DİJİTAL SİTOLOJİDE KANSER TANIMA İÇİN ANALİTİK VE ÖNGÖRÜSEL YARI-GÜDÜMLÜ ÖĞRENME

ANALYTICAL AND PREDICTIVE QUASI-SUPERVISED LEARNING FOR CANCER RECOGNITION IN DIGITAL CYTOLOGY

Bilge Karaçalı

Elektrik-Elektronik Mühendisliği Bölümü
İzmir Yüksek Teknoloji Enstitüsü
bilgekaracali@iyte.edu.tr

ÖZETÇE

Bu çalışmada yarı-güdümlü öğrenme ile dijital sitoloji verilerinde kanser tanıma gerçekleştirilmiştir. Tanıma uygulanan veri kesin referans olarak sadece kansersiz örneklerden oluşan işaretlenmiş bir veri kümesini içermekte, kanserli örnekler ise kansersiz örneklerle beraber işaretlenmemiş bir karışık veri kümesinde verilmektedir. Bu kapsamda yarı-güdümlü öğrenme algoritması kullanılarak hem eldeki karışık veri kümesindeki kanserli örnekleri saptayacak analitik bir yöntem ile birlikte daha sonradan karşılaşılan örnekleri eldeki bu veriye dayanarak kanserli veya kansersiz olarak ayırt edecek öngörüsel bir yöntem kurgulanmıştır. Deneylerde yarı-güdümlü öğrenme algoritmasına dayanan yöntemleri, güdümlü destek vektör makinası sınıflandırıcılarına dayalı alternatif yaklaşımlara göre her iki durumda da daha yüksek bir tanıma başarısı elde etmiştir. Bu sonuçlar yarı-güdümlü öğrenmenin, istatistiksel öğrenme için sadece işaretlenmiş kansersiz örneklerin mevcut olduğu durumlarda hem analitik hem öngörüsel tanımada geçerli olan tek yaklaşım olduğunu göstermektedir.

ABSTRACT

In this work, cancer recognition in digital cytology data was carried out using quasi-supervised learning. The data subject to recognition contained ground-truth data only in the form of a labeled set of cancer-free samples and the cancerous samples were provided along with cancer-free samples in an unlabeled mixed dataset. In this framework, a predictive method was derived to label future samples as cancerous or cancer-free based on this data at hand together with an analytical method to label the cancerous samples in the mixed dataset. In the experiments, the methods based on the quasi-supervised learning algorithm achieved higher recognition performance in both cases than the alternative approaches based on supervised support vector machine classifiers. These results indicate that the quasi-supervised learning is the only valid approach in both analytical and predictive recognition when only labeled cancer-free samples are available for statistical learning.

1. GİRİŞ

Biyomedikal verilerin bilgisayarlı analizinde güdümlü öğrenme yaklaşımını gerçekleyen sınıflandırıcı temelli yöntemler ağırlıklı olarak kullanılmaktadır [1]. Bu şekilde oluşturulan karar sistemleri sorgu altındaki örnekleri, birtakım ayırıcı özelliklerini içeren özniteliklerine göre olası kategorilerden en uygun olanına atayabilmektedir. Güdümlü sınıflandırıcıların kurgulanmasındaki en kritik bileşen ise, uygulamada karşılaşılacak örneklere benzer ve hangi kategoriye veya sınıfa ait oldukları bilinen örnekleri içeren eğitim kümesidir.

Biyomedikal veri analizi uygulamalarında yukarıdaki amaca yönelik eğitim kümeleri tipik olarak eldeki çok sayıda veri örneğinin alanın uzmanı bir kişi tarafından elle işaretlenmesiyle oluşturulur. Bu ise yüksek miktardaki sayısal veride elle işaretlemenin zahmetli olması, uzun sürmesi ve işaretlemeyi yapan kişinin kişisel yorumlarına açık olması sebebiyle problemli bir süreçtir.

