iki ölçümün birlikte verildiği YSA’nın performansı %95,39 bulunmuştur. 125 sağlıklı 125 glokomlu bireyin verileri ile SAP ve OCT ölçümlerinin teker teker giriş verisi olarak kullanıldığında elde edilen performanstan daha iyi bir neticeye ulaşılmıştır [2].

Işık ve Hacıfendioğlu tarafından 2012 yılında gerçekleştirilen çalışmada, söz konusu hastalığın göz sinirleri hasar görmeden önce teşhis edilebilmesi ve dünyada körlük nedenleri arasında ilk sıralarda yer alan hastalığın tahmin edilebilmesi amaçlanmıştır. Çalışmada Pamukkale Üniversitesi Göz Hastalıkları Anabilim Dalından alınan hasta bilgileri kullanılmıştır. Bu çalışmada makine öğrenmesi sınıflandırma yöntemlerinden 3 önemli yöntem olan SVM, YSA ve Karar Ağaçları kullanılarak glokom hastalığı başlangıç safhasında teşhisi için sınıflandırma yapılmış ve birbirleri ile karşılaştırılmıştır. Adı geçen makine öğrenmesi yöntemlerinin performansları arasında en yüksek sınıflandırma başarısının SVM ile elde edildiği görülmüştür [2].

Belghith ve ark. tarafından 2013 yılında yapılan çalışmada, bir glokom ilerlemesi algılama tekniği, HRT (Heidelberg Retina Tomografi) görüntüleri kullanarak ileri sürülmüştür. Değişim haritasında konumsal piksel bağımlılığına entegre olmayan mevcut metotların aksine MRF (Markov Rastgele Alan) kullanımı böyle bir bağımlılığın üstesinden gelmek için ileri sürülmüştür. Model parametrelerini hesaplamak için Monte Carlo Markov Zinciri prosedürü kullanılmıştır. Daha sonra mevcut glokom ilerleyişini algılama metotları için önerilen çerçevenin teşhis performansı karşılaştırılmıştır [6].

Bu çalışmadaki amaç, az nitelik kullanarak glokom verilerinde verimli sınıflandırma yapılması ve söz konusu hastalığın göz sinirleri hasar görmeden önce tahmin ve teşhis edilebilmesidir. Çalışmada Hacıfendioğlu'nun [2] yüksek lisans tezinden alınan veri kümesi kullanılmıştır. Bu veride 84 glokom hastası ve 84 sağlıklı olmak üzere toplam 168 kişi bulunmaktadır. Veriler öncelikle kaba kümeler yöntemleri ile indirgenmiş ve indirgenen bu veriler Karar Ağaçları (C4.5, ID3 ve CART) ve Yapay Sinir Ağları teknikleri ile sınıflandırılmıştır. Bu çalışma, dört bölüm olarak organize edilmiştir. 1.bölümde, glokom tanımı ve literatürdeki çalışmalardan bahsedilmiştir. 2. bölümde, çalışmada kullanılan metotlar anlatılmış, 3. bölümde ise gerçekleştirilen deneyler açıklanmıştır. Çalışma neticesinde ulaşılan sonuç ve değerlendirmelere ise 4. Bölümde yer verilmiştir.

2. Yöntem

2.1. Kaba kümeler

Kaba kümeler (Rough Sets), Zdzislaw Pawlak tarafından 1980'li yıllarda geliştirilmiştir. Kaba kümeler (KK) kesin olmayan, eksik veya belirsiz veri kümelerinde bilgi keşfi, gizli örüntülerin ortaya çıkarılmasında güçlü matematiksel yaklaşımlar sunar. KK'ler özellik seçimi, özellik çıkarımı, değişken indirgeme, karar kurallarının çıkarılması ve desen çıkarımı için kullanılabilirler [7].

- Bilgi sistemi

Kaba kümelerde bilgi sistemi $S = (U, Q, V)$ şeklinde tanımlanır. Burada $U = \{X_1, X_2, \dots, X_n\}$ sonlu boş olmayan evrendir. Evren, elde edilen gözlemler kümesidir. $Q = AUd$ sonlu boş olmayan özellik kümesidir. Özellik kümesi durum (A) ve (d) karar özelliklerinin birleşiminden oluşmaktadır. $V = UV_a, a (a \in A)$ özelliğine ait değer kümesidir [7].

