

Destek Vektör Makinesi Algoritması ile Kalp Hastalıklarının Tahmini

İrem ÖZCAN¹, Beyda TAŞAR^{1*}, Ahmet Burak TATAR¹, Oğuz YAKUT¹

¹ Mekantronik Mühendisliği, Fırat Üniversitesi, Elâzığ, Türkiye (irem.ozcan093@gmail.com, btasar@firat.edu.tr, atatar@firat.edu.tr, oyakut@firat.edu.tr)

Received: Jan.6, 2019

Accepted: Mar.6, 2019

Published: Dec.1, 2019

Özetçe— Son yıllarda yapılan araştırmalar ve istatistikler ani yaşam kayıplarının yüzde 46,2'si (17,5 milyon) kalp ve damar hastalıkları nedeni yaşandığını göstermektedir. Bu nedenle erken teşhis ve müdahalenin önemli olduğu bu tür hastalıkların basit ve kolay bir takım testler ile hastanın rutin kontrolü son derece önemlidir. Bu çalışmada destek vektör makinesi ve ileri yayımlı YSA tahmin algoritmalarının kalp hastalığının erken teşhisinde kullanılabilir olduğu önerilmiş ve kullanılmıştır. 170 adet kişiden alınan veriler kişisel ve tıbbi veriler kullanılarak yapılan çalışmada İleri yayımlı YSA ile % 83.33, destek vektör makinesi algoritması ile de % 91.67'lik bir doğrulukta kalp hastalığının varlığının tespiti başarılmıştır.

Anahtar kelimeler: yapay sinir ağları, destek vektör makinesi, kalp rahatsızlığı tahmini.

Abstract— Research and statistics in recent years show that 46.2% of sudden life losses (17.5 million) are caused by cardiovascular diseases. For this reason, early diagnosis and intervention of these diseases is very important and simple and easy tests of the patient are very important. In this study, support vector machine, forward propagation ANN, classification algorithms have been proposed and used in the early diagnosis of heart disease. In the study conducted using data from 170 people, the predictive success was achieved with 83.33% with forward spread ANN and with an accuracy of 91.67% with support vector machine algorithm.

Keywords: artificial neural networks, support vector machine, heart disease prediction.

1.Giriş

Eğitim, gelir düzeyi ve bulaşıcı hastalıklara yönelik etkin kontrol yöntemleri sayesinde bu yüzyılda yaşam süresinin uzadığı görülmektedir. Bu durum beraberinde yaşlı nüfusun artışı, beslenme düzeninin bozulması ile birlikte kalp damar rahatsızlığı başta olmak üzere bulaşıcı olmayan hastalıklar (BOH) grubundaki kanser, kronik hava yolu hastalıkları vb. artışını beraberinde gerdi. Yaşam süresinin artışı pozitif bir durum olmasına karşın BOH alınan tedbirler ile sınırlandırılması gereken bir problem olarak toplum sağlık örgütlerinin önüne çıkmıştır (T.C. Sağlık Bakanlığı, 2010). En son yapılan 2012 istatistik verilerine göre dünyada meydana gelen 56 milyon ölüm vakasının 38 milyonu özellikle kalp ve damar hastalıkları, kanser, kronik hava yolu hastalıkları ile meydana gelmiştir (Global Status Report, 2014). Gelişmiş ve gelişmekte olan ülkelerin; BOH'leri ile mücadelede harcamak zorunda kaldıkları bütçe gün geçtikçe artmakta ve hasta yoğunluğu nedeni ile sağlık hizmeti sunumu zorlaşmaktadır. Düşük gelirli ülkelerde ise kişilerin bu sağlık hizmetine bakım ve tedaviye erişimi onları fakirlik sınırını altına itmektir (T.C. Sağlık Bakanlığı, 2010; Global Action Plan, 2013)

Türkiye Kalp ve Damar Hastalıkları Önleme ve Kontrol Programı Eylem Planı (2015-2020)'na göre; Türkiye'de özellikle istemik kalp hastalıkları ve serebrovasküler hastalıklar olmak üzere kalp ve damar hastalıkları tüm ölüm nedenleri arasındadır (T.C. Sağlık Bak., 2010). Kalp ve damar hastalıkları açısından olumlu olarak değerlendirilebilecek unsur ise büyük ölçüde “önlenebilir” ve kontrol edilebilir olmalarıdır. Kan basıncı, obezite, kolesterol kontrolü ile kalp ve damar hastalığı görülme sıklığının yarıya indirilebileceğini bildirmektedir. Türkiye Cumhuriyeti Sağlık Bakanlığı 2013-2017 Stratejik Plan ve Eylem Planı'nda da (T.C. Sağlık Bakanlığı,2012) kronik hastalıkların önlenmesine geniş yer verilmiş olup “Bulaşıcı olmayan hastalıkların görülme sıklığını ve risk faktörlerini azaltmak ve izlemek” hedefi yer almıştır.

