

Uzman Sistemlerin Tiroit Teşhisinde Kullanılması

Ramazan Solmaz, Mücahid Günay, Ahmet Alkan

Elektrik Elektronik Mühendisliği Bölümü, KSÜ, Kahramanmaraş
rsolmaz@ksu.edu.tr, gunay@ksu.edu.tr, aalkan@ksu.edu.tr

Özet: Uzman sistemlerin gelişme ve uygulanması güncel bir konu olup, özellikle medikal tanılamada birçok uygulaması vardır. Bu çalışmada bazı belirleyici kan değerleri kullanılarak uzman karar destek sistemi geliştirilmiştir. Bu amaçla ön-işlenmiş veri setine üç sınıflama ve iki kümeleme yöntemi uygulanmıştır. Uygulanan sınıflama yöntemleri destek vektör makineleri, lineer diskriminant analizi ve ileri beslemeli yapay sinir ağlarıdır. Aynı zamanda kıyaslama yapabilmek için bulanık c-ortalama ve k-ortalama kümeleme yöntemleri kullanılmıştır. Tüm yöntemlerin kabul edilebilir başarımları olmasına rağmen, uygulama sonuçlarından sınıflama yöntemlerinin (%93,48 ile %96,57 arasında) kümeleme yöntemlerinden daha yüksek başarıma (%88,83 ile %90,68 arasında) sahip olduğu görülmüştür. Yapılan çalışmada elde edilen sonuçlar, önerilen analiz sisteminin tiroit hastalığı için tanı koymada karar destek sistemi olarak kullanılabileceğini göstermiştir.

Anahtar Sözcükler: Tiroit Hastalığı, Uzman Sistemler, YSA, K-Ortalama, BCO, DVM.

Expert system for thyroid disease diagnosis

Abstract:

Development and application of expert systems are actual and have a various applications especially in medical diagnosis. In this study, an expert decision support system is developed for a thyroid diagnosis by using some decisive blood values. For this aim, three classification and two clustering techniques are applied to the pre-processed data set. The employed classification techniques are support vector machines (SVM), linear discriminant analysis (LDA) and feed-forward artificial neural networks. To have a comparison two clustering techniques, namely fuzzy c-means and k-means have also been used for the analysis. Although all methods gave acceptable accuracy rates, application results exposed that classification techniques have higher achievement rates (changing between 93.48% to 96.57%) than clustering techniques (changing between 88.83% to 90.68%). Obtained results show that the proposed analysis system can be used as an expert system for thyroid decision support system.

Keywords: Thyroid Diseases, Expert Systems, ANN, K-Means, FCM, SVM

1. Giriş

Tiroit bezi boyunda, soluk borusunun her iki yanında yer alan, kelebeği andıran bir salgı bezidir. Önden bakıldığında çene altında yer alan ve erkeklerde daha belirgin bir şekilde görülen bu bez, halk arasında “âdemelması” olarak adlandırılan kıkırdaksı çıkıntının hemen altında yer alır ve yutkunmakla hareket eder [7].

Tiroit bezi iki tane hormon üreterek dolaşım sistemine salgılar. Bu hormonlardan birine triiyodotironin (T3) ve diğerine tiroksin (T4) adı verilir. Bu hormonlar vücut metabolizmasını düzenleyerek metabolizma-nın hızını kontrol ederler. Fazla hormon salgılanırsa metabolizma hızlanır ve hipertiroidi hastalığı gelişir. Bu durumda kalp hızı artarak titreme, kas güçsüzlüğü, sinirlilik, uykusuzluk, kilo kaybı, sıcağa tahammülsüzlük, aşırı terleme, ishal, guatr, artmış kalp hızı, çarpıntı gibi şikâyetler oluşabilir. Tiroit hormonu az salgılandığında ise hipotiroidi oluşur. Bu durumda halsizlik, kas güçsüzlüğü, yorgunluk, soğuğa karşı tahammülsüzlük, kalın ‘puffy’ cilt, kabızlık, donuk duygu durumu, guatr, zayıf ve yavaşlamış kalp hızı,

hatırlama gücü gibi durumlara sebep olabilmektedir[6,7].

