



Examination of various classification strategies in classification of lung nodule characteristics

Aydin Kaya^{ID}, Ali Seydi Keceli^{ID}, Ahmet Burak Can^{ID}

Department of Computer Engineering, Hacettepe University, Ankara, 06800, Turkey

Highlights:

- Classification of pulmonary nodule characteristics
- Comparing the deep and traditional features
- Effect of different types of classifiers and data balancing on the results

Keywords:

- Nodule characteristics
- Pulmonary nodules
- Transfer learning
- Deep features
- Image processing

Graphical/Tabular Abstract

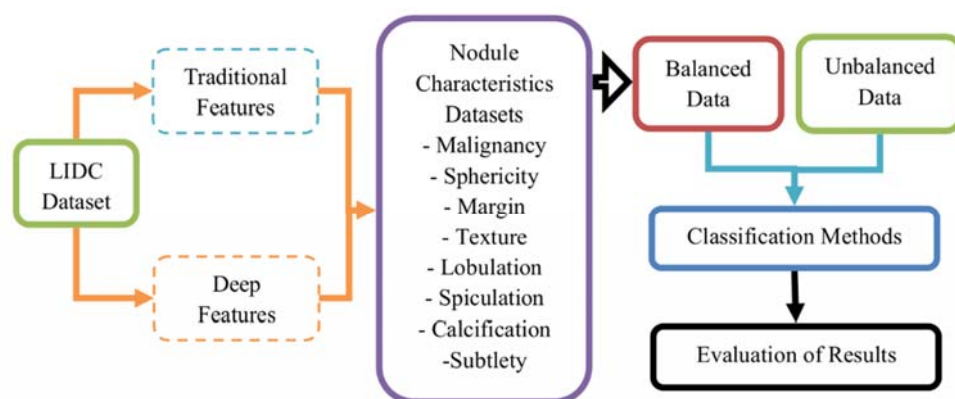


Figure A. Experiment design pattern for Lung nodule characteristics.

Article Info:

Research Article

Received: 30.10.2017

Accepted: 06.03.2018

DOI:

10.17341/gazimmfd.416530

Purpose: Nodule characteristics used in the evaluation of lung nodules are generally subjective assessments of expert opinions. Among these characteristics, the most known and used one for prediction is the degree of malignancy. In the classification studies in the literature, deep features are used besides the traditional features extracted from the nodule appearance and morphological structure. In this study, we proposed an experiment design (Figure A) to examine the combinations of various feature, balancing, and classification methods.

Theory and Methods: Traditional features (morphological features), deep features (fc7 and fc8 layers of Alexnet architecture), and the combinations both used in predicting the nodule characteristics. Four classification algorithms with different structures (ensemble-based, tree-based, instance-based, function-based) are evaluated for predicting nodule characteristics. Reference data were obtained from publicly available LIDC dataset, by means of majority voting from subjective assessments of radiologists. ADASYN balancing method was used to balance the unbalanced datasets.

Correspondence:

Author: Aydin Kaya

e-mail:

aydinkaya@cs.hacettepe.edu.tr

phone: +90 312 297 7500

Results: The effects of different classification models on the classification accuracy, sensitivity and specificity are examined. Balancing the datasets has improved the sensitivity score, but a little effect on classification accuracy and specificity. In average, Random Forest classifier achieved the best classification accuracy and specificity score; the SVM classifier achieved the best sensitivity. However, we cannot conclude that a specific method or feature set were achieved the best result in all experiments.

Conclusion: The results of the experiments shown that the classification strategy needs to be specifically determined starting from the used features to the classification algorithm according to the performance criterion to be achieved.



Akciğer nodül özelliklerinin tahmininde çeşitli sınıflama stratejilerinin incelenmesi

Aydın Kaya*^{ID}, Ali Seydi Keçeli^{ID}, Ahmet Burak Can^{ID}

Hacettepe Üniversitesi, Bilgisayar Mühendisliği Bölümü, Ankara, 06800, Türkiye

Ö N E Ç I K A N L A R

- Akciğer nodül özelliklerinin sınıflandırılması
- Derin öznitelikler ve geleneksel özniteliklerin karşılaştırılması
- Farklı sınıflayıcı türlerinin ve veri dengelemenin sonuçlara etkisi

Makale Bilgileri

Araştırma Makalesi

Geliş: 30.10.2017

Kabul: 06.03.2018

DOI:

10.17341/gazimmfd.416530

Anahtar Kelimeler:

Nodül özellikleri,
akciğer nodülleri,
nakil öğrenme,
derin öznitelikler,
görüntü işleme

ÖZET

Nodül özellikleri, akciğer nodüllerinin değerlendirilmesinde kullanılan, genelde uzman görüşüne göre öznel değerlendirmelerdir. Bunlardan en bilineni ve çalışmalarda tahmin için en çok kullanılanı habislik derecesidir. Literatürde yapılan sınıflama çalışmalarında, nodül görünümü ve morfolojik yapısı üzerinden çıkarılan geleneksel özniteliklerin yanında derin öznitelikler de kullanılmaktadır. Bu çalışmada geleneksel öznitelikler, derin öznitelikler ve bunların kombinasyonları nodül özelliklerinin tahmininde kullanılmıştır. Dört farklı yapıdaki sınıflama algoritması nodül özelliklerinin tahmini için değerlendirilmiştir. Nodül özelliklerinin referans veri kümeleri, doktorların öznel değerlendirmelerinden çoğunluk oylama yöntemi ile oluşturulmuştur. Oluşturulan bu veri kümeleri büyük oranda dengesiz sınıf dağılımına sahiptir. Bu durumun sınıflama sonuçlarına etkisini incelemek için veri dengeleme prosedürü uygulanmıştır. Bu tüm yöntemlerin kombinasyonlarıyla, farklı sınıflama modellerinin sınıflama başarımına, duyarlılığa ve belirliliğe etkileri incelenmiştir. Deneyler sonucunda, ulaşılmak istenen başarımların kriterine göre, kullanılacak özniteliklerden sınıflama algoritmasına kadar, sınıflama stratejisinin özel olarak belirlenmesi gerektiği tespit edilmiştir.

Examination of various classification strategies in classification of lung nodule characteristics

H I G H L I G H T S

- Classification of pulmonary nodule characteristics
- Comparing the deep and traditional features
- Effect of different types of classifiers and data balancing on the results

Article Info

Research Article

Received: 30.10.2017

Accepted: 06.03.2018

DOI:

10.17341/gazimmfd.416530

Keywords:

Nodule characteristics,
pulmonary nodules,
transfer learning,
deep features,
image processing

ABSTRACT

Nodule characteristics used in the evaluation of lung nodules are generally subjective assessments of the expert opinions. Among these characteristics, the most known and used one for prediction is the degree of malignancy. In the classification studies in the literature, deep features are used besides the traditional features extracted from the nodule appearance and morphological structure. In this study, traditional features, deep features, and the combinations both used in predicting the nodule characteristics. Four classification algorithms with different structures are evaluated for predicting nodule characteristics. Reference data sets of nodule characteristics were generated by means of majority voting from subjective assessments of radiologists. These generated data sets generally have large unbalanced class distributions. Data balancing procedure has been applied to examine the effect of this condition on classification results. With the combinations of these methods, effects of different classification models on the classification accuracy, sensitivity and specificity are examined. The results of the experiments shown that the classification strategy needs to be specifically determined starting from the used features to the classification algorithm according to the performance criterion to be achieved.

*Sorumlu Yazar/Corresponding Author: aydinkaya@cs.hacettepe.edu.tr, aliseydikeceli@gmail.com, ahmetburakcan@gmail.com /

Tel: +90 312 297 7500

1. GİRİŞ (INTRODUCTION)

Akciğer kanseri, dünyadaki kansere bağlı ölümlerin önde gelen nedenlerinden biridir [1]. Küçük boyutlu bir akciğer nodülü, yaklaşık 2-30 mm çapında ve belirsiz sınırlardaki bir akciğer lezyonudur. Bu nodüller genelde tomografi taramalarında tesadüfen bulunmaktadır [2]. Tarama teknolojilerindeki ilerlemeler, nodüllerin tespit oranını artırmaktadır. Tespit sonrası nodüllerin değerlendirilmesi amacıyla, Bilgisayar Destekli Tanı (BDT) Sistemleri ikinci bir okuyucu olarak radyologlara yardımcı olmak için geliştirilmiştir. BDT sistemlerinde iki temel adım vardır: Nodüllerin saptanması ve sınıflandırılması. Bir hastanın nodülünün değerlendirilmesindeki zorluk, iyi huylu veya kötü huylu olup olmadığını belirlemektir. Radyologlar tarafından konulan teşhisler, son derece öznel olabilmekte ve radyologların deneyim düzeyine bağlı olarak farklılıklar gösterebilmektedir.