Mevcut sınıfları örnekleyen eğitim kümelerinin elle oluşturulmasında karşılaşılan sorunların aşılması için yarı-güdümlü öğrenme yöntemi geliştirilmiştir [2]. Güdümlü sınıflandırmanın aksine yarı-güdümlü öğrenme, normal duruma ait verilerin toplandığı bir kontrol veri kümesine ek olarak verilen ve hem normal hem hedef duruma ait verileri içeren, ancak veri işaretlemesi yapılmamış karışık bir veri kümesi üzerinden tanıma gerçekleştirmektedir. Bu şekilde bilgisayarlı tanıma için hedef sınıfa ait verilerin elle işaretlenmesi gerekliliği ortadan kaldırılmış olmaktadır. Özellikle biyomedikal uygulamalarda geniş veri kümelerinde hedef sınıfa ait verilerin bulunup bulunmadığı kolaylıkla belirlenebildiği ancak bu verileri teker teker işaretlemenin zorluğu göz önüne alındığında yarı-güdümlü yaklaşım, istatistiksel öğrenme için gereken uzman girdisini en az düzeye indirmektedir.

Bu çalışmada yarı-güdümlü öğrenme yaklaşımı, dijital sitoloji verilerinde kanser tanıma problemine uygulanmıştır. Çalışılan senaryo, kanserli örnek içermeyen bir kontrol kümesi oluşturulduktan sonra derlenen ve herhangi bir biçimde işaretlenmemiş veriler içeren karışık veri kümesinin içerisinden kanserli olanların tanınması üzerine kurgulanmıştır. Bu analitik tanıma senaryosunda yarı-güdümlü öğrenmenin tanıma başarısı, karışık veri kümesindeki kanserli örneklerin değişen oranı için belirlenmiştir. Öngörüsel tanıma senaryosunda ise kontrol ve karışık kümelere dayanarak daha sonra gelen veriler üzerinde elde edilen tanıma başarısı değerlendirilmiştir. Deneysel sonuçlarda yarı-güdümlü öğrenme yönteminin, literatürdeki güdümlü sınıflandırıcılara dayalı olarak türetilebilecek

978-1-4673-0056-8/12/$26.00 ©2012 IEEE

alternatif yaklaşımlara kıyasla daha yüksek bir tanıma başarısı sağladığı gözlenmiştir.

Çalışmada kullanılan dijital sitoloji verisi ve analitik ve öngörüsel yarı-güdümlü öğrenme ile olası alternatif yaklaşımlar aşağıda özetlenmiştir. Deneysel Sonuçlar Bölümünde yarı-güdümlü yaklaşımın eldeki veriler üzerine yürütülen kontrollü deneylerde sağladığı kanserli örnekleri tanıma başarısı, alternatif yaklaşımların tanıma başarıları ile beraber sunulmuştur. Elde edilen bulgular Tartışma Bölümünde yorumlanmıştır.

2. MATERYAL VE METOT

Bu bölümde öncelikle kullanılan dijital sitoloji veri kümesi tanıtılmıştır. Daha sonra, bu veri kümesi üzerine kurgulanan analitik ve öngörüsel kanser tanıma problemleri irdelenmiş ve her biri için yarı-güdümlü öğrenme algoritmasına dayanan çözümler ve olası alternatifler formüle edilmiştir.

2.1. Dijital Sitoloji Veri Kümesi

Bu çalışmada, daha önce birçok kez kanser tanıma yöntemlerinin geliştirilmesinde ve sınanmasında kullanılmış olan Wisconsin Tanısal Meme Kanseri veri kümesi kullanılmıştır [1]. Bu veri kümesinde, 357 tanesi normal ve 212 tanesi de kanserli olan toplam 569 dijital sitoloji örneği yer almaktadır. Her örnek için çekilen dijital fotoğraflardaki hücre çekirdeklerinin yapısal özelliklerini yansıtan 30 adet öznitelik hesaplanarak bir vektör halinde kaydedilmiştir.

