

T.C.  
ERCIYES ÜNİVERSİTESİ  
BİLİMSEL ARAŞTIRMA PROJELERİ  
KOORDİNASYON BİRİMİ

**KORONER ARTER HASTALIĞINDA KARDİYAK OLAYLARIN  
ÖNGÖRÜLMESİNDE ROLÜ OLAN YENİ BİYOKİMYASAL PARAMETRELERİN  
YAPAY SİNİR AĞLARI İLE BELİRLENMESİ**

**Proje No: TSA-11-3373**

Proje Türü:  
NAP

**SONUÇ RAPORU**

**Proje Yürütücüsü:**

Doç. Dr. Nihat KALAY  
Kardiyoloji Ana Bilim Dalı

Doç. Dr. M. Tuğrul İnanç  
Kardiyoloji Ana Bilim Dalı

Arş. Gör. Gülhan TOĞA  
Endüstri Mühendisliği Bölümü

2014

KAYSERİ



## İçindekiler

ÖZET .....	4
ABSTRACT .....	5
TEŞEKKÜR .....	6
1. GİRİŞ .....	8
2. GEREÇ ve YÖNTEM.....	11
2.1. YSA.....	12
2.2. KARAR AĞAÇLARI .....	14
2.3. LOJİSTİK REGRESYON.....	15
3. BULGULAR .....	16
3.3. LOJİSTİK REGRESYON.....	18
5. KAYNAKLAR .....	21

## ÖZET

Koroner Arter Hastalarının daralan damarlarına Perkutan Koroner Girişim olarak ifade edilen stent, balon ve anjiyoplasti gibi çeşitli yollarla müdahale edilmektedir. Bu operasyonlar sonrasında damar tıkanıklıkları giderilen hastaların hastane içi ve taburcu sonrası sağlıklı ömürleri farklılık göstermektedir. Bu çalışmada, koroner arter hastalığının ortaya çıkmasında ve ilerlemesinde etkisi olduğu düşünülen

- Brain natriüretik peptit (BNP)
- Cystatin C
- Co peptin
- Soluble CD40 ligand
- Faty asit bağlayıcı protein
- Lipoprotein associated phospholipase
- Ischemia-modified albumin
- Hyoloronik Asit biyokimyasallar ile rutin biyokimyasalların hastaların PKG sonrası 1 yıllık süreçte gözlemlenen ölüm-sağ kalma durumları için Yapay Sinir Ağları, Karar Ağaçları ve Lojistik Regresyon ile modelleme yapılmıştır.

**Anahtar Kelimeler:** Yapay Sinir Ağları; Stent; Karar Ağacı; Lojistik Regresyon

## **ABSTRACT**

Percutaneous Coronary Intervention such as stent implantation, balloon and angiography are applied to narrowed arteries in coronary artery patients. In hospital mortality and long term mortality alter in patients who undergo PCI. In this study,

- Brain natriuretic peptit (BNP)
- Cystatin C
- Co peptin
- Soluble CD40 ligand
- Faty asid associated protein
- Lipoprotein associated phospholipase
- Ischemia-modified albümin and Hyoloronik Asit biomarkers were used to analyze long-term mortality after stent implementation with Artificial Neural Networks, Decision Trees and Logistic Regression.

**Key words:** Artificial Neural Networks, Stent, Decision Trees, Logistic Regression

## **TEŐEKKÖR**

BAP birimine projemize olan desteklerinden dolayı teőekkÖr ederiz.



## 1. GİRİŞ

Proje kapsamında hastaların işlem sonrası 1 yıllık süreçleri takip edilmiştir. Hastaların sağ kalma- ölme durumları üzerinde etkisi olduğu düşünülen parametreler girdi olarak kullanılarak, çıktı değişkeni (ölü-sağ) üzerindeki etkisi farklı teknikler yardımı ile analiz edilmiştir. Literatürde uzun vadede hasta sağlığı üzerine çeşitli metotların kullanıldığı birçok çalışma bulunmaktadır.

Garg *et al.* PKG olmuş ST segmentli MI hastalarının 1 yıllık klinik bulgularını Syntax skoru ve PAMI skoruna göre kıyaslamışlardır. Sonuç olarak kardiyak olayların tahminlemede SYNTAX skorunun daha başarılı olduğu belirtilmiştir [1]. Holmes *et al.* 4 yıllık süreçte PKG hastalarından sadece stent takılan 3761 kişilik hasta grubunda hastane içi ölümleri, hastane sonrası uzun vadede ölümleri ve hastanın by pass yapılma durumunu tahminleyen bir regresyon yöntemi kullanmışlardır. 1.2 yıl  $\pm$  1 aylık takibi yapılan hastaların 6 farklı girdiye (diyabet, şok, önceki kardiyak ameliyatları, renal yetmezlik, New York Scorü ve üfürme oranı) göre uzun dönemde ölüm ihtimalleri modellenmiştir. Çalışma PKG hastalarının sadece stent grubunu ele alma yönünden farklılık göstermiştir [2]. Halkin *et al.* lojistik regresyon ve jakknife yöntemi ile Cadillac skorunu kullanarak PKG sonrası 30 gün ve sonraki bir yıl olmak üzere hastane içi ve uzun dönemde ölüm ihtimallerini modellemiştir. Girdi olarak yaş, Killip sınıfı, anemi ve renal yetmezlik, üçlü damar hastalıkları, TIMI akış derecesini kullanarak öngörülen bu yöntem ile ölümlerin tahminlemesini yapmışlardır. Ayrıca çalışmada farklı risk skorlarının (Cadillac, Pami, Zwolle ve TIMI) birbiri ile kıyaslaması yapılmış ve Cadillac skorunun daha başarılı olduğu vurgulanmıştır [3]. Borges *et al.* Anjiyo olan, ilaç kullanan ve ameliyat olmuş 611 hastayı ele almış ve bu hastaların 5 yıllık klinik bulgularını takip ederek diyabetli hastaların ölüm oranlarının diğerlerine göre daha yüksek olduğunu belirtmişlerdir [4]. Peterson *et al.* PKG olan 588 398 hasta için operasyon öncesinde, hastane içi ve 30 günlük ölüm tahminlemesi yapabilmek için lojistik regresyon modeli oluşturmuşlardır. Çalışmada sadece 30 günlük bir mortalite tahmininin yetersiz olduğu, zaten hastane içi ve 30 günlük tahminlerin yüksek korelasyon içerdiği belirtilmiştir. Çalışma içerisinde, yönetsel açıdan bakıldığında yapılan işlemin başarısına göre ödeme sistemi geliştirilmesi için bu gibi risk skorlarının ve modellemelerin geliştirilmesinin önemi vurgulanmıştır [5]. PKG müdahalesi yapılan koroner arter hastalarının ömürleri üzerinde hastanın fizyolojik özellikleri, biyotanımlayıcı değerleri, metabolik durumu, aile öyküsü ve klinik geçmişi gibi birçok parametrenin etkisi olmaktadır [6]. Koroner arter hastalığının seyrini etkileyen parametrelerin ve ayrıca operasyon esnasındaki koşulların hastanın hastane içi ölüm ve daha sonraki süreç