Dünyada yaklaşık 200 milyon insanda tiroit hastalığı bulunduğu ve bu hastalığın ülkemizde her 10 kişiden 3’ünü etkilediği bilinmektedir. Tiroit hastalıklarının bayanlarda daha sık görüldüğü de ifade edilmektedir [6, 7].

Tiroit hastalıklarının teşhisi için kan testleri, tiroit ultrasonografisi, tiroit sintigrafisi, tiroit ince iğne aspirasyon biyopsisi gibi yöntemler kullanılmaktadır.

Keleş, A., ve Keleş A., (2008 çalışmalarında [http://ftp.ics.uci.edu/pub/](http://ftp.ics.uci.edu/pub/machine-learning-databases) machine-learning-databases veri tabanından aldıkları tiroit hastalığına ait 215 veriyi kullanarak NEFCLASS-J ile doğruluk oranı %95,33 olan bir başarı elde ettiğini belirtmiştir [9].

Şenol, C., ve Yıldırım, T., “Bulanık-Sinir Ağı Yapısı İçin Yeni Bir Karma Yaklaşım” konulu çalışmasında yapay sinir ağları ile bulanık mantığın bir arada kullanıldığı karma bir yapı tasarlamıştır. Bulanık-YSA

karma yapısı YSA yapısından daha başarılı olduğunu belirtmiştir [14].

Temurtaş, F., çalışmasında tiroit verilerinin doğru yorumlanması yoluyla yapay sinir ağlarının tiroit hastalığında başarıyla kullanılabilceğini, olasılıksal YSA ile (%94,81) en iyi başarıyı bu sınıflama yöntemi elde ettiğini belirtmiştir. [16]

Şenol, C., Yıldırım T., çalışmalarında ANFIS, Fuzzy-MLP, MLP, Fuzzy-RBF, RBF, Fuzzy-CSFNN, CSFNN gibi uzman sistemlerin başarılarını karşılaştırılmış %92,93 ile en iyi başarıyı Fuzzy-CSFNN ile elde ettiklerini belirtmiştir [15].

Doğantekin, E ve arkadaşları yapmış oldukları çalışmalarında tiroit hastalıklarının teşhisi için genelleştirilmiş diskriminant analizi ve dalgacık destek vektör makinesi sistemi yöntemi kullanmışlardır. Tiroit hastalıklarının teşhisi için bu uzman sistem sınıflandırma doğruluğunun yaklaşık % 91,86 olarak elde edildiğini ifade etmişlerdir [3].

Polat, K., Şahan, S., Güneş, S., Yapay bağışıklık sistemleri yapay zekanın yeni ama etkili bir dalı olduğunu belirtmişlerdir. Tiroit hastalığının sınıflama problemini çözmek için gelişmiş bulanık ağırlıklı ön işleme ile AIRS karma yapısını kullanarak %85 sınıflama doğruluk oranına ulaştıklarını belirtmişlerdir [12].

Kodaz, H., arkadaşları çalışmalarında bilgi kazancı tabanlı yapay bağışıklık tanıma sistemi adında yeni bir yapay bağışıklık sistemi (IG-AIRS) ile %95,90 doğru sınıflama oranına ulaştıklarını belirtmişlerdir [11].

Polat ve Güneş yapmış oldukları çalışmada üç aşamadan oluşan bir sistem kullanılmıştır. PCA ile 5 özellik olan veri setini 2 özelliğe; ikinci aşamada k-NN ile ön işleme yapılmıştır. Üçüncü aşamada ise adaptif sinirsel-bulanık çıkarım sistemi (ANFIS) ile %100 doğru sınıflama başarısına ulaştıklarını belirtmişlerdir [13].