2.2. Yarı-Güdümlü Perspektifte Analitik Tanıma

Literatürde tanıma problemleri, temel olarak, ilgilenilen gruplardan hangisine ait olduğu belli olan örnekleri sayısal öznitelik vektörleri cinsinden birbirinden ayırt eden sınıflandırıcılar üzerinden çözülmeye çalışılmıştır. Farklı gruplardaki örnekleri birbirleriyle karşılaştırarak öznitelik uzayının farklı bölgelerini ilgili gruplarla eşleştirmeye çalışan en yakın komşuluk [3, 4] ve destek vektör makinaları [5, 6] gibi birçok sınıflandırma yöntemi mevcuttur. İkili bir sınıflandırma probleminde amaç, sorgulanan bir örneğin özniteliklerini ifade eden $x \in X$ vektörünü C_0 ve C_1 ile ifade edilen iki gruptan birine,

$$I(x) = \begin{cases} 0 & x \in C_0 \text{ ise} \\ 1 & x \in C_1 \text{ ise} \end{cases} \tag{1}$$

belirleyici fonksiyonunu yaklaşıklayarak atayan bir $f(x)$ fonksiyonu oluşturmaktır. Böyle bir fonksiyon, bu grupların X gözlem uzayındaki $p_1(x)$ ve $p_2(x)$ olasılık dağılımlarının bilindiği durumlarda

$$f(x) = \begin{cases} 0 & p_0(x) > p_1(x) \text{ ise} \\ 1 & p_0(x) < p_1(x) \text{ ise} \end{cases} \tag{2}$$

ifadesiyle en az sınıflandırma hatasını sağlayacak şekilde oluşturulabilir [7]. Olasılık fonksiyonlarının yokluğunda benimsenen yaklaşım ise, gerçek sınıf bilgisi bilinen bir $\{x_i, y_i\}$, $x_i \in X$, $y_i \in \{0, 1\}$, $i = 1, 2, ..., \ell$, eğitim kümesindeki

$$E(f) = \sum_{i=1}^{\ell} \left| f(x_i) - y_i \right| \tag{3}$$

ile ifade edilen sınıflandırma hatasını, bir takım düzenlilik şartlarını da sağlayarak mümkün olduğunca en aza indiren bir fonksiyon kurgulamaktır. Bu çerçevede en temel yöntem, herhangi bir düzenlilik şartına bakılmaksızın x örneğini eğitim setinde d: $X \times X \rightarrow \mathbb{R}$ metriği cinsinden kendisine en yakın olan örneğin sınıfına atayan ve

$$f^{\text{eyk}}(x \mid \{x_i, y_i\}) = y_{i_0}, \quad i_0 = \arg\min_i d(x, x_i) \tag{4}$$

ile tanımlanan en yakın komşuluk sınıflandırıcısıdır [3].

Gerçek uygulamalarda bu yaklaşımın en kritik noktası, uygun bir sınıflandırıcı oluşturmak için gereken ve her iki gruba ait gerçek sınıf bilgisi içeren örneklerden oluşan zengin bir eğitim kümesidir. Ancak birçok uygulamada bu şekilde bir kesin referans veri kümesi elde etmek, yukarıda da değinildiği gibi çeşitli zorluklar içermektedir.

Yarı-güdümlü öğrenme, eğitim seti oluşturmadaki sorunların aşılması için önerilmiş olan esasen analitik bir tanıma yaklaşımıdır [2]. Bu yaklaşım, C_0 ile C_1 gruplarından sadece birinin, uzlaşımlı olarak C_0 grubunun tek bir kontrol sınıfına ait örneklerden oluşmuş olmasını gerektirmektedir. Bunun yanında C_1 grubu, C_0 grubundakilerle aynı sınıftan bilinmeyen sayıda örneğin yanında ikinci bir hedef sınıfa ait örneklerden derlenmekte ve bunlardan hedef sınıfa ait olanlar otomatik olarak işaretlenmektedir. Yaklaşımın analitik doğası, C_1 grubundaki örnekler için tanıma yapılırken bütün örneklerin birden değerlendirilmesinden kaynaklanmaktadır.