içerisinde ölme durumu üzerinde önemli etkisi vardır. Bu sebeple koroner arter hastalığına etki eden biyotanımlayıcıların yanı sıra operasyon ile ilgili olan parametreler de literatürde taranmış ve modelleme girdisi olarak kullanılmasına uzmanların da görüşleri doğrultusunda karar verilmiştir. 1970'lerden beri koroner arter hastalığına neden olan birçok parametre ortaya atılmış ve bunlarla sonuca ulaşılmaya çalışılmıştır. Yapılan çalışmalara göre koroner arter hastalığını tetikleyen birçok faktör bulunmaktadır. Vasküler Risk Çalışmasında hipertansiyon bulunan hastalarla birlikte diyabet ve/veya koroner arter hastalığı bulunan hastalar araştırılmıştır. Vasküler Risk Çalışması grubunda yer alan hipertansiyonlu 4506 hastadan 2664'ünün takipleri sonucu diyabet ve koroner arter hastalığından herhangi biri daha bulunan hastalarda ölüm oranlarında anlamlı artış görülmüştür. Hem diyabet hem de koroner arter hastalığı olan hastalarda ölüm oranı % 18.4'e çıkmaktadır [7]. Türk kardioloji derneğinin yayınladığı "Ulusal Kılavuz"da; erkeklerin 45 yaşından ve kadınların 55 yaşından büyük olmasının, kadınların erken menopoza girmesinin, ailede birinci derece akrabalarda infarktüs olmasının, sigara içmenin, yüksek tansiyon varlığının ve kolesterol yüksekliğinin KAH oluşumunu hızlandırdığı belirtilmiştir [8]. Mobley *et al.* özellikle erkeklerde önemli sayılabilecek damar tıkanıklığının modellenmesi için: Yaş, Boy, Kilo, Diyabet, Hipertansiyon, Stres Testi Sonuçları ve Sürekli Ağrı verilerini girdi parametreleri olarak kullanılmıştır. Sözel olan Var-Yok veya Evet-Hayır gibi dilsel değişkenlerin modellenmesi için 0-1 şeklinde kodlamaları yapılmıştır [9]. Kurt *vd.* lojistik regresyon, sınıflandırma, regresyon ağacı ve yapay sinir ağları tekniklerinin koroner arter hastalığını tahminlemedeki başarısı kıyaslanmıştır. Girdi parametreleri: Yaş, Cinsiyet, Aile geçmişi, Sigara kullanma, Diyabet, Hipertansiyon, Kolesterol, Vücut kütle indeksi'dir. 1245 veri kullanılmış olup, örneklem hasta ve sağlıklı bireyleri ayıran kontrol ve deney grubundan oluşmaktadır. 865 veri koroner arter hastalığı olan kişilerdir, diğer 380 veri ise tamamen sağlıklı bireylerden oluşmaktadır. Koroner arter hastalığı olan deney grubu, birçok çalışmada belirtildiği gibi tıkanıklığın herhangi bir damar için %50'den büyük olduğu hastalardan oluşmaktadır. Bu durumun tespitinde koroner anjiyografi, stres elektrokardiyografi, sintigrafi veya patolojik dinlenme elektrokardiyografi testleri kullanılmıştır. Kullanılan yapay sinir ağları mimarisi Çok Katmanlı (MLP) YSA'dır. Her bir metot için hassaslık, pozitif tahmin edebilme, negatif tahmin edebilme ve doğru sınıflandırma yüzdeleri bulunmuş ve böylece hangi metodun daha verimli olduğuna karar verilmeye çalışılmıştır. Yapılan kıyaslama sonucunda MLP'nin en iyi sonuç verdiği gözlenmiştir [10]. Çolak *vd.* girdi olarak kullandığı değişkenler: Yaş, Cinsiyet, Şeker Hastalığı, Hipertansiyon, Aile Öyküsü, Sigara Kullanımı, Obezite, Stres, Fiziksel

Aktivite Yapıp Yapmama, Trigliserid, LDL, HDL, Ürik Asit, Beyaz Küre, Hemoglobin, Direk Bilirubin'dir. Bağımlı değişken Koroner arter hastalığının varlığı olarak verilmiştir. Çalışmada, anjin ile hastaneye başvuran 273 bireyden 124'ü deney grubuna alınmış, kalan 113 birey ise kontrol grubuna dahil edilmiştir. Olabilirlik oran istatistiği kullanılarak yapılan karşılaştırma sonucunda KAH tahminlemede lojistik regresyon modeli için Adımsal yöntemlerin kullanımının Enter yönteminden daha iyi olduğu sonucuna varılmıştır [11]. Morrow *et al.* hs-CRP, Hemoglobin A1C, Brain Natriüretik Peptit, Mikroalbuminerya ve Troponin gibi biyotanımlayıcıların KAH üzerinde önemli etkisi olduğunu ve hastalığın teşhisinde gelecekte kullanılacak en önemli parametrelerin bunlar olacağını vurgulamıştır [12]. M. Karaolis *et al.* yaptığı veri madenciliği çalışması ile MI , PKG ve bypass ameliyatı için en önemli parametrelerin neler olduğunu araştırmış ve PKG öncesi hastanın durumu için aile öyküsü, yaş, hipertansiyon, diyabet verilerinin en önemli belirleyici olduğunu vurgulamıştır. Operasyon sonrası riskin arttığı hastaların en önemli ortak özelliklerinin ise diastolik kan basıncı, LDL ve operasyon sonrası sigara kullanımı olduğunu vurgulamıştır [13]. Kincl *et al.* 1254 kişi üzerinde yaptıkları çalışmada lojistik regresyon kullanarak koroner arter hastalığını etkileyen laboratuvar tanımlayıcılarını lökosit, CRP, fibrinojen, ürik asit, kreatinin, HDL ve LDL kolesterol, trigliserid ve glikoz olarak tespit etmiştir. Çalışmada hastanın fizyolojik özellikleri ile ilgili cinsiyet, yaş, boy, kilo, vücut kitle indeksi % EF (üfürme) oranı belirleyici özellikler olarak verilmiştir [14]. Ramasamy, kardiyak troponin, fatty asit, iskemi modifiye albumin, sol damar performansı göstergesi olan brain natriüretik peptit ve c-reaktif proteinin koroner arter hastalarının risk durumunu belirlemedeki önemini göstermiştir. Bu ve benzeri biyotanımlayıcıların hastaların risk durumunu tespit için sıklıkla kullanıldığını ancak hala risk tahminlemede kullanılması gereken biyotanımlayıcılar konusunda bazı çelişkilerin olduğunu da vurgulamıştır. Lactate dehydrogenase (LDH) and creatine kinase (CK) ile miyoglobinin de uzun zamandır risk modellemede başarılı şekilde kullanıldığını ifade etmiştir [9]. Grivaa *et al.* bazı biyotanımlayıcıların seviyeleri ve koroner arter hastalığının ilerlemesi arasındaki ilişkiyi ortaya koyan çalışmalarında 128 koroner arter hastasından aldıkları kanları Elisa testine tabi tutarak MCP-1(Macrophage chemoattractant protein-1), MMP-3 (matrix-metalloproteinase-3), soluble CD 40 Ligand ve sTNFR2 (soluble tumour necrosis factor receptor-2) biyotanımlayıcılarının önemini araştırmışlardır. Her hastanın Syntax skorunu tespit ederek bu parametreler ile Syntax skoru arasındaki ilişkiyi incelemişlerdir. Sadece KAH'ın ilerlemesi ile MMP-3 arasında bir ilişki tespit ederken diğer biyotanımlayıcılarda herhangi bir ilişki ortaya koyamamışlardır. Çalışma için hastaların demografik yapısında yaş

