

Histopatolojik Görüntülerde Uzamsal Öznitelikler Vasıtasıyla Meme Kanseri Tespiti ve Derecelendirilmesi

Detection and Grading of Breast Cancer via Spatial Features in Histopathological Images

M. Emin Bağdigen ve Gokhan Bilgin
Yıldız Teknik Üniversitesi
Bilgisayar Mühendisliği Bölümü
34220 İstanbul, Türkiye
emin@mavis.com.tr, gbilgin@yildiz.edu.tr

Özetçe —Bu çalışmada, çok yaygın bir kanser türü olan meme kanserinin tespit edilmesi ve derecelendirilmesi üzerine çalışılmıştır. Etiketli örnek sayısı 4812 olan eğitim verisi üzerinde yerel ikili örüntü histogramları çıkarılarak ve Gabor filtreleri kullanılarak öznitelik matrisleri oluşturulmuştur. Sonrasında ise, k-en yakın komşu, karar ağacı ve topluluk öğrenmesi yöntemleri olan bagging, adaboost, rastgele ormanlar gibi öğreticili sınıflandırma yöntemleri kullanılarak sınıflandırma işlemi gerçekleştirilmiştir. Test verisi olarak ayrıştırılan 491 adet görüntü ile yapılan sınıflandırmaların başarıları hesaplanmıştır. Çalışmanın sonunda öznitelik çıkarma yöntemleri ve sınıflandırıcıların başarıları karşılaştırılmış, elde edilen sonuçlar tablolar halinde sunulmuştur.

Anahtar Kelimeler—Histopatolojik görüntüler, öznitelik çıkarımı, yerel ikili örüntüler, Gabor filtreleri, sınıflandırma.

Abstract—In this study, the detection and grading of breast cancer, which is a very common type of cancer, are studied. In feature extraction phase, feature matrices are created by using local binary pattern histograms and Gabor filters on a dataset with 4812 train data. Afterwards, the classification process are carried out using k-nearest neighbor, decision trees and several ensemble classification methods such as bagging, adaboost, random forests. The accuracies of the classifications are measured with 491 test image which are selected from dataset. At the end of the study, success of feature extraction methods and classifiers are compared and the results were presented in tables.

Keywords—Histopathological images, feature extraction, local binary patterns, Gabor filters, classification.

I. GİRİŞ

Günümüzde sağlık açısından en büyük ve yaygın problemlerden biri kanserdir. Yapılan istatistiksel bir çalışmaya göre, 2019 yılında, Amerikada 1,762,450 yeni kanser vakası tespit edileceği öngörülmüştür [1]. Aynı yıl içerisinde de ortalama günlük olarak 1,700 kişi olmak üzere toplam 606,808 kişinin kanser sebebi ile hayatını kaybedeceği beklenmektedir. 2030

yılında ise hayatını kanser sebebi ile kaybedecek insan sayısı 12,000,000 olacağı öngörülmektedir. Kadınlarda bu vakaların yaklaşık olarak %30'unu da meme kanseri vakaları oluşturmaktadır. Yapılan diğer bir çalışmada Dünya üzerinde 3,500,000 kadınının meme kanseri olduğu tespit edilmiştir. Verilen bu istatistiksel bilgilerin de gösterdiği üzere kanser büyük bir problem haline gelmiştir [2].

Kadınlarda en çok görülen kanser türü olan meme kanseri hastalarında, erken teşhis ve başarılı tedavi sonucu hayatta kalma oranı %90'lara varmaktadır. 2016 yılında yapılan araştırmaya göre erken teşhis ve tedavi sonrasında toplam 3,560,570 kanser hastasının iyileştiği tespit edilmiştir. Erken teşhis yöntemlerinin ve teknolojinin gelişmesi gibi sebeplerden dolayı 2026 yılında iyileşecek kanser hastası sayısının 4,571,210'a kadar yükseleceği öngörülmektedir [3].