Bu araştırma alanındaki en büyük sıkıntılardan birisi iyi düzenlenmiş, birçok uzman tarafından etiketlenmiş, yüksek miktarda veriye sahip, açık erişimli veri tabanlarının eksikliğidir. 2001 yılında Birleşik Devletler Ulusal Kanser Enstitüsü'nün (U.S. National Cancer Institute) talebi ile oluşturulmaya başlanan ve 2010 yılında tamamlanan Lung Image Database Consortium (LIDC) veri tabanı [3] ile bu eksiklik büyük ölçüde giderilmiştir. LIDC veri tabanı, habislik derecesinin yanı sıra farklı nodül özelliklerinin değerlendirmesini içeren açık erişimli bir veri tabanıdır. Benzer bir veri tabanı Belçika ve Hollanda hükümetlerinin girişimi ile NELSON [4] adıyla oluşturulmuştur. Bu girişimler, görece olarak daha çok veri ile gerçekleştirilen, literatürle karşılaştırılabilirliği olan çalışmaların yapılmasına olanak sağlamıştır. Ayrıca, LIDC veri tabanında olduğu gibi sadece habislik derecesinin değil, farklı değerlendirme özelliklerinin de sağlanması bu veri tabanlarının değerini artırmaktadır. Literatürdeki 2010 öncesi çoğu çalışmada, nodül özellikleri sınıflama aşamasında değerlendirmeye katılmamıştır. Nadir örneklerden biri olan Jabon vd. [5] çalışmasında, görüntü öznelikleri ile beraber, LIDC veri kümesindeki tanımlara yakın özellikler (radyografik tanımlayıcılar) kullanılarak, içerik ve anlamsal tabanlı görüntü erişim sistemi önerilmiştir. Bir başka içerik tabanlı erişim sistemi de Junior vd. tarafından sunulmuştur [6]. LIDC veri kümesinin yayınlanması ile birlikte, nodül özelliklerinin sınıflamada doğrudan kullanıldığı ya da bu özelliklerin sınıflanmasına yönelik yöntemlerin sunulduğu çalışmalar ortaya çıkmıştır. Vinay vd. [7] çalışmalarında, takım sınıflayıcılar ile nodül habislik derecesini, görüntü öznelikleri ve nodül özelliklerini beraber kullanarak habislik sınıflandırması yapmışlardır. Zinovev vd. [8, 9] nodül özelliklerini doğrudan görüntü öznelikleri ile sınıflamışlardır. Bu çalışmada, etkin öğrenme (active learning), tekil/takım sınıflayıcılar ve derin kanı ağları (deep belief networks) kullanılmış ve sonuçları karşılaştırılmıştır. Nodül özelliklerini katlı bir sınıflama mantığında –öncelikle bu özelliklerin sınıflandığı, çıktılarının ise habislik tahmini için kullanıldığı- araştırmalar da

mevcuttur. Kaya ve Can [10, 11] çalışmalarında, bulanık tabanlı ve oylama tabanlı takım sınıflayıcılar kullanarak önce nodül özelliklerini, daha sonra ise bu sonuçlardan habislik derecesini tahmin etmişlerdir. Li vd. [12], radyologlar arası değerlendirmedeki özneliği azaltmak amacıyla, üç adet radyografik tanımlayıcı ile katlı sınıflama yaklaşımını sunmuşlardır. Literatürdeki çalışmalarda [7-12] özneliklere bakıldığında, derin öğrenme yaklaşımlarının kullanılmasından önce geleneksel olarak tabir ettiğimiz şekil ve dokuya bağlı (morfolojik) özneliklerin kullanımı yaygındır [13-15]. Derin öğrenme yaklaşımlarının kullanıldığı çalışmalar 2014 sonrası yaygınlaşmaya başlamıştır. Tıbbi görüntüler üzerine çalışmaların ilk örneklerinde, derin öğrenme aday nodüllerin tespitinde kullanılmıştır. Akciğer nodüllerinde [16], kolonik poliplerde [17], beyin MR görüntülerinde [18], göğüs tomografisi görüntülerinde [19] nodüllerin tespiti için 2B ve 3B Evrişimsel Sinir Ağları (Convolutional Neural Networks) kullanılmıştır. Evrişimsel Sinir Ağlarının (ESA) eğitiminde yüksek miktarda etiketlenmiş veriye sahip olmak gerekmektedir. Bu yüzden, birçok probleme göre daha az verinin olduğu tıbbi görüntülerin sınıflanmasında doğrudan ESA'ları kullanmak etkili sonuç vermemektedir. Bu yüzden önceden eğitilmiş ESA'ların öznelik çıkarma için kullanıldığı transfer öğrenme yaklaşımı ortaya atılmıştır [20]. Transfer öğrenme yaklaşımının kullanıldığı habislik tahmin çalışmaları da literatürde artmaya başlamıştır [21]. Bu çalışmada, görüntü işleme yöntemleri ile akciğer nodülü görüntülerinden elde edilen nodüllerin şekil ve dokusal yapısına bağlı geleneksel öznelikler ile birlikte, önceden eğitilmiş bir ESA'dan elde edilen gözetimsiz derin özneliklerin nodül özelliklerini sınıflamadaki başarımları incelenmiştir. Önceden eğitilmiş evrişimsel sinir ağı için Imagenet [22] veri tabanı üzerinden eğitilmiş Alexnet [23] kullanılmıştır. Sınıflama için dört farklı özellikteki temel sınıflayıcı kullanılacaktır. Bu sınıflayıcılar, destek vektör makinaları (DVM) [24], rastgele orman (RO) [25], k-En Yakın Komşuluk (KEYK) [26] ve Adaboost [27] yöntemleridir. Nodül özelliklerine ait veri kümelerinin aşırı dengesiz dağılıma sahip olması nedeniyle, veri dengeleme uygulamasının sonuçlara etkisi de incelenmiştir. Veri dengeleme için ADASYN (Adaptive Synthetic Sampling) [28] yaklaşımı kullanılmıştır. Makalenin geri kalanı şu şekilde düzenlenmiştir. Kesim 2'de deneyde kullanılan veri kümesi, çıkarılan öznelikler, referans verinin oluşturma aşamaları, veri kümesi dengeleme yaklaşımı ve sınıflama yaklaşımları anlatılacaktır. Kesim 3'te deneysel sonuçlar verilecek ve yorumlanacaktır. Kesim 4'te ise çalışmanın çıktılarını değerlendirilerek, olası gelecek çalışmalardan bahsedilecektir.

2. DENEYSEL YÖNTEM VE ARAÇLAR (EXPERIMENTAL METHOD AND TOOLS)

2.1. Veri Kümesi (Dataset)

Bilgisayar destekli tanı sistemleri, radyologların değerlendirmelerinde yardımcı olmak için yaygın olarak

kullanılan sistemlerdir. Bu sistemler için en önemli eksiklik, etiketlenmiş, yeterli sayıda ve genelleştirme için uygun verinin bulunmamasıdır.

Bu çalışmada kullanılan Lung Image Database Consortium (LIDC) veri tabanı [3], ABD Kanser Araştırmaları Enstitüsü'nün girişimi ile başlatılmıştır. 1010 hastaya ait bilgisayarlı tomografi (BT) görüntüleri ve değerlendirmelerini içermektedir. BT akciğer görüntülerinin yanı sıra; cihazın parametreleri, modeli ve yapılandırma algoritması, hasta demografik bilgileri, sigara geçmişi ve ölçümü yapılmış herhangi klinik tanı bilgisi de veri tabanına eklenmiştir.

Veriler “*blinded*” ve “*unblinded*” olarak adlandırılan iki aşamalarda değerlendirilmiştir. İlk aşamada, radyologlar birbirlerinin değerlendirmelerinden habersiz ilk değerlendirmelerini yapmışlardır. İkinci aşamada ise radyologlara, diğer paydaşların fikirlerini de inceleyip gerekli görürlerse kararlarında güncelleme yapabilme imkânı verilmiştir. Veri tabanında sadece ikinci aşamaya ait veriler dâhil edilmiştir.

LIDC veri tabanının birçok açık erişimli veri tabanından farkı, sadece habislik derecesinin (malignancy) değil, farklı nodül özelliklerine ait değerlendirmelerin de sağlanmasıdır. Tablo 1’de tüm nodül özellikleri, kısa tanımları ve dereceleri verilmiştir. Veri tabanındaki nodüllere ait temel istatistik bilgileri Tablo 2’de verilmiştir. Şekil 1’de, bir hastanın bir

nodülüne ait değerlendirmeyi temsili olarak gösteren görsel bulunmaktadır.

2.2. Öznitelikler (Features)

Çalışmada kullanılan öznitelikler iki ana sınıfta ele alınmıştır. Literatürde sıkça kullanılan şekle, boyuta ve dokuya bağlı (morfolojik) öznitelikler “geleneksel” öznitelikler adı altında toplanmıştır[10]. Toplamda 155 tane olan bu özniteliklerden bazıları, gri-seviye birliktelik matrisi (gray-level cooccurrence matrix) ve Haralick öznitelikleri, en-boy oranı (aspect ratio), sarmalayan kutunun alanı, Zernike momentleri, dış merkezlilik (eccentricity), dairesellik (circularity), katılık (solidity), korelasyon, zıtlık, otokorelasyon, küme belirginliği, küme tonu, benzemezlik, entropi, enerji, homojenlik ve maksimum olasılık gibi gri-seviye özniteliklerdir [29].

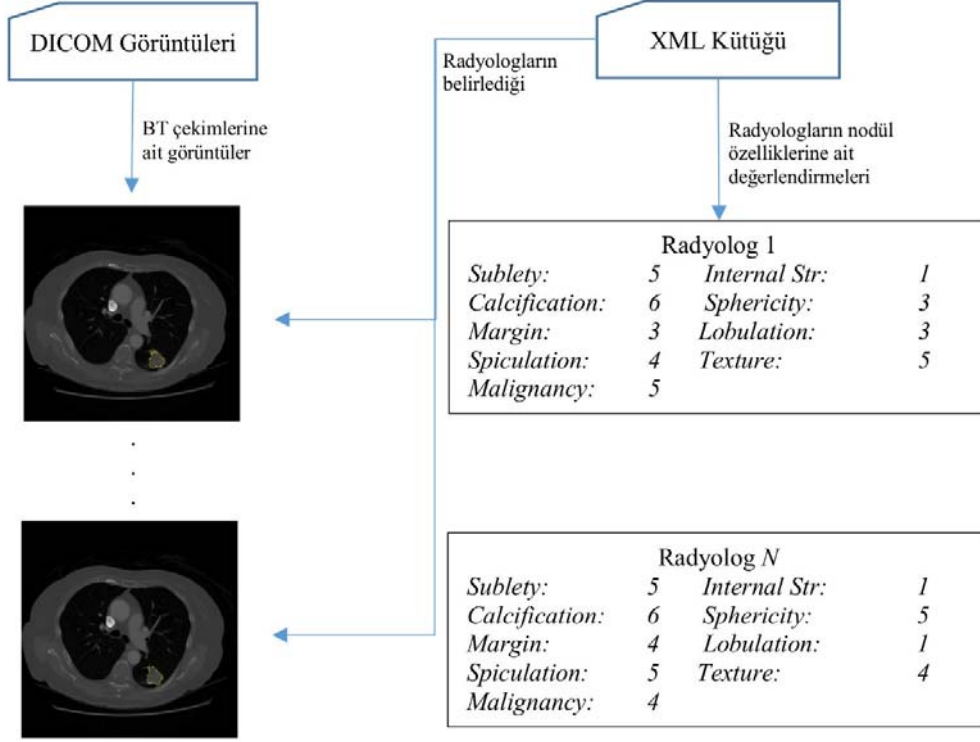
Son yıllarda birçok alanda çalışılan derin öğrenme yöntemleri ile elde edilen öznitelikler, derin öznitelikler olarak adlandırılmaktadır. Derin öğrenme, yüksek düzeyde soyutlamalarla verileri temsil etme yöntemidir. Genel olarak ham veriler çeşitli katmanlar vasıtasıyla işlenir ve piksel başına yoğunluk vektörü içinde kenarların, belirli şekillerin, bölgelerin sayısal olarak ifade edilmesi sağlanır. Derin öğrenmede en çok kullanılan mimarilerden birisi de Evrişimsel Sinir Ağlarıdır (ESA). Bir ESA üzerinde birden çok katman bulunmaktadır. Bunlar genel olarak giriş katmanı, evrişim (convolution) katmanları, havuz katmanları