Bu çalışma kapsamında BOH hastalıklarından biri olan kalp hastalığının bazı kolay ve hızlı tetkikler sonucu elde edilen tıbbi veriler ile tahmini (erken öngörümü) gerçekleştirilmiştir. Böylece hekimlerce konulacak teşhise yardımcı bilgisayar destekli tanı destek algoritması geliştirilmiştir.

2. Önceki Çalışmalar

Araştırmacılar, kalp ve damar rahatsızlıklarının tahmini ve hastalıkların sınıflandırılması alanında son yıllarda pek çok çalışma gerçekleştirilmiş ve yayınlanmıştır. Yuehjen ve diğ. (Shao, Y.E. et al, 2014), kalp hastalıklarının tahmini için çok sayıda hibrit model önermiş ve kullanmışlardır bunlardan bir kaç lojistik regresyon (LR), çok değişkenli uyarlanabilir regresyon (MARS), yapay sinir ağı (YSA) dır. Kahramanli, H. ve ark. (2008) ise diyabet ve kalp hastalıklarını öngörmek için yapay sinir ağı (YSA) ve bulanık sinir ağı (FNN) içeren bir hibrit sinir ağı önerdiler. Önerdikleri hibrit sistem diyabet ve kalp hastalıklarının tahmininde ortalama % 86.8'lik bir doğruluk oranına sahipti. Kumari M. ve ark. (2011) kardiyovasküler hastalığı tespit etmek için Yapay Sinir Ağı (YSA), Destek Vektör Makinesi (SVM), Karar Ağacı ve RIPPER sınıflandırıcı gibi birçok algoritmayı test ettiler ve bu algoritmalar ile sırası ile %80.06, % 84.12,% 79.05 ve% 81.08' öngörüm başarıları elde ettiler. Mazurowski ve ark. (2008) tıbbi tanı tayini için Standart geri yayımlı YSA (BPNN) ve parçacık sürüsü optimizasyonu (PSO) algoritmalarını kullanmışlardır. Çalışmalarında BP'nin PSO'dan daha iyi performans gösterdiğini ifade etmişlerdir. Das R., ve ark. (2009), kalp hastalığını öngörmek için sinir ağlarına dayalı bir algoritma geliştirmişlerdir ve sistemin doğruluk oranı yaklaşık % 89.01'dir. Purwar, A. Ve Singh, S.K. (2015), eksik değer imputasyonu (HPM-MI) olarak isimlendirdikleri yeni bir hibrit tahmin algoritmasını geliştirmişlerdir. K-means kümeleme prensibine dayanan bu algoritma melez bir modeldir ve algoritmanın başarıları Makine Öğrenme UCI veri bankasından alınan üç ayrı dataset üzerinde test edilmiştir. Bu veri setler; Pima Kızılderilileri Diyabet, Wisconsin Meme Kanseri ve Hepatittir. Sonuçların veri setinde eksik veri olması durumunda bile iyi öngörüm sonuçları vermektedir. Bagaj ve ark. (2015) ise veri madenciliği sınıflandırma algoritmasını kalp hastalığı tahmininde kullanmışlardır. Önerilen sistem, tıbbi hataları azaltabilir yetenekte olup ve yüksek doğrulukta tahmin sonucu vermekteydi. Beheshti ve ark. (2014), hızlandırılmış parçacık sürüsü optimizasyonu adlı yeni bir meta-heuristik yaklaşım önerdiler. Önerilen sistemin performansı, dokuz standart tıbbi veri seti üzerinden değerlendirildi ve sonuçlar ümit vaat ediyordu. Turabieh, H. ise kalp hastalığının tahmini için GWO-YSA hibrit algoritması önermiştir (Turabieh, H., 2016).