Yarı-güdümlü öğrenmenin esası, en yakın komşuluk sınıflandırıcısının uçdeğer davranış özellikleri kullanılarak her bir x örneği için $p(C_0 \mid x)$ ve $p(C_1 \mid x)$ sonsal olasılıklarının

$$p(C_0 \mid x) = 1 - p(C_1 \mid x) \cong \sum_{R_n} \mathbf{1}\left(f^{\text{eyk}}(x \mid R_n) = 0\right) p(R_n)$$

ile yaklaşıklanmasına dayanır. Yukarıdaki ifadede R_n, rastlantısal olarak her gruptan n tane olmak üzere toplam $2n$ örnek içeren bir referans kümesini, $\mathbf{1}(\cdot)$ ise argümanı doğru olduğunda 1, aksi halde 0 değerini veren fonksiyonu göstermektedir. Yarı-güdümlü öğrenme algoritması bu olasılıkları, bir takım çözümlemeler kullanarak

$$P_1(x) = \frac{1}{\sum_{R_n \subset \{x_i, y_i\}} 1} \sum_{R_n \subset \{x_i, y_i\}} f^{\text{eyk}}(x \mid R_n) \tag{5}$$

ve $P_0(x) = 1 - P_1(x)$ ile hesaplar. Sonsal olasılıkların bu şeklide kestirimi için en uygun olan n parametresi ise ilgili bir enerji fonksiyonelinin en küçültülmesi ile bulunur.

Bu formülasyona bağlı olarak yarı-güdümlü tanıma, her x_i örneği için $P_0(x_i)$ ve $P_1(x_i)$ olasılıklarını, bir-eksik çapraz sağlama çerçevesinde $\{x_j | \ j \neq i\}$ örneklerini kullanarak hesaplar. Böylelikle C_1 grubundaki

$$\{x_i \in C_1 | \ P_1(x_i) \geq P_c\}$$

şartını sağlayan örnekler, C_0 grubundaki örneklerden farklı ve kurgu gereği hedef sınıfa ait örnekler olarak belirlenir. Tanıma için aranan eşik değer P_c de C_0 grubundaki örneklerin kabul edilebilir bir oranı tanıma kriterini sağlayacak şekilde (sabit yanlış kabul oranlı tanıma kapsamında) seçilebilir.

Alternatif olarak C_1 grubundaki hedef sınıfına ait örneklerin belirlenmesinde C_0 ve C_1 gruplarını ayrıştırmak üzere kurgulanmış sınıflandırıcılardan da faydalanılabilir. Örneğin, her x_i için $\{x_j | j \neq i\}$ veri kümesini ayrıştıran

$$h_i(x) = \sum_{j \neq i} \alpha_j (2y_j - 1) K(x, x_j) + \beta \tag{6}$$

olacak şekilde

$$f_i(x) = \begin{cases} 0 & h_i(x) < 0 \text{ ise} \\ 1 & h_i(x) > 0 \text{ ise} \end{cases} \tag{7}$$

destek vektör makinası sınıflandırıcı fonksiyonu oluşturulup $h_i(x_i)$ değerleri belirlenen bir eşikten büyük olan örnekler, yine sabit yanlış kabul oranlı tanıma kapsamında hedef sınıfa

atanabilir. Bu ifadede K: $X \times X \rightarrow \mathbf{R}$ simetrik bir kesin artı çekirdek fonksiyonunu göstermektedir. Sınıflandırıcının $\{\alpha_j\}$ ve β parametreleri destek vektör makinası eğitimi sırasında çözümlenmekte ve katsayıları

