

Histopatolojik İmgelerin Sınıflandırılmasında Etkili Özellik Seçimi

Efficient Feature Selection in the Classification of Histopathological Images

Ömer Osman DURSUN¹, Suat TORAMAN², İbrahim TÜRKOĞLU³

¹Sivli Havacılık Yüksek Okulu, Dicle Üniversitesi, Diyarbakır
omerdursun23@gmail.com

²Enformatik Bölümü, Fırat Üniversitesi, Elazığ

³Yazılım Mühendisliği Bölümü, Teknoloji Fakültesi, Fırat Üniversitesi, Elazığ
storaman@firat.edu.tr, iturkoglu@firat.edu.tr

Özet

Bu çalışmada, histopatolojik imgelerin sınıflandırılmasında, imge-veri işleme tekniklerinde kullanılan ve birinci dereceden istatistiksel özellik olarak adlandırılan özelliklerden hangisinin veya hangilerin seçilmesinin daha uygun olacağını belirlemek hedeflenmiştir. Çalışmada kullanılan birinci dereceden istatistiksel özellikler imgenin beklenen değer, standart sapma, yamukluk ve basıklık özellikleridir. Bu özelliklerden hangisi kullanılırsa, imge sınıflandırma işleminin daha başarılı olacağı karşılaştırmalı olarak tespit edilmiştir. Karşılaştırma işleminde etkili özellik seçimi için, sıralı ileri seçim algoritması kullanılmıştır. Sonuç olarak kullanılan histopatolojik imgelere göre, basıklık özelliği tek başına kullanıldığında %90.38 sınıflama başarımları ile diğer özelliklerden ve bu özelliklerin bileşimlerinden daha etkili olduğu görülmüştür.

Abstract

In this study, the classification of histopathological images, image-data processing techniques and used in the so-called first-order statistical properties of the features is to determine which one or ones would be feasible for the selection. First-order statistical properties of the image used in this study are expected value, standard deviation, skewness and kurtosis properties. These features are used in which the image was found to be more successful classification process. The sequential forward selection algorithm was used in comparison of effective feature selection process. As a result, images used in histopathology, according to kurtosis feature, when used alone, the classification performance of 90.38% and is more effective than other features and combinations these factors.

1. Giriş

İmge (görüntü, fotoğraf), bir nesnenin sunumudur. İmge işleme ise bir imgeden başka bir imgenin elde edilmesi için kullanılan tekniklerin bütünüdür [1]. İmgelerin kullanılabilecek birinci dereceden istatistiksel özellikleri; beklenen değer, standart sapma, yamukluk, basıklık vb. gibi

birçok özelliği bulunmaktadır [2]. Bu özelliklerden hangisinin imgeleri sınıflandırmada kullanılacak en uygun özellik veya özellikler olduğunun belirlenmesi gerekmektedir. En uygun özelliklerin belirlenmesi literatürde “Etkili Özellik Seçimi” olarak ifade edilmektedir. Özellik seçimi aşaması, ayırt edici özellik bilgisini kaybetmeden hangi özelliklerin ayırt edicilik bakımından en iyi olduğunun belirlendiği süreçtir. Özellik seçimi için kullanılan sıralı ileri seçim algoritması (SFS-Sequential Forward Selection) değişik alanlarda başarı ile uygulanmaktadır. Tıp alanında; değişik özelliklere dayanan bölgelerde SUMO protein tahmininde [3], Hidroloji alanında, aylık akarsu akışının tahmin verilerin seçimi ve indirgemesinde [4], Örtü tanıma alanında, makine öğrenmesinde boyut indirgeme işleminde [5], Robot ve sistem enformatiği alanında, makine öğrenmesindeki hatanın azaltılmasında [6], Elektrik ve elektronik mühendisliği alanında, elektrik güç dağılım ağlarında akım ya da voltaj dalga şeklinin karışması olayı olan güç kalite sınıflamasının tespitinde, destek vektör makinesi ve bayes sınıflandırıcılarla etkili özellik seçiminin yapılmasında sıralı ileri seçim algoritması başarıyla uygulanmıştır [7]. Etkili özellik seçimi ile belirlenen özellik ne kadar iyi seçilirse imgenin tanınması o derece etkili olur. Bu işlem sayesinde imgelerin sınıflandırılmasında hız ve maliyet açısından önemli oranda kazanç elde edilmiş olmaktadır.

2. Teorik bakış

Bu bölümde, hepatitli imgelerin sınıflandırılmasında etkili özellik seçimi işleminin alt bölümleri sunulmaktadır.