- Ayırt edilemezlik ilişkisi

Bir veri kümesinin gereğinden büyük olması veya elde edilen gözlemlerin birbirine benzemelerinden ya da aynı olmalarından dolayı gözlemler birbirinden ayırt edilememektedir. $B \subseteq A$ olmak üzere B özellik için ayırt edilemezlik ilişkisi $IND(B)$;

$$IND(B) = \{(x_1, x_2) \in U \times U : a \in B, a(x_1) = a(x_2)\} \quad (1)$$

şeklinde gösterilir. Burada $IND(B)$, B -ayırt edilemezlik bağıntısıdır. Eğer x_1 ve x_2 $IND(B)$ kümesine ait ise B özellik kümesi ile x_1 ve x_2 birbirinden ayırt edilememektedir. Gözlem kümesi ($U =$ evren), $U/IND(B)$ biçiminde B -ayırt edilemezlik bağıntısına göre birkaç denklik sınıfına bölünebilir. Bu denklik sınıfları $[x]_{IND(B)}$ şeklinde gösterilir. $IND(B)$ 'nin tüm denklik sınıfları B 'nin temel kümesini oluşturmaktadır. Evrenin karar özelliğine göre denklik sınıfları karar özelliğinin değer sınıflarını oluşturur [7].

- Küme yaklaşımları

Kaba kümelerde temel amaç $IND(B)$ ikili bağıntısı ile yaklaşımların oluşturulmasıdır. $X \subseteq U$ olmak üzere X 'in ayırt edilemezlik bağıntısı kullanılarak X 'e kesin olarak ait olan kümelerin birleşimidir ve,

$$\underline{BX} = U\{x_i \in U \mid [x_i]_{IND(B)} \subseteq X\} \quad (2)$$

biçiminde gösterilebilir, üst yaklaşım ise,

$$\underline{BX} = U \{x_i \in U | [x_i]_{IND(B)} \cap X \neq \Phi\} \quad (3)$$

biçiminde gösterilir [7].

$X \subset U$ 'ya ait alt ve üst yaklaşımlar evreni (U) $POS(X)$ pozitif bölge, $NEG(X)$ negatif bölge ve $BND(X)$ sınır bölgesi şeklinde üç bölgeye ayırmaktadır. Bu bölgelere ait kümeler,

$$POS(X) = \underline{BX}$$

$$NEG(X) = U - \overline{BX} \quad (4)$$

$$BND(X) = \overline{BX} - \underline{BX}$$

şeklinde hesaplanmaktadır.

- Özelliklerin önemi

Özellik seçiminde, $a \in A$ gibi bir özelliğin önemi $P \subseteq A$ ve $R \subseteq A$ gibi R ve P özellik kümeleri arasındaki bağımlılık ile açıklanabilir. Çünkü eğer bir P özellikler kümesinden olan özellikler tek olarak bir R özellikler kümesinden özelliklerin değerleri tarafından belirleniyorsa, P tamamı ile R 'ye bağımlıdır ve $P \Rightarrow R$ şeklinde gösterilir P ve R özellik kümeleri arasındaki bağımlılık,

$$\gamma_R(P) = \frac{|POS_R(P)|}{|U|} \quad (5)$$

şeklinde hesaplanmaktadır. Burada $POS_R(P)$ $U/IND(B)$ bölgelerinde R özellik kümesine ait tüm denklik sınıfların pozitif bölgelerinin birleşimini belirtmektedir. $\gamma_R(P)$, değeri 0 (sıfır) ile 1 (bir) arasında bir değerdir. $\gamma_R(P) = 1$, çıkması durumunda P , R 'ye tamıyla bağımlıdır. Eşitlik 5'ten görüldüğü gibi a gibi bir değişkenin önemi $POS_R(P)$ 'ye bağlıdır. $a \in R$ olmak üzere P ve R özellik kümeleri için, $POS_R(P) = POS_{R-\{a\}}(P)$ ise a özelliği indirgenebilir. Diğer bir deyişle R özellik kümesindeki a özelliği P özellik kümesine göre gereksizdir. $B \subset A$ olmak üzere, eğer $POS_B(d) = POS_A(d)$ ise d karar özelliği için B özellik kümesinde bulunan özellikler önceliklidir. Bir bilgi sistemi için birden fazla indirgenmiş küme söz konusu olabilir [7].