3. Yöntem ve Metot

Bu çalışmada kalp rahatsızlığının tahmini için iki ayrı algoritma önerilmiş ve sonuçları karşılaştırılmalı olarak analiz edilmiştir. Seçilen iki algoritmanın da (İleri yayımlı yapay sinir ağları ve destek vektör makinası sınıflandırma algoritması) özellikle tıbbi verilerin sınıflandırmasında yüksek başarıya sahip olduğu bilinmektedir.

3.1. İleri yayımlı YSA

İleri yayımlı YSA algoritması doğrusal olmayan modelleri herhangi bir sürekli fonksiyona veya türevlerine yakınlaştırma yeteneğine sahiptir ve tıbbi verilerin sınıflandırılmasına uygundur (Demuth H.,2014). Tasarlanan YSA iki katmanlı olup her bir katmanında sırası ile 3ve 2 nöron bulunmaktadır. Katmanın ağırlığı, üst simgesi sayısı, ağırlık matrisi ve aktivasyon fonksiyonu sırasıyla w , k , $w_{ij}(k)$ ve $g(1)$ ile ifade edilmiş olup iki katmanlı YSA' nin çıkış eşitliği Denk. (1)'de yer almaktadır (Demuth H.,2014).

$$y_m = g^{(2)} \left(\sum_{q=0}^Q \omega_{mq}^{(2)} g^{(1)} \left(\sum_{i=0}^N \omega_{qi}^{(1)} x_i \right) \right) \quad (1)$$

3.2. Destek Vektör Makinası (DVM)

Destek vektör makineleri (DVM), sınırlı sayıda öğrenme örüntüsü üzerinden iyi bir genelleme düzeyi elde etmek amacıyla Yapısal Risk Minimizasyonu (YRM) tümevarım prensibini uygulayan bir öğrenme makinesidir. DVM, karmaşık veri setlerinde, çözümlemesi zor örüntülerin tanımlanmasında kullanışlı bir öğrenme algoritmasını uygulamaktadır. Algoritma, önceden gözlenmemiş verilerin sınıflandırma kestirimi için örneklerden ayırt edebilen bir sınıflandırma öğrenmesini gerçekleştirmektedir (Pal, M. & Foody, G.M., 2010).

4. Deneysel Sonuçlar

Problem için önerilen YSA ve DVM algoritmaları, MATLAB 2017b kullanılarak modellenmiştir.

4.1. Problemin Tanımı

Bu çalışmada UCI Machine Learning Repository veri bankasına ait Statlog (Heart) Data Set veri tabanından kalp hastalığı ile ilgili bir tıbbi veri kümesi kullanılmıştır. Cleveland veri tabanı, kişileri iki sınıfa (Normal (0), kalp hastası (1)) ayırmak için kullanılmıştır. Veri kümesi her bir kayıt için 13 adet özellikten oluşmaktadır. Ve toplamda 270 kişiye ait kayıtları barındırmaktadır. Tablo 1’de her bir kayıt(kişi) için var olan ve kullanılan 13 adet veri kümesi (öznitelik) ayrıntılı olarak yer almaktadır. Tablo 2’de ise sınıflandırma kümesine yer verilmiştir.

Tablo 1. Kalp hastalığı ile ilgili bir tıbbi veri kümesi

Öznitelik	Aralığı
Yaş	Sürekli
Cinsiyet	0: erkek 1:bayan
Göğüs ağrısı tipi	1 ile 4 arası
Dinlenme durumunda kan basıncı (tansiyon)	Sürekli
Serum kolesterol (mg/dl)	Sürekli
Tokluk şekeri düzeyi > 120 mg/dl	1=doğru 0 =yanlış
Dinlenme durumunda Elektrokardiyografi düzeyi (0,1,2)	1 ile 2 arası
Maksimum kalp atış ritmi	Sürekli
Egzersiz ile oluşan göğüs ağrısı	1=evet 0:=hayır
Dinlenme durumunda ST değeri	Sürekli
Pik egzersiz durumunda ST segmentinin eğimi	1 ile 2 arası
Büyük damarların sayısı (0-3)	Sürekli
Hasar Oranı: 3 = normal; 6 = kalıcı hasar; 7 = geri çevrilebilir hasar	3,6,7

Tablo 2. Sınıflandırma kümesi

Sınıf Tipi	Nümerik Değeri
Sınıf 1 (Normal)	1
Kalp Hastası	2

4.2. Sınıflandırma Algoritmaları için Parametre Tayini

YSA ve DVM algoritmalarının başarılı sonuçlar vermesi için deneysel olarak parametreler test edilmiş ve her bir algoritma için optimal iterasyon sayısı, popülasyon sayısı parametreleri seçilmeye çalışılmıştır. Bazı ön deneylerden sonra önerilen YSA, ve DVM için kullanılan parametre ayarları Tablo 3-4’te gösterilmektedir.