ve cinsiyet: risk faktörleri olarak kilo, BMI, bel çevresi, eski sigara içici, aktif sigara içici, metabolik sendromu, hipertansiyonu: var olan karakteristikleri sistolik ve diastolic kan basıncı, sol damar üfürümü, Syntax skoru: biyokimyasal değişkenlerinde ise LDL, HDL, toplam kolesterol, trigliserid, ApoB/ApoA1, fasting glisemia, HbA1c, Creaktif protein, MMP-3, MCP-1, sCD40L, sTNFR2 verilerini kullanmışlardır [15].

## 2. GEREÇ ve YÖNTEM

Yılmaz-Mehmet ÖZTAŞKIN Kalp Hastanesinde Perkutan Koroner Girişimde bulunulan ve sadece stent takılarak müdahale edilen 400 hasta numunesi toplanmıştır. Literatürde koroner arter hastalığının seyrinde önemli etkisi olduğu tespit edilmiş biyolojik tanımlayıcılar ile bazı klinik bulgular her hasta için kaydedilmiştir.

Stent takılan hastaların Acil Servis'te gerekli miktarda alınan kanları, pıhtılaşmaya zaman tanımadan serum elde edilmesini kolaylaştıran tüplere konulmuş ve daha sonra Erciyes Üniversitesi Merkez Laboratuvarında incelenmesi istenilen biyotanımlayıcılar Elisa testi ile kan serumundan analiz edilmiştir. Alınan kanlar sarı kapaklı 13x100'lük 5 mL BD Vacutainer plastik SST jelli tüp içerisine konulmakta ve bunlar acilen santrifüj makinesinde oda sıcaklığında dakikada 1500 devir ile 10 dakika santrifüj edilerek ve -75 derecede dondurucuya alınmaktadır. Santifüj için kullanılan tüpler içinde bulunan jel, santrifüj sonrasında serum ve kan hücreleri arasında fiziksel bir engel oluşturmakta, tüp çeperinde bulunan silika partikülleri sayesinde serum pıhtılaşmasını hızlandırmaktadır. Numuneler analizlerden önce kesinlikle çözdürülmemiş ve laboratuvar analizlerine kadar dondurucuda saklanmıştır. Modellemenin girdilerini oluşturan bağımsız değişkenler, literatür taraması ve uzman görüşü doğrultusunda belirlenmiş olup, şu şekildedir:

Demografik Özellikler:

Hastanın yaşı, Cinsiyet, Vücut Kütle İndeksi (BMI) (kg/m<sup>2</sup>), Sigara Kullanımı, HT (Hipertansiyon), DM (Diabetes mellitus- Şeker hastalığı), GGT (Gama Glutamil Transferaz): MPV (Mean Platelet Volume), CRP(C-reactive protein)(mg/l), Ürik Asit (mg/dl), LDL (Low Density Lypoprotein) (mg/dl), HDL (High Density Lypoprotein) (mg/dl), Trigliserid (mg/dl), Hemoglobin (mg/dl), WBC-Beyaz Küre (x10<sup>3</sup>/ µL), Platelet (/mm<sup>3</sup>), Rdw-cv (Red cell distribution width-coefficient variation)(fl), Rdw-sd (Red cell distribution width; SD, standart deviation) (fl), Serum glucose (mg/dl), Creatinin (mg/dl), Ck (Creatine Kinase)(ng/ml), Ck-MB (Creatine Kinase Miyocardial Band) (ng/ml), CK-MB, Kardiyak Troponin (ng/ml): cTn , Total Kolesterol (mg/dl), Direct Bilirubin (mg/dl), Total bilirubin(mg/dl), Reflow, No reflow,

LAD, RCA, CX , Ana koroner müdahalesi, Tek- Çok damar müdahalesi, Müdahale edilen damar sayısı, ve yeni biyotanımlayıcılar olan

- Brain natriüretik peptit (BNP)
- Cystatin C
- Co peptin
- Soluble CD40 ligand
- Fatty asit bağlayıcı protein
- Lipoprotein associated phospholipase
- Ischemia-modified albümin
- Hyaluronik Asit parametreleri girdi olarak kullanılmıştır.

Toplamda 43 girdi ve 1 çıktılı modeller oluşmuştur. Bu projede modellemeler YSA, Karar Ağaçları ve Lojistik Regresyon kullanılarak yapılmıştır. 400 hastaya ait numunelerin toplanması işlemi tamamlandıktan sonra ayrıştırılan serumlar biyotanımlayıcıların alımı ile birlikte analiz edilmiştir. 4'er kutu olarak alınan biyotanımlayıcıların içerisinde test numuneleri çıkınca ortalama 82'şer adet kit kullanılabilir halde bulunmaktadır. Bu nedenle 328 hasta analize dahil edilmiş bunların da değerlerinde yüksek varyans gözlemlenenler outlier analiz yapılarak elenmiştir. Toplamda 289 hasta ile çalışma yapılabilmektedir.

## **2.1. YSA**

Regresyona yönelik yapılan Yapay sinir ağları gibi modellemeler lineer olmayan girdi setleri ile çıktılar arasındaki bağıntıyı göstererek modelleme yapmaktadır. Bu gibi çalışmalar teşhis, finans, bilim, biyoloji, görüntü işleme gibi alanlarda literatürde sıklıkla kullanılmaktadır [16].

Sınıflandırma modelleri, çıktı değerleri kategorik olarak belirlendi ise her girdi setinin bir kategorik çıktı değerine atanması için kullanılmaktadırlar. Sınıflandırma tekniklerini lineer ve nonlineer olarak ikiye ayırmak mümkündür. Lineer sınıflandırma tekniklerinin başlıcaları multiple regresyon, lineer discriminant analizi ve lojistik regresyondur. Sınıflar lineer olarak ayrılabilir olduğunda bu teknikleri kullanmak oldukça anlamlı ve başarılıdır [17]. Ancak gerçek hayat problemlerinde veriler lineer olmayabilir ve sınıfların sınırları nonlineer şekilde belirlenmiş olabilir. Nonlineer sınıflandırma yapan modeller kuadratik diskriminant analizi, K-en yakın komşuluk ve YSA 'dır [18].

Sınıflandırma iki adımdan oluşur: bir eğitime setinin danışmanlı öğrenme ile bir model oluşturması ve bu model doğrultusunda her girdiyi bir sınıfa atamak. En çok kullanılan sınıflandırma yöntemlerinden bazıları; Bayes sınıflandırma, Karar ağaçları, YSA, K-nearest neighbour sınıflandırması ve genetik algoritmadır [19,20]. Veri madenciliği amacı ile kullanılan çeşitli modelleme tekniklerinin ayrıntıları ile ilgili çok sayıda yayın yapılmıştır [21,22,23].

Yapay sinir ağları, sinir sisteminin işleyişini örnek alarak nonlineer haritalama yapan modelleme tekniğidir [24]. YSA, eldeki bilgileri zaman serileri ve istatistik veri işlem yöntemlerinde olduğu gibi sıra ile değil, tüm girdilerin birbirine paralel olarak işlendiği ve birleştirilerek istenen çıktılarının yine paralel olarak elde edildiği bir yöntemdir [25]. YSA ile literatürde çok farklı problem türlerine çözüm getirilmeye çalışılmaktadır. Bilinmeyen bir girdi örüntüsünün eğitmenli öğrenme ile belirli özelliklerinden dolayı ait olduğu sınıfı bulmaya yarayan “örüntü sınıflandırma”, girdileri benzerlik ve benzemezliklerine göre eğitmensiz öğrenme ile kümelere ayıran “kümeleme”, girdi ve çıktı arasındaki gizli ilişkiyi çözerek bunların arasındaki bağlantıya göre “Fonksiyon yakınlaştırma”, bir zaman serisini dikkate alarak belirli bir senaryo doğrultusunda başka bir zamana ait “tahminleme”, belirli kısıtlar altında çalışan bir amaç fonksiyonuna ait en küçük ya da en büyük noktayı bulmak olarak bilinen “optimizasyon” problemlerinde Yapay Sinir Ağları kullanılabilir [26]. Yapay sinir hücreleri sinir sistemini taklit ederek modellenmiştir. Yapay nöronlar kurdukları bağlarla yapay sinir ağlarını oluşturur. Yapay nöronlar giriş sinyallerini alırlar, bu sinyalleri işler ve toplarlar; daha sonra ise bu sinyallere uygun çıktı değerlerini üretirler. Bir yapay sinir hücresinin çıktı değerini oluşturabilmesi için beş bölüme ihtiyacı vardır. Girdiler: Dış kaynaklıdır ve yorumlanmak üzere ağa gelen sinyalleri içerir. Bunu, gerçek sinir sisteminde duyu organlarından alınan sinyallere veya diğer sinir hücresinden gelen sinyallere benzetebiliriz. Bu tabakada ağa girdi olarak gelen veriler bulunur. Giriş birimlerinde veri işleme yapılmaz, yalnızca bu tabakada bulunan veri saklı tabaka birimlerine aktarılır. Ağırlıklar: İlgili sinyalin gücünü belirler. Çıktı üzerinde algılanan girdilerin ne kadar etkisi olması gerektiğini ayarlamak için kullanılır. Bu ağırlıkların değerleri pozitif, negatif veya sıfır olabilir. Ağırlıklar nöronlar arası bağlantıyı sağlayan dentritler gibidir. Birleştirme fonksiyonu: Duyu organlarından gelen sinyallerin bir araya toparlanmasını sağlar. Taşıyıcı sinirlerden geçen sinyallerin birleştirilmesi ile eşdeğerdedir. NET fonksiyonu birleştirilmiş sinyalleri, X girdileri ve W ağırlıkları ifade ederse; birleştirme fonksiyonu şu şekilde yazılabilir:

$$\text{NET} = \sum_{i=1}^N X_i W_i + \text{bias}$$

Aktivasyon fonksiyonu: Merkezi sinir sistemine gelen bu birleştirilmiş sinyallerin algılanıp, yorumlanması ile eş değerdedir. Genellikle doğrusal olmayan bir fonksiyon seçilerek yapay sinir ağlarına “doğrusal olmama” özelliği verilir [26]. Çok tabakalı YSA’larda neronlarda bilgi aktarımını sağlamak için kullanılan farklı işlemciler vardır. Aktivasyon fonksiyonu tüm sinyallerin ağırlık toplamını düğümlere iletmek ve böylece ateşleme yoğunluğunu hesaplamak için kullanılır.

## 2.2. KARAR AĞAÇLARI

Karar ağaçları veri madenciliği algoritmalarında var olan durumdan yola çıkarak, parametrelere dayalı sınıf belirleme yöntemidir. Karar ağaçlarının oluşturulması için kullanılan en yaygın algoritmalar ID3, C4.5, C5.0, CART, QUEST ve CHAID gibi algoritmalarlardır.