2.1. Histopatolojik imgeler

Histopatoloji, doku ve organların doğal yapısında oluşan lezyonların mikroskopisini inceleyen bir bilim dalıdır [8]. Hızla gelişen teknolojiye tıp bilimi de faydalanmaktadır. Bilgisayar teknolojisinin, özellikle bazı işlemleri hızlı yapması ve zamandan tasarruf sağlaması önemini arttırmıştır. Tıbbi imgelerin sayısal ortama aktarılmasında, mikroskoplara bağlanan sayısal kameralar kullanılmaktadır. Mikroskop altında çekilen resimlere, histopatolojik imge denilmektedir. Histopatolojik imgelere literatürde, hepatitli imgelerin tespit

edilmesi [9], gözdeki damar çatlaklarını belirlenmesi [10] gibi çalışmalarda karşılaşılmaktadır.

2.2. İstatistiksel özellikler

Beklenen değer (mean), standart sapma, yamukluk (skewness) ve basıklık (kurtosis) istatistikte kullanılan en önemli kavramlardır. Bu kavramlar, dağılım tanımlamada, tahmin edicilerin performanslarını değerlendirmede, hipotez testlerinde test istatistiklerinin elde edilmesinde ve birçok uygulamada kullanılır. Beklenen değer(μ), dağılımın merkeze eğiliminin ölçüsünü verir. Beklenen değer, x değişkenin ağırlıklı olasılıklarının aldığı değerlerin toplamıdır [11].

$$\mu = \sum_{i=1}^{\infty} x_i f(x_i) \quad (1)$$

Standart Sapma(σ), varyansın karekökü standart sapma olarak adlandırılır. Varyans ise, istatistiksel yayılımın, mümkün bütün değerlerin beklenen değer veya ortalama uzaklıklarının karelerinin ortalaması şeklinde bulunan bir ölçüdür.

$$\sigma = \sum_{i=1}^{\infty} ((x_i - \mu)^2 f(x_i))^{1/2} \quad (2)$$

Yamukluk, verinin örnek ortalama etrafındaki asimetrik ölçüsüdür. Eğer yamukluk negatif ise, veri ortalama değerine soluna doğru dağılmıştır. Eğer yamukluk pozitif ise, veri ortalama değerine sağına doğru dağılmıştır. Normal dağılımın yamukluğu 0'dır.

$$s = \frac{E(x - \mu)^3}{\sigma^3} \quad (3)$$

Basıklık, bir dağılımın aykırı eğilim değerinin bir ölçüsüdür (Bir dağılımın tepesinin genişliği ya da düzlüğünün merkeze yakınlığının derecesidir.). Normal dağılımın basıklığı 3'tür.

$$k = \frac{E(x - \mu)^4}{\sigma^4} \quad (4)$$

2.3. Özellik seçimi

Özellik seçimi aşaması, ayırt edici özellik bilgisini kaybetmeden hangi özelliklerin ayırt edicilik bakımından en iyi olduğunun belirlendiği süreçtir. Özelliklerin ayırt edicilik bilgisi düşük ise, buna bağlı olarak bilginin sınıflandırılma performansı düşük olacaktır. Bir başka deyişle, etkili özellik bilgisi seçilirse bilginin sınıflandırılması daha kolay olur. Öte yandan özellik vektör uzayında, sınıf içindeki varyansın küçük ve sınıflar arası mesafenin büyük olması, özellik seçiminde hedeflenen amaçtır. Özelliklerin farklı sınıflara olan mesafe değerleri alınır ve en yakın sınıfa yerleştirilir. Bu aşamada farklı yöntemler de kullanılabilir. Bunlardan biri özelliklerin ayrı ayrı incelenmesi ve ayırt edici bilgisi düşük olan özelliğin çıkarılmasıdır. Diğerisi ise, bütün bu özelliklerin beraber incelenmesidir. Bazen özellik vektörüne lineer ya da lineer olmayan dönüşüm uygulamak daha iyi bir ayırt edici özellik elde edilmesini sağlayabilir [2].

2.3.1. Sıralı özellik seçimi

Özellik seçiminin en yaygın yöntemlerinden birisi sıralı özellik seçimidir. Bu yöntem iki bileşenden oluşur.

Bütün uygun özellik altkümelerini minimum yapmaya çalışan yöntem olan amaç fonksiyonu kriter olarak adlandırılır. Ortak kriter, regresyon modelleri için ortalama karesel hata ve sınıflandırma modelleri için ise yanlış sınıflama oranıdır.