Bu çalışmada kan tahlillerinde ölçülen değerlerden tanı konulmuş veriler kullanarak uzman sistemlerle sınıflama işlemini veya direkt veriler kullanarak kümeleme işlemi yapılacaktır. Yapılacak bu işlemlerde farklı ve bu alanda uygulanmamış metotlar kullanılmıştır. Bu metotlarla teşhiste hekimlere yardımcı olabilecek başarılı sınıflama sonuçları elde edilmesi amaçlanmıştır.

2. Materyal ve Metot

2.1 Materyal

Bu çalışmada kullanılacak tiroit verileri UCI makine öğrenmesi (<ftp://ftp.ics.uci.edu/pub/machine-learning-databases>) veri tabanından alınmıştır. Veri tabanından alınan hastalığa ait veriler farklı sınıflama ve

kümeleme yöntemleri kullanılarak metotlarının benzetim sonuçları elde edilmiştir.

Veri tabanından alınan veri seti üç sınıfa ait 215 örnek içermektedir.

Sınıf 1: normal (150 adet)

Sınıf 2: hiper (35 adet)

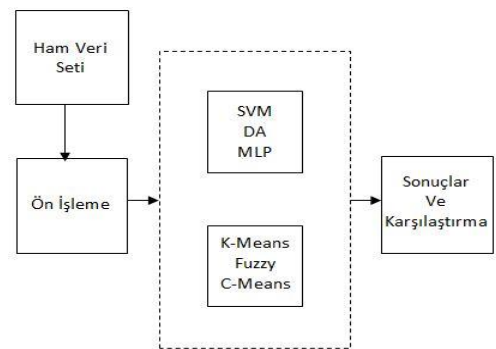
Sınıf 3: hipo (30 adet)

Her örnek tiroit hastalığına ait beş özellik taşımaktadır. Bunlar:

- Özellik 1: T3-resin uptake test (yüzdelerik değer).
- Özellik 2: İzotopik deplasman yöntemiyle ölçülen toplam serum thyroxin miktarı (T4).
- Özellik 3: Radioimmuno assay yöntemiyle ölçülen toplam serum triiodothyronine (T3).
- Özellik 4: Radioimmuno assay yöntemiyle ölçülen bazal tiroit-uyarıcı hormonu (TSH).
- Özellik 5: 200 mg thyrotropin-releasing hormon enjeksiyonundan sonra TSH değerinin bazal değerle kıyaslandığında en yüksek mutlak fark değeri.

2.2. Metot

Bu çalışmada veri setine literatürde yaygın olarak kullanılan ve iyi bilinen, ancak bu veri setine daha önce uygulanmamış sınıflama ve kümeleme teknikleri uygulanmıştır. Sınıflama algoritmaları olarak MLP (Multi layer perceptron) yapay sinir ağları, Destek Vektör Makineleri (SVM), Lineer Diskriminant Analiz yöntemleri kullanılmıştır. Ayrıca aynı veri seti eğitici öz öğrenme yöntemleri olarak bilinen kümeleme algoritmalarından FCM (fuzzy c-means) ve k-means kümeleme yöntemleri ile de incelenmiştir. Şekil.1'de kullanılan yöntemler ve akış diyagramı verilmiştir. Belirtilen yöntemler literatürde yaygın olarak bulunabileceğinden, ayrıntılı matematiksel anlatımlarına yer verilmemiş, özet bilgiler sunulmuştur.