Tablo 3. YSA Parametreleri

Parametre	Değeri
İterasyon Sayısı	1000
Popülasyon Sayısı	190
Eğitim süresi	8s

Tablo 4. DVM Parametreleri

Parametre	Değeri
Kernel Fonksiyonu	Lineer
Metot	SMO
Popülasyon Sayısı	190

4.3. ROC Analizi Sınıflandırma Başarı Tayini

Kullanılan sınıflandırma algoritmalarının ayırt ediciliğini belirlemek amacıyla ROC (Alıcı işlem karakteristikleri, Receiver Operating Characteristic) yöntemi kullanılmıştır.

Tablo 5. YSA ve DVM için ROC Tablosu

YSA			DVM		
A=35	B=6	41	A=38	B=3	41
C=9	D=30	39	C=6	D=33	39
59	41	80	17	73	80
S= 0.8010; ACC=0.8333			S= 0.8636, ACC=0.9167		

Tabloda;

A: Gerçekte hasta olup, tanı testi sonucuna göre de hasta olarak belirlenen olgulardır (Doğru pozitif, DP)

B: Gerçekte sağlam oldukları halde testin hatalı olarak hasta dediği olgulardır (Yanlış pozitif, YP)

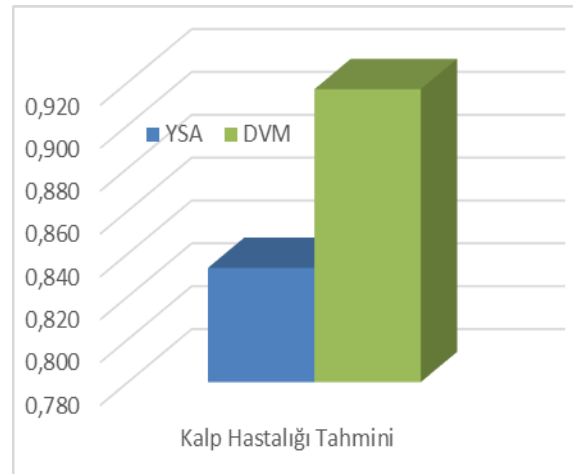
C: Gerçekte hasta olup, tanı testi sonucuna göre sağlam olarak belirlenen olgulardır (Yanlış negatif, YN)

D: Gerçek tanı sonucuna uygun olarak testinde sağlam dediği gerçek negatif olgulardır (Doğru negatif, DN)

İfade etmektedir. Hassasiyet ve Seçicilik eşitlik 2-3 ile hesaplanmıştır.

$$\text{Hassasiyet (S)} = A / (A+C) \quad (2)$$

$$\text{Doğruluk (ACC)} = A+B+C+D / (A+ D) \quad (3)$$



Şekil 1. Sınıflandırma algoritmalarının doğruluk başarısının karşılaştırma grafiği

5. Sonuçlar

Bu çalışmada, Cleveland veri setinde yer alan 170 kişiye ait, 13'er adet tıbbi veri kullanılarak kişilerin kalp hastalığına sahip olup olmadıkları öngörüsünün gerçekleştirilmesi çalışması gerçekleştirilmiştir. 170 adet kişiye ait verasetinin 90 âdeti eğitim 80 âdeti test veri seti olarak random şekilde ayrılmıştır. YSA ve DVM Algoritmalarının tahmin (ön görü) başarı yüzdesi ROC Analiz metodu ile sırası ile % 83.33 ve % 91.67 olarak hesaplanmıştır. Tasarlanan YSA sınıflandırıcısının

başarım kabiliyeti literatüre denk iken DVM algoritması kalp hastalığının tahmininde kayda değer oranda yüksek bir başarıya sahiptir. DVM algoritmasının; basit tıbbi veriler ile kalp hastalığının tanısının konulmasında hekimlere destek olabilecek; bilgisayar destekli bir karar mekanizması olarak kullanılmaya uygun olduğu görülmektedir. Gelecek çalışmalarda çok sınıflı DVM algoritması ile kalp hastalığının çeşitlerinin tahmini konusunda algoritmanın geliştirilmesi faydalı görülmektedir.