Sıralı arama algoritması, kriter fonksiyonunu değerlendirirken aday altkümeden özellikler eklenir ya da çıkarılır. N özellikli veri kümesinin 2^N altkümelerinin hepsinde kriter değerinin tam karşılaştırması genellikle yapılamaz (N 'in büyüklüğüne ve amaç çağırma maliyetine bağlı olarak), sıralı arama aday kümeyi daima genişleterek yada küçültürken sadece tek bir yönde hareket eder [12].

Sıralı özellik seçimi iki değişik yönetime sahiptir:

- Sıralı ileri seçimde (SFS- Sequential Forward Selection); eklenen özellikler, kriter fonksiyonunu azaltmayana kadar sırasıyla boş aday kümeye özellikleri ekler.

Boş kümeden başlayarak sırasıyla x^+ özelliği daha önceden seçilmiş olan Y_k özellikleri ile birleştirildiğinde $J(Y_k + x^+)$ maksimum yapar.

Algoritma:

- $Y_0 = \{\emptyset\}$ boş kümeyle başla
- $x^+ = \operatorname{argmax} J(Y_k + x^+) \quad x \notin Y_k$
gelecek en iyi özelliği seç
- $Y_{k+1} = Y_k + x^+; k = k + 1$ güncelle
- ii. adıma git

- Sıralı geri seçimde (SBS - Sequential Backward Selection); çıkarılan özellikler, kriter fonksiyonunu arttırmaya kadar sırasıyla tüm aday kümeden özellikleri çıkarır.

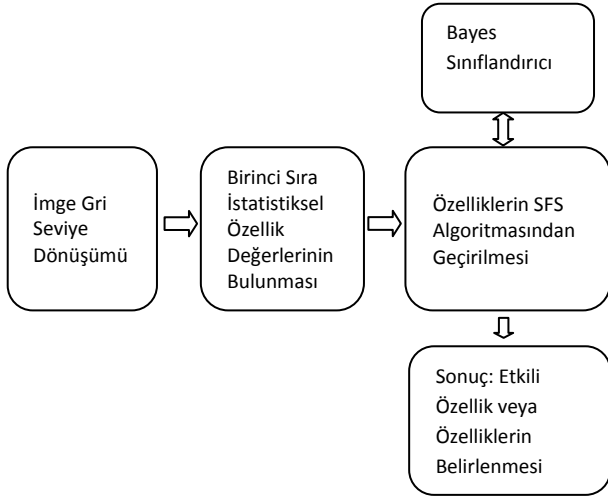
Tam kümeden başlayarak $J(Y + x^-)$ amaç fonksiyonunun değerini en aza indirmek için x^- özellikleri sırayla çıkarılır[13].

Algoritma:

- $Y_0 = X$ tam kümeyle başla
- $x^- = \operatorname{argmax} J(Y_k - x^-) \quad x^- \in Y_k$ en kötü özelliği çıkar
- $Y_{k+1} = Y_k - x^-; k = k + 1$ güncelle
- ii. adıma git

3. Önerilen yöntem

Şekil 1'de imgelerin sınıflandırılmasında kullanılacak olan etkili özellik seçimi yöntemi sunulmuştur. Sunulan yöntemin üç aşaması bulunmaktadır. Birinci aşama, histopatolojik imgelerin alınması ve imgelerin gri seviyeye dönüştürülmesi aşamasıdır. İkinci aşama, histopatolojik imgelerin birinci dereceden istatistiksel özellik değerlerinin elde edilmesi aşamasıdır. Son aşamada ise, belirlenen birinci dereceden istatistiksel özellik değerlerine göre SFS algoritması ile imgelerin etkili özellik veya özelliklerinin belirlenmesi aşamasıdır.



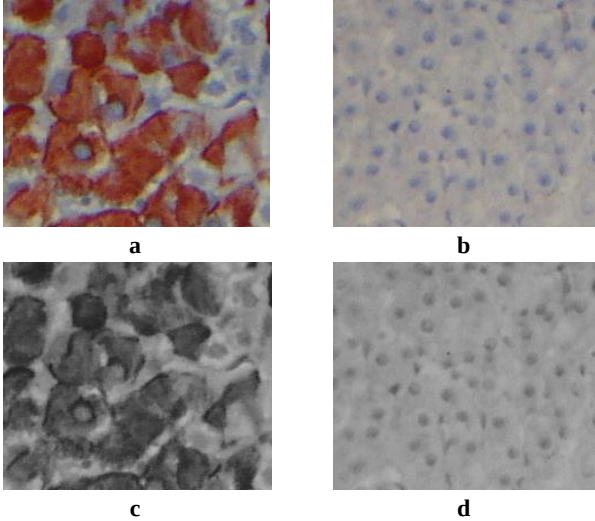
Şekil 1: İmgelerin sınıflandırılmasında etkili özellik seçim sistemi.