$$\alpha_j = 0$$

olan x_j vektörler de destek vektörleri olarak tanımlanmaktadır. Çekirdek fonksiyonu bilinen vektör iç çarpımı olarak

$$K(x, x_j) = x^T x_j \tag{8}$$

seçildiğinde en yüksek kenar boşluklu doğrusal sınıflandırıcı,

$$K(x, x_j) = \exp\left(-\frac{1}{2\sigma^2} \left\| x - x_j \right\|^2 \right) \tag{9}$$

ile gösterilen Gauss çekirdeği olarak seçildiğinde ise doğrusal olmayan bir sınıflandırıcı elde edilmektedir. Bu çekirdeğin genliğini belirleyen σ parametresi mümkün olan en az destek vektörü oluşacak şekilde sıralı arama yöntemiyle bulunur.

2.3. Yarı-Güdümlü Öğrenmeyle Öngörüsel Tanıma

Yarı-güdümlü öğrenme perspektifinde analitik tanıma, tektürel C_0 kümesi ve karışık C_1 kümesine dağılmış $\{x_i\}$ örneklerinden hangilerinin C_1 kümesine özgün olduklarını saptamayı amaçlar. Buna karşın öngörüsel tanımada amaç, C_0 ve C_1 kümelerindeki örneklerden yararlanarak gelecekte karşılaşılacak örnekler için tanıma kuralları belirlemektir.

Öngörüsel tanımanın gerektirdiği ayrıştırma kuralının temeli, örneklerin C_0 ve C_1 gruplarına özgünlüklerinin eldeki $\{x_i\}$ kümesindeki örnekler kullanılarak belirlenmesidir. Yukarıda anlatılan analitik yaklaşımda C_1 içinde $P_1(x_i) > P_c$ şartını sağlayan örneklerin hedef sınıfa atanması, verilen yeni bir x örneği için de

$$f(x) = \begin{cases} 0 & P_1(x) < P_c \text{ ise} \\ 1 & P_1(x) > P_c \text{ ise} \end{cases} \tag{10}$$

tanıma kuralının işletilmesini önermektedir. Bu ifadede $P_1(x)$, analitik yarı-güdümlü öğrenmede belirlenen n kullanılarak eldeki bütün $\{x_i\}$ verisi üzerinden hesaplanmaktadır. Burada $\{x_i\}$ örnekleri eğitim kümesi olarak değerlendirilmekte, tanıyıcının eğitimi de en başarılı ayrıştırma için kullanılacak n parametresinin $\{x_i\}$ üzerinde saptanmasını içermektedir.

Bu noktada dikkat edilmesi gereken husus, C_1 kümesindeki örneklerin gerçekte hangi sınıfa ait oldukları bilinmediği için güdümlü bir sınıflandırma yönteminin eğitimine olanak sağlamamasıdır: C_0 kümesindeki verinin kontrol sınıfını örneklemesine karşın hedef sınıfını örnekleyen veri bulunmamaktadır. Bu durumda yarı-güdümlü öğrenmeye alternatif olarak yapılabilecek tek şey, her ne kadar gerçeği yansıtmasa da C_1 kümesindeki örnekler tümden ayrı bir sınıfa aitmiş gibi bir güdümlü sınıflandırıcı oluşturarak bu sınıflandırıcının C_1 kümesine atadığı x örneklerini hedef sınıfa ait örnekler olarak kaydetmektir.

3. DENEYSEL SONUÇLAR

Yarı-güdümlü öğrenme probleminin zorluğu üzerine belirleyici olan unsur, hedef sınıfa ait örneklerin C_1 kümesindeki oranıdır. Bu oran 1'e yaklaştıkça problem güdümlü tanımaya yaklaşmakta ve buna bağlı olarak da literatürdeki sınıflandırıcı temelli yöntemler tarafından çözülebilir olmaktadır. Buna karşın bu oran azaldıkça problem güdümlü tanımadan uzaklaşmakta ve hedef sınıfa ait örneklerin tespiti de aynı oranda zorlaşmaktadır.