Şekil.1. Çalışmada Kullanılan Yöntemler Ve Akış Diyagramı

2.2.1 Sınıflandırma Yöntemleri

2.2.1.1. Ön-İşleme

Veri tabanından tiroit hastalığına ait 215 adet örnek alındı. Alınan bu veriler aşağıdaki normalizasyon ifadesi kullanılarak ön işleme yapılmıştır.

$$U_N = 0.8 \times \left(\frac{U_R - U_{MIN}}{U_{MAX} - U_{MIN}} \right) + 0.1 \quad (1)$$

U_R: Normalize edilecek veriler
U_{Min}: En küçük değerdeki veri
U_{Max}: En büyük değerdeki veri

2.2.1.2. Destek Vektör Makineleri (DVM)

DVM'nin amacı lineer olarak ayrılabilen veri setleri için en uygun ayırıcı düzlemin oluşturulmasıdır. Lineer olarak ayrılamayan veri setleri öncelikle bir dönüşüm tekniği uygulanarak başka boyuta taşınır ve lineer ayrılabilir hale getirilir. Bu dönüşüm tekniği uygulandıktan sonra DVM lineer en iyi ayrımı yapmaktadır. Oluşturulan bu en uygun ayrımı yapan düzleme hiper düzlem adı verilir. Çalışılan veri seti n boyutlu ise oluşturulan hiper düzlemin boyutu n-1 olacaktır [8, 4].

2.2.1.3 Diskriminant Analizi (DA)

Diskriminant Analizi, veri setindeki değişkenlerin iki veya daha fazla gerçek gruplara ayrılmasını belirlemek amacıyla yararlanılan bir yöntemdir. DA veri setini en az iki gruba ayırabilen danışmanlı bir sınıflandırma yöntemidir. DA ile elde edilen doğrusal bileşen ile p adet özellik ayrılır. Burada doğrusal bileşen elde edilirken gruplar arası kareler toplamının maksimum, gruplar içi kareler toplamının minimum olması amaçlanır [1].

2.2.1.4 İleri Beslemeli YSA (FFNN)

YSA modellenirken sinir sisteminin ve sinir hücrelerinin yapısından esinlenilmiştir. YSA'da işlem birimleri nöronlara benzetilmiştir. Tecrübe ederek öğrenme, önceki bilgiler yardımıyla tümevarım yapabilme, bunlara bağlı olarak yeni sonuçlar üretebilme gibi insan beyni fonksiyonlarını taklit eden bir yapıdadır [2].

2.2.2 Kümeleme Yöntemleri

2.2.2.1. K-Ortalama

K-Ortalama yönteminin amacı veri setini girişte belirlenen k adet kümeye ayırmaktır. Ayırma işlemi gerçekleştirilirken küme içi benzerliklerinin en fazla, kümeler arası benzerliğin ise en az olması amaçlanmaktadır. Burada küme benzerliği, ağırlık merkezi ile veri setindeki tüm nesnelerin arasındaki uzaklıkların ortalamasıdır [5].

2.2.2.2. Bulanık C-Ortalama (BCO)

BCO algoritmasının sonuçları, K-Ortalama gibi sadece 0(İlgili kümeye ait değil)-1(İlgili kümeye ait) değerleri olmayıp 0-1 arası tüm değerleri alabilir. Yani BCO, kesin aitlik belirtmek yerine aitlik derecesi belirtir. Dolayısıyla bulanık yöntemler klasik yöntemlere göre

çıkış verisinde daha çok bilgi taşır [10].

Yapılan çalışma ile farklı yöntemlerle incelenen tiroit veri seti üzerinde elde edilecek başarılar karşılaştırılmış ve tipta teşhise yönelik kullanılabilecek uzman sistemler konusunda kullanılmıştır.

3. Uygulama

Kullanılan tüm sınıflama yöntemlerinde eğitim ve test verilerini belirlemek için 10-kat Çapraz Doğrulama kullanılmıştır. Materyal bölümünde de bahsedildiği gibi veri setimiz 215 elemandan oluşmaktadır. 10-kat çapraz doğrulama bu veri için uygulandığında işlem adımları şu şekilde gerçekleşmektedir.