Chi-squared Automatic Interaction Detector (CHAID) algoritması en eski karar ağacı oluşturma yöntemlerinden birisidir. CHAID ile diğer ağaçlardan farklı olarak ikiden fazla dallanmış düğümler oluşturabilmektedir. Ağacın oluşturulabilmesi için, hedef ve sınıflandırıcı değişkenlerin kategorik veya nümerik olması fark etmez. Sürekli değişkenler algoritma bünyesinde kategorize edilmektedir. CHAID, Chi- Square testi vasıtası ile farklılık gözlemlenen kategorilere göre sınıflandırma yapar ve nümerik değişkenler için F testi kullanır. Denetimli öğrenme modellerinden olan CHAID, ilk defa Kass tarafından ortaya atılmıştır. Exhaustive CHAID algoritması ise CHAID temelli olup, her bir dal için tüm mümkün bölünmeleri araştırır, bu sebeple hesaplaması biraz daha uzun sürer. Bu algoritma dallandırmada çoklu sınıflardan başlayarak en az dallı ve en çok açıklayıcı modeli elde edene kadar devam eder. CHAID bazen bir değişken için optimal bölmeyi belirleyemez, bu durumda kalan tüm kategoriler istatistiksel anlamlılık verse de birleştirmeyi durdurur. Exhaustive CHAID ise bu durumda optimal iki kategori ile dallandırmaya devam edip, en son olarak çıktı ile en güçlü birlikteliği veren dalı seçer. Pearson Ki Kare testi aracılığı ile kategorik değişkenlerin sağ ve ölü gruplarda anlamlı bir farklılığının olup olmadığı kontrol edilir. Pearson katsayısını hesaplarken, rassal X ve C sayılarından oluşan, x değerli X özelliğinin ve c değerli C sınıfın olduğunu göz önünde bulundurursak aşağıdaki formül kullanılmalıdır:

$$\rho(X, C) = \frac{\sum_i (x_i - \bar{x}_i)(c_i - \bar{c}_i)}{\sqrt{\sum_i (x_i - \bar{x}_i)^2 \sum_j (c_j - \bar{c}_j)^2}}$$

Pearson Ki Kare testi iki bağılı değişkenin olasılık dağılımları arasındaki farklılığı ölçer. Eğer bir özellik gereksiz ise bu özellik ve belirlenen diğer özellik arasındaki olasılık farkı çok az olacaktır. İki rassal değişken olan X ve X' için Pearson Ki Kare yapabilmek için eğitim setinde en az 100 adet veri olmalıdır. Bir özelliğin tekrarlanma sıklığına göre Pearson Ki kare testi şu şekilde hesaplanır:

$$X^2(X, X') = \sum_{i=1}^k \frac{(F_i - F_i')^2}{F_i'}$$

Büyük değerli X2 bu özelliklerin fazlalık olmadığını göstergesidir. p değeri  $\alpha$ 'dan büyük olduğunda iki değişkenin dağılımlarının  $\alpha$  önem derecesinde eşit olduğunu bu sebeple özelliklerden birinin fazlalık olduğunu ifade eder. En uygun p değeri her bir sınıf için çapraz doğrulama yapılarak bulunmaktadır. Karar ağacı oluşturulmasında, popüler bir veri madenciliği programı olan SPPS Clementine kullanılmıştır.

### 2.3. LOJİSTİK REGRESYON

Lojistik Regresyon genellikle teşhis amaçlı çalışmalarda kullanılan çıktı değişkeni ikili değerler aldığımda başvurulan oldukça başarılı bir regresyon yöntemidir. Çıktı değişkeni 0-1 şeklinde kodlanarak regresyon denklemi oluşturulur.

Genellikle uygulanması gereken adımlar şunlardır:

- Veriler eğitilir.
- Logit denklem oluşturulur.
- Tüm vakalara ait logit değerler bulunur.
- Her vakanın hangi sınıfa ait olacağını belirleyen olasılık değerleri hesaplanır.

$$p = \frac{1}{1 + e^{-\text{logit}}}$$

. Bu olasılık vakanın o sınıfa ait olma olasılığıdır.

- 4. Adımda bulunan değerler 0.5 kesme değeri ile kıyaslanır.
- Eğer  $p > 0.5$  ise, 1 nolu sınıfa atama yapılır; eğer  $p < 0.5$  ise, 0 nolu sınıfa atama yapılır.

### 3. BULGULAR

#### 3.1. YSA

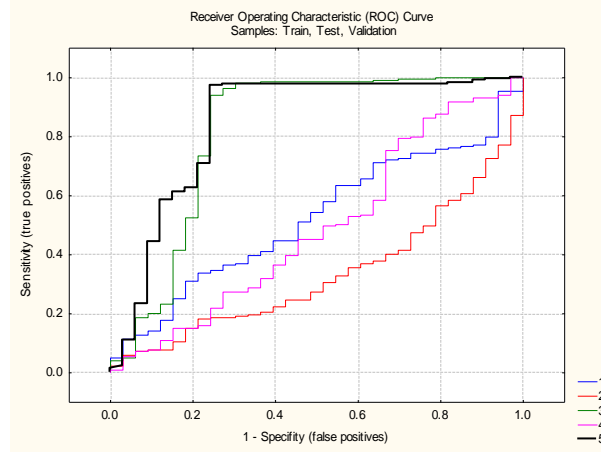
Uzun vadede yapılan takip sonrası stent takılan hastaların 33' ünün öldüğü bilgisine ulaşılmıştır. 43 girdili ve 1 çıktılı YSA modellemesi için deneme yanılma yolu ile optimum ağ oluşturulmuştur. Bu ağın oluşturulmasında kısa vade ölüm modellemesinde yapılan model ayarları değiştirilmeden kullanılmıştır. Sadece veri setindeki azalma nedeni ile bölümlendirme eğitime, test ve doğrulama aşamaları için; %70, %30 olarak yapılmıştır. Kategorik değişkenlerin her bir kategorisi girdi olarak alındığı için 80 girdili ve 2 çıktılı ağlar oluşturulmuştur. Seçilen en başarılı 5 ağ Tablo 1'de verilmiştir.

Tablo 1. YSA sonuçları

Ağ Adı	Eğitime Test Doğrulama Ara tabaka			
	Perf.	Perf.	Perf.	Akt.
MLP 80-19-2	87	81	91.9	Logistic
MLP 80-20-2	87	81	91.9	Logistic
MLP 80-8-2	100	72.9	86.9	Tanh
MLP 80-16-2	87	81	91.9	Exponential
MLP 80-23-2	100	75.6	86.9	Identity

En yüksek doğru sınıflandırma başarısı toplamda % 94.5 doğru sınıfa atama ile 5. ağıdır. 80 girdi, 23 gizli düğümlü bir ara tabaka ve 2 çıktıdan oluşan ağ, ölümlerin % 75.8'ini ve sağ kalma durumlarının % 97.3'ünü doğru sınıfa atanmıştır. Bu ağ mimarisi oluşturulurken, gözlemlenen 33 ölümün 23'ü eğitime aşamasında kullanılmış ve tamamı doğru sınıfa atanmıştır. Test aşamasında kullanılan 10 ölüm verisinden 2'i doğru sınıfa atanmıştır. Sınıflama başarısı açısından çok da yeterli olmayan bir model oluşturulmuştur. Sağ sınıflandırmasında ise toplam 6 yanlış atama yapmıştır. ROC eğrilerinde de görüldüğü üzere en başarılı ağ olan MLP-80-23-2 ağının eğri altında kalan alanı 0.84 değeri ile en yüksektir.





Şekil1. YSA ROC eğrisi

### 3.2. KARAR AĞAÇLARI

Veriler aynen YSA'da bölümlendiği gibi 70-30 eğitim ve test olarak ayrılmıştır. Eğitime aşamasında CHAID ile bir karar ağacı oluşturulmuştur ve testte ağacın başarısı ölçülmüştür. Oluşturulan ağaç çok büyüktür. Buradan ağacın genelleme yeteneğinin zayıf olduğu anlaşılmaktadır. Yapılan sınıflandırma için risk tahmini 0.012 ve standart hata 0.007 olarak bulunmuştur. Yapılan tahmin sonucu 3 ölüm doğru sınıfa atanamamıştır.

Tablo 2. CHAID eğitim sonuçları

Eğitim Sonuçları					
Tahmin			% Doğruluk		
CHAID	Gözlemlenen	0	1	%94.5	
		0	178		5
		1	6		14

Ağacın ayırım noktası YAŞ girdisidir ve oluşan kurallardan bazıları şu şekildedir:

- Eğer Yaşı  $\leq 61$  ve CRP  $> 14$  ve Cinsiyet = Erkek ve Sigara =Kullanmıyor ve Hg  $\leq 13.8$  ise ÖLÜ
- Eğer  $61 < \text{Yaşı} \leq 75$  ve Ürik asit  $\leq 6.6$  ve Ck-mb  $\leq 66$  ve Cr  $> 1.5$  ve LDL  $> 125.6$  ise ÖLÜ

- Eğer  $61 < \text{Yaşı} \leq 75$  ve  $\text{Ürik asit} \leq 6.6$  ve  $\text{Ck-mb} > 66$  ve  $\text{LAD} = \text{Var}$  ve  $\text{Sigara} = \text{Kullanıyor}$  ve  $\text{HT} = \text{Var}$  ise ÖLÜ

Kuralların hepsine ağacın büyüklüğü nedeni ile yer verilmemiştir.

Tablo 3. CHAID test sonuçları

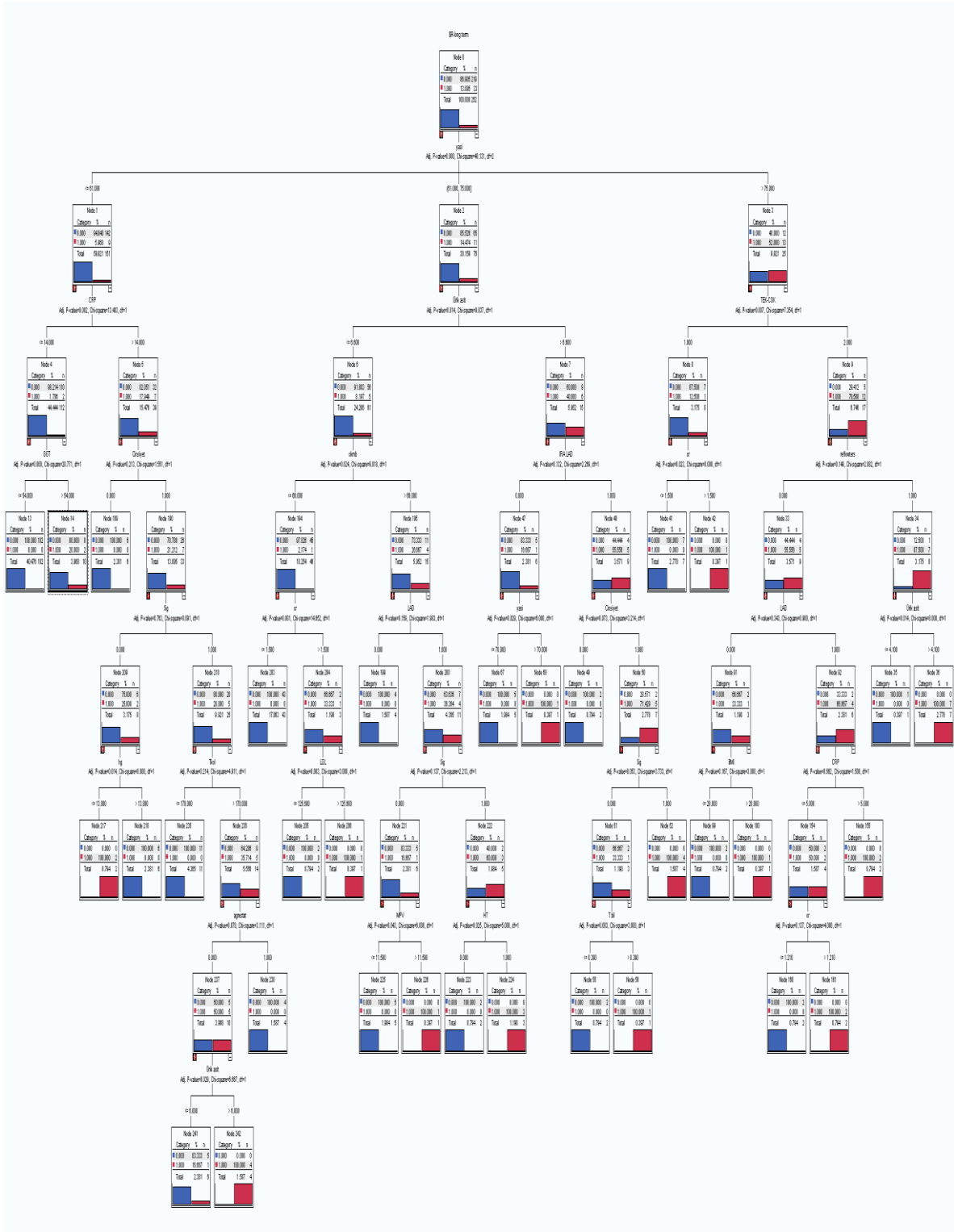
Test Sonuçları					
Tahmin				% Doğruluk	
CHAID	Gözlemlenen		0	1	%93
		0	66	0	
		1	6	14	

### 3.3. LOJİSTİK REGRESYON

SPSS paket programı yardımı ile lojistik regresyon denklemi oluşturulmuştur. Bunun için veri eleme yöntemlerinden Enter, Backward Stepwise ve Forward Stepwise teknikleri ele alınmıştır. Veri setleri diğer 2 yöntemdeki gibi parçalanarak kullanılmıştır. Aşağıda test ve eğitime aşamasına ait sonuçlar verilmektedir.

Tablo 4. Lojistik Regresyon Eğitime Sonuçları

Eğitime Sonuçları						
		Tahmin	Doğruluk %		Açıklayıcı Girdiler	
Enter methodu	Gözlenen	0	1		Tüm Girdiler	
		0	183	0		
		1	0	20		100%
Backward stepwise methodu	Gözlenen	0	182	1	Yaş, Ürik Asit, TG, BNP, Hyolonik Asit, Faty Asit Bağlayıcı Protein	
		1	6	14		96.6%
		0	178	5		91.6%
1	12	8				

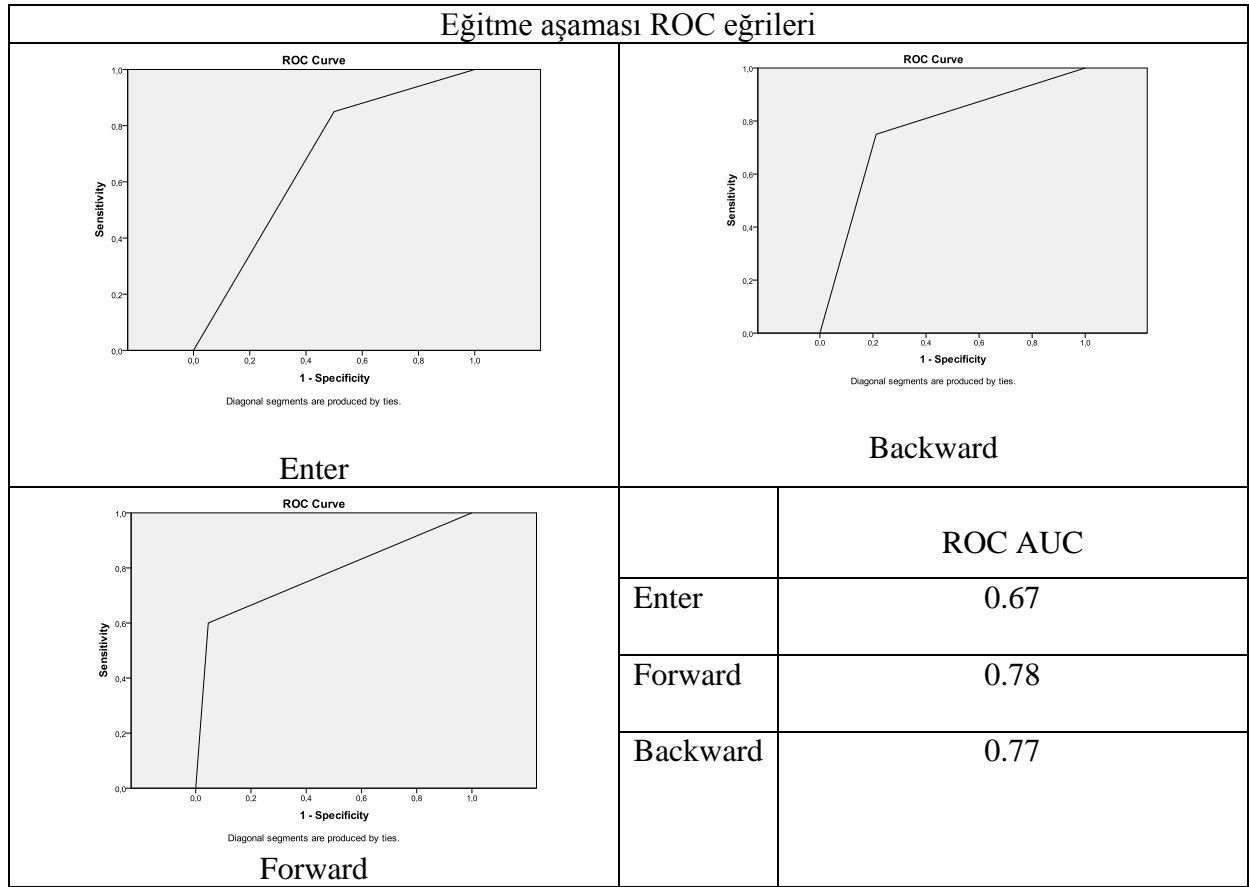


Şekil 2. CHAID karar Ağacı

Tablo 5. Lojistik Regresyon Test Sonuçları

		Eğitme Sonuçları			Açıklayıcı Girdiler
		Tahmin	Doğruluk %		
Enter methodu		0	1		Tüm Girdiler
		0	33	33	
		1	3	17	58%
Backward stepwise methodu	Gözlenen	0	53	13	Yaş, Ürik Asit, TG, BNP, Hyolonik Asit, Faty Asit Bağlayıcı Protein
		1	5	15	
Forward stepwise methodu		0	63	3	CRP, Ürik Asit, Ck-mb, BNP, Hyolonik Asit,
		1	8	12	

Bu modellere ait ROC eğrilerine bakılacak olursa Forward Stepwise metodu en başarılı olandır.



Şekil 3. Lojistik Regresyon için ROC eğrileri ve AUC

#### 4. TARTIŞMA ve SONUÇLAR

Yapılan modellemeler göstermiştir ki YSA kalp hastalıklarında ölümlerin modellenmesi üzerinde oldukça başarılıdır. Diğer iki yöntemle göre en iyi performans YSA ile elde edilmiştir. BNP ve Hyoloronik asit çalışmaya dahil edilen biyokimyasal tanımlayıcılar arasında en önemlileri olarak YSA'ya uygulanan hassasiyet analizi ile tespit edilmiştir. Çalışma birçok farklı tekniğin bir arada kullanıldığı ve yeni biyokimyasalların kalp hastalarının ömürleri üzerine olan etkisinin araştırıldığı başarılı ve orijinal bir çalışmadır.

Proje 400 hasta üzerinden hedeflenmiş olup, istenilen hedefe ulaşamamıştır. Nedeni ise sarf malzemesi olan kitlerin kutularından çıkan çalışır kitlerin beklenenden az olmasıdır. Kalan numunelerin yeni bir çalışmada değerlendirilmesi planlanmaktadır. Projenin kalan numunelerinin analizlerinin yapılması durumunda daha başarılı sonuçlar elde edileceği düşünülmektedir.

Hasta numunelerinin toplanması aşamasında beklenenden daha büyük sıkıntılar çıkmıştır. Bu sıkıntılar bir personelin ücret karşılığı tutularak numune alımına katkıda bulunması sağlanarak aşılmıştır. BAP projelerinde kısmi ücretli çalışacak personel alımına (kayıt tutma işlemleri gibi işler için) imkan verilmesi projelerin işleyişini kolaylaştırabilecektir.