Tablo 1. Nodül özelliklerinin tanımı ve dereceleri (Definitions and ratings of the nodule characteristics)

Nodül Özelliği	Tanımı	Dereceleri
Kalsifikasyon (KL-Calcification)	Nodüldeki kalsifikasyon gözlemi.	1 - 6
İç Yapısı (İY-Internal Structure)	Nodülün beklenen içyapısı	1 - 4
Lobülasyon (LO-Lobulation)	Lobüler şekilde sınırdan belirginlik oranı.	1 - 5
Habislik (HA-Malignancy)	Habislik derecesi.	1 - 5
Sınır Belirginliği (SIB-Margin)	Sınırların ne kadar belirgin olduğu bilgisi.	1 - 5
Yuvarlaklık (YUV-Sphericity)	Yuvarlaklık bakımından boyutsal şekil değeri.	1 - 5
Spikül Uzanımları (SPU-Spiculation)	Spikül sergileme derecesi.	1 - 5
Belirginlik (BLG-Sublety)	Nodül ve çevresi arasındaki zıtlık.	1 - 5
Doku Yoğunluğu (DY-Texture)	Nodülün iç yoğunluğu.	1 - 5

Tablo 2. LIDC veri tabanındaki nodüllere ait istatistikler. (Statistics of the nodules in LIDC database)

İstatistik	Sayı	Birim
Vaka sayısı – (aynı hastaya ait birden fazla vaka olabilir.)	1010	Adet
Radyografik tanımlayıcı sayısı	9	Adet
Ortalama nodül hacmi	618,90	mm ³
Nodül hacmi standart sapma	3650,65	
En büyük nodül hacmi	27067,00	mm ³
En küçük nodül hacmi	4,39	mm ³



Şekil 1. LIDC veri tabanında bir hastaya ait nodül hakkında N adet radyoloğun değerlendirmesinin (nodül sınırları ve özellikleri) temsili gösterimi

(Representation of a nodule that is evaluated by N radiologists (nodule boundaries and characteristics) in LIDC database)

(pooling), doğrultulmuş doğrusal ünite (ReLU) ve tam bağlı (fully-connected) katmanlardır. Giriş katmanı, görüntünün ESA'ya uygun olarak belirlenen ham girdi verilerini tutar. Evrişim katmanları, hesaplama ağırlıklı işleri yapan ağına katmanlardır. Evrişim katmanı parametreleri, katmana girdi olan değerler üzerinde hareket eden filtrelerdir. Her bir filtre, uzamsal olarak girdiden küçüktür ve girdi verisi üzerinde dolaştırılır. Bu filtrenin her mekânsal konumdaki yanıtlarını veren aktivasyon değerleri üretilir. Tam bağlı katmanlar ise, basitçe önceki katmanlardan gelen girdiyi belli boyuttaki vektörlere çevirir. Genelde bunlar, yüksek seviye öznelik oluşturan katmanlar olarak değerlendirilebilir.

Derin ağların iyi sonuçlar üretmesi için genelde büyük miktarda eğitim verisine ihtiyaç vardır. Eğer bir alanda yeterince eğitim verisi yoksa transfer öğrenme yaklaşımı iyi sonuçlar verebilmektedir. Transfer öğrenme, bir alana ait veri üzerinden eğitilmiş derin ağların, farklı bir alandaki görüntü verisi üzerinde kullanılarak öznelik oluşturma yaklaşımıdır [30]. Bu çalışmada, önceden Imagenet verileri [21] üzerinde eğitilmiş bir ESA olan Alexnet [23] ile transfer öğrenme yaklaşımı kullanılarak BT görüntülerinden nodüllere ilişkin derin öznelikler çıkarılmıştır. Alexnet, sekiz katmandan oluşan (5 evrişim katmanı, 3 tam bağlı katman), Imagenet üzerine tanımlanmış bir ESA modelidir. Transfer öğrenme yaklaşımı bağlamında, Alexnet'in son iki katmanı olan Fc7 ve Fc8 tam bağlı katmanlarından elde edilen derin öznelikler veri kümesine eklenmiştir. Şekil

2'de Alexnet mimarisi ve derin özneliklerin çıkarımı için temsili gösterim verilmiştir.

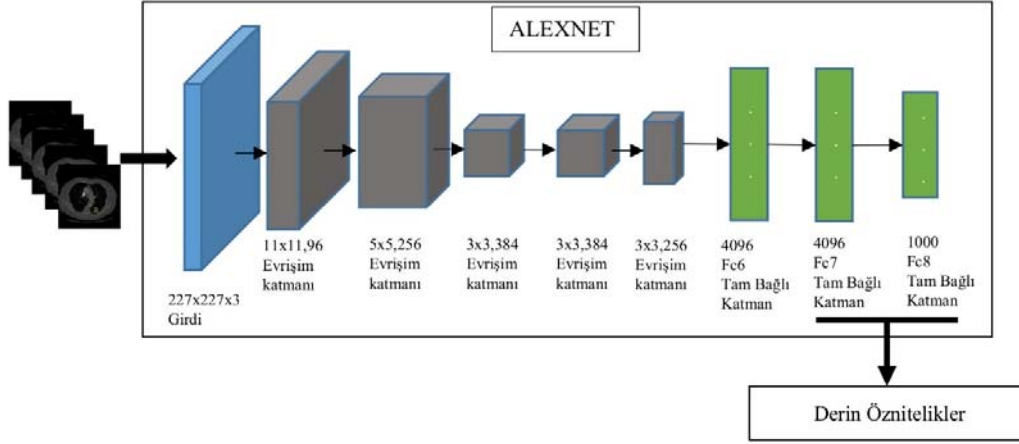
2.3. Referans Verinin Oluşturulması

(Determining the Reference Data)

LIDC veri kümesindeki bütün örnekler, üç veya dört radyolog tarafından değerlendirilmiştir. Bir nodülü farklı uzmanlar, farklı derecelerle değerlendirebildiğinden dolayı, nodül özellikleri için kesin referans veri doğrudan elde edilememektedir. Bu sebeple, çoğunluk oylama yöntemi ile kesin referans verisi oluşturulmuştur. Bu yöntem, tüm değerlendiricilerin konu üzerinde eşit derecede iyi ve tecrübeli olduğunu varsaymaktadır [31]. Bir örnek üzerinde, değerlendiricilerin yüzde elliden fazlasının doğru kabul ettiği sonuç kesin referans olarak kabul edilmiştir.

$$y_i = \begin{cases} 1, & \left(\frac{1}{R}\right) \sum_{j=1}^R y_i^j > 0.5 \\ 0, & \left(\frac{1}{R}\right) \sum_{j=1}^R y_i^j < 0.5 \end{cases} \quad (1)$$

İkili bir veri kümesi için, çoğunluk oylamanın genel tanımı Eş. 1'deki gibi yapılabilir. $y_{i,i}$ örnek için sınıf değerini temsil ederken, y_i^j i . örnek üzerinde j . değerlendiricinin yaptığı sınıf değeridir. Çoğunluk oylama ile her bir nodül özelliği için ayrı bir veri kümesi oluşturulmuştur. Bir nodül örneğini bir özellik kümesine eklemek için, en az üç radyoloğun değerlendirmesinin tutarlılığı gerekmektedir. Bu işlem, her



Şekil 2. Alexnet mimarisi ve derin özniteliklerin elde edildiği evrişim katmanların gösterimi
(Representation of Alexnet architecture and the convolutional layers which deep features are obtained)

Tablo 3. Bir nodül üzerinde dört radyoloğun değerlendirme sonuçları. Bu nodül örneği için, kalın yazı tipiyle belirtilen değerlendirmeler ilgili özelliğin veri kümesine eklenmiştir
(Evaluation results of four radiologists on a nodule. For this nodule sample, the evaluations specified in bold font have been added to the related nodule characteristic's dataset.)

	Belirginlik	Kalsifikasyon	Küresellik	Sınır Bel.	Lobülasyon	Spikül Uza.	Doku Yoğ.	Habislik
Radyolog A	5	6	3	3	3	5	5	5
Radyolog B	5	6	3	4	4	5	5	5
Radyolog C	5	6	5	4	1	5	4	4
Radyolog D	4	3	4	4	3	5	5	5

bir nodül özelliği için farklı örnek sayılarına sahip veri kümelerinin oluşmasına neden olur. Tablo 3’de örnek olarak bir nodül için dört farklı radyoloğun değerlendirmesi verilmiştir. Bu verilen tabloya göre, bahsi geçen nodül örneği küresellik ve lobülasyon özelliklerinin veri kümeleri hariç, diğer tüm özelliklerin veri kümelerine eklenmiştir.