Görüntüleme tekniklerinin gelişmesi ve görüntülerin bilgisayar ortamına kolay ve maliyeti çok düşük bir şekilde aktarılabilmesi, bilgisayarın bu alanda kullanılabilirliğini artırmıştır [4]. Erken tedavinin de çok büyük önem arzettiği kanser vakalarında kanser tespiti yapabilecek bilgisayar destekli sistemler üzerine çalışmalar yapılmaya başlanmıştır [5]. Bu çalışmalar sonucunda da bilgisayar destekli tespit ya da teşhis sistemleri, görüntülerdeki anormal durumları tespit edebilir düzeye gelmiş, kanser evreleri ile ilgili sınıflandırmalar yapıp, erken teşhis etme konusunda kilit rol alacak duruma gelmişlerdir [6]. Bu çalışmalar çok daha fazla araştırmacının dikkatini çekmiş, bu konuda çalışan araştırmacı sayısı artmıştır [7]–[9].

Bilgisayar destekli sistemlerle ilgili literatürde daha önce birçok çalışma yapılmıştır. Son yıllarda popüler konu haline gelen Derin Öğrenme yöntemini kullanarak meme kanseri için histopatolojik görüntülerde sınıflandırmalar yapılmış ve çalışma sonucunda başarı oranı %91'e varan sonuçlar elde edilmiştir [10]. Yapılan bir çalışmada da elde edilen histopatolojik görüntülerde sınıflandırma yapmak için evrimsel sinir ağları (ESA) kullanılmış, yaklaşık olarak %90 başarı oranı sağlanmıştır [11]. Başka bir çalışmada da özellik çıkarma yöntemlerinden yerel ikili örüntü (YİÖ), bileşik yerel ikili örüntü

(BYİÖ) ve gri-seviyeli eş-oluşum matrisleri (GSEOM) kullanılarak öz nitelikler çıkarılmış, k-en yakın komşu (k-EYK), destek vektör makinesi (DVM) ve rastgele ormanlar (RO) gibi sınıflandırma algoritmaları ile sınıflandırmalar yapılmış, %98-99'luk başarı oranları sağlanmıştır [12], [13]. Diğer bir çalışmada yine ESA kullanılarak %82 başarı oranları elde edilmiştir [14]. Bu alandaki başka bir çalışmada ise, doku tabanlı öz nitelikler çıkarılarak, DVM yöntemi ile kanser ve kanser olmayan doku olmak üzere sınıflandırma yapılmış, daha sonra yapılan sınıflandırmaya göre de kanser tespit edilmiş ise yüksek ve düşük olmak üzere kanser seviye tespiti de yapılmıştır. Kanser tespitinde %93 başarı oranı elde edilirken, kanser seviyesi tespitinde %90 başarı oranı elde edilmiştir [15].

II. KULLANILAN ÖZ NİTELİK ÇIKARIM YÖNTEMLERİ

Çalışmada görüntüler üzerindeki dokusal farklılıkları ortaya çıkaran öz nitelikler çıkarılması amacıyla yerel ikili örüntüler ve Gabor filtrelerinden yararlanılmıştır.

A. Yerel İkili Örüntü

YİÖ, ilk olarak 1994 yılında sunulmuştur [16]. Bu algoritma, gri seviye değerlerinden bağımsız bir şekilde, dokusal özellik çıkarılmasını sağlamaktadır. Algoritmada, görüntüdeki her bir piksel için, komşu piksel değerleriyle karşılaştırma sonucu bir değer elde edilmektedir. Örnekl olarak Şekil 1'de gösterilen görüntüde elde edilen değer, ikili kod olarak 01101101; onluk tabanında ise 109'dur. Bu şekilde tüm görüntüdeki pikseller için bulunan değerlerle oluşturulan histogram değerleri, görüntüye ait doku tabanlı öz nitelik matrisini ifade etmektedir. Bu algoritma, parlaklık değerinin değişmesine karşı da dayanıklı olmasının yanında hesaplama hızı da yüksek olan bir algoritmadır [17], [18].

154	161	159
165	157	140
153	158	170

⇒

0	1	1
1		0
0	1	1

Şekil 1: Yerel ikili örüntü hesaplamasına dair örnek

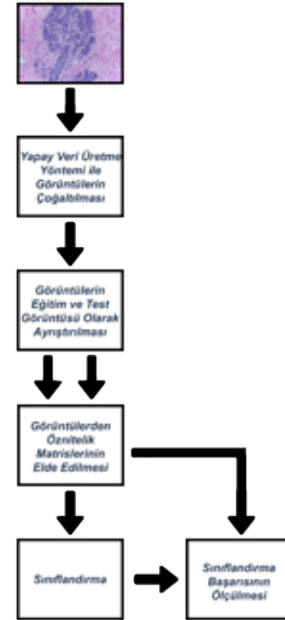
Algoritmada, değeri hesaplanacak pikselin komşu sayısı ve örneklem yarıçapı değiştirilerek, dokusal farklılığın en fazla olacağı değerler seçilebilmektedir [19].

B. Gabor Filtreleri

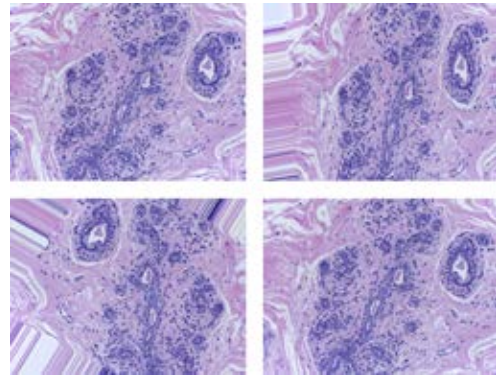
Gabor filtresi, doku tanıma ve sınıflandırma, kenar algılama gibi problemler için kullanılabilir doğrusal bir filtredir. İki boyutlu Gabor filtreleri, memeli görsel korteksine çok benzer olmasından dolayı, dokusal farklılıkları ayırt etmesi konusunda oldukça uygundur. Farklı açı ve frekanslarla oluşturulacak Gabor filtreleri [20] ile istenilen desene ait etkili ayırt edici özellikler, başarılı bir şekilde elde edilebilmektedir [21], [22]. Bu sebepten dolayı, dokusal sınıflandırma, plaka tespiti, yüz tanıma, parmak izi tanıma gibi uygulamalarda yoğun olarak kullanılmaktadır.

III. DENEYSEL SONUÇLAR

Bu çalışmada dört farklı kanser evresine sahip histopatolojik görüntülerden oluşan, meme kanseri hastalığına ait görüntüler üzerinde çalışılmıştır [14]. Çalışmanın akış şeması Şekil 2'de gösterilmiştir. Bioimaging Challenge 2015 Breast Histology [14] veri kümesine'ne ait görüntüler üzerinde öncelikle yapay veri üretimi yapılarak veri kümesinde bulunan görüntü sayısı artırılmıştır. Orjinal görüntüden üretilen yapay verilere ait görüntüler, Şekil 3'de gösterilmiştir. Oluşturulan veriseti, %90 eğitim, %10 test verisi olacak şekilde ayrıştırılmış, çalışma öncesi elde edilen yapay verilerle birlikte oluşturulan veriseteine ait sayısal veriler de Tablo 1'de gösterilmiştir. Yapay veri üretme işlemi Keras Kütüphanesi kullanılarak yapılmıştır.



Şekil 2: Önerilen çalışmanın akış şeması



Şekil 3: Üretilen yapay görüntü verisi örnekleri

Çalışmada ilk yöntem olarak, dokusal farklılıkları ortaya koyan, YİÖ kullanılmıştır. Algoritmada, örneklem yarıçapı 3 seçilirken, komşu sayısı 24 seçilmiştir. Algoritma sonu-

	Orjinal Görüntüler	Eğitim Görüntüleri	Test Görüntüleri
Benign	69	1221	130
In Situ	63	1252	127
Invasive	62	1236	124
Normal	55	1103	110

Tablo I: Veri kümesine ait sayısal veriler

cunda her bir görüntü için 26 adet veri içeren histogram elde edilmiştir. İkinci yöntem olarak Tablo II'deki parametreler kullanılarak Gabor filtreleri oluşturulmuştur. Oluşturulan bu Gabor filtreleri, görüntülere uygulanmış ve elde edilen sonuç matrislerinin ortalama değerleri ve değişinti değerleri hesaplanmıştır. Böylece bir görüntü üzerine uygulanan bir filtreden, iki adet öznelik çıkarımı yapılmıştır. Tablo II'te gösterilen parametreler kullanılarak oluşturulan 16 farklı filtre kullanılarak, her bir görüntü için toplamda 32 öznelik çıkarılmıştır. Üçüncü yöntem olarak, kullanılan öznelik çıkarma yöntemleri sonucu elde edilen öznelik matrisleri birleştirilmiş ve yeni bir öznelik matrisi oluşturulmuştur. Böylece her bir görüntü için 58 adet öznelik çıkarımı yapılmıştır. Üç farklı yöntemle çıkarılan öznelik matris boyutları Tablo III'te gösterilmiştir. Kullanılan öznelik çıkarma yöntemleri, Scikit-Learn Kütüphanesi kullanılarak gerçekleştirilmiştir.

theta	sigma	frekans
0	1	0.05
		0.25
	3	0.05
		0.25
0.785	1	0.05
		0.25
	3	0.05
		0.25
1.570	1	0.05
		0.25
	3	0.05
		0.25
2.356	1	0.05
		0.25
	3	0.05
		0.25

Tablo II: Farklı Gabor filtreleri oluşturmak için kullanılan parametreler

	Matris Boyutu
GABOR	4812 × 32
LBP	4812 × 26
GABOR + LBP	4812 × 58

Tablo III: Oluşturulan öznelik matris boyutları

Oluşturulan öznelik matrisleri ile eğitici sınıflandırma yöntemlerinden olan k-EYK, Karar Ağaçları (KA) ve seçilen Topluluk Öğrenmesi (TÖ) yöntemleri (Bagging, Adaboost, RO) ile sınıflandırmalar yapılmıştır. Test verisi olarak ayrıştırılan görüntüler kullanılarak öznelik çıkarma yöntemlerinin başarıları ve sınıflandırma başarıları ölçülmüştür.

Yapılan çalışmada eğitim verisi olarak ayrıştırılan görüntüler ile eğitici sınıflandırma işlemi yapılmış ve test verisi olarak ayrıştırılan görüntüler kullanılarak sınıflandırmaların başarıları test edilmiştir. Elde edilen sonuçlara ait, kesinlik (precision), hassasiyet (recall) ve f1-skoru (f1-measure)

değeri hesaplanmıştır. Oluşturulan üç farklı öznelik matrisi ile de sınıflandırma yöntemleri uygulanmış ve başarı sonuçları alınmıştır. Alınan sonuçlar k-EYK için Tablo IV'te; KA için Tablo V'te; Bagging için Tablo VI'da; Adaboost için Tablo VII'de ve RO için Tablo VIII'de sırasıyla sunulmuştur.

GABOR			
n_neighbors	kesinlik	hassasiyet	f1-skoru
5	0.98	0.98	0.98
10	0.96	0.96	0.96
20	0.93	0.93	0.93
50	0.71	0.7	0.71
100	0.61	0.61	0.61
LBP			
n_neighbors	kesinlik	hassasiyet	f1-skoru
5	0.95	0.95	0.95
10	0.89	0.89	0.89
20	0.82	0.84	0.84
50	0.68	0.67	0.67
100	0.65	0.62	0.62
GABOR + LBP			
n_neighbors	kesinlik	hassasiyet	f1-skoru
5	0.97	0.97	0.97
10	0.95	0.95	0.95
20	0.92	0.92	0.92
50	0.71	0.7	0.7
100	0.61	0.61	0.6

Tablo IV: k-EYK sınıflandırma başarıları

	kesinlik	hassasiyet	f1-skoru
GABOR	0.95	0.95	0.94
LBP	0.92	0.92	0.92
GABOR + LBP	0.96	0.96	0.96

Tablo V: KA sınıflandırma başarıları

GABOR			
n_estimators	kesinlik	hassasiyet	f1-skoru
0.2	0.92	0.92	0.92
0.4	0.96	0.96	0.96
0.6	0.97	0.97	0.97
0.8	0.98	0.98	0.98
LBP			
n_estimators	kesinlik	hassasiyet	f1-skoru
5	0.83	0.82	0.82
10	0.92	0.92	0.92
20	0.98	0.98	0.98
50	0.99	0.99	0.99
GABOR + LBP			
n_estimators	kesinlik	hassasiyet	f1-skoru
5	0.93	0.92	0.92
10	0.96	0.96	0.96
20	0.96	0.96	0.96
50	0.67	0.67	0.97

Tablo VI: Bagging sınıflandırma başarıları

IV. SONUÇ

Yapılan çalışmada eğitim verisi üzerinden, üç farklı yöntem ile çıkarılan öznelikler vasıtasıyla, farklı sınıflandırma yöntemleri de kullanılarak sınıflandırmalar yapılmıştır. RO algoritması ile, farklı yöntemlerle oluşturulan üç farklı öznelik matrisiyle de başarılı ve birbirine yakın sonuçlar elde edilmiştir. Bu durumun tam tersi olarak da Adaboost algoritması ile, çıkarılan hiçbir öznelik matrisi ile başarılı sonuç elde

GABOR			
n_estimators	kesinlik	hassasiyet	f1-skoru
5	0.38	0.48	0.42
10	0.57	0.51	0.47
20	0.58	0.57	0.56
50	0.61	0.59	0.59
100	0.59	0.59	0.59
LBP			
n_estimators	kesinlik	hassasiyet	f1-skoru
5	0.49	0.48	0.48
10	0.58	0.55	0.55
20	0.56	0.55	0.55
50	0.53	0.53	0.53
100	0.60	0.59	0.59
GABOR + LBP			
n_estimators	kesinlik	hassasiyet	f1-skoru
5	0.57	0.54	0.54
10	0.54	0.53	0.52
20	0.63	0.63	0.62
50	0.65	0.63	0.63
100	0.6	0.61	0.6

Tablo VII: Adaboost sınıflandırma başarıları

GABOR			
n_estimators	kesinlik	hassasiyet	f1-skoru
5	0.98	0.98	0.98
10	0.98	0.98	0.98
20	0.98	0.98	0.98
50	0.98	0.98	0.98
100	0.98	0.98	0.98
LBP			
n_estimators	kesinlik	hassasiyet	f1-skoru
5	0.95	0.95	0.95
10	0.97	0.97	0.97
20	0.98	0.98	0.98
50	0.98	0.98	0.98
100	0.98	0.98	0.98
GABOR + LBP			
n_estimators	kesinlik	hassasiyet	f1-skoru
5	0.97	0.97	0.97
10	0.97	0.97	0.97
20	0.98	0.98	0.98
50	0.97	0.97	0.97
100	0.98	0.98	0.98

Tablo VIII: RO sınıflandırma başarıları

edilememiştir. Kullanılan k-EYK algoritmasında ise bakılan komşu sayısı 10'un altında olduğu sürece başarılı sonuçlar elde edilirken komşuluk sayısının 20'nin üzerindeki değerlerde başarısız sonuçlar alınmıştır. Genel anlamda en başarılı sınıflandırıcı Bagging algoritması olmasa da, YİÖ özelliklerini kullanarak ve 50 örneklem üzerinden alınan kesinlik, hassasiyet ve f1-skoru açısından 0.99 başarı oranı ile çalışmadaki en iyi sonuçlar elde edilmiştir. Sonuç olarak, sınıflandırma yöntemine bağlı olarak, kullanılan tüm uzamsal öznitelik çıkarma yöntemlerinin başarılı olduğu görülmüştür.

KAYNAKÇA

- [1] R. L. Siegel, K. D. Miller, and A. Jemal, "Cancer statistics, 2019," *CA: A Cancer Journal for Clinicians*, vol. 69, no. 1, pp. 7–34, 2019.
- [2] C. E. DeSantis, J. Ma, A. Goding Sauer, L. A. Newman, and A. Jemal, "Breast cancer statistics, 2017, racial disparity in mortality by state," *CA: A Cancer Journal for Clinicians*, vol. 67, no. 6, pp. 439–448, 2017.
- [3] K. D. Miller *et al.*, "Cancer treatment and survivorship statistics, 2016," *CA: A Cancer Journal for Clinicians*, vol. 66, no. 4, pp. 271–289, 2016.
- [4] A. Albayrak and G. Bilgin, "Automatic cell segmentation in histopathological images via two-staged superpixel-based algorithms," *Medical & Biological Engineering & Computing*, vol. 57, no. 3, pp. 653–665, 2019.
- [5] J. Tang, R. M. Rangayyan, J. Xu, I. El Naqa, and Y. Yang, "Computer-aided detection and diagnosis of breast cancer with mammography: recent advances," *IEEE Transactions on Information Technology in Biomedicine*, vol. 13, no. 2, pp. 236–251, 2009.
- [6] M. P. Sampat, M. K. Markey, A. C. Bovik, *et al.*, "Computer-aided detection and diagnosis in mammography," *Handbook of Image and Video Processing*, vol. 2, no. 1, pp. 1195–1217, 2005.
- [7] R. M. Rangayyan, F. J. Ayres, and J. L. Desautels, "A review of computer-aided diagnosis of breast cancer: Toward the detection of subtle signs," *Journal of the Franklin Institute*, vol. 344, no. 3-4, pp. 312–348, 2007.
- [8] M. L. Giger, "Computer-aided diagnosis of breast lesions in medical images," *Computing in Science & Engineering*, vol. 2, no. 5, p. 39, 2000.
- [9] J. Tang, R. M. Rangayyan, J. Xu, I. El Naqa, and Y. Yang, "Computer-aided detection and diagnosis of breast cancer with mammography: Recent advances," *IEEE Trans. on Information Technology in Biomedicine*, vol. 13, no. 2, pp. 236–251, 2009.
- [10] N. Bayramoglu, J. Kannala, and J. Heikkilä, "Deep learning for magnification independent breast cancer histopathology image classification," in *2016 23rd International conference on pattern recognition, ICPR'16*. IEEE, 2016, pp. 2440–2445.
- [11] N. Hatipoglu and G. Bilgin, "Cell segmentation in histopathological images with deep learning algorithms by utilizing spatial relationships," *Medical & Biological Engineering & Computing*, vol. 55, no. 10, pp. 1829–1848, 2017.
- [12] I. O. Sigirci, A. Albayrak, and G. Bilgin, "Detection of mitotic cells using completed local binary pattern in histopathological images," in *IEEE 23rd Signal Processing and Communications Applications Conference, SIU'15*, 2015, pp. 1078–1081.
- [13] F. A. Spanhol, L. S. Oliveira, C. Petitjean, and L. Heutte, "A dataset for breast cancer histopathological image classification," *IEEE Transactions on Biomedical Engineering*, vol. 63, no. 7, pp. 1455–1462, 2015.
- [14] T. Araújo, G. Aresta, E. Castro, J. Rouco, P. Aguiar, C. Eloy, A. Polónia, and A. Campilho, "Classification of breast cancer histology images using convolutional neural networks," *PloS One*, vol. 12, no. 6, p. e0177544, 2017.
- [15] S. Doyle, S. Agner, A. Madabhushi, M. Feldman, and J. Tomaszewski, "Automated grading of breast cancer histopathology using spectral clustering with textural and architectural image features," in *IEEE 5th Int. Symp. on Biomedical Imaging: From Nano to Macro*. IEEE, 2008, pp. 496–499.
- [16] T. Ojala, M. Pietikainen, and D. Harwood, "Performance evaluation of texture measures with classification based on kullback discrimination of distributions," in *IEEE Proc. of 12th Int. Conf. on Pattern Recognition*, vol. 1. IEEE, 1994, pp. 582–585.
- [17] G. Zhao, T. Ahonen, J. Matas, and M. Pietikainen, "Rotation-invariant image and video description with local binary pattern features," *IEEE Transactions on Image Processing*, vol. 21, no. 4, pp. 1465–1477, 2011.
- [18] M. D. Kumar, M. Babaie, S. Zhu, S. Kalra, and H. R. Tizhoosh, "A comparative study of cnn, bow and lbp for classification of histopathological images," in *IEEE Symposium Series on Computational Intelligence*. IEEE, 2017, pp. 1–7.
- [19] T. Ojala, M. Pietikainen, and T. Maenpää, "Multiresolution gray-scale and rotation invariant texture classification with local binary patterns," *IEEE Trans. on Pattern Analysis & Machine Intelligence*, no. 7, pp. 971–987, 2002.
- [20] G. H. Granlund, "In search of a general picture processing operator," *Computer Graphics and Image Processing*, vol. 8, no. 2, pp. 155–173, 1978.
- [21] F. Riaz, A. Hassan, and S. Rehman, "Texture classification framework using gabor filters and local binary patterns," in *Science and Information Conference*. Springer, 2018, pp. 569–580.
- [22] Y. Chen, L. Zhu, P. Ghamisi, X. Jia, G. Li, and L. Tang, "Hyperspectral images classification with gabor filtering and convolutional neural network," *IEEE Geoscience and Remote Sensing Letters*, vol. 14, no. 12, pp. 2355–2359, 2017.