Yukarıda anlatılan analitik ve öngörüsel yarı-güdümlü öğrenme yaklaşımlarının kanser sitoloji örneklerinde hedef sınıf olan kanserli örnekleri tanımadaki başarısı, C_1 kümesindeki hedef sınıfa ait örneklerin $\lambda = 0.50$, 0.25 ve 0.10 arasında değişen oranları için ayrı ayrı belirlenmiştir. Bunun için kanserli olmayan örneklerden rastlantısal olarak seçilen 100 tanesi ile C_0 kümesi, $100(1 - \lambda)$ tane kansersiz örnek ile 100λ tane kanserli örnek birleştirilerek de C_1 kümesi oluşturulmuştur. Elde edilen veri kümesine her öznitelikte gözlenen ortalama değer 0.0 ve standart sapma 1.0 olacak şekilde doğrusal normalleme uygulandıktan sonra analitik yarı-güdümlü öğrenme ile her x_i örneği için $P_1(x_i)$ değerleri bulunmuş ve değişen P_c eşik değerleri için C_1 kümesindeki örnekler üzerinde

$$P_{DT}(P_c) = \frac{1}{100\lambda} \sum_{\substack{x_i \in C_1 \\ x_i \text{ kanserli}}} \mathbf{1}\left(P_1(x_t) > P_c\right) \tag{11}$$

ve

$$P_{YT}(P_c) = \frac{1}{100(1-\lambda)} \sum_{\substack{x_i \in C_1 \\ x_i \text{ kansersiz}}} \mathbf{1}\left(P_1(x_i) > P_c\right) \tag{12}$$

ile hesaplanan doğru tanıma ve yanlış tanıma oranlarına ait tanıma eğrileri bulunmuştur. Bu eğrinin altında kalan alanın büyüklüğü de başarılı analitik tanımanın göstergesi olarak kaydedilmiştir. Bu işlem her λ için birbirinden bağımsız 50 kez tekrarlanarak ortalama alanlar hesaplanmıştır.

Kıyaslama amacıyla aynı problem için doğrusal ve Gauss çekirdeği kullanan destek vektör makinaları daha önceden belirlendiği şekilde kurgulanarak tanıma başarıları hesaplanmıştır. Bunun için ilgili doğru tanıma ve yanlış tanıma oranları değişen T eşik değerleri için

$$P_{DT}(T) = \frac{1}{100\lambda} \sum_{\substack{x_i \in C_1 \\ x_i \text{ kanserli}}} \mathbf{1}\left(h_i(x_t) > T\right) \tag{13}$$

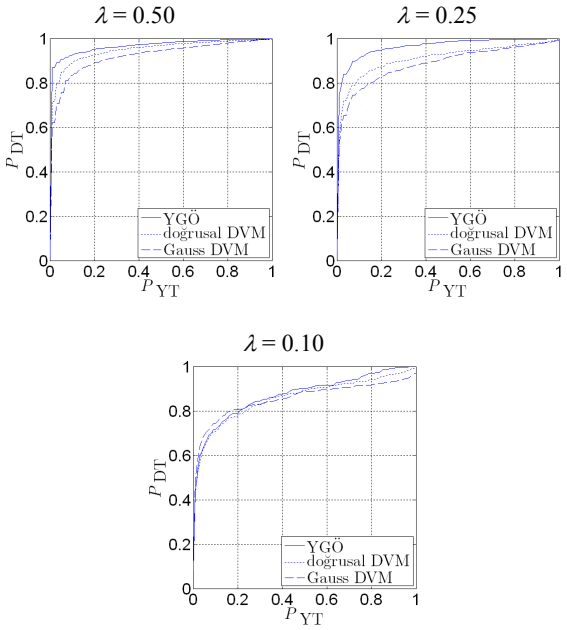
ve

$$P_{YT}(T) = \frac{1}{100(1-\lambda)} \sum_{\substack{x_i \in C_1 \\ x_i \text{ kansersiz}}} \mathbf{1}\left(h_i(x_i) > T\right) \tag{14}$$