2.4. Veri Kümesi Dengeleme (Dataset Balancing)

Birçok nodül özelliği için veri kümesi sınıf dağılımı yüksek derecede dengesizdir. Örneğin *Belirginlik* veri kümesinde nodüllerin %71’i “5” olarak, %19’u “4” olarak, %9’u “3” olarak ve geri kalan %1’lik kesimi “1” ve “2” olarak işaretlenmiştir. Benzer durum diğer bazı nodül özelliklerinin veri kümelerinde de bulunmaktadır. Az veriye sahip sınıfların örnek sayısını artırmak için veri dengeleme yöntemi uygulanmıştır.

Veri dengeleme yaklaşımı olarak ADASYN (Adaptive Synthetic Sampling) yaklaşımı kullanılmıştır [28]. ADASYN öncelikle azınlıkta kalan sınıfın genel örnek sayısına göre oranını belirler. Bu oran belirlenen eşik değerinin altında kalıyorsa, oluşturulması beklenen sayıda (örneğin; azınlıkta olan sınıf örneklem sayısını, çoğunlukta olan sınıf ile aynı sayıya getirme) yapay örnek oluşturulur. Yapay örnekler, azınlıkta olan sınıfın rastgele bir elemanı alınarak bu elemanın özniteliklerine göre k en yakın

komşuluğunda bulunan örneklerin özniteliklerinden elde edilen oransal bir yaklaşımla elde edilir. ADASYN mekanizması rastgele eleman belirleme ile örneklem çoğaltmaktadır. Veri dengeleme öncesinde önemli olan özniteliklerin seçilmesi ile daha ayırt edici bir veri kümesi oluşturulabilir [32-34].

2.5. Sınıflama Yaklaşımları (Classification Approaches)

Bu çalışmada, Adaboost, Rastgele orman (RO), k-En Yakın Komşuluk (kEYK), ve Destek Vektör Makinası (DVM) algoritmaları ile sınıflama yapılmıştır. Bu kesimde bu algoritmaların genel çalışma prensipleri hakkında bilgi verilmiştir.

Adaboost algoritması [27] en önemli bütünleşik algoritmalarından birisidir. Sınıflama yaklaşımlarında, bütünleşik yöntemlerin genelleştirmeleri tekil yöntemlerden daha başarılı olabilmektedir. Adaboost algoritması, kolay gerçekleştirimi ve sağlam teorik altyapısı ile başarılı uygulamaları olan bir yaklaşımdır.

- X örneklem uzayı ve $Y = \{-1, +1\}$ sınıf etiketleri olmak üzere,
- Eğitim kümesi $\{(x_1, y_1), \dots, (x_m, y_m)\}$ ve t 'nci öğrenme adımında ağırlıkların dağılımları D_t olarak ifade edilsin. Adaboost algoritması öncelikle tüm eğitim örneklerine,

(x_i, y_i) ($i \in \{1, \dots, m\}$), eşit ağırlık atamaktadır. Temel öğrenme algoritması, eğitim kümesi ve D_t 'den zayıf öğrenici (*weak learner*) $h_t : X \rightarrow Y$ elde etmektedir.

- Bu aşamadan sonra hatalı sınıflanan örneklerin ağırlıkları artırılır ve D_{t+1} hesaplanır. Bu işlemler T adım kadar ağırlıklar güncellenerek devam eder.
- Sonuç model, T adet zayıf öğrenicinin aldığı ağırlıklı oy çoğunluğu değerlendirilerek oluşturulur.

Karar ağaçları sınıflama için sıkça kullanılan yaklaşımlardır. Rastgele orman (RO) algoritması ise bünyesinde birçok karar ağacı yaratan ve bunlar içinden en uygununu belirleyen bütünlük bir yaklaşımdır [25]. Rastgelelik hem öznelik seçme, hem de deneme/eğitim veri kümelerini oluşturmada uygulandığı için ayrı bir genelleştirme yaklaşımına gerek duymaz. RO'da oluşturulan her bir karar ağacı, orijinal veri kümesinden elde edilmiş farklı örneklem (*bootstrap sample*) üzerine oluşturulur. Veri kümesinin bir kısmı deneme için ayrılır ve her bir ağacın oluşturulması esnasında eğitim setine dâhil edilmez.

Karar ağaçlarının oluşturulması sırasında, ağacın her bir düğümü belirlenirken; tamamı X kadar olan öznelik kümesi içinden rastgele seçilen x ($x < X$) öznelik kullanılır. Düğüme dâhil edilecek öznelikler *gini katsayısı* denen bir istatistiki ölçüm ile belirlenir. Öznelik önemliliği (*variable importance*) değerlendirmesi yine bu aşamada yapılabilir.

k-En Yakın Komşuluk (kEYK) yaklaşımı, eğitim kümesindeki her bir elemanı, en yakın k komşusunun sınıf değerlerini ve özneliklerini ele alarak sınıflar. Bu işlemin sonucunu üç kavram etkiler: eğitim kümesindeki verilerin etiketleri, k değeri ve özneliklerin uzaklığının hesaplandığı ölçü birimi. k değerinin küçük seçilmesi sınıflama aşamasında verinin gürültüden (*noise*) etkilenmesine sebep olabilir. Büyük seçilirse örneğin ait olduğu sınıf dışından da birçok komşu değerlendirmeye katılabilir. Uzaklık ölçümü ise yapılan sınıflandırmadaki veri türüne göre farklı yaklaşımlar arasında belirlenebilir. Destek Vektör Makinaları (DVM) veri madenciliği yöntemleri içinde güvenilir sonuçlar veren, az sayıda örneklem ile eğitilebilen, dengesiz veriden olumsuz etkilenen, verinin boyutuna duyarsız olan bir algoritmadır [24]. DVM, verinin doğrusal ayrılabilir olduğu durumlarda daha başarılı sonuçlar üretmektedir. Verinin doğrusal ayrılabilir olduğu durumları daha iyi ifade etmek için aşağıdaki örnek verilebilir. İki sınıflı bir küme -1 ve $+1$ etiketleri ile sınıflansın. Veri $X = \{x^t, r^t\}$, eğer $x^t \in C_1$ ise $r^t = +1$, eğer $x^t \in C_2$ ise $r^t = -1$ olsun. Öznelikler arasındaki en iyi ayırmayı sağlayan hiper düzlem w ve w_0 Eş. 2 ve Eş. 3 gibi ifade edilebilir.

$$r^t = +1 \text{ için } w^t x^t + w_0 \geq +1 \quad (2)$$

$$r^t = -1 \text{ için } w^t x^t + w_0 \leq -1 \quad (3)$$

Yukarıdaki formüller aşağıdaki gibi genelleştirilebilir (Eş. 4).

$$r^t (w^t x^t + w_0) \geq +1 \quad (4)$$

DVM ilk olarak doğrusal olarak ayrılabilen veri kümelerinde çalışır olarak tasarlanmışsa da, çekirdek işlevleri (*kernel functions*) ile doğrusal ilişkiler olmayan örneklem üzerinde de uygulanabilmektedir.

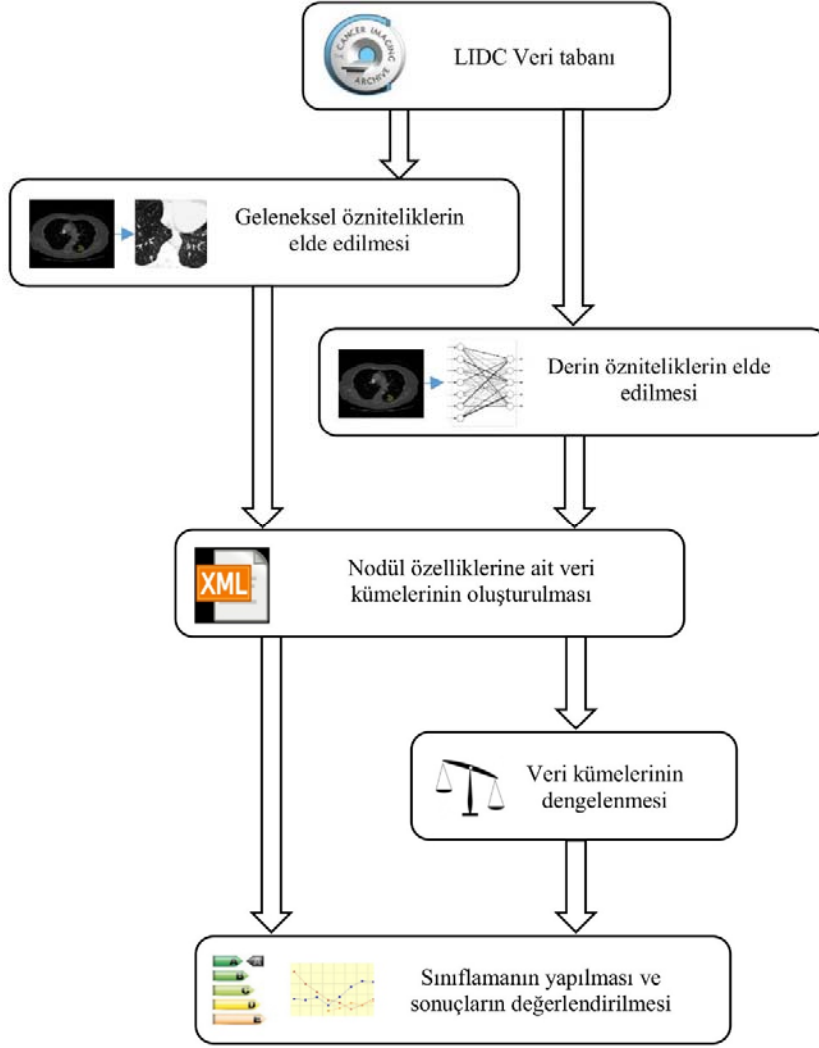
3. SONUÇLAR VE TARTIŞMALAR (RESULTS AND DISCUSSIONS)

Bu çalışmada oluşturulan deneysel yaklaşım, Şekil 3'te gösterilmiştir. Öncelikle LIDC veri kümesi işlenmiş, DICOM görüntülerinden her bir nodül için sayısal öznelikler (geleneksel ve derin öznelikler) elde edilmiştir. Yine LIDC'den elde edilen bilgiler ışığında, her bir nodül özelliği içinde yer alacak nodüller belirlenmiş ve ayrı veri kümeleri oluşturulmuştur. Nodül özelliklerine ait veri kümelerindeki aşırı dengesiz dağılımın sınıflama başarımındaki etkisini incelemek için bu veri kümelerine ek olarak; ADASYN veri dengeleme yaklaşımı kullanılarak, dengeli veri kümeleri oluşturulmuştur. Sınıflama için dört farklı özellikte sınıflayıcı seçilmiş ve destek vektör makinası (DVM) [24], rastgele orman (RO) [25], k-En Yakın Komşuluk (kEYK) [26] ve Adaboost [27] sınıflayıcıları kullanılmıştır. Genelleştirme yaklaşımı olarak tek örneklem dışarıda çapraz geçerlilik (*leave one out*) kullanılmıştır. Bu genelleştirme yaklaşımında her bir örnek (nodül) bir kez deneme verisi olacak şekilde ve kalan örnekler eğitim kümesinde olacak şekilde ayrı ayrı sınıflayıcı modeller eğitilmektedir. Eğitim ve deneme aşaması örnek sayısı kadar gerçekleştirilmektedir. Bu genelleştirme yaklaşımı, özellikle dengesiz dağılıma sahip verilerde, yöntemlerin başarımlarını incelemek için tercih edilen bir genelleştirme yaklaşımıdır.