- Çekirdek özellik kümesi

Bir bilgi sisteminden elde edilen indirgenmiş kümelerin kesişiminden elde edilen kümeye çekirdek (Core) özellik kümesi olarak isimlendirilmektedir. Çekirdek özellik kümesi ayırt edilebilirlik matrisinden de elde edilebilir [7].

- Ayırt edilebilirlik matrisi

S bilgi sisteminde A durum özellikleri için ayırt edilebilir matris $M(A) = (m_{ij})_{n \times n}$ dir. $M(A)$,

$$M(A) = \begin{cases} \Phi \\ \{a \in A : a(x_i) \neq a(x_j)\} \end{cases} \quad (6)$$

şeklinde gösterilir. $M(A)$ ayırt edilebilir matris simetrik özelliğine sahiptir. $M(A)$ 'nın her elamanı x_i ve x_j değerlerini farklı kılan özellikler kümesinden oluşmaktadır [7].

- Özellik seçim metodu

A özellik kümesinin ayırt edilebilirlik matrisine bağlı olarak yeni bir bilgi sistemi S' ,

Evren : $U' = \{(x_i, x_j) \in U \times U : d(x_i) \neq d(x_j)\}$ şeklinde oluşturulabilir [7].

- Kural çıkarımı

Durum özelliklerine göre kural sayısı ciddi anlamda artmaktadır. Ancak özelliklerin kategorik olması ve yaklaşımların iyi olması durumunda kural sayısı orantılı olarak artmamaktadır [7].

2.2. Karar ağaçları

Sınıflama modelleri içerisinde yer alan karar ağaçları yöntemleri tahmin edici ve tanımlayıcı özelliklere sahiptir. Karar ağaçları, yorumlanmalarının kolay olması, veri tabanı sistemleri ile kolayca entegre edilebilmeleri, güvenilirliklerinin daha iyi olması nedenlerinden dolayı sınıflama modelleri içerisinde en yaygın kullanıma sahip olan yöntemlerden birisidir. Ayrıca yapılandırması ve anlaşılması daha kolay bir yöntem olması, model şeffaflığını sağlaması ve görsel bir sunuma sahip olması da yaygın kullanımına sebep olarak gösterilebilir [2].

Karar ağacı tekniğini kullanarak verinin sınıflanması, öğrenme ve sınıflama olmak üzere iki basamaklı bir işlemdir. Öğrenme basamağında önceden bilinen bir eğitim verisi, model oluşturmak amacıyla sınıflama algoritması tarafından analiz edilir. Öğrenilen model, sınıflama kuralları veya karar ağacı olarak gösterilir. Sınıflama basamağında ise test verisi, sınıflama kurallarının veya karar ağacının doğruluğunu belirlemek amacıyla kullanılır. Eğer doğruluk kabul edilebilir oranda ise kurallar, yeni verilerin sınıflanması amacıyla kullanılır. Eğitim verisindeki hangi alanların hangi sırada kullanılarak ağacın oluşturulacağı belirlenmelidir. Bu amaçla en yaygın olarak kullanılan ölçüm, entropi ölçümüdür. Entropi ölçüsü ne kadar fazla ise o alan kullanılarak ortaya konulan sonuçlar da o oranda belirsiz ve kararsızdır. Bu nedenle, karar ağacının kökünde entropi ölçüsü en az olan alanlar kullanılır. Verilen bir A_k alanının entropi ölçüsünü bulan formüller şu şekildedir [8].

$$E(C|A_k) = \sum_{j=1}^{M_k} p(a_{k,j}) \left[- \sum_{i=1}^N p(c_i|a_{k,j}) \log_2 p(c_i|a_{k,j}) \right] \quad (7)$$

Bu formülde;

$E(C|A_k) = A_k$ alanının sınıflama özelliğinin Entropi ölçüsü,

$p(a_{k,j}) = a_k$ alanının j değerinde olma olasılığı,

$p(c_i|a_{k,j}) = a_k$ alanı j.değerindeyken sınıf değerinin c_i olma olasılığı,

$M_k = a_k$ alanının içerdiği değerlerin sayısı; $j = 1, 2, \dots, M_k$,

N = farklı sınıfların sayısı; $i = 1, 2, \dots, N$,

K = alanların sayısı; $k = 1, 2, \dots, K$.

Eğer bir S kümesindeki elemanlar, kategorik olarak $C_1, C_2, C_3, \dots, C_i$ sınıflarına ayrıştırılırlarsa, S kümesindeki bir elemanın sınıfını belirlemek için gereken bilgi şu formülle hesaplanmaktadır:

$$I(S) = -(p_1 \log_2(p_1) + p_2 \log_2(p_2) + \dots + p_i \log_2(p_i)) \quad (8)$$

Bu formülde p_i, C_i sınıfına ayrılma olasılığıdır.

Entropi denklemini şu şekilde de ifade edilebilir:

$$E(A) = \sum_{i=1}^n \frac{|S_i|}{|S|} x I(S_i) \quad (9)$$

Bu durumda A alanı kullanılarak yapılacak dallanma işleminde, bilgi kazancı şu formülle hesaplanmaktadır:

$$Kazanç(A) = I(S) - E(A) \quad (10)$$

Başka bir deyişle Kazanç (A), A alanının değerini bilmekten kaynaklanan entropideki azalmadır.

Karar ağaçlarında kullanılan birçok algoritma mevcuttur. ID3, C4.5, C5.0, CART, CHAID ve QUEST bunlara örnek olarak gösterilebilir [8]. Bu çalışmada ID3, C4.5, CART algoritmaları kullanılmıştır.

- ID3 algoritması

ID3 algoritması ilk olarak J.Ross Quinlan tarafından Sydney Üniversitesinde geliştirilmiştir. Entropiye dayalı bir algoritma olup sadece ayrık veriler üzerinde çalışabilmektedir. Ancak ID3, eksik verilerle çalışabilme yeteneğine sahip değildir

Veri tabanında çok sayıda değişken varsa ve eğitim kümesi çok fazla kayıt içeriyorsa, fakat fazla hesaplama yapmadan uygun bir karar ağacı oluşturulmak isteniyorsa ID3 algoritması kullanılabilir. ID3 yinelemeli yapıya sahip bir algoritmadır [2].

- C4.5 algoritması

C4.5 algoritması, en çok bilinen karar ağacı algoritmalarından biridir. C4.5 algoritmasında, test öznitelik seçim ölçütü olarak bilgi kazancı oranı kullanılmakta ve her bir set için, en yüksek bilgi kazancı oranına sahip öznitelik seçilmektedir. C4.5 algoritması, ID3 algoritmasına dayanan ve bu algoritmanın bazı kısıtlarını ortadan kaldıran bir yöntemdir. C4.5 algoritması hem sürekli hem ayrık öznitelikler ile çalışabilmektedir. Buna ek olarak, eksik öznitelik değerleri içeren eğitim veri setleri ile çalışabilmektedir. Bunun yanı sıra, karar ağacı oluşturma sırasında ya da sonrasında bazı düğümlerin ya da alt ağaçların silinmesi ile aşırı uygunluk problemini ortadan kaldırmakta, eğitim setindeki istisnai ve gürültülü değerlerin çıkarılmasını sağlamaktadır [9].

- CART algoritması

Morgan ve Sonquist'in AID (Automatic Interaction Detection) adlı karar ağacı algoritmasının devamı niteliğine Breiman ve diğerleri tarafından 1984 yılında önerilmiştir. Hem sayısal hem de nominal veri türlerini, girdi ve kestirimsel değişken olarak kabul edebilen CART algoritması, sınıflandırma ve regresyon problemlerinde bir çözüm olarak kullanılabilir. CART karar ağacı, ikili olarak özyinelemeli biçimde bölünen bir yapıya sahiptir. Dallanma kriteri olarak gini indeksinden yararlanan CART ağacı, kuruluş aşamasında herhangi bir durma kuralı olmaksızın sürekli olarak bölünerek büyümektedir. Artık yeni bir bölünmenin gerçekleşmeyeceği durumda, bu sefer, uçtan

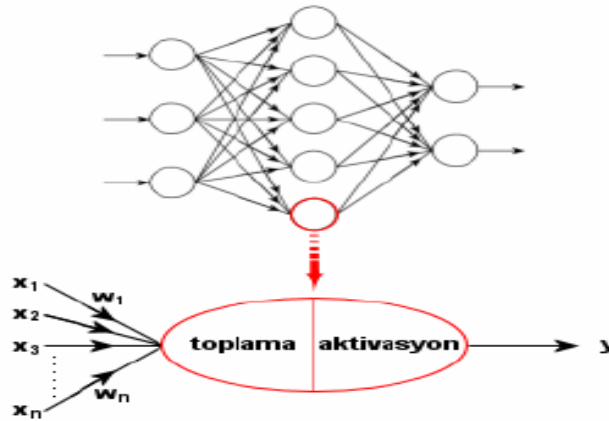
köke doğru budama işlemi başlatılır. Olası en başarılı karar ağacı, her budama işlemi sonrası bağımsızca seçilmiş bir test verisi ile değerlendirme yapılarak tespit edilmeye çalışılır [8].

2.3. Yapay sinir ağları – YSA

Yapay sinir ağları, örnekler aracılığı ile olayları öğrenebilen, öğrendiği bilgileri kullanarak yeni bilgiler türetebilen ve keşfedebilen, böylelikle kazandığı bilgi, birikim ve tecrübeler ile çevreden gelen etkilere insanlarınkine benzer tepkiler verebilen sistemlerdir. Örneklerden elde ettikleri bilgiler ile kendi deneyimlerini oluşturur ve benzer konularda kararlar verebilirler. Bu özellikleri sayesinde ilişkilendirme, sınıflandırma ve optimizasyon konularında başarılı bir şekilde uygulanmaktadırlar [2].

- Yapay sinir hücresinin yapısı

Yapay sinir ağlarının temel birimi sinir hücreleridir. Mühendislik disiplinlerinde, yapay sinir hücrelerine, proses elemanları da denilmektedir. Her bir proses elemanı temel unsura sahiptir. Bunlar, girişler, ağırlıklar, toplama fonksiyonu, aktivasyon fonksiyonu ve çıkıştır. Şekil 4.9.'da görüldüğü gibi her bir hücre çok-girişli tek çıkışlı yapıdadır [2].



Şekil 3. Yapay sinir hücresinin yapısı[2]

Girişler (x_1, x_2, \dots, x_n): Giriş katmanındaki hücreler için, kullanıcı tarafından örnekler ile oluşturulmuş veri kümesidir.

Ağırlıklar (w_1, w_2, \dots, w_n): Girişlerin, çıkışa ne oranda aktarılacağını gösterir. Örneğin w_1 ağırlığı, x_1 girişinin, çıkışa olan etkisini göstermektedir.

Toplama Fonksiyonu: Bir hücrenin net girdisini hesaplamak için kullanılır. Bu amaç ile değişik fonksiyonlar kullanılmaktadır. En fazla tercih edilen, ağırlıklı toplam fonksiyonudur. Bu fonksiyonda her giriş kendi ağırlığı ile çarpılır ve bu değerler toplanır. x_i girişleri, w_i ağırlıkları ve n hücrenin giriş sayısını göstermek üzere ağırlıklı toplam fonksiyonu;

$$NET = \sum_{i=1}^n x_i w_i \quad (11)$$

Yapay sinir ağındaki bütün hücrelerin toplama fonksiyonlarının aynı olması gerekmez. Her hücre bağımsız olarak farklı bir toplama fonksiyonuna sahip olabilir [2].

Aktivasyon Fonksiyonu: Hücrenin net girdi değerine karşılık üretilecek çıktı değerinin hesaplanmasında kullanılır. Çok katmanlı algılayıcılarda olduğu gibi, bazı sinir ağı modellerinde aktivasyon fonksiyonunun türevlenebilir olması şartı vardır. Toplama fonksiyonunda olduğu gibi, bütün hücrelerin aynı aktivasyon fonksiyonunu kullanması gerekmez. Her hücre bağımsız olarak farklı bir aktivasyon fonksiyonuna sahip olabilir [2].

- YSA'ların sınıflanması

YSA'lar yapılarına göre, ileri beslemeli (feedforward) ve geri beslemeli (feedback) ağlar olmak üzere iki şekilde sınıflandırılırken, öğrenme algoritmalarına göre de, danışmanlı, danışmansız ve takviyeli öğrenme olmak üzere üç farklı şekilde sınıflandırılırlar. İleri beslemeli ağlar genellikle örüntü tanıma problemlerinde, geri beslemeli ağlar da optimizasyon problemlerinde yaygın olarak kullanılmaktadır [2].

3. Bulgular

Bu çalışmada kullandığımız veri kümesinde, 84 hasta ve 84 sağlıklı olmak üzere toplam 168 kişi bulunmaktadır. Bu kişilerden 90'ı kadın 58'i erkektir ve yaşları 26-80 aralığında değişmektedir. Glokom hastalığının teşhisinde kullanılan 8 adet nitelik ve normal değer aralıkları Tablo 1'de verilmiştir.

Tablo 1. Glokom parametreleri ve normal değerleri [2]

Parametre	Değer Aralığı
Göz İçi Basıncı	10-21 mmHg
Santral Kornea Kalınlığı	545 um
Disk Area	1,63-2,43mm ²
Cup Area	0,11-0,68 mm ²
Rim Area	1,31-1,96 mm ²
C/D Lineer	0,36-0,80
Mean RNFL Thickness	0,18-0,41
Yaş	40 altı

3.1. Göz içi basıncı - GİB

Glokomun en kolay tanınan bulgusu, göz içi basıncı yüksekliğidir. Ön ve arka kamarayı dolduran aköz hümanın kornea ve skleraya yaptığı basıncıdır [2].

3.2. Santral kornea kalınlığı - SKK

Kornea, saydam ve esnek yapısı olan damarsız dokudur. Korneanın göz içi yapılarını koruma, hava ve gözyaşı tabakası sayesinde gözün en önemli kırıcılık yüzeyini oluşturma görevleri vardır. GİB'nin doğru olarak ölçülmesi glokom tanı ve tedavisinin en önemli basamaklarından birisidir [2].

3.3. Disk alanı

Kontür çizgisi içindeki disk alanıdır. Normal popülasyonda dik alanı çok önemli varyasyonlar gösterir [2].

3.4.Çukurluk alanı

Optik disk çukurluğunu ifade eder ve kontür çizgisi ile çevrelenmiş, referans planının altındaki kısımdır [2].

3.5.Nöroretinal halka alanı (Rim area)

Kontür çizgisi ile çevrelenmiş referans planı üstündeki bölgedir [2].

3.6. Çukurluk/Disk oranı (C/D Lineer)

Çukurluk alanının disk alanına oranıdır. Çukurluk çapının disk çapına bölünmesi ile elde edilen ondalık değerdir. Değer 1'e ne kadar yakın olursa, kayıp o oranda büyüktür [2].

3.7. Ortalama retina sinir lifi tabakası kalınlığı

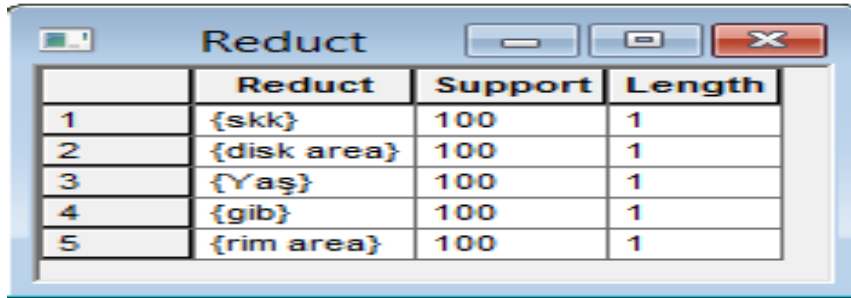
Kontür çizgisi boyunca ortalama retina sinir lifi tabakası kalınlığıdır. Yani referans planı ile retinal yüzey arasındaki mesafedir [2].

3.8. Yaş

Yaşla GİB bariz olarak artış gösterir. 30–39 yaş arasındaki popülasyonda 23 mmHg ve üzerindeki GİB oranı % 1,2-5 iken 70-79 yaş arasında %10.9'a çıkmaktadır [2].

3.9. Uygulama

168 adet hasta bilgisinden oluşan veri kümesi, 84 eğitim ve 84 de test olmak üzere ikiye ayrılmıştır. Ayrılan bu kümelerden eğitim kümesi, literatürde en çok kullanılan Johnson's indirgeme algoritmasıyla indirgenmiştir. İndirgenmiş veri tablosu Şekil 4'te gösterilmektedir.



	Reduct	Support	Length
1	{skk}	100	1
2	{disk area}	100	1
3	{Yaş}	100	1
4	{gib}	100	1
5	{rim area}	100	1

Şekil 4. İndirgenmiş veri tablosu

8 nitelikten oluşan veri kümesi, 5 niteliğe indirgenmiştir. İndirgenen bu veri kümesi ise Rapidminer aracılığıyla sınıflandırıp performanslar ölçülmüştür. Tablo 2 ve Tablo 3'te sırasıyla Karar Ağaçları ve YSA kullanılarak elde edilen bu ölçüm ve değerlendirme sonuçlarına yer verilmektedir.

Tablo 2'de görüldüğü gibi performans ölçümü için çapraz doğrulama metodu kullanılmıştır. Çapraz doğrulama için 2'den 10'a kadar kümeler kullanılmıştır. Karar Ağaçları için kullanılan 3 algoritmadan en yüksek başarı oranı C4.5 algoritmasının 6'lı kümesine ait olup %93.45 olarak elde edilmiştir. Bu değer, 8 nitelik kullanılarak gerçekleştirilen Haciefendioğlu'nun tez çalışmasında [2] ID3 algoritmasında 7'li küme için %94.05 olarak elde edilmiştir. Çalışmamızda nitelikler 8'den 5'e indirgeniği için %93.45'lik bir oran oldukça iyi bir başarı oranı olarak görülebilir.

Tablo 2. Karar Ağaçları kullanılarak elde edilen ölçüm sonuçları

X- Validation Doğrulama Sayısı	Karar Ağaçlarında Kullanılan Fonksiyon	%Performans
2	gain_ratio(C4.5)	89.29% +/- 2.38%
	information_gain(ID3)	89.88% +/- 2.98%
	gini_index(CART)	88.10% +/- 4.76%
3	gain_ratio(C4.5)	91.67% +/- 4.45%
	information_gain(IDE3)	91.67% +/- 4.45%
	gini_index(CART)	89.29% +/- 2.92%
4	gain_ratio(C4.5)	89.88% +/- 4.25%
	information_gain(ID3)	90.48% +/- 2.38%
	gini_index(CART)	89.29% +/- 4.91%
5	gain_ratio(C4.5)	88.20% +/- 7.64%
	information_gain(IDE3)	87.61% +/- 8.15%
	gini_index(CART)	88.20% +/- 7.64%
6	gain_ratio(C4.5)	93.45% +/- 2.45%
	information_gain(IDE3)	92.86% +/- 2.92%
	gini_index(CART)	91.67% +/- 3.95%
7	gain_ratio(C4.5)	93.45% +/- 4.37%
	information_gain(IDE3)	91.67% +/- 4.98%
	gini_index(CART)	90.48% +/- 6.57%
8	gain_ratio(C4.5)	92.86% +/- 5.83%
	information_gain(ID3)	92.86% +/- 5.32%
	gini_index(CART)	91.07% +/- 5.55%
9	gain_ratio(C4.5)	89.77% +/- 7.18%
	information_gain(ID3)	89.86% +/- 5.82%
	gini_index(CART)	87.43% +/- 7.32%
10	gain_ratio(C4.5)	92.79% +/- 5.27%
	information_gain(IDE3)	90.40% +/- 6.19%
	gini_index(CART)	90.99% +/- 6.77%

Yapay Sinir Ağları ile oluşturduğumuz ağda ileri beslemeli sinir ağı (A feed-forward neural network) ve geriye yayılım (back propagation) algoritması kullanılmıştır.

Karar Ağaçlarında olduğu gibi YSA'da da performans ölçümü için çapraz doğrulama metodu kullanılmıştır. YSA yöntemine ait parametreler için hata (epsilon) 1.0E-5 olarak, momentum da 0,2 olarak sabit tutulmuştur. Rapidminer öğrenme hızını (learning rate) 0 ile 1 arasında değiştirilerek en iyi sonucu 0.3 olarak bulmuştur. Çapraz doğrulama sayısı değiştirilerek elde edilen performans tablosu da aşağıda verilmiştir. Buna göre elde edilen en iyi performans Tablo 3'de koyu renkle gösterilen 4 doğrulama sayısı ile 91.67% +/- 2.66% olmuştur.

Tablo 3. YSA kullanılarak elde edilen ölçüm sonuçları

X-Validation Doğrulama Sayısı	%Performans
2	87.50% +/- 1.79%
3	88.10% +/- 5.52%
4	91.67% +/- 2.66%
5	89.32% +/- 2.91%
6	90.48% +/- 6.41%
7	88.69% +/- 5.77%
8	89.29% +/- 5.71%
9	89.80% +/- 6.12%
10	91.65% +/- 3.95%

4. Sonuç ve Tartışma

Bu çalışmada, geç kalındığında körlüğe neden olabilecek glokom hastalığının teşhisi için sınıflandırma tekniklerinden Karar Ağaçları (C4.5, ID3, CART) ve YSA algoritmaları kullanılmıştır. Çalışmada kullanılan veri kümesi Hacıfendioğlu'nun tez çalışmasından [2] alınmıştır. Ancak bu çalışmadaki fark, veri kümesindeki nitelik sayısının 8'den 5'e indirgenmiş olmasıdır. İlgili sınıflandırma teknikleri ise bu indirgenmiş veri kümesi üzerinde uygulanmıştır. Değerlendirme amacıyla çapraz doğrulama metodu kullanılmış ve 2'den 10'a kadar kümeler üzerinde ölçümler gerçekleştirilmiştir. Gerçekleştirilen ölçümlere dair genel sonuçlar Tablo 4'de gösterilmektedir.