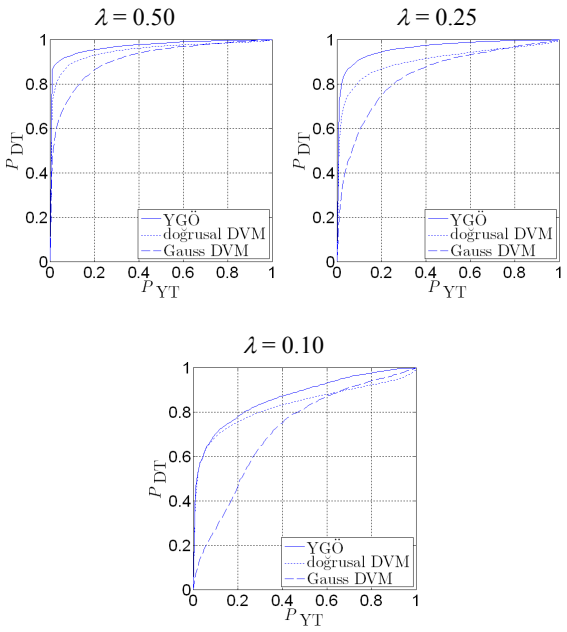
ile hesaplanmış ve yine bunlara ilişkin tanıma eğrisinin altında kalan alanlarla tanımadaki başarı ölçülmüştür.

Yarı-güdümlü öğrenmenin öngörüsel tanımadaki başarısının hesaplanması için ise C_0 ve C_1 kümelerinin dışında kalan örnekler üçüncü bir C_{test} kümesinde toplanarak C_0 ve C_1 kümelerindeki örneklere uygulanan doğrusal normalleme uygulanmıştır. Daha sonra bu kümedeki her örnek için P_1 değerleri C_0 ve C_1 kümelerindeki örnekler üzerinden hesaplanmış ve yukarıdakine benzer bir yaklaşımla tanıma eğrileri oluşturulmuştur. Bu eğrilerin altında kalan alanlar da başarı ölçütü olarak hesaplanmıştır. Yine kıyaslama amacıyla C_0 ve C_1 kümelerindeki örnekler eğitim kümesi olarak değerlendirilerek doğrusal ve Gauss çekirdekli destek vektör makinası sınıflandırıcıları kurgulanmış ve ilgili tanıma eğrileri ile eğri altındaki alanlar hesaplanmıştır.

Değişen λ oranları altındaki deneylerden elde edilen ortalama tanıma eğrileri Şekil 1 ve 2'de gösterilmiştir. Bu eğrilerde, yarı-güdümlü öğrenmenin çalışılan dijital sitoloji verisinde hem analitik hem öngörüsel kanser tanımada destek vektör makinası temelli alternatif yöntemlere kıyasla daha başarılı olduğu görülmektedir. Tanıma başarıları arasındaki fark $\lambda = 0.25$ için en belirgin düzeydedir. Tanıma eğrisi altındaki alanların Tablo 1'de verilen ortalama değerleri de bu gözlemleri desteklemektedir.



Şekil 1: Analitik tanıma eğrileri. Her λ için birbirinden bağımsız 50 tekrarda gözlemlenen eğrilerin ortalaması gösterilmiştir.



Şekil 2: Öngörüsel tanıma eğrileri. Her λ için birbirinden bağımsız 50 tekrarda gözlemlenen eğrilerin ortalaması gösterilmiştir.

