



Karaciğer Kanseri Teşhisinde Destek Vektör Makinesi Tabanlı Uzman Sistem Tasarımı

Support Vector Machine based Expert System Design for Liver Disorder Diagnosis

Aysegul Alaybeyoglu
İzmir Katip Çelebi Üniversitesi
Bilgisayar Mühendisliği Bölümü
Çiğli, İzmir
aysegul.alaybeyoglu@ikc.edu.tr

Naciye Mulayim
İzmir Katip Çelebi Üniversitesi
Biyomedikal Mühendisliği Bölümü
Çiğli, İzmir
aysegul.alaybeyoglu@ikc.edu.tr

Özetçe—Bu çalışmada karaciğer kanseri hastalığının teşhisi için Destek Vektör Makineleri yöntemi kullanılarak uzman sistem tasarımı geliştirilmiştir. Hasta verileri için Indian Liver Patient Dataset (ILPD) veri tabanından yararlanılmıştır. Önerilen Sistemlerin performans değerlendirilmesi için Doğruluk, Pozitif Kestirim Değeri, Negatif Kestirim Değeri, Duyarlılık, Özgüllük Hassaslık ve F-ölçütü kullanılmıştır.

Anahtar Kelimeler — destek vektör makinesi; karaciğer kanseri; uzman sistem.

Abstract—In this study, in order to diagnose liver disorder, a support vector machine based expert system is developed. Indian Liver Patient Dataset (ILPD) is used as patient data. For performance evaluations of the proposed system Accuracy, Positive Predictive Value, Negative Predictive Value, Sensitivity, Specificity, Precision and F-Measure metrics are used.

Keywords — support vector machine; liver disorder; expert system.

I. GİRİŞ

Günümüzde uzman sistemler birçok gerçek dünya problemlerinin çözümünde yaygın olarak kullanılmaktadır. Uzman sistemlerde, yapay zekâ teknikleri insanların düşünce sistemlerini ve davranışlarını taklit edebildiği için, günlük problemlerin çözümü için zamanında ve doğru kararların elde edilmesi için kullanılır. Özellikle medikal alanda, herhangi bir hastalığın teşhisi için, çeşitli yapay zekâ tekniklerine dayanan uzman sistemleri kullanan birçok araştırma ve uygulamalar vardır.

Karaciğerin insan vücudu için hayati bir iç organ olması ve diğer organlarla ilişkili olan birçok fonksiyonu yerine getiriyor olması, bu organda oluşabilecek bir hastalığın teşhisinin önemini çok net ortaya koymaktadır. Karaciğer bir hasarla karşılaştığı zaman kendini yenileyebilme yeteneğine sahip

olmasına rağmen, büyük hasarlar karaciğer fonksiyonlarının sürekliliği için engel teşkil eder. Bu yüzden karaciğer hastalıklarında erken teşhis çok önemlidir.

Bu çalışmada karaciğer kanser hastalığının teşhisi için C# programlama dili ve “Indian Liver Patient Dataset (ILPD)” veri tabanı kullanılarak Destek Vektör Makinesi tekniklerine dayalı uzman sistem geliştirilmesi hedeflenmiştir. Önerilen Sistemlerin performans değerlendirilmesi için Doğruluk, Pozitif Kestirim Değeri, Negatif Kestirim Değeri, Duyarlılık, Özgüllük Hassaslık ve F-ölçütü kullanılmıştır.

Çalışmanın ikinci bölümünde mevcut çalışmalardan, üçüncü bölümünde çalışmanın gerçekleştirilmesi için kullanılan yöntem olan destek vektör makinesinden, dördüncü bölümde geliştirilen sistemin detaylarından ve son olarak beşinci bölümde ölçüm sonuçlarında elde edilen verilerden bahsedilmektedir.

II. İLGİLİ ÇALIŞMALAR

Yapay zeka tekniklerinin hastalık teşhisinde kullanılması üzerine geliştirilen çok sayıda çeşitli çalışmalar mevcuttur. Bu çalışmalardan özellikle de destek vektör makinesi tekniğinin karaciğer kanseri hastalığı teşhisinde kullanıma dair çalışmalar da geliştirilmiştir [1-5]. [1]’de destek vektör makinesi, C4.5 ve yapay sinir ağları yöntemleri karaciğer hastalığı teşhisinde kullanılmış olup, performans analizleri gerçekleştirilmiştir. [2]’de karaciğer hastalığı teşhisinde, destek vektör makinesi, bayesian ve K-En Yakın Komşuluğu yöntemleri kullanılmıştır. [3]’te destek vektör makinesi ve Naive Bayes Algoritmaları, [4]’te destek vektör makinesi ve Random Forest algoritmaları kullanılmış olup performans analizleri gerçekleştirilmiştir. [5]’te karaciğer kanseri veri setleri destek vektör makinesi kullanılarak sınıflandırılmıştır.

III. DESTEK VEKTÖR MAKİNESİ

Destek Vektör Makineleri, bir öğrenme algoritması olup istatistiksel öğrenme teorisine dayanmaktadır[6]. Destek Vektör Makinelerinin kullanımındaki amaç iki sınıfa ait verileri birbirinden en uygun şekilde ayırmaktır. Bu amaçla karar sınırları belirlenir. Günümüzde yüz tanıma sistemlerinden, ses analizine kadar birçok sınıflandırma probleminde kullanılmaktadırlar. Destek Vektör Makineleri, veri setinin doğrusal olarak ayrılabilme ve ayrılamama durumuna göre ikiye ayrılmaktadır [6-9].

Doğrusal Destek Vektör Makineleri ile sınıflandırmada, iki sınıfa ait örneklerin doğrusal olarak dağıldığı varsayıldığında, bu iki sınıfın eğitim verisi kullanılarak elde edilen bir karar fonksiyonu yardımı ile birbirinden ayrılması amaçlanmaktadır. Veri setini ikiye ayıran doğru karar doğrusu olarak tanımlanmaktadır. Amaç en uygun karar doğrusunu belirlemektir [6-9].

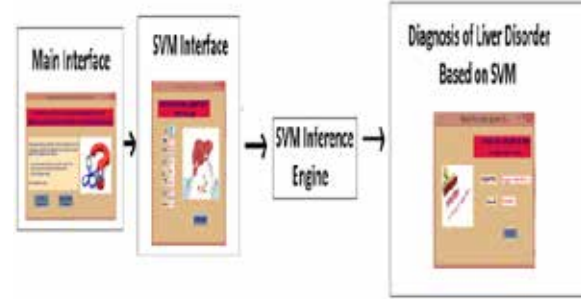
Doğrusal Olmayan Destek Vektör Makineleri ile sınıflandırmada Doğrusal olmayan bir veri kümesinde doğrusal bir karar sınırları çizilemez. Bu nedenle çekirdekler tanımlanır. Çekirdek yöntemi, doğrusal olmayan verilerde makine öğrenimini yüksek oranda arttırmaktadır. Polynomial Kernel ve Gaussian RBF (Radial Basis Function) Kernel en çok kullanılan çekirdek yöntemlerindedir [6-9].

IV. GELİŞTİRİLEN SİSTEM

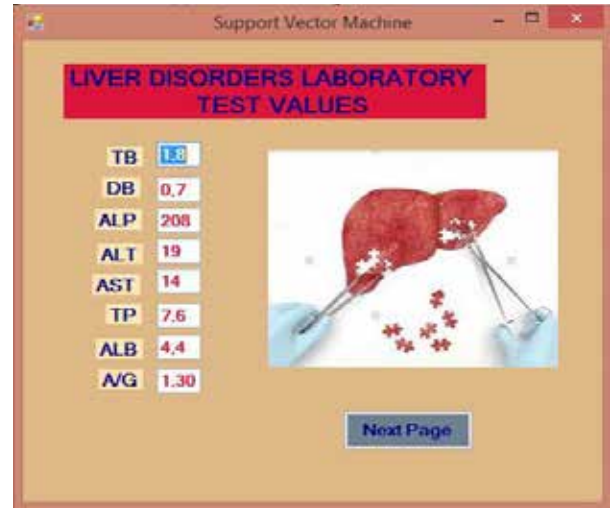
Bu çalışmada karaciğer kanseri hastalığının teşhis edilebilmesi için Destek Vektör Makinesi tabanlı Uzman Sistem Tasarımı gerçekleştirilmiştir. Gerekli olan hasta verileri için ILPD veri tabanından yararlanılmıştır. Kullanıcının geliştirilen sistemi kullanabilmeleri için ara yüzler tasarlanmıştır. Kullanıcılar laboratuvar test sonuçlarını bu ara yüzler aracılığı ile sisteme girmekte ve sistem, girilen verileri destek vektör makinesi tekniğini kullanarak hasta ile ilgili sonuç çıkarsamasında bulunmaktadır.

Şekil 1'de geliştirilen sistemin genel şeması sunulmaktadır. Şekil 1'de de görüldüğü üzere, kullanıcı sistemi çalıştırdığında ana arayüz olarak tanımlanan başlangıç sayfası ile karşılaşmakta ve sistemin çalışma prensibi ile ilgili bilgi almaktadır. Devam tuşuna bastığında, laboratuvar test sonuç verilerini girebileceği bir ara yüz ile karşılaşmaktadır. Kullanıcı veri giriş işlemini tamamladıktan sonra Destek Vektör Makinesi tekniği ile girilen veriler işlenmekte ve karaciğer hastalık durumu ile ilgili sonuç verilmektedir.

Şekil 2'de ILPD veri seti için Karaciğer Kanserli Laboratuvar Test Sonuçlarının Girildiği Ekran Ara yüzü gösterilmektedir. Kullanıcı test sonucundaki TB, DB, ALP, ALT, AST, TP, ALB ve A/G değerlerini Şekil 2'de ara yüz aracılığı ile sisteme girmekte ve devam tuşuna bastıktan sonra destek vektör makinesi bu veriler üzerinde işlem yaparak Şekil 3'te belirtilen sonucu ekranda göstermektedir.



Şekil1. Önerilen Sistemin Genel Şeması



Şekil2. ILPD Veri Seti için Karaciğer Kanserli Laboratuvar Test Sonuçlarının Girildiği Ekran Ara yüzü



Şekil 3. Destek Vektör Makinesinin Laboratuvar Verini İşlemesi Sonucunda Elde Edilen Sonucun Ekranda Gösterilmesi

V. ÖLÇÜM SONUÇLARI

Sistem gerçekleştiriminde C# programlama dili kullanılmış olup, veri seti olarak da Indian Liver Patient Dataset (ILPD) kullanılmıştır. Veri setinde 583 hasta veri örneği bulunmakta olup, bu verilerden 167 adedi karaciğer hastası olmayan, 416 adedi de karaciğer hastası olan verilerden oluşmaktadır. 142



adedi bayan 441 adedi erkek verisine aittir. Veri setinde kullanılan özelliklerin açıklamaları:

TB: Total Bilirubin, DB: Direct Bilirubin, ALKPHOS: Alkaline Phosphatase, SGPT: Alanine Aminotransferase, SGOT: Aspartate Aminotransferase, TP: Total Proteins, ALB: Albumin, A/G: Ratio Albumin ve Globulin Ratio [10]. Performans analizlerinde kullanılan metrikler [11]:

True positive (TP-Gerçek Pozitif): Doğru teşhis konulan hasta bireyler.

False positive (FP- Yanlış Pozitif): Hasta teşhisi konulan sağlıklı bireyler.

True negative (TN-Gerçek Negatif): Doğru teşhis konulan sağlıklı bireyler.

False negative (FN-Yanlış Negatif): Sağlıklı teşhisi konulan hasta bireyler.

Ölçümlerde kullanılan formüller Tablo I'de sunulmuştur.

TABLO I. ÖLÇÜMLERDE KULLANILAN FORMÜLLER

Doğruluk	$(TP+TN)/(TP+TN+FP+FN)$
Pozitif Yordama Değer	$= TP / (TP + FP)$
Negatif Yordama Değeri	$= TN / (FN + TN)$
Duyarlılık	$= TP / (TP + FN)$
Özgüllük	$= TN / (FP + TN)$

Tablo II. de önerilen sistemin uygulamasından elde edilen sonuçlar verilmiştir.

TABLO II. KARACİĞER KANSERİ TEŞHİSİ İÇİN DESTEK VEKTÖR MAKİNESİ TABANLI DEĞERLENDİRME SONUÇLARI

Karaciğer Kanseri Teşhisi için Destek Vektör Makinesi Algoritması tabanlı değerlendirme sonuçları	Gerçek pozitif değer (Datasetteki)	Gerçek negatif değer (Datasetteki)
Destek Vektör Makinesi Algoritması tabanlı sistemin pozitif değerlendirmesi	True Positive (TP) = 163	False Positive (FP) = 24
Destek Vektör Makinesi Algoritması tabanlı sistemin negatif değerlendirmesi	False Negative (FN) = 41	True Negative (TN) = 72

Tablo 2'de görüldüğü üzere, sistem 163 adet hastayı gerçek pozitif, 41 adet yanlış negatif, 24 adet hastayı yanlış pozitif, 72 adet hastayı gerçek negatif olarak belirlemiştir.

TABLO III. KARACİĞER KANSERİ TEŞHİSİ İÇİN DESTEK VEKTÖR MAKİNESİ TABANLI DEĞERLENDİRME SONUÇLARI

Negatif Yordama Değer (%)	Pozitif Yordama Değer (%)	Özgüllük (%)	Duyarlılık (%)	Doğruluk (%)
63.71	87.16	75	79.9	78.33

Tablo III'te görüldüğü üzere Negatif yordama değeri %63.71, Pozitif Yordama değeri %87.16, Özgüllük %75, Duyarlılık %79.9 ve Doğruluk %78.33 olarak belirlenmiştir. Pozitif yordama değerinin 87.16 olması önemli bir sonuç olarak kabul edilebilir çünkü sistem pozitif sonuçları değerlendirmede yüksek güvenilirlikle bir sonuç vermiştir.

VI. SONUÇ

Bu çalışmada destek vektör makinesi algoritması tabanlı karaciğer kanseri teşhisi amaçlı bir uzman sistem tasarımı gerçekleştirilmiştir. Önerilen sistem uzmanlar ve sağlık personeli tarafından zamanında ve doğru teşhis için yardımcı bir sistem olarak kullanılabilir. Ayrıca destek vektör makinesi algoritmasının sistemin tahmin mekanizmasında kullanılması daha doğru tahminlerin yapılmasında katkı sağlayıcı olmuştur.

KAYNAKLAR

- [1] Ramana B.V., Babu M.S.P., Venkateswarlu N. B., "A Critical Study of Selected Classification Algorithms for Liver Disease Diagnosis", International Journal of Database Management Systems (IJMS), Vol.3, No.2, May 2011
- [2] Ribeiro R., Marinho R., Velosa J., Ramalho F., Sanches J.M., " Diffuse liver disease classification from ultrasound surface characterization, clinical and laboratorial data", Pattern Recognition and Image Analysis pp 167-175 2011
- [3] Gulia A., Dr.Vohra R., Rani P., " Liver Patient Classification Using Intelligent Techniques", Volume 4, Issue 2, ISSN: 2277 128X, 2014.
- [4] Kahramanli H., Allahverdi N., " Mining Classification Rules for Liver Disorders", Issue 1, and Volume 3, 2009.
- [5] Banati H., Bajaj M., "Performance analysis of firefly algorithm for data clustering", Int. J. Swarm Intelligence, Vol. 1, No. 1, 2013
- [6] Akpınar H., "DATA -Veri Madenciliği Veri Analizi", Papatya Bilim, 2014
- [7] Kavzoğlu, T., Çölkesen, İ., "Destek Vektör Makineleri ile Uydu Görüntülerinin Sınıflandırılmasında Kernel Fonksiyonlarının Etkilerinin İncelenmesi" Harita Dergisi Temmuz 2010 Sayı 144.
- [8] Osuna, E.E., Freund, R., Girosi, F., "Support Vector Machines: Training and Applications", A.I. Massachusetts Institute of Technology and Artificial Intelligence Laboratory, Massachusetts, 1997.
- [9] Ülgen, E.K., "Makine Öğrenimi-Destek Vektör Makineleri", <https://medium.com/@k.ulgen90/makine-%C3%B6%C4%9Frenimi-b%C3%B6l%C3%BCm-4-destek-vekt%C3%B6r-makineleri-2f8010824054>
- [10] <http://archive.ics.uci.edu/ml/datasets/ILPD+%28Indian+Liver+Patient+Dataset%29>
- [11] Badaru G., Badaru V., Ionita I., " Possible Inference Engine Part For An Expert System In Materials Selection ", ISSN 2307-5376, 2009.