Tablo 4. Genel sonuçlar

ALGORİTMA	PERFORMANS (%)
C4.5	93.45% +/- 2.45%
ID3	92.86% +/- 5.32%
CART	91.67% +/- 3.95%
YSA	91.67% +/- 2.66%

Tablo 4'deki veriler Karar Ağaçları için 6'lı küme üzerinde gerçekleştirilen çapraz doğrulama sonuçlarıdır. YSA için ise 4 doğrulama sayısı ile elde edilen doğrulama sonucudur. Buna göre en düşük başarı oranı CART algoritmasına ait olup, 91.67% +/- 3.95% değerindedir. En yüksek başarı oranı ise C4.5 algoritmasına ait olup %93.45 değerindedir. Hacıfendioğlu'nun tez çalışmasında [2] ise Karar Ağaçları kullanılarak elde edilen en yüksek başarı oranı %94.05, YSA için ise bu oran %96.43 değerindedir. Gerçekleştirilen çalışmada ulaşılan başarı oranlarının bu değerlere yakın olduğu görülmektedir. Ayrıca bu çalışmada, nitelik sayısı yaklaşık yarısına, 8'den 5'e düşürüldüğü için %90'ların üzerinde elde edilen bu oranlar hastalığın teşhisinde kabul edilebilir değerler olarak yorumlanabilmektedir.

Kaynaklar

- [1] Bhowmik, D., Sampath Kumar, K.P., Deb L., Paswan S., Dutta, A.S., (2012). Glaucoma -A Eye Disorder Its Causes, Risk Factor, Prevention and Medication, *The Pharma Innovation*, 1(1), 66-81
- [2] Hacıfendioğlu Ş., (2012). Makine Öğrenmesi Yöntemleri ile Glokom Hastalığının Teşhisi, *Yüksek Lisans Tezi*, Selçuk Üniversitesi
- [3] Kumarasamy, N.A., Lam, F.S., Wang, A.L., Theoharides, T.C., (2006). Glaucoma: Current And Developing Concepts For Inflammation, Pathogenesis And Treatment, *European Journal Of Inflammation*, 4(3), 129-137
- [4] Bock, R., Meirer, J., Michelson, G., NY'U, L. G., Hornegger, J. (2007), Classifying glaucoma with image-based features from fundus photographs, *University of Erlangen-Nuremberg, Erlangen*, 355-364.
- [5] Bizios, D., Heijl, A., Hougaard, J.L., Bengtsson, B., (2010), Machine learning classifiers for glaucoma diagnosis based on classification of retinal nerve fibre layer thickness parameters measured by Stratus OCT, *Acta Ophthalmologica*, 88(1), 44-52.
- [6] Belghith, B., Bowd, C., Balasubramanian, M., Weinreb, R.N., Zangwill, L.M., (2013), A Bayesian framework for glaucoma progression detection using Heidelberg Retina Tomograph images, *(IJACSA) International Journal of Advanced Computer Science and Applications*, 4(9), 223-229.

- [7] Kaya, Y., Yeşilova, A., Tekin, R., (2011). E-Posta Kullanım Alışkanlığının Kaba Kümeler ile İncelenmesi, *6th International Advanced Technologies Symposium (IATS'11)*, 62-65.
- [8] Çalış, A., Kayapınar, S., Çetinyokuş, T., (2014). Veri Madenciliğinde Karar Ağacı Algoritmaları ile Bilgisayar ve İnternet Güvenliği Üzerine Bir Uygulama, *Endüstri Mühendisliği Dergisi*, 25(3-4), 2-19.
- [9] Onan, A., (2015). Şirket İflaslarının Tahmin Edilmesinde Karar Ağacı Algoritmalarının Karşılaştırmalı Başarım Analizi, *Bilişim Teknolojileri Dergisi*, 8(1), 9-19.