3.1. Histopatolojik imgelerin alınması

Uygulamada kullanılan histopatolojik imgeler, Fırat Üniversitesi, Fırat Tıp Merkezi, Patoloji bölümünden alınmıştır. Bu histopatolojik imge görüntüleri, hepatitli hastalardan alınan karaciğer dokularının mikroskop altında çekilen resimleridir. İmgelerin alınmasında Olympus BX-50 ışık mikroskobuna bağlı, Olympus C-4000Z sayısal fotoğraf makinesi kullanılmıştır. Alınan sayısal imgelerin çözünürlük değerleri 2288x1712 piksel ve 72 dpi dir. Çözünürlük değerinin yüksek olması bölütleme işleminin daha iyi olmasını sağlamaktadır.

3.2. Ön işlem süreci

İmge gri seviyeye dönüştürülür.



Şekil 2: a. Hepatitli imge. b. Hepatitsiz imge. c-d. Gri seviye dönüşümü yapılmış imgeler.

3.3. Birinci dereceden istatistiksel özellik değerlerinin bulunması

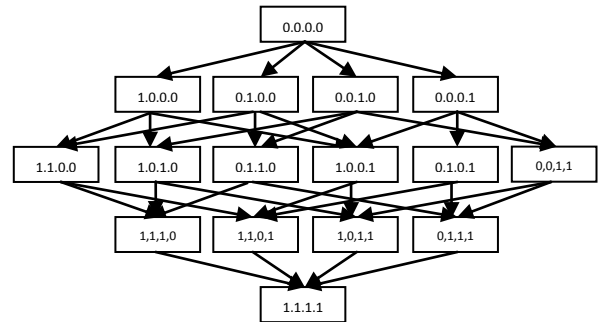
İmgenin birinci dereceden istatistiksel özelliklerinden olan beklenen değer, standart sapma, yamukluk ve basıklık değerleri elde edilir.

Çizelge 1: Bir imgenin birinci dereceden istatistiksel özellik değerleri.

Birinci dereceden istatistiksel özellikler	Hepatitli İmge	Hepatitsiz İmge
Beklenen Değer (μ)	14,4651	21,8535
Standart Sapma (σ)	6,3372	3,5235
Yamukluk(s)	0,1914	-1,2387
Basıklık(k)	2,0029	5,3645

3.4. Özelliklerin SFS algoritmasından geçirilmesi

İmgeden elde edilen birinci dereceden istatistiksel özellikleri kullanarak, imgenin sınıflandırılmasında kullanılacak olan hangi özelliğin etkili özellik olduğu belirlenir. Etkili özellik belirleme işlemi için sıralı ileri seçim algoritması kullanılır. N elemanlı bir kümede SFS algoritması, bütün özellikleri bir küme olarak kabul eder. SFS algoritması işleme boş küme ile başlar. Rastgele bir özellik seçilir. Seçilen bu özellik herhangi bir sınıflandırıcıdan geçirilerek etkinliği belirlenir. Bu çalışmada sınıflandırıcı olarak, Bayes kullanılmıştır. Bu işlem sonucunda en yüksek etkinliğe sahip özellik alınır. Süreç, bu özelliğin en yüksek etkinliğe sahip ikili, üçlü, dördü v.b küme bileşimlerini karşılaştırarak tam küme sayısına kadar devam eder. Şekil 3 de, SFS algoritmasının çalışma yapısı şematik olarak gösterilmiştir [13].



Şekil 3: Sıralı ileri seçim

Çizelge 2: Histopatolojik imgelerinin test değerlendirme sonuçları.

		Hepatitli	Hepatitsiz
Toplam Örnek Sayısı		26	26
Beklenen Değer(μ)	Doğru Sınıflandırma	23	22
	Yanlış Sınıflandırma	3	4
	Ortalama Doğruluk yüzdesi (%)	86,53	
Standart Sapma(σ)	Doğru Sınıflandırma	17	25
	Yanlış Sınıflandırma	9	1
	Ortalama Doğruluk yüzdesi (%)	80,76	
Yamukluk(s)	Doğru Sınıflandırma	24	14
	Yanlış Sınıflandırma	2	12
	Ortalama Doğruluk yüzdesi (%)	73,07	
Basıklık(k)	Doğru Sınıflandırma	21	26
	Yanlış Sınıflandırma	5	0
	Ortalama Doğruluk yüzdesi (%)	90,38	