215 eleman 10 eşit gruba ayrılmak istenmektedir. 215, 10'a tam bölünemeyeceğinden 5 grup 21 elemanlı olurken diğer 5 grup 22 elemanlı olacaktır. Her bir grubun elemanları sıralı değil rastgele olarak seçilir. Örnek olarak 1. grup 215 elemandan 7., 24., 25., 88., 108., ..., 204. elemanlar seçilerek oluşturulur. Diğer guruplar da bu şekilde oluşturulduktan sonra sırasıyla her bir grup test için ayrılır ve geri kalan 9 gruba uygulanacak olan yöntemin algoritması eğitilir. Daha sonra ayrılan grupla test edilir. 10 grup olduğundan dolayı 10 ayrı test sonucu olacaktır. Bu testlerin ortalama başarısı ise genel başarıyı yani uygulanacak olan yöntemin başarısını gösterecektir. Çapraz doğrulamanın kullanıldığı yöntemlerde başarı her zaman aynı olmayabilir. Bunun sebebi elemanların rastgele seçilmesidir. Bu rastgelelikten dolayı eğitim ve test verileri farklı olacağından başarı oranları bir birine çok yakın çıksa da aynı olmaz.

Tablo 1'deki başarı oranlarına bakıldığında tiroit hastalığı tanısında sınıflama yöntemi -rinin kümeleme yöntemlerine göre daha başarılı olduğunu söyleyebiliriz. Burada en yüksek başarı %96,57 sınıflama doğruluğu DVM ile elde edildiği görülecektir. Uygulama yapılırken %95,34 doğruluk derecesiyle FFNN'in eğitimi yaklaşık 10 sn sürdüğü buna karşılık test işlemini çok kısa sürede yapabilmektedir. Diagquadratic diskriminant analizi ile çok kısa bir sürede %96,27 lik doğruluk derece ile DVM'ye yakın bir başarı elde edilmiştir.

Kullanılan yöntemler		Başarı oranı (%)	İşlem süresi (sn)
DVM		96,57	0,101
İleri Beslemeli YSA		95,34	9,72
Diskriminant Analizi	Linear	93,95	0,663
	Diaglinear	93,95	0,019
	Quadratic	95,81	0,020
	Diagquadrati	96,27	0,020
	c		
Mahalanobis		93,48	0,018
BCO		90,69	0,457
K-Ortalama		88,83	0,926

Tablo 1. Uygulanan Yöntemlerin Başarıları

Kümeleme yöntemleri danışmansız öğrenme yöntemlerinden olduğundan eğitim söz konusu değildir. Kümelemenin sonuçlarına baktığımızda BCO'nun %90,69 doğruluk derecesiyle başarı oranı %88,83 olan K-Ortalama'dan daha başarılı bir kümeleme başarısı göstermiştir.

4. Sonuç ve Öneriler

Uzman sistemler her geçen gün yeni alanlarda kullanılmaktadır. Tıp alanında teşhis ve tedavilerde daha doğru ve kesin sonuçlar elde etmek, insan kaynaklı hataları yok etmek ve hekime yardımcı olacak ve kolaylık sağlayacak bir karar destek sistemi tasarlanabilir. Hekimin olmadığı veya yetersiz olduğu yerlerde hastalığa ön tanı veya tanı koymak için geliştirilecek uzman sistemlerin insan hayatı için çok önemli olduğu açıktır. Çalışmada kullanılan tüm yöntemlerin kabul edilebilir başarımları olmasına rağmen, uygulama sonuçlarından sınıflama yöntemlerinin (%93,48 ile %96,57 arasında) kümeleme yöntemlerinden daha yüksek başarıma (%88,83 ile %90,68 arasında) sahip olduğu görülmüştür.

Çalışma devam etmekte olan bir çalışmanın başlangıç aşaması olup, karma yapılu uzman sistemler kullanılarak daha başarılı yöntem-lerin geliştirilebileceği ve bu yöntemlerin daha fazla veriler kullanılarak başarılı doğru- luk derecesi elde edilip gelecekte kullanıla-bilecek bir tanılama sisteminin geliştirilmesi amaçlanmaktadır.