#### 5. KAYNAKLAR

1. Garg, S., Sarno, G., Serruys, P. W., Rodriguez, A.E., Bolognese, L., Anselmi, M., Cesare, N.D., Colangelo, S., Moreno, R., Gambetti, S., Monti, M., Bristot, L., Bressers, M., Garcia-Garcia, H. M. Parrinello, G., Campo G., Valgimigli, M., 2011. Prediction of 1-year clinical outcomes using the syntax score in patients with acute st-segment elevation myocardial infarction undergoing primary percutaneous coronary intervention **JACC: Cardiovascular Interventions**, **4(1)**: 66-75.
2. Holmes, D.R., Berger, P.B., Garratt, K.N., Mathew, V., Bell, M. R., Barsness, G.W., Higano, S. T., Grill, D. E., Hammes, L.N., Rihal, C. S., 200. Application of the New York State PTCA mortality model in patients undergoing stent implantation. **Circulation**, **102**: 517–522.

3. Halkin, A., Singh, M., Nikolsky, E., Grines, C.L., Tchong, J.E., Garcia, E., Cox, D.A., Turco, M., Stuckey, T.D., Na, Y., Lansky, A.J., Gersh, B.J., O'Neill, W.W., Mehran, R., Stone, G.W., 2005. Prediction of mortality after primary percutaneous coronary intervention for acute myocardial infarction: the CADILLAC risk score. **J. Am. Coll. Cardiol.**, **45**:1397–1405.
4. Borges, J.C., Lopes, N., Soares, P. R., Góis, A.F.T., Stolf, N.A., Oliveira, S.A., Hueb, W.A., Ramires J.A.F., 2010. Five-year follow-up of angiographic disease progression after medicine, angioplasty, or surgery. **Journal of Cardiothoracic Surgery**, **5**: 91
5. Peterson ED, Dai D, DeLong ER, Brennan JM, Singh M, Rao SV, Shaw RE, Roe, M.T., Ho, K.K., Klein, L.W., Krone, R.J., Weintraub, W.S., Brindis, R.G., Rumsfeld, J.S., Spertus, J.A., 2010. Contemporary mortality risk prediction for percutaneous coronary intervention: results from 588,398 procedures in the National Cardiovascular Data Registry. **J. Am. Coll. Cardiol.**, **55**: 1923–1932.
6. Romic, Z., Mayer, L., Kirin, M., 2004. Biochemical markers in acute coronary syndrome. **Acta. Med. Croatica**, **58 (2)**:111–4.
7. Anonim, Vasküler risk çalışmasının sonuçları açıklandı, 2009, (Web Sayfası: <http://www.sdplatform.com/Haber.aspx?HID=2510>), (Erişim tarihi Aralık 2011).
8. Türk Kardiyoloji Derneği, Ulusal klap sağlığı eylem planı, (Web sayfası: [www.tkd-online.org/UKSP/UKSP\\_EylemPlani.pdf](http://www.tkd-online.org/UKSP/UKSP_EylemPlani.pdf)), (Erişim tarihi: Kasım 2011).
9. Mobley, B., Schechter, E., Moore, W., McKee, P. ve Eichner, J., 2005. Neural network predictions of significant coronary artery stenosis in men, **Artif. Intell. Med.**, **34 (2)**: 151–161.
10. Kurt, I., Ture, M. ve Kurum, A.T., 2008. Comparing performances of logistic regression, classification and regression tree, and neural networks for predicting coronary artery disease, **Expert Syst. Appl.**, **34** : 366–374.
11. Çolak, C., Çolak, M.C., Orman, M. N., 2007. Koroner arter hastalığının tahmininde lojistik regresyon modeli seçim yöntemlerinin karşılaştırılması, **Anadolu Kardiyol Dergisi**, **7**: 6-11.

12. Morrow, D.A., Braunwald, E., 2003. Future of biomarkers in acute coronary syndromes moving toward a multimarker strategy. **Circulation**, **108(3)**: 250-252.
13. Karaolis, M. A., Moutiris, J. A., Hadjipanayi, D., Pattichis, C. S., 2010. Assessment of the risk factors of coronary heart events based on data mining with decision trees. **IEEE Trans. Inf. Technol. Biomed.**, **14 (3)**: 559–566.
14. Kincl, V., Panovsky, R., Meluzin, J., Semenka, J., Groch, L., Tomcikova, D., Jarkovsky, J., Dusek, L., 2010. Association between laboratory markers and presence of coronary artery disease. **Biomed Pap Med Fac Univ Palacky Olomouc Czech Repub**, **154**: 227-233.
15. Grivaa, M., Naplavaa, R., Spendlikovab, M., Jarkovskyc, J., Hlinomazd, O., Cihalika, C., 2010. Potential role of selected biomarkers for predicting the presence and extent of coronary artery disease. **Biomed Pap Med Fac Univ Palacky Olomouc Czech Repub.**, **154(3)**: 219-225.
16. Khashei, M., Hamadani, A. Z., 2012. Bijari, M. A novel hybrid classification model of artificial neural networks and multiple linear regression models. **Expert Systems with Applications**, **39 (3)**: 2606-2620.
17. Satapathy, S., Murthy, J., Prasad R.P., Misra, B., Dash, P., Panda, G., 2009. Particle swarm optimized multiple regression linear model for data classification. **Applied Soft Computing**, **9**: 470–476.
18. Benoit, G., 2002. Data Mining [Chapter 6, pp 265-310]. In Cronin, B. (Ed.), Annual Review of Information Science and Technology, 36: 265-310. Silver Spring: American Society for Information Science and Technolog.
19. Dunham, M.H., 2003. Data Mining Introductory and Advanced Topics. Upper Saddle River, Pearson Education Inc.,NJ.
20. Witten, I.H., Frank, E., 2005. Data Mining: Practical Machine Learning Tools And Techniques. Morgan Kaufmann, San Francisco, USA, pp.607.
21. Hastie, T, Tibshirani R, Friedman J., 2001. The Elements of Statistical Learning. Data Mining, Inference and Prediction. Springer, pp.737.

22. Ur-Rahman, N., Harding, J.A., 2011. Textual data mining for industrial knowledge management and text classification: A business oriented approach. **Expert Systems with Applications, In press.**
23. Ngai, E.W.T., Xiu, L., Chau, D.C.K.,2009. Application of data mining techniques in customer relationship management: A literature review and classification. **Expert Systems with Applications, 36(2): 2592-2602**
24. Co, H. C., 2008. Confirmation testing of the Taguchi methods by artificial neural-networks simulation. **International Journal of Production Research, 46(17): 4671-4685.**
25. Şen, Z.,2004. Yapay Sinir Ağları İlkeleri. Su vakfı, İstanbul,
26. Basheer, I.A., Hajmeer, M.,200. Artificial neural networks: fundamentals, computing, design, and Application. **Journal of Microbiological Methods, 43: 3-31.**