4. Sonuç

Uygulanan SFS algoritmasıyla, histopatolojik imgeleri sınıflandırmada kullanılacak etkili özellik veya özelliklerin seçimi gerçekleştirilmiştir. Kullanılan birinci dereceden istatistiksel özelliklerden olan basıklığın, histopatolojik imgeleri sınıflandırmada etkili bir özellik olduğu görülmektedir. Kullanılan sıralı ileri seçim algoritması ile birden fazla özellik kullanılmaktansa, etkili özellik kullanmanın zamandan ve veri maliyetinden kazanç sağladığı gösterilmektedir. Kullanılan test verisinin büyüklük ve kalitesine bağlı olarak elde edilen etkili özellik sayısı değişiklik gösterebilir. Farklı imgeleri sınıflandırma işleminde, etkili özellikler bulunurken sadece bir özelliğin değil, bazen birden çok özelliğinde etkili özellik olarak kullanılabileceği görülmektedir. Kullanılan birinci dereceden istatistiksel özellikler, imgelerin gri seviye değerleri hakkında bilgi vermektedir. Uygulanan bu sistem ile histopatolojik imgelerin değerlendirilmesinin yanı sıra, tıbbi alanda benzer imgelerin (mikrobiyoloji, biyokimya, v.s.) analizinde de ikinci dereceden istatistiksel özellikler ve farklı özellik seçim algoritmaları (sıralı geri seçim, tam arama) kullanılarak daha etkili özellik seçimi ve sınıflandırma işlemi yapılmasına alt yapı hazırlanmıştır.

Teşekkür

Histopatolojik imgeleri bize sağlayarak değerlendirmesine yardımcı olduğu için Prof. Dr. İbrahim ÖZERCAN'a (Fırat Üniversitesi, Fırat Tıp Merkezi, Patoloji Bölümü) teşekkür

ederiz. Bu çalışma Fırat Üniversitesi tarafından desteklenmiştir (Proje No: FÜBAP-985).

5. Kaynaklar

- [1] Toraman, S., ve Türkoğlu, İ., "Histopatolojik İmgelerin Akıllı Yapılar İle Bölütlenmesi", *ASYU-Akıllı Sistemlerde Yenilikler ve Uygulamaları Sempozyumu*, 81-84, İstanbul, 2006.
- [2] Theodoridis, S., Koutroumbas, K., *Pattern Recognition*, Academic Press- Elsevier, California, 2009.
- [3] Boshu, L., Sujun, L., Yinglin, W., Lin, L., Yixue L., Yudong, C., "Predicting the protein SUMO modification sites based on Properties Sequential Forward Selection (PSFS)", *Biochemical and Biophysical Research Communications*, 358., 136-139, 2007.
- [4] Noori, R., Karbassi, A.R., Moghaddamnia, A., Han, D., Zokaei-Ashtiani, M.H., Farokhnia, A., Ghafari Gousheh, M., "Assessment of input variables determination on the SVM model performance using PCA, Gamma test, and forward selection techniques for monthly stream flow prediction", *Journal of Hydrology*, 401., 177-189, 2011.
- [5] Casado Yusta, S., "Different metaheuristic strategies to solve the feature selection problem", *Pattern Recognition Letters*, 30., 525-534, 2009.
- [6] Romero, E., Alquézar, R., "Comparing error minimized extreme learning machines and support vector sequential feed-forward neural networks", *Neural Networks*, 25., 122-129, 2012.
- [7] Gunal, S., Gerek, O. N., Ece, D. G., Edizkan, R., "The search for optimal feature set in power quality event classification", *Expert Systems with Applications*, 36., 10266-10273, 2009.
- [8] Yenerman M., *Genel Patoloji*, Nobel Tıp Kitabevleri, İstanbul, 1994.
- [9] Türkoğlu, İ., ve Toraman, S., "Karar Ağaçları ve Fraktal Analiz Kullanarak Histopatolojik İmgelerin Sınıflandırılması", *Gazi Üniversitesi Mühendislik-Mimarlık Fakültesi Dergisi*, 22 (4), 753 - 758, Aralık-2007.
- [10] Giansanti R., Fumelli P., Passerini G., et al., "Imaging system for retinal change evaluation", *IPA97 Conference*, Pub.No. 443, 1997.
- [11] Martinez, W. L., Martinez, A. R., *Computational Statistics Handbook with MATLAB*, CHAPMAN & HALL/CRC, Florida, 2002.
- [12] <http://www.mathworks.com>
- [13] http://courses.cs.tamu.edu/rgutier/csce666_f11