4. TARTIŞMA

Bu çalışmada yarı-güdümlü öğrenme yaklaşımı dijital sitoloji verilerindedi kanser tanıma problemine uygulanmıştır. Gerçek veriler üzerindeki analitik tanıma deneylerinde yarı-güdümlü yaklaşım, işaretlenmemiş bir kümedeki örneklerin arasında kanserli olanları sadece kansersiz örneklerin bulunduğu bir kontrol veri kümesi kullanarak başarıyla saptamıştır. Öngörüsel tanıma deneylerinde de benzer bir tanıma başarısı gözlenmiştir. Bu sonuçlar dijital sitolojide kanser tanımanın kansersiz örneklere ek olarak sadece kanserli örnekler içeren bir eğitim kümesi olmadan yapılabildiğini göstermektedir.

Tablo 1: Analitik ve öngörüsel tanıma deneylerinde gözlenen tanıma eğrileri altında kalan alanların ortalama değerleri.

	$\lambda = 0.50$	$\lambda = 0.25$	$\lambda = 0.10$
Analitik tanıma			
YGÖ	0.9933	0.9833	0.8901
doğrusal DVM	0.9704	0.9254	0.8485
Gauss DVM	0.9317	0.8573	0.7352
Öngörüsel tanıma			
YGÖ	0.9736	0.9645	0.8701
doğrusal DVM	0.9509	0.9077	0.8317
Gauss DVM	0.9119	0.8439	0.7190

Kontrol veri sınıfına ek olarak hedef sınıfın da örneklendiği bir eğitim kümesinin olmadığı durumlarda bilgisayarlı tanıma için yarı-güdümlü öğrenmenin tek geçerli seçenek olarak ortaya çıkmaktadır. Güdümlü öğrenmeye dayanan sınıflandırıcı tabanlı yaklaşımlar bu probleme uyarlanabilse de yukarıdaki sonuçlarda da görüldüğü gibi yarı-güdümlü öğrenmeye kıyasla başarılı olamamaktadırlar.

Son olarak burada çalışılan tanıma probleminde doğrusal destek vektör makinası yaklaşımının Gauss çekirdeği kullanan destek vektör makinasından daha başarılı olması da dikkat çekicidir. Bu, tanıma probleminin güçlü bir doğrusal bileşeni olduğunun göstergesidir. Yarı-güdümlü öğrenmenin bu doğrusallık bilgisini kullanmadan her ikisinden de daha başarılı olması ise bu yaklaşımın sunduğu yüksek istatistiksel öğrenme gücüne tanıklık etmektedir.

Bu çalışma, PIRG03-GA-2008-230903 numaralı Avrupa Birliği Projesi ile kurulan İYTE Biyomedikal Bilgi İşleme Laboratuarının işlemsel altyapısı üzerine gerçekleştirilmiştir.

5. KAYNAKÇA

[1] O. L. Mangasarian, W. N. Street, and W. H. Wolberg, "Breast-Cancer Diagnosis and Prognosis Via Linear-Programming," *Operations Research,* vol. 43, pp. 570-577, Jul-Aug 1995.

[2] B. Karacali, "Quasi-Supervised Learning for Biomedical Data Analysis," *Pattern Recognition,* vol. 43, pp. 3674-3682, 2010.

[3] T. M. Cover and P. E. Hart, "Nearest Neighbor Pattern Classification," *Information Theory,* vol. 13, pp. 21-27, 1967.

[4] K. Fukunaga and L. D. Hostetler, "K-Nearest-Neighbor Bayes-Risk Estimation," *Ieee Transactions on Information Theory,* vol. 21, pp. 285-293, 1975.

[5] V. N. Vapnik, *Statistical Learning Theory*: Wiley-Interscience, 1998.

[6] C. Cortes and V. Vapnik, "Support-Vector Networks," *Machine Learning,* vol. 20, pp. 273-297, Sep 1995.

[7] R. O. Duda, P. E. Hart, and D. G. Stork, *Pattern Classification*, 2 ed.: Wiley-Interscience, 2000.