Yapılan çalışmada elde edilen sonuçlar, önerilen analiz sisteminin tiroit hastalığı için tanı koymada karar destek sistemi olarak kullanılabilceğini göstermiştir.

5. Kaynaklar

[1] Alkan, A., Günay, M., "Identification of EMG signals using discriminant analysis and SVM classifier", *Expert Systems with Applications* 39, (2012) 44-47.

[2] Bozkurt, M.R. 2007. "EMG İşaretlerinin Modern Yöntemlerle Önişlenmesi ve Sınıflandırılması". S.Ü. Doktora Tezi, Sakarya, 112s.

[3] Dogantekin, E., Dogantekin, A., Avcı, D., "An expert system based on Generalized Discriminant Analysis and Wavelet Support Vector Machine for diagnosis of thyroid diseases", *Expert Systems with Applications*, vol. 38, pp. 146-150, 2011.

[4]Günay,M.,"EMG İşaretlerinin Sınıflandırılması ve Öbikleştirilmesi", Kahramanmaraş Sütçü İmam Üniversitesi Fen Bilimleri Enstitüsü yüksek lisans tezi, 18s. 2011.

[5] Günay, M., Alkan, A. 2009. "EMG İşaret- lerinin K-Ortalama Algoritması Kullanılarak

Öbikleştirilmesi." *KSÜ Mühendislik Bilimleri Dergisi*, 12(2).

[6]<http://www.mahirakyildiz.com/tiroidhs.html>/08.09.2012.

[7]http://www.rehberdergisi.com/Rehber.asp?DergiDetay=Saife_Gozlem&RehberOzelNo=1293/10.08.2012

[8] Karagülle, F. 2008. Destek Vektör Makinelerini Kullanarak Yüz Bulma. T.Ü. Yüksek Lisans Tezi, Edirne, 75s.

[9] Keleş A.,Keleş A., "Expert system for thyroid diseases diagnosis", *Expert Systems with Applications*, vol. 34, pp.242-246, 2008.

[10] Kocyigit, Y., Korurek, M. 2005. "EMG İşaretlerini Dalgacık Dönüşümü ve Bulanık Mantık Sınıflayıcı Kullanarak Sınıflama." *ITÜ dergisi/d Mühendislik*, 4(3):25-31.

[11] Kodaz, H., Özşen S., Arslan, A., Güneş, S., "Medical application of information gain based artificial immune recognition system (AIRS):Diagnosis of thyroid disease", *Expert System With Applications*, vol. 36 pp. 3086-3092, 2009.

[12] Polat, K., Şahan, S., Güneş, S., "A novel hybrid method based on artificial immune recognition system (AIRS) with fuzzy weighted pre-processing for thyroid disease diagnosis", *Expert System With Applications*, vol. 32 pp. 1141-1147, 2007.

[13] Polat, K., Güneş, S., "A hybrid medical decision making system based on principles component analysis, k-NN based weighted pre-processing and adaptive neuro-fuzzy inference system", *Digital Signal Processing*, vol. 16, pp. 913-921, 2006.

[14] Şenol,C., Yıldırım, T., "Bulanık-Sinir Ağı Yapısı İçin Yeni Bir Karma Yaklaşım", *Elektrik - Elektronik ve Bilgisayar Mühendisliği Sempozyumu (Eleco'2008)*, 26-30 Kasım 2008, Bursa, Türkiye.

[15] Şenol, C., Yıldırım T., "Thyroid and Breast Cancer Disease Diagnosis using Fuzzy-Neural Networks", 6th International Conference on Electrical and Electronics Engineering (Eleco'2009), Bursa, Turkey, December, 2009.

[16] Temurtaş F, "A comparative study on thyroid disease diagnosis using neural networks" , *Expert System With Applications*, vol. 36 pp. 944-949,2009.