Kaynaklar

Bajaj, P., Choudhary, K. and Chauhan, R. (2015) Prediction of Occurrence of Heart Disease and Its Dependability on RCT Using Data Mining Techniques. *Advances in Intelligent Systems and Computing*, Springer India, 340, 851-858.

Beheshti, Z., Shamsuddin, S.M., Beheshti, E. and Yuhaniz, S.S. (2014) Enhancement of Artificial Neural Network Learning Using Centripetal Accelerated Particle Swarm Optimization for Medical Diseases Diagnosis. *Soft Computing*, 18, 2253-2270. <http://dx.doi.org/10.1007/s00500-013-1198-0>

Das, R., Turkoglu, I. and Sengur, A. (2009) Effective Diagnosis of Heart Disease through Neural Networks Ensembles. *Expert Systems with Applications*, 36, 7675-7680. <http://dx.doi.org/10.1016/j.eswa.2008.09.013>

Howard D., Mark B., (2014). *Neural Network Toolbox For Use with MATLAB®*, Issue-6, The MathWorks, Inc. 3 Apple Hill Drive Natick, MA 01760-2098

Global Action Plan for The Prevention and Control of NCDs 2013-2020 WHO 2013. <http://www.who.int/nmh/publications/ncd-action-plan/en/> (Global Action Plan 2013-2020) (Erişim Haziran 2014)

Global Status Report on Noncommunicable Diseases 2014, WHO, http://apps.who.int/iris/bitstream/10665/148114/1/9789241564854_eng.pdf?ua=1 (Erişim tarihi Mayıs 2015)

Kahramanli, H. and Allahverdi, N. (2008) Design of a Hybrid System for the Diabetes and Heart Diseases. *Expert Systems with Applications*, 35, 82-89. <http://dx.doi.org/10.1016/j.eswa.2007.06.004>

Kumari, M. and Godara, S. (2011) Comparative Study of Data Mining Classification Methods in Cardiovascular Disease Prediction 1. *International Journal of Computer Science and Technology*, 2, 304-308.

Mazurowski, M.A., Habas, P.A., Zurada, J.M., Lo, J.Y., Baker, J.A. and Tourassi, G.D. (2008) Training Neural Network Classifiers for Medical Decision Making: The Effects of Imbalanced Datasets on Classification Performance. *Neural Networks*, 21, 427-436. <http://dx.doi.org/10.1016/j.neunet.2007.12.031> T.C. Sağlık Bakanlığı Temel Sağlık Hizmetleri Genel Müdürlüğü, Türkiye Kalp ve Damar Hastalıklarını Önleme ve Kontrol Programı 2010-2014, Basım 2010 Ankara.

Pal, M., Foody, G.M., (2010), Feature selection for classification of hyperspectral data by SVM, *IEEE Transactions on Geoscience and Remote Sensing*, 48(5), 2297-2307.

Peker, M. , (2016) A decision support system to improve medical diagnosis using a combination of k-medoids clustering based attribute weighting and SVM, *Journal of Medical Systems*, 40: 116. doi:10.1007/s10916-016-0477-6

Purwar, A. and Singh, S.K. (2015) Hybrid Prediction Model with Missing Value Imputation for Medical Data. *Expert Systems with Applications*, 42, 5621-5631. <http://dx.doi.org/10.1016/j.eswa.2015.02.050> Shao, Y.E., Hou, C.-D. and Chiu, C.-C. (2014) Hybrid Intelligent Modeling Schemes for Heart Disease Classification. *Applied Soft Computing*, 14, 47-52. <http://dx.doi.org/10.1016/j.asoc.2013.09.020>

Turabieh, H. (2016) A Hybrid ANN-GWO Algorithm for Prediction of Heart Disease. *American Journal of Operations Research*, 6, 136-146. <http://dx.doi.org/10.4236/ajor.2016.62016>

T.C. Sağlık Bakanlığı, Stratejik Plan 2013-2014 (Basım 2012)

UCI, Machine Learning Repository, Statlog (Heart) Data Set,
[http://archive.ics.uci.edu/ml/datasets/statlog+\(heart\)](http://archive.ics.uci.edu/ml/datasets/statlog+(heart))