Sınıflama yöntemi olarak 100 adet ağaçlı, ağaç derinliği kısıtlanmamış RO sınıflayıcılar; niteliklerin normalleştirildiği, 5 komşuluklu, ağırlıksız Öklid uzaklığı kullanılan kEYK sınıflayıcılar; doğrusal işlevli, 1000 adımlı sınırlı DVM sınıflayıcılar ve 20 kestiricili Adaboost sınıflayıcılar kullanılmıştır. ADASYN yönteminde azınlıkta olan sınıflar, çoğunlukta olan sınıflarla aynı örneklem sayısına çoğaltılmıştır. Sınıflama sonuçları Sınıflama Başarımı (SB), Duyarlılık (DUY), Belirlilik (BEL) ölçütleri kullanılarak karşılaştırılmıştır. Yapılan deneylerde Nodül Özellikleri (NÖ), Sınıflama Yaklaşımları (SY), Öznelik Grupları (ÖG) ve Veri Kümesi Dengeleme (VKD) arasındaki tüm farklı kombinasyonlar değerlendirmeye alınmıştır. Bu deney deseninin tüm çıktılarını doğrudan ele almak karmaşık olacağı için, üstten aşağı doğru bir yaklaşımla sonuçların özetleri incelenecek ve daha önce özet olarak verilen sonuçlar adım adım daha ayrıntılı olarak verilecektir.

3.1. Tüm Nodül Özellikleri için Ortalama Sınıflama Sonuçlarının Değerlendirilmesi (Evaluation of Average Classification Results for All Nodule Characteristics)

Tablo 4'te tüm nodül özellikleri için dengeli ve dengelenmemiş veriler üzerinde ortalama sınıflama sonuçları (tüm sınıflayıcılardan elde edilen sonuçların ortalaması) gösterilmiştir.



Şekil 3. Çalışmada kullanılan deney yaklaşımının temsili gösterimi. (Representation of the experimental approach used in the study.)

Sınıflama başarımları (SB) 75 ± 16 aralığında değişirken, duyarlılık değerleri (tahmin edilmek istenen sınıfın doğru tahmin edilmesi oranı – Gerçek Olumlu (*True Positive*)) %35 civarında gözlemlenmiştir. Baskın olan sınıfın doğru tahmin edilmesi genel sınıflama başarımlarını oranını yükseltirken, baskın olmayan sınıfların daha az tanınması duyarlılık (DUY) değerini düşürmektedir. Belirlilik (BEL) değeri ise tüm nodül özellikleri için birbirine yakındır. Bunun nedeni tahmin edilmek istenen sınıfa ait olmayan örneklerin (Gerçek Olumsuz – *True Negative*) daha yüksek olasılıkla tespit edilmesidir. LIDC veri kümesinin dengesiz veri dağılımının sonuçları bu genel tabloda açıkça belli olmaktadır. Tablo 5’te sonuçlar dengeli ve dengelenmemiş veri olarak ikiye ayrılarak verilmiştir. Nodül özelliklerinin çoğunluğunda dengeleme sonrasında sınıflama başarımları düşerken, duyarlılık değeri artış göstermiştir. Bu durum, baskın olan sınıfın etkisinin azalması ve azınlıkta olan sınıfların daha iyi ifade edilmesi olarak yorumlanabilir. Kalsifikasyon ve doku yoğunluğu özelliklerinde değişiklik olmamış, Habislik özelliğinde ise olumsuz yönde bir

değişiklik olmuştur. Geri kalan 5 adet nodül özelliğinde ise, dengeleme işlemi duyarlılık değerini olumlu yönde değiştirmiştir.

Tablo 4. Nodül özellikleri (NÖ) için yapılan sınıflamaların ortalama sonuçları
(Average classification results for each nodule characteristics.)

NÖ	SB	DUY	BEL
Belirginlik (BLG)	0,7225	0,3742	0,8565
Kalsifikasyon (KL)	0,9178	0,3541	0,8508
Yuvarlaklık (YUV)	0,4981	0,4470	0,8188
Sınır Belirginliği (SIB)	0,6535	0,3739	0,8341
Lobülasyon (LO)	0,8386	0,3094	0,8323
Spikül Uzunlukları (SPU)	0,8914	0,4018	0,8586
Doku Yoğunluğu (DY)	0,9094	0,2696	0,8466
Habislik (HA)	0,6228	0,4864	0,8861

3.2. Sınıflama Yöntemlerinin Başarımlarının Değerlendirilmesi

(Performance Evaluation of the Classification Methods)

Tablo 6’da farklı sınıflama yaklaşımları için, dengelenmiş ve dengelenmemiş veriler üzerinde tüm öznel gruplarıyla ortalama sınıflama başarımları verilmiştir. RO ve Adaboost sınıflayıcılar, dengelenmiş verilerde en yüksek başarımları verirken, dengelenmiş veride RO sınıflayıcı ile diğer sınıflayıcılar arasındaki başarımların farkı yüksektir. DVM sınıflayıcı ise (dengelenmemiş veride) duyarlılık değerinde en yüksek başarımları göstermiştir. Diğer deneylerde olduğu gibi sınıflama başarımları dengelenmemiş verilerde genel olarak daha yüksek iken, duyarlılık değeri dengeli verilerde %4 civarında daha yüksektir. Belirlilik değerinde veri dengelemenin bir etkisi olmadığı gibi, sınıflayıcılar arasında da bu değerlendirme metriğinde diğer başarımlar ölçütleri kadar fark yoktur.

Tablo 7’de tüm sınıflama yaklaşımlarının nodül özellikleri üzerindeki başarımlarının incelendiği deneylerin sonuçları gösterilmiştir. Belirginlik (BLG) özelliği için en yüksek sınıflama başarımlarını ve belirliliği RO verirken, en yüksek duyarlılığı DVM sınıflayıcı vermektedir. Kalsifikasyon (KL) özelliği için Adaboost ve RO en yüksek sınıflama başarımlarını, DVM en yüksek duyarlılığı, RO ve DVM ise en yüksek belirliliği vermektedir. Yuvarlaklık (YUV) özelliği için tüm ölçümlerde RO en yüksek başarımları sağlamıştır. Sınır Belirginliği (SIB) için RO en yüksek sınıflama başarımlarını, DVM en yüksek duyarlılığı, DVM ve RO en yüksek belirliliği sağlamaktadır. Lobülasyon (LO) özelliği için RO, KEYK ve Adaboost en yüksek sınıflama başarımlarını, DVM en yüksek duyarlılığı, DVM, RO ve Adaboost en yüksek belirliliği vermiştir. Spikül Uzanımları (SPU) özelliği için KEYK ve RO en yüksek sınıflama başarımlarını, DVM en yüksek duyarlılığı, RO en yüksek belirliliği

vermiştir. Doku Yoğunluğu (DY) özelliği için: RO ve Adaboost en yüksek sınıflama başarımlarını, RO ve DVM en yüksek duyarlılığı, DVM en yüksek belirliliği vermiştir. Habislik (HA) özelliği için: RO en yüksek sınıflama başarımlarını ve duyarlılığı, RO ve DVM en yüksek belirliliği vermiştir. Bu sonuçların grafiksel gösterimi Şekil 4’te verilmiştir.

3.3. Öznel Gruplarının Sınıflama Başarımlarına Etkilerinin İncelenmesi

(Evaluation of the Effect of the Feature Groups on Classification Performance)

Tablo 8’de tüm öznel grupları için ortalama sınıflama sonuçları gösterilmiştir. Dengelenmemiş ve dengelenmiş veri kümeleri üzerinde yapılan deneylerde, geleneksel öznel gruplar (G) ile birleştirilen derin öznel gruplar (fc7, fc8), sınıflama başarımları ve belirlilik ölçümlerinde en yüksek başarımları elde etmiştir. Duyarlılık ölçümünde, dengelenmiş verilerde tekil derin öznel gruplar daha başarılı sonuçlar elde ederken, dengelenmemiş verilerde öznel grupların kombinasyonu daha yüksek başarımlara ulaşmıştır. Genel olarak başarımlar incelendiğinde, geleneksel ve derin öznel grupların birlikte kullanımı ile sınıflama başarımları ve duyarlılık sınıflama çıktılarında yaklaşık %5 artış olmuştur.

Tablo 9’de ve Şekil 5’te her bir sınıflama yaklaşımının farklı öznel grupları üzerinde başarımları gösterilmiştir. Burada amaç, hangi sınıflayıcının, hangi öznel grupları üzerinde daha yüksek başarımları sağladığını incelemektir. RO sınıflayıcı, geleneksel öznel gruplar ile en yüksek başarımlarına ulaşmıştır. DVM sınıflayıcıda RO sınıflayıcının aksine geleneksel öznel gruplarla başarımlar çok daha düşük olmasına rağmen, derin öznel gruplarla yapılan deneylerde %22 civarında sınıflama başarımları artışı, %5 civarında duyarlılık artışı ve %2 civarında belirlilik düşüşü

Tablo 5. Nodül özellikleri (NÖ) için yapılan sınıflamaların dengeli ve dengelenmemiş veri ile ortalama sonuçları
(Average classification results of balanced and unbalanced dataset for each nodule characteristics)

NÖ	Dengelenmemiş Veri			Dengeli Veri		
	SB	DUY	BEL	SB	DUY	BEL
BLG	0,7341	0,3461	0,8666	0,7110	0,4023	0,8465
KL	0,9200	0,3559	0,8515	0,9155	0,3523	0,8501
YUV	0,5146	0,4288	0,8230	0,4817	0,4651	0,8145
SIB	0,6615	0,3428	0,8333	0,6455	0,4050	0,8350
LO	0,8553	0,2753	0,8234	0,8219	0,3436	0,8412
SPU	0,9079	0,3510	0,8427	0,8749	0,4526	0,8744
DY	0,9254	0,2697	0,8466	0,8935	0,2694	0,8465
HA	0,6852	0,5524	0,8987	0,5605	0,4203	0,8735

Tablo 6. Sınıflama yaklaşımları için dengelenmiş ve dengelenmemiş veriler ile ortalama sonuçlar
(Average results with balanced and unbalanced data for classification approaches)

SY	Dengelenmemiş Veri			Dengeli Veri		
	SB	DUY	BEL	SB	DUY	BEL
RO	0,8350	0,4173	0,8689	0,8287	0,4583	0,8741
DVM	0,7012	0,4412	0,8696	0,6585	0,4268	0,8593
KEYK	0,7643	0,2934	0,8177	0,7065	0,3258	0,8176
Adaboost	0,8015	0,3091	0,8367	0,7586	0,3444	0,8398

Tablo 7. Sınıflama yöntemlerinin her bir nodül özelliği üzerindeki ortalama başarımları. "D." eki dengelenmiş veri kümesi üzerinde yapılan deneyleri göstermektedir
(The average performances of the classification methods on each nodule characteristics. "D." shows the experiments on the balanced data set)

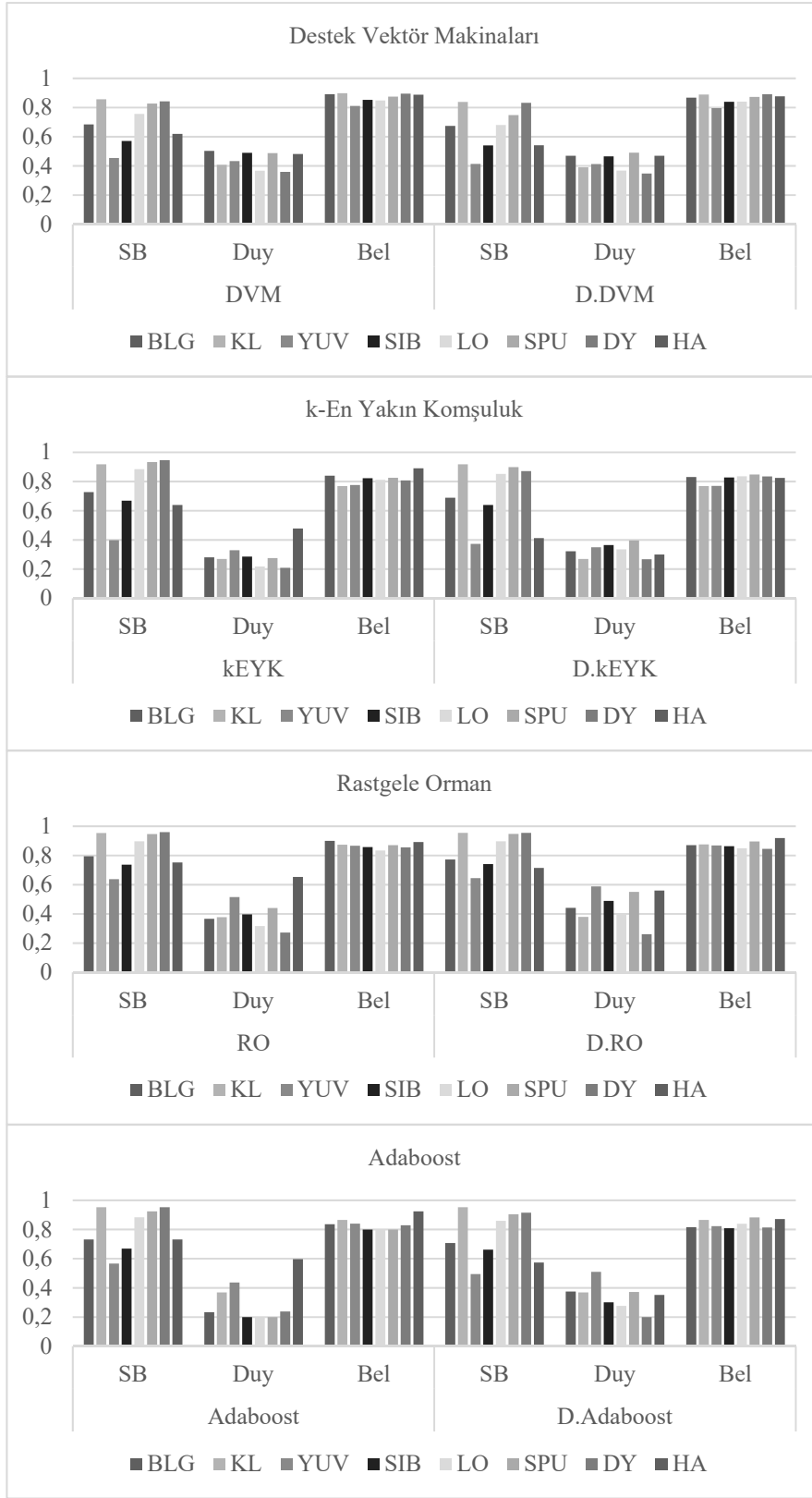
		RO	D. RO	DVM	D. DVM	kEYK	D. kEYK	Adaboost	D. Adaboost
BLG	SB	0,7945	0,7726	0,6836	0,6741	0,7271	0,6895	0,7310	0,7079
	DUY	0,3671	0,4421	0,5032	0,4693	0,2812	0,3231	0,2329	0,3749
	BEL	0,9003	0,8717	0,8907	0,8672	0,8397	0,8308	0,8358	0,8162
KL	SB	0,9544	0,9552	0,8565	0,8377	0,9175	0,9175	0,9517	0,9518
	DUY	0,3777	0,3801	0,4074	0,3905	0,2696	0,2696	0,3688	0,3688
	BEL	0,8740	0,8765	0,8976	0,8894	0,7690	0,7690	0,8654	0,8654
YUV	SB	0,6382	0,6450	0,4545	0,4138	0,3986	0,3726	0,5670	0,4953
	DUY	0,5153	0,5889	0,4331	0,4132	0,3297	0,3487	0,4370	0,5097
	BEL	0,8660	0,8695	0,8109	0,7958	0,7753	0,7698	0,8399	0,8231
SIB	SB	0,7376	0,7415	0,5708	0,5406	0,6685	0,6385	0,6691	0,6615
	DUY	0,3961	0,4887	0,4898	0,4652	0,2861	0,3654	0,1991	0,3007
	BEL	0,8576	0,8635	0,8529	0,8401	0,8233	0,8278	0,7993	0,8085
LO	SB	0,8965	0,8964	0,7561	0,6803	0,8843	0,8512	0,8844	0,8598
	DUY	0,3163	0,3937	0,3666	0,3679	0,2183	0,3358	0,2000	0,2769
	BEL	0,8347	0,8498	0,8482	0,8405	0,8109	0,8357	0,8000	0,8387
SPU	SB	0,9467	0,9486	0,8271	0,7479	0,9332	0,8993	0,9248	0,9040
	DUY	0,4406	0,5513	0,4884	0,4907	0,2753	0,3964	0,1998	0,3720
	BEL	0,8713	0,8963	0,8740	0,8724	0,8255	0,8474	0,7998	0,8817
DY	SB	0,9603	0,9554	0,8422	0,8322	0,9465	0,8716	0,9525	0,9148
	DUY	0,2723	0,2620	0,3594	0,3485	0,2087	0,2675	0,2382	0,1999
	BEL	0,8550	0,8461	0,8948	0,8911	0,8072	0,8350	0,8292	0,8140
HA	SB	0,7517	0,7149	0,6189	0,5412	0,6386	0,4118	0,7315	0,5740
	DUY	0,6530	0,5600	0,4817	0,4694	0,4781	0,2997	0,5967	0,3522
	BEL	0,8921	0,9198	0,8877	0,8774	0,8910	0,8255	0,9241	0,8712

Tablo 8. Öznitelik grupları (ÖG) için yapılan sınıflamaların dengelenmiş ve dengelenmemiş veriler ile ortalama sonuçları.
(The average results of the classifications made for the feature groups (ÖG) with balanced and unbalanced data.)

ÖG	Dengelenmemiş Veri			Dengeli Veri		
	SB	DUY	BEL	SB	DUY	BEL
G	0,7272	0,3466	0,8444	0,7031	0,3664	0,8455
fc7	0,7872	0,3693	0,8437	0,7334	0,4066	0,8456
fc8	0,7840	0,3620	0,8401	0,7287	0,4024	0,8448
G+fc7	0,8001	0,3776	0,8586	0,7617	0,3807	0,8492
G+fc8	0,7790	0,3706	0,8544	0,7634	0,3881	0,8534

gerçekleşmiştir. kEYK sınıflayıcı da DVM sınıflayıcıya benzer şekilde, derin özniteliklerde geleneksel özniteliklere nazaran, sınıflama başarımında %5, duyarlılıkta %15, belirlilikte ise %3 civarında başarımların artışı gözlemlenmiştir. Adaboost sınıflayıcıda ise geleneksel özniteliklerin dâhil olduğu tüm gruplar birbirine yakın çıkmıştır. Duyarlılık

değerlerinde derin öznitelikler kısmen daha başarılı iken, geleneksel öznitelikler sınıflama başarımında daha yüksek sonuç vermiştir. Önceki deneylerde görüldüğü üzere, duyarlılık ölçümünde dengelenmiş veri kümelerinde daha yüksek başarımlar elde edilmişken, sınıflama başarımı ölçümünde başarımlar düşmüştür.

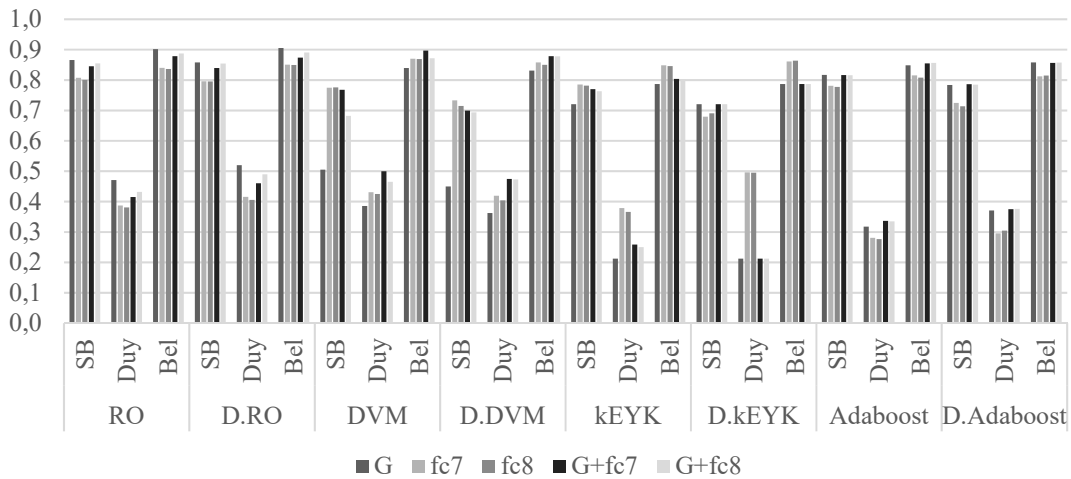


Şekil 4. Sınıflama yöntemlerinin her bir nodül özelliği üzerindeki ortalama başarımlarını gösteren grafikler (Graphs showing the average performances of the classification methods on each nodule characteristics)

Tablo 9. Sınıflama yaklaşımlarının her bir öznelik grubu üzerinde elde ettiği ortalama sınıflama başarımları. "D." eki dengelenmiş veri kümesi üzerinde yapılan deneyleri göstermektedir (The average classification performance obtained by the classification approaches on each feature group. "D." shows the experiments on the balanced data set)

		G	fc7	fc8	G+fc7	G+fc8
RO	SB	0,8659	0,8078	0,8006	0,8456	0,8549
	DUY	0,4711	0,3873	0,3810	0,4152	0,4320
	BEL	0,9018	0,8402	0,8365	0,8784	0,8876
D.RO	SB	0,8580	0,7959	0,7955	0,8396	0,8545
	DUY	0,5199	0,4157	0,4057	0,4603	0,4902
	BEL	0,9054	0,8506	0,8499	0,8741	0,8908
DVM	SB	0,5050	0,7752	0,7759	0,7680	0,6820
	DUY	0,3855	0,4305	0,4248	0,5000	0,4652
	BEL	0,8396	0,8703	0,8692	0,8974	0,8716
D.DVM	SB	0,4501	0,7337	0,7149	0,6998	0,6939
	DUY	0,3624	0,4194	0,4041	0,4749	0,4733
	BEL	0,8313	0,8580	0,8501	0,8791	0,8777
KEYK	SB	0,7208	0,7850	0,7817	0,7705	0,7634
	DUY	0,2125	0,3791	0,3660	0,2588	0,2505
	BEL	0,7875	0,8489	0,8461	0,8041	0,8020
D.kEYK	SB	0,7208	0,6794	0,6905	0,7208	0,7208
	DUY	0,2125	0,4961	0,4952	0,2125	0,2125
	BEL	0,7875	0,8615	0,8641	0,7875	0,7875
Adaboost	SB	0,8170	0,7809	0,7777	0,8163	0,8157
	DUY	0,3175	0,2804	0,2762	0,3365	0,3349
	BEL	0,8488	0,8152	0,8085	0,8547	0,8563
D.Adaboost	SB	0,7836	0,7246	0,7138	0,7866	0,7844
	DUY	0,3707	0,2952	0,3045	0,3752	0,3764
	BEL	0,8579	0,8123	0,8149	0,8563	0,8577

Sınıflama Yöntemleri ve Nitelik Grupları Arasındaki Başarım İlişkisi

**Şekil 5.** Sınıflama yaklaşımlarının her bir öznelik grubu üzerinde elde ettiği ortalama sınıflama başarımlarının grafiksel gösterimi. (Graphical representation of the average classification performances obtained by classification approaches on each feature group.)

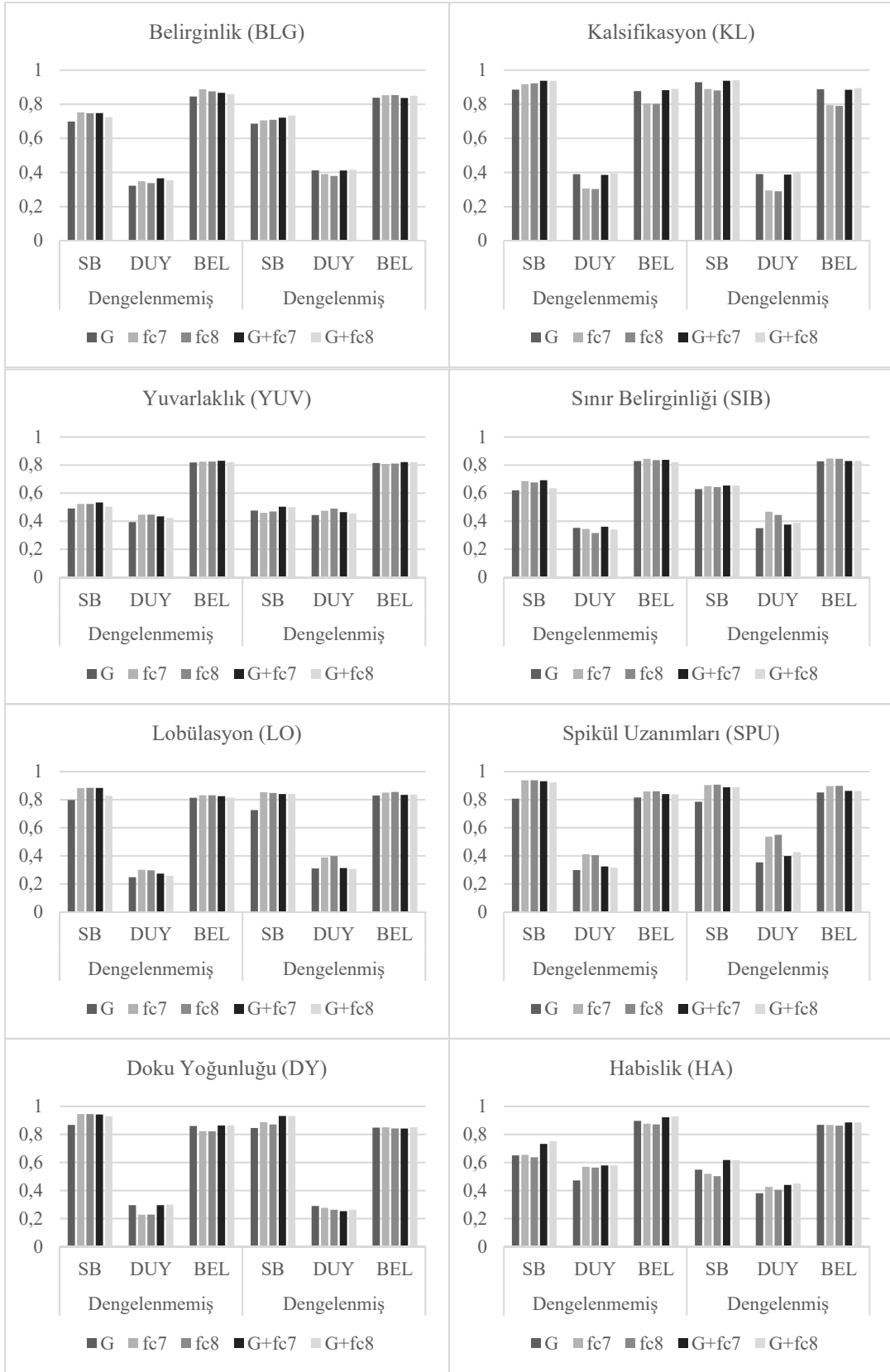
Tablo 10 ve Şekil 6'da tüm nodül özellikleri için hangi öznelik grubunun daha başarılı olduğunu tespit etmek için yapılan deneylerin sonuçları gösterilmiştir. Sınıflama başarımı bağlamında Belirginlik (BLG), Sınır Belirginliği (SIB), Lobülasyon (LO) ve Yuvarlaklık (YUV) özellikleri için fc7 ve G+fc7 öznelik grupları; Kalsifikasyon (KL) ve Habislik (HA) özellikleri için G+fc7 ve G+fc8 öznelik grupları; Spikül Uzanımları (SPU) ve Doku Yoğunluğu (DY) özellikleri için fc7 ve fc8 öznelik grupları en yüksek başarıma ulaşmıştır. Duyarlılık bağlamında BLG, DY ve HA özellikleri için G+fc7 ve G+fc8 öznelik grupları; YUV, LO ve SPU özellikleri için fc7 ve fc8 öznelik grupları; KL

özellikleri için G ve G+fc8 öznelik grupları; SIB özellikleri için G ve G+fc7 öznelik grupları en yüksek başarıma ulaşmıştır. Belirlilik bağlamında BLG, LO ve SPU özellikleri için fc7 ve fc8 öznelik grupları; KL, DY ve HA özellikleri için G+fc7 ve G+fc8 öznelik grupları; YUV ve SIB özellikleri için fc7 ve G+fc7 öznelik grupları en yüksek başarıma ulaşmıştır.

Dengelenmiş veri kümesi üzerinde yapılan çalışmalarda, önceki deneylerde karşılaştığı üzere sınıflama başarımında düşüş olurken, duyarlılık değeri birçok öznelik grubu için yükselmiştir. Belirlilik değerinde ise değişim az olmuştur.

Tablo 10. Öznelik gruplarının her bir nodül özelliği üzerindeki ortalama başarımları
(Average performances of feature groups on each nodule characteristics)

NÖ	ÖG	Dengelenmemiş Veri			Dengeli Veri		
		SB	DUY	BEL	SB	DUY	BEL
BLG	G	0,6983	0,3225	0,8448	0,6868	0,4129	0,8383
	fc7	0,7523	0,3484	0,8876	0,7050	0,3908	0,8530
	fc8	0,7473	0,3383	0,8748	0,7083	0,3793	0,8538
	G+fc7	0,7481	0,3659	0,8665	0,7210	0,4123	0,8369
	G+fc8	0,7243	0,3554	0,8593	0,7341	0,4163	0,8504
KL	G	0,8857	0,3901	0,8772	0,9283	0,3910	0,8877
	fc7	0,9172	0,3062	0,8050	0,8895	0,2941	0,7936
	fc8	0,9224	0,3035	0,8024	0,8808	0,2900	0,7898
	G+fc7	0,9374	0,3862	0,8828	0,9379	0,3882	0,8850
	G+fc8	0,9374	0,3935	0,8901	0,9412	0,3981	0,8944
YUV	G	0,4907	0,3942	0,8179	0,4760	0,4437	0,8143
	fc7	0,5228	0,4466	0,8237	0,4605	0,4734	0,8077
	fc8	0,5223	0,4463	0,8238	0,4692	0,4894	0,8103
	G+fc7	0,5338	0,4344	0,8303	0,5030	0,4643	0,8208
	G+fc8	0,5033	0,4224	0,8194	0,4998	0,4549	0,8197
SIB	G	0,6197	0,3517	0,8291	0,6290	0,3497	0,8260
	fc7	0,6858	0,3448	0,8441	0,6479	0,4672	0,8462
	fc8	0,6754	0,3155	0,8353	0,6430	0,4442	0,8446
	G+fc7	0,6913	0,3608	0,8371	0,6540	0,3753	0,8283
	G+fc8	0,6352	0,3410	0,8207	0,6538	0,3887	0,8296
LO	G	0,7974	0,2486	0,8140	0,7256	0,3101	0,8295
	fc7	0,8826	0,3001	0,8318	0,8532	0,3884	0,8504
	fc8	0,8849	0,2964	0,8316	0,8481	0,3987	0,8556
	G+fc7	0,8840	0,2741	0,8252	0,8407	0,3136	0,8339
	G+fc8	0,8279	0,2573	0,8146	0,8421	0,3071	0,8363
SPU	G	0,8081	0,2989	0,8162	0,7858	0,3533	0,8520
	fc7	0,9383	0,4111	0,8588	0,9042	0,5357	0,8968
	fc8	0,9380	0,4043	0,8597	0,9069	0,5495	0,8986
	G+fc7	0,9317	0,3246	0,8410	0,8881	0,3983	0,8630
	G+fc8	0,9237	0,3163	0,8375	0,8897	0,4262	0,8618
DY	G	0,8676	0,2955	0,8596	0,8453	0,2901	0,8484
	fc7	0,9445	0,2281	0,8229	0,8879	0,2766	0,8496
	fc8	0,9443	0,2285	0,8222	0,8711	0,2627	0,8429
	G+fc7	0,9417	0,2963	0,8635	0,9320	0,2538	0,8413
	G+fc8	0,9288	0,2999	0,8646	0,9313	0,2640	0,8505
HA	G	0,6498	0,4717	0,8966	0,5485	0,3801	0,8679
	fc7	0,6544	0,5692	0,8753	0,5191	0,4265	0,8676
	fc8	0,6373	0,5633	0,8706	0,5021	0,4052	0,8624
	G+fc7	0,7329	0,5786	0,9225	0,6173	0,4402	0,8847
	G+fc8	0,7516	0,5792	0,9288	0,6154	0,4495	0,8847



Şekil 6. Öznitelik gruplarının her bir nodül özelliği üzerindeki ortalama başarımlarını gösteren grafikler (Graphs showing the average performances of feature groups on each nodule characteristics)

Tablo 11. Literatürdeki çalışmaların sonuçları. (Results from the Literature)

	Kesin Referans Yöntemi	SB	DUY	BEL
RO (Ortalama)	Çoğunluk Oylama (5 sınıflı)	0,75	0,65	0,89
RO (En iyi)	Çoğunluk Oylama (5 sınıflı)	0,82	0,64	0,94
T-RO[10]	Çoğunluk Oylama (5 sınıflı)	0,83	0,58	0,94
T-DVM[10]	Çoğunluk Oylama (5 sınıflı)	0,81	0,56	0,94
T-RO[10]	Çoğunluk Oylama (3 sınıflı)	0,84	0,83	0,92
EoE (Ortalama)[7]	Belirsiz (5 sınıflı)	0,78	-	-
DİA [8]	Belirsiz (5 sınıflı)	0,63	-	-
EÖ (Ortalama) [9]	Ortanca Oylama (5 sınıflı) – (Farklı radyolog anlaşma oranları da incelenmiş)	0,78	-	-
FESA (Ortalama) [36]	1-2 derecesiyle 4-5 birleştirilmiş, 3. Derece atılmış. İkili veri kullanılmış.	0,82	0,74	0,88

3.4. Deneysel Sonuçlarının Literatürdeki Çalışmalarla Karşılaştırılması

(Comparing the Experimental Results with the Studies on Literature)

LIDC veri tabanı, açık erişimli bir kaynak olması sebebiyle birçok farklı çalışmada kullanılmıştır. Böyle bir veri tabanı kullanılmasına rağmen, farklı kesin referans yaklaşımları ve farklı varsayımlar yüzünden literatürdeki çalışmaları doğrudan karşılaştırmak zordur [35].

Tablo 11’de, çalışmalarında nodül özelliklerini de sınıflama aşamasına ekleyen çalışmaların sonuçları ve kesin referans oluşturma yaklaşımları verilmiştir. Tabloda ilk iki satırda, habislik özelliğinde en yüksek başarıyı sağlayan RO sınıflayıcının tüm öznelik grupları için ortalama sonuçları ve en yüksek değeri dâhil edilmiştir. Zinovev vd.[8, 9] çalışmalarında Etkin Öğrenme (EÖ) ve Derin İnanç Ağları (DİA) yöntemleri kullanılmışlardır. Vinay vd.[7] çalışmalarında, takım sınıflayıcılar ile nodül habislik derecesini, görüntü öznelikleri ve nodül özelliklerini beraber kullanarak habislik sınıflandırması yapmışlardır (EoE). Kaya ve Can [10] çalışmalarında RO (T-RO) ve DVM (T-DVM) tabanlı topluluk sınıflayıcı yaklaşımı önermişlerdir. Xie vd.[36] farklı ESA’lardan gelen bilgileri karar aşamasında birleştirmişlerdir (FESA). Bu yöntemi denemek için LIDC veri kümesindeki habislikte 1-2 ve 4-5 derecesini birleştirmiş, 3 derecesini veriden atmışlardır.

Tablo 11’e bakıldığında, bu çalışmada en yüksek başarıyı gösteren RO sınıflayıcı, sınıflama başarımında diğer çalışmalarla karşılaştırılabilir bir değere ulaşmıştır. Duyarlılık değeri olan çalışmalarda ise (5 sınıflı veride) en yüksek sonucu elde etmiş, belirlilikte ise diğerleriyle benzer başarıya ulaşmıştır. FESA çalışmasında veri kümesi 2 sınıflı değerlerden, T-RO (3 sınıflı) ise 3 sınıflı değerlerden oluşmuştur. Belirsizliğin daha az olduğu 2 ve 3 sınıflı veri kümelerinde duyarlılık değerinin 5 sınıflı veri üzerinde yapılan çalışmalara göre daha yüksek olduğu gözlemlenmektedir.

4. SONUÇLAR (CONCLUSIONS)

Bu çalışmada LIDC veri tabanından elde edilen uzman değerlendirme sonuçları kullanılarak, nodül özellikleri için farklı sınıflayıcı ve öznelik gruplarının dengelenmiş ve dengelenmemiş veri kümeleri üzerinde sınıflama başarımları incelenmiştir. Sekiz nodül özelliği, beş farklı öznelik grubu, dört farklı sınıflayıcı ve bir veri dengeleme yaklaşımı ile $8 \times 5 \times 4 \times 2 = 320$ deney içeren bir desen oluşturulmuştur. Deneysel sonuçları farklı açılardan ele alınarak değerlendirilmiştir.

Literatürde sıkça kullanılmış geleneksel öznelikler, önceden eğitilmiş bir ESA’dan transfer öğrenme ile elde edilen derin öznelikler ve bunların kombinasyonları değerlendirmeye alınmıştır. Deneyler sonucunda, farklı sınıflama yaklaşımlarının ve farklı öznelik gruplarının her nodül özelliği için ayrı ayrı ele alınması gerektiği çıkarımı yapılmıştır. Hatta sınıflayıcıların başarımlarını ölçütleri düzeyinde değişkenlikler gözlemlenmekte ve belli bir sınıflayıcıyı tüm ölçütlerde öne çıkaran sonuçlar elde edilememektedir.

LIDC veