

Meme Kanseri Tanısı İçin Özniteliklerin Öneminin Değerlendirilmesi Üzerine Bir Çalışma

Kemal AKYOL(0000-0002-2272-5243)

Kastamonu Üniversitesi Mühendislik ve Mimarlık Fakültesi Bilgisayar Mühendisliği Bölümü, kakyol@kastamonu.edu.tr

Geliş Tarihi: 23.06.2017

Kabul Tarihi: 31.05.2018

Öz

En yaygın kanser türlerinden biri olan meme kanseri kadınları etkileyen ölümcül bir hastalıktır. Önerilen çalışmada, Wisconsin meme kanseri veriseti üzerinde öznitelik seçimine dayalı Özyinelemeli Özellik Seçimi metodu kullanılarak özniteliklerin önemliliği araştırılmış ve sonrasında Rastgele Orman ve Lojistik Regresyon sınıflandırıcı algoritmaları kullanılarak makine öğrenmeleri gerçekleştirilmiştir. Eğitim ve test aşamalarını içeren öğrenme süreci 5 katlı çapraz doğrulama tekniği kullanılarak gerçekleştirilmiştir. Deneysel çalışmalar, Rastgele Orman algoritması kullanılarak en iyi sınıflandırma başarısı (%98 doğruluk) elde edildiğini göstermiştir.

Anahtar Kelimeler: Meme kanseri risk tahmini, özniteliklerin önemliliği, Özyinelemeli Özellik Eleme, Rastgele Orman, Lojistik Regresyon.

A Study on Assessing the Importance of Attributes for Breast Cancer Diagnosis

Kemal AKYOL

Kastamonu University Faculty of Engineering and Architecture Computer Engineering Department

Abstract

Breast cancer, one of the most common types of cancer, is a deadly disease affecting women. The importance of attributes was investigated by using the Recursive Feature Selection based on feature selection on Wisconsin breast cancer dataset, and then the machine learnings were performed by utilizing Random Forest and Logistic Regression classifier algorithms in the proposed study. The learning process involving training and testing phases was performed by utilizing the 5-fold cross-validation technique. Experimental studies showed that the best classification performance (98% accuracy) was achieved by applying the Random Forest algorithm.

Keywords: Logistic Regression, Breast cancer risk prediction, importance of attributes, Recursive Feature Elimination, Random Forest.

1. GİRİŞ

Meme dokusundaki süt kanalları veya bezlerinde gelişen [1] ve kadın nüfusunu etkileyen yaygın kanser tiplerinden bir tanesi olarak bilinen ve dünya genelinde her yıl 1.38 milyon yeni vakanın tahmin edildiği bir hastalıktır [2]. Erken dönemde meme kanseri tespiti ve teşhisi, başarılı tedavinin ve hastanın iyileşmesinin başarı şansını artırır. Pürüzsüz ve iyi tanımlanmış bir sınıra sahip kitle iyi huylu olabilir. Ancak, düzensiz bir sınıra sahip kitle kanserli olabilir [3].

Özyinelemeli Özellik Eleme (ÖÖE, Recursive Feature Elimination [4]) ile özniteliklerin öneminin araştırıldığı ve sonrasında sırasıyla [5-6]'da tanıtımları yapılan Lojistik

Regresyon (LR) ve Rastgele Orman (RO) sınıflandırıcı algoritmaları ile eğitim ve test aşamaları yürütülerek sınıflandırma sürecinin tamamlandığı bu çalışmada meme kanseri hastalığı için önemli olan özniteliklerin araştırması yapılmıştır. Üzerinde çalışılan veriseti, Wisconsin Meme Kanseri Verisetidir (WMKV). Odaklanılan yöntemler çerçevesinde bu çalışmanın hibrid bir yaklaşım içerdiği ve %98 civarında doğru sınıflandırma oranı ile başarılı bir performans sergilediği ifade edilebilir.

Çalışmanın ilerleyen bölümleri şu şekildedir: 2. bölümde önceki çalışmalara yer verilmiştir. 3. bölümde, üzerinde çalışılan veriseti hakkında detaylı bilgi verilmiştir. 4. bölümde deneysel sonuçlar ayrıntılı bir şekilde

*Sorumlu Yazar: Kastamonu Üniversitesi Mühendislik ve Mimarlık Fakültesi Bilgisayar Mühendisliği Bölümü, kakyol@kastamonu.edu.tr

sunulmuştur. Son bölümde ise, çalışmanın sonuçlarına ve alan uzmanının faydalanabileceği bilgilere yer verilmiştir.

2. ÖNCEKİ ÇALIŞMALAR

Meme kanseri her yıl milyonlarca hastada sıklıkla görülen bir hastalıktır [7]. Ve bu hastalığa dair yüksek risk içeren kadınların tespiti için öne sürülmüş birçok çalışma vardır. Powell ve ark [8], Kaliforniya eyaleti Marin ilçesindeki yüksek oranda meme kanseri, nulliparite ve gecikmiş doğum bilgisi içeren kadın topluluğu verilerini kullanarak Breast Cancer Risk Assessment Tool [9], BRCAPRO [10] ve International Breast Intervention Study [11] meme kanseri risk değerlendirme modellerini karşılaştırmışlardır. Al-Diab ve ark, sundukları çalışmada, Sudi Arabistandaki meme kanserinin tüm kanserli kadın hastaların içinde %19.8 bir orana sahip olduğunu ortaya koymuşlardır [12].

Alharbi ve Tchier, bulanık tabanlı sistemler ve evrimsel genetik algoritmalar kullanarak meme kanserinin erken tanısı için hekimlere yardımcı olacak otomatik bilgisayar tanı sistemi üzerine odaklanmışlardır. Hibrid bir çalışmayı içeren çalışmalarında %97 başarımın elde edildiği optimize edilmiş yüksek performanslı bir tanı sistemi gerçekleştirmişlerdir [13].

Porto-Mascarenhas ve ark, tükürük biomarkerlarının meme kanserini karakterize etmekte yardımcı olabilecekleri düşüncesi doğrultusunda bu biomarkerların yeteneklerini değerlendirmek için gerçekleştirdikleri çalışmada, tükürük biomarkerlarının meme kanserinin erken evreden ziyade ileri aşamalarını tanımladıklarını görmüşlerdir [14].

Ramalho ve ark, meme kanseri hastalarındaki kanser tedavileri ile bilişsel olay bozuklukları arasındaki ilişkiyi, tedaviden önceki kaygı düzeylerini hesaba katarak değerlendirmeyi amaçlayan bir yaklaşım sunmuşlardır. Çalışmalarında, yalnızca başlangıçta kaygısı olmayan kadınlar üzerinde kimyasal tedavi ve bilişsel bozukluk arasındaki ilişkinin istatistiksel olarak önemli olduğunu ortaya koymuşlardır [15].

Peters ve ark, çalışmalarında özellikle erken evre yumurtalık kanseri teşhisi konan BRCA1/2 taşıyıcıları için düzenli meme kanseri verilerinin toplanmasını ve ilaç

ve/veya profilaktik mastektomi yani iki memenin de ameliyatla alınması yöntemleri ile risk azaltmayı önermişlerdir [16].

Pudkasam ve ark, egzersiz programlarının meme kanserini iyileştiren yağ(adipozite) ve immünolojik belirteçler üzerindeki bilinen etkilerini araştırmışlardır [17].

Cimpean ve ark, çalışmalarında Mast hücrelerinin meme kanserinin adjuvan tedavisi için bir önemi olabileceği üzerine çalışmışlardır. Mast hücreleri, belirli tümörleri çevreleyen stromada birikirler ve tümörün çevresinde oluşan enflamasyon reaksiyonuna katılırlar. Meme kanseri de dâhil olmak üzere katı tümörlerle ilişkili anjiyojenezde bu hücrelerden artmış sayıda vardır [18].

Almutlaq ve ark, yaş, ilk doğum yaşı, erken menarş, cinsiyet, beslenme faktörleri, tütün içilmesi, alkol tüketimi, düşük doz radyasyon, obezite, fiziksel aktivite, emzirme, hormonal, hormon replasman tedavisi, steroid hormon reseptörleri, mamografik yoğunluk, iyi huylu meme hastalığı ve genetik faktörlerini inceleyerek Suudi Arabistan'daki meme kanserine neden olan faktörleri araştırmışlardır [19].

Lukong, tüm dünyada ciddi bir sağlık sorunu olan bu hastalığın erken tanısı, tedavisi ve önlenmesi için daha fazla hedefin karakterize edildiği ve son derece yenilikçi yaklaşımların geliştirildiği daha özel tedavileri içeren çalışma üzerine yoğunlaşmışlardır [20].

Coleman, yetkin klinik ve kendi kendine meme muayenesinin dikkatli bir kullanımı ile küçük tümörlerin tespitini ve morbiditenin azalmasını sağlayacağı yönünde bir çalışma sunmuştur [21].

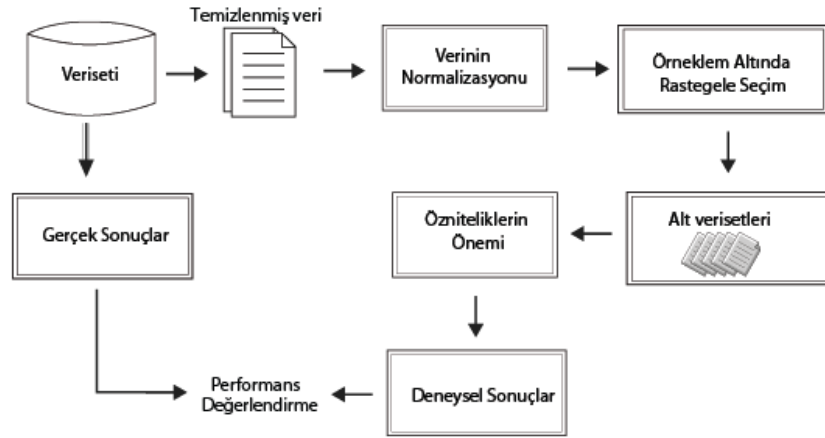
3. VERİSETİ

Bu çalışmada kullanılan veriseti ⁽¹⁾ Wisconsin Üniversitesi Hastanelerinden temin edilen halka açık bir verisetidir. Toplam 699 örnek veri içeren bu verisetindeki sonuç değişkeninde iyi huylu hastalık için 2, kötü huylu hastalık için 4 değeri vardır. Tablo 1'de sunulduğu üzere veri setinde örnek kayıt numarası haricinde toplam 10 öznitelik vardır.

Tablo 1. Öznitelikler ve değer aralıkları

Sıra No	Öznitelik adı	Kısaltması	Değer aralığı
1	Clump Thickness (Kapanma Kalınlığı)	KK	1 - 10
2	Uniformity of Cell Size (Hücre Boyutunun Eşbiçimliliği)	HBE	1 - 10
3	Uniformity of Cell Shape (Hücre Şekil Eşbiçimliliği)	HŞE	1 - 10
4	Marginal Adhesion (Marjinal Yapışma)	MY	1 - 10
5	Single Epithelial Cell Size (Tek Epitelyal Hücre Boyutu)	TEHB	1 - 10
6	Bare Nuclei (Çıplak Çekirdekler)	ÇÇ	1 - 10
7	Bland Chromatin (Yumuşak Kromatin)	YK	1 - 10
8	Normal Nucleoli (Normal Nükleoli)	NN	1 - 10
9	Mitoses (Mitoz)	M	1 - 10
10	Class	Class	2: iyi huylu, 4: kötü huylu

4. DENEYSSEL SONUÇLAR

**Şekil 1.** Önerilen yaklaşıma dair iş akışı.

Önerilen yaklaşıma dair iş akışı Şekil 1’de sunulmuştur. Bu yaklaşım çerçevesinde ÖÖE + LR ve ÖÖE + RO olmak üzere iki ayrı melez kombinasyonu içeren deneysel çalışmalar “scikit-learn” makine öğrenmesi kütüphanesinin dâhil edildiği Python 2.7 platformunda gerçekleştirilmiştir. Bu yaklaşım sırasıyla şu adımları içermektedir:

- Boş değer içeren kayıtlar işleme alınacak olan verisetinden çıkartılmıştır. Böylece, verisetindeki veri sayısı 683 olmak üzere iyi huylu ve kötü huylu hastalık sayıları sırasıyla 458 ve 241’dir. Şekildeki iş akışından görüleceği üzere ilk olarak veriseti 0 ve 1 değerleri aralığında normalize edilmiştir.

- Dengesiz veriler ile çalışılması öğrenmeden ziyade sistemin ezberlemesine neden olmaktadır. Verisetindeki pozitif ve negatif hasta sayısı arasındaki büyük farklılıklardan ötürü dengesiz sınıf içeren çoğunluk verilerinin olduğu gruba Örnekleme Altında Rastgele Seçim (ÖARS, Random Under Sampling) metodu uygulanmıştır. Böylece, iyi huylu ve kötü huylu hastalık sayıları sırasıyla 239 ve 239 olmak üzere her iki sınıf için dengelenmiş veriseti elde edilmiştir. Elde edilen sonuçlar Tablo 2’de sunulmuştur.

Tablo 2. Dengesiz ve dengeli alt verisetlerine dair bilgiler.

Dengesiz Verisetleri				Dengeli Verisetleri					
		Train	Test			Train	Test		
		True	192	47			True	199	40
		False	354	90			False	183	56
		True	190	49			True	193	46
		False	356	88			False	189	50
WMKV				WMKV					
True	239	True	189	50	True	239	True	189	50
False	444	False	357	87	False	239	False	193	46
		True	191	48			True	193	46
		False	355	89			False	189	50
		True	195	44			True	190	49
		False	351	93			False	192	47

c) Önerilen yaklaşım ile yüksek performans elde etmek için k katlı çapraz doğrulama kuralları çerçevesinde dengeli verisetinden alt verisetleri elde edilmiştir. Bu teknik sınıflandırma çalışmalarında kararlı ve doğru bir şekilde sınıflandırma yapmak için yaygın bir şekilde kullanılır. Eğitim ve test verileri çapraz bir şekilde yer değiştirilerek rastgele örnekleme ile ilgili hatalar en aza indirilmiştir [22]. Bu çalışmada 5 katlı çapraz doğrulama işlemi uygulanmış olup her bir sınıflandırma analizi için veriseti %80-20 oranında eğitim ve test kümeleri olarak bölünmüştür. Tablo 4'teki deneysel sonuçlar için sunulan karışıklık matrislerindeki veriler incelendiğinde görüleceği üzere eğitim ve test kümelerindeki veri dağılımı da dengeli bir

yapıya sahiptir. d) Özniteliklerin değerlendirilmesi sürecinde sonuç değişkeni ile ilişkili olan özniteliklerin tespiti yapılarak bilginin keşfi gerçekleştirilir. Ayrıca, bu teknik alakasız verilerden doğan hataların indirgenmesini sağlar. Bu düşünceyle, bu alt verisetlerindeki hedef değişkeni olan "Class" için önemli özniteliklerin tespiti için ÖÖE öznitelik seçim metodu kullanılmıştır. Her bir alt verisetindeki özniteliklerin önemliliği Şekil 2'de ve bu değerlerin ortalaması Tablo 3'te sunulmuştur. Tüm özniteliklerin önemli olduğunun görüldüğü bu tablodaki değerler 5 alt verisetinden elde edilen özniteliklerinin önemliliği değerlerinin ortalamasıdır.

Tablo 3. Özniteliklerin ortalama önemlilik değerleri

Test değişkeni	Kısaltması	Ortalama önemlilik değeri
Kapanma Kalınlığı	KK	0.93
Hücre Boyutunun Eşbiçimliliği	HBE	0.95
Hücre Şekil Eşbiçimliliği	HŞE	0.96
Normal Nükleoli	NN	0.96
Çıplak Çekirdekler	ÇÇ	0.97
Yumuşak Kromatin	YK	0.97
Mitoz	M	0.98
Marjinal Yapışma	MY	0.98
Tek Epitelyal Hücre Boyutu	TEHB	0.98

e) Eğitim aşamasında verisetindeki eğitim verileri ile modelin inşa edildiği ve test aşaması olan ikinci aşamada test verilerinin değerlendirildiği veri sınıflandırması iki aşama içerir. Bu aşamaları içeren işlemler LR ve RO sınıflandırıcı algoritmaları ile her bir alt veriseti üzerinde yürütülmüştür. Yani, hastalığın tanısı için bu alt verisetleri bu sınıflandırıcı algoritmalarına girdi olarak verilmiş ve sistemin öğrenmesi gerçekleştirilmiştir. Sırasıyla Eşitlik 1 ve 2'de verilen Doğru Sınıflandırma Oranı (Dso) ve

Duyarlılık (Duy) performans ölçüm metrikleri çerçevesinde elde edilen deneysel sonuçlar karışıklık matrisi yapısında Tablo 4'te sunulmuştur.

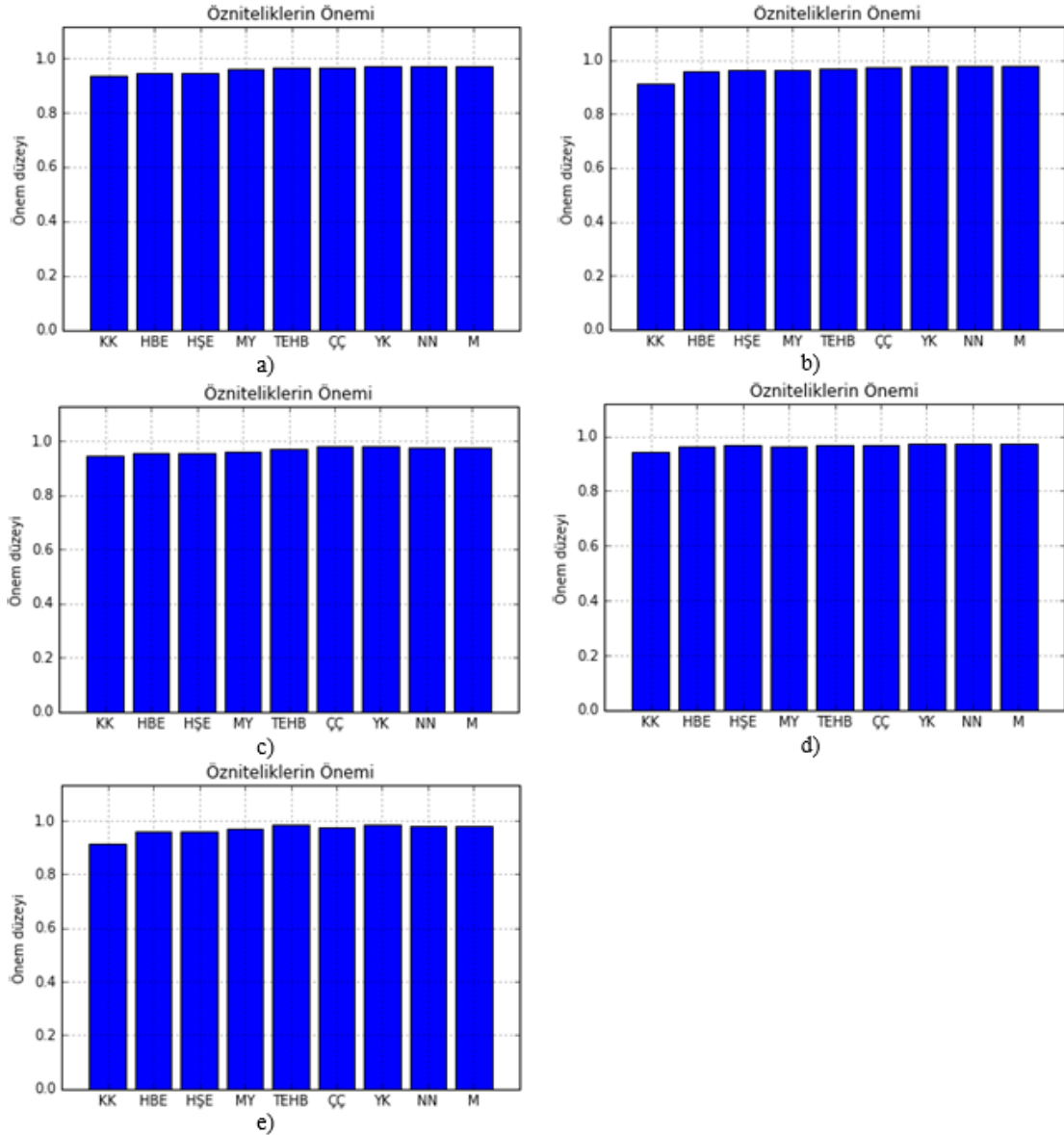
$$Duy = GP / (GP + GN) \quad (1)$$

$$Dso = (GP + GN) / (GP + YP + GN + YN) \quad (2)$$

Eşitlik 1'deki Duy, gerçekte hasta olanlar arasından gerçek doğru bulunan hasta sayısının oranıdır. Model başarısının ölçülmesinde oldukça

çok kullanılan belirleyici ölçüt olan Eşitlik 2'deki Dso değeri, doğru tanı konmuş örnek sayısının toplam örnek sayısına oranı şeklinde ifade edilebilir [23]. Örneğin, 1 nolu alt veriseti için LR ile yürütülen sınıflandırma sonuçlarına göre, gerçekte pozitif olan 40 verinin hepsi pozitif bulunmuştur. Dahası, negatif olan 56 veriden 55 tanesi negatif olarak bulunmuş, 1 tanesini yanlış tahmin yaparak pozitif bulunmuştur. Böylece, bu

veriseti için Dso ve Duy metrikleri sırasıyla %98.96% ve %100 olarak elde edilmiştir. Bu doğrultuda, deneysel sonuçların ortalaması ilgili bölümün en altında verilmiştir. Yaklaşık olarak %98 civarında başarılı sınıflandırma oranı ile her iki sınıflandırıcı algoritmasının başarılı olduğunu ifade etmekle beraber, ÖÖE ve RO hibrit modelinin daha başarılı olduğu ifade edilebilir.



Şekil 2. Özniteliklerin önemi: a) Veriseti no. 1, b) Veriseti no. 2, c) Veriseti no. 3, d) Veriseti no. 4, e) Veriseti no. 5.

Tablo 4. Tüm alt veri setleri için sınıflandırma sonuçları.

Dengeli alt verisetleri	Sınıflandırma Algoritmaları					
	LR			RF		
Veriseti no. 1		TEN	TEP		TEN	TEP
	GN	55 (GN)	1 (YP)	GN	55	1
	GP	0 (YN)	40 (GP)	GP	0	40
		Dso: %98.96		Dso: %98.96		
		Duy: %100		Duy: %100		
Veriseti no. 2		TEN	TEP		TEN	TEP
	GN	49	1	GN	49	1
	GP	2	44	GP	1	45
		Dso: %96.88		Dso: %97.92		
		Duy: %95.65		Duy: %97.83		
Veriseti no. 3		TEN	TEP		TEN	TEP
	GN	45	1	GN	45	1
	GP	3	47	GP	1	49
		Dso: %95.83		Dso: %97.92		
		Duy: %94.0		Duy: %98.0		
Veriseti no. 4		TEN	TEP		TEN	TEP
	GN	50	0	GN	50	0
	GP	3	43	GP	1	45
		Dso: %96.88		Dso: %98.96		
		Duy: %93.48		Duy: %97.93		
Veriseti no. 5		TEN	TEP		TEN	TEP
	GN	46	1	GN	46	1
	GP	3	46	GP	1	48
		Dso: 95.83%		Dso: %97.92		
		Duy: 93.88%		Duy: %97.96		
Ortalama Dso		%96.88		%98.34		
Ortalama Duy		%95.4		%98.34		
* GN: Gerçek Negatif, GP: Gerçek Pozitif, TEN: Tahmin Edilen Negatif, TEP: Tahmin Edilen Pozitif, YN: Yanlış Negatif, YP: Yanlış Pozitif						

5. SONUÇ

En yaygın kanser türlerinden biri olan meme kanseri kadınları etkileyen ölümcül bir hastalıktır. Kliniksel çalışmalara ve alan uzmanlarına faydalı olacağı düşüncesi çerçevesinde gerçekleştirilmiş olan bu çalışmada ÖÖE metodu ile Wisconsin meme kanseri verisetindeki özniteliklerin önemliliği araştırılmıştır. Böylelikle, ilgili verisetindeki sonuç değişkeni için önemli olan veya olmayan değişkenlerin tespiti yapılmıştır. Uygulanan özniteliklerin değerlendirilmesi yöntemine göre verisetindeki tüm öznitelikler oldukça anlamlı bulunmuştur. Bu bilgiler, alan uzmanının verisetindeki ilgili öznitelik bilgisinin önemi bilgisine ulaşmasını sağlayacaktır. Ardından, sonuç değişkeni için önemli olan öznitelikler kullanılarak LR ve RO sınıflandırma algoritmalarının her biri ile eğitim ve test aşamalarını içeren makine öğrenmesi gerçekleştirilmiştir. Bu süreçte öğrenmenin daha iyi olması için 5 katlı çapraz doğrulama tekniğinden faydalanılmıştır. Deneysel sonuçlara göre, LR ve RO algoritmalarının her ikisi ile başarılı sınıflandırmanın gerçekleştirilmiş olduğu, bunun yanında RO'nun LR'ye göre biraz daha iyi olduğu sonucuna varılabilir. Yazar, bu çalışmanın gerçek vaka

kontrol verileri üzerinde yapılacak olan meme kanseri araştırmalarına ve onkolojistlere yardımcı olabileceğini düşünmekte ve ileride bu çerçevede bir çalışma planlamaktadır. Bu yönüyle çalışmanın literatüre katkı sağlayacağını düşünmektedir. Ayrıca yazar, bu yaklaşımın hastalık tanı işlemlerinde oldukça etkili olacağını düşünmekte ve alternatif öznitelik seçim yaklaşımlarını da araştırmaktadır.

TEŞEKKÜR

Yazar, çalışmada kullanılan verisetini araştırmacıların kullanımına sunan Breast Cancer Wisconsin Data Set geliştiricilerine teşekkür eder.

KAYNAKÇA

[1] Breast cancer basics, URL: <http://www.webmd.com/breast-cancer/guide/understanding-breast-cancer-basics>, (Erişim zamanı; Temmuz, 20, 2017).

- [2] A. Jemal, F. Bray, M.M. Center, J. Ferlay, E. Ward, D. Forman, "Global cancer statistics", *CA Cancer J. Clin.* 61(2), 69–90, 2011.
- [3] E.I. Papageorgiou, J. Subramanian, A. Karmegam, N. Papandrianos, "A risk management model for familial breastcancer: A new application using Fuzzy CognitiveMap method", *Computer Methods and Programs in Biomedicine*, 122(2), 123–135, 2015.
- [4] I. Guyon, J. Weston, S. Barnhill and V. Vapnik, "Gene selection for cancer classification using support vector machines", *Mach Learn* 46(1-3), 389-422, 2002.
- [5] D.W. Hosmer, and S. Lemeshow, "Applied logistic regression, 2nd ed.", *Wiley Series in Probability and Statistics*, Canada, 2000.
- [6] L. Breiman, "Random forests", *Machine Learn*, 45, 5-32, 2001.
- [7] Globocan 2012: Estimated Cancer Incidence, Mortality and Prevalence Worldwide 2012, URL: http://globocan.iarc.fr/Pages/fact_sheets_cancer.aspx. (Erişim zamanı; 20 June 2017).
- [8] M. Powell, F. Jamshidian, K. Cheyne, J. Nititham, L.A. Prebil, R. Ereman, "Assessing Breast Cancer Risk Models in Marin County, a Population With High Rates of Delayed Childbirth", *Clinical Breast Cancer*, 14(3), 212-220, 2014.
- [9] Breast Cancer Risk Assessment Tool for Health Care Providers. Office of Cancer Communication. Bethesda (MD): National Cancer Institute; 1998.
- [10] N. Elsayegh, A.M. Gutierrez Barrera, K.I. Muse, H. Lin, H.M. Kuerer, M. Helm, J.K. Litton, and B.K. Arun, "Evaluation of BRCAPRO Risk Assessment Model in Patients with Ductal Carcinoma In situ Who Underwent Clinical BRCA Genetic Testing", *Front Genet*, 7, 1-7, 2016.
- [11] The International Breast Cancer Intervention Study, URL: <https://www.ibis-trials.org/thetrials/ibistrials/ibis-1>, (Erişim zamanı; Nisan, 8, 2018).
- [12] A. AlDiab, S. Qureshi, K. Al Saleh, F.H. AlQahtani, A. Aleem, A. AlSaif, V. Qureshi, M.R. Qureshi, "Studies on the methods of diagnosis and biomarkers used in early detection of breast cancer in the Kingdom of Saudi Arabia", *World Journal of Medical Sciences*, 8(1), 36-47, 2013.
- [13] A. Alharbi, F. Tchier, "Using a genetic-fuzzy algorithm as a computer aided diagnosis tool on Saudi Arabian breast cancer database", *Mathematical Biosciences*, 286, 39-48, 2017.
- [14] E.C. Porto-Mascarenhas, D.X. Assad, H. Chardin, D. Gozal, G.D. Luca Canto, A.C. Acevedo, E.N.S. Guerra, "Salivary biomarkers in the diagnosis of breast cancer: A review", *Critical Reviews in Oncology/Hematology* 110, 62–73, 2017.
- [15] M. Ramalho, F. Fontes, L. Ruano, S. Pereira, N. Lunet, "Cognitive impairment in the first year after breast cancer diagnosis: A prospective cohort study", *Breast*, 32, 173-178, 2017.
- [16] M.L. Peters, J.E. Garber, N. Tung, "Managing hereditary breast cancer risk in women with and without ovarian cancer", *Gynecologic Oncology*, 146(1), 205–214, 2017.
- [17] S. Pudkasam, K. Tangalakis, N. Chinlumprasert, V. Apostolopoulos, L. Stojanovska, "Breast cancer and exercise: The role of adiposity and immune markers", *Maturitas*, 105, 16-22, 2017.
- [18] A.M. Cimpean, R. Tamma, S. Ruggieri, B. Nico, A. Toma, D. Ribatti, "Mast cells in breast cancer angiogenesis", *Critical Reviews in Oncology/Hematology* 115, 23–26, 2017.
- [19] B.A. Almutlaq, R.F. Almuazzi, A.A. Almuhayfir, A.M. Alfouzan, B.T. Alshammari, H.S. AlAnzi, H.G. Ahmed, "Breast cancer in Saudi Arabia and its possible risk factors", *Journal of Cancer Policy*, 12, 83–89, 2017.
- [20] K.E. Lukong, "Understanding breast cancer - The long and winding road", *BBA Clinical*, 7, 64–77, 2017.
- [21] C. Coleman, "Early Detection and Screening for Breast Cancer, *Seminars in Oncology Nursing*", 33(2), 141–155, 2017.
- [22] R. Kohavi, "A study of cross-validation and bootstrap for accuracy estimation and model selection", *Proceedings of the 14th international joint conference on Artificial intelligence*, 2, 1137-1143, 1995.
- [23] A. Baratloo, M. Hosseini, A. Negida and G.E. Ashal, "Part 1: Simple Definition and Calculation of Accuracy, Sensitivity and Specificity", *Emerg (Tehran)*, 3(2), 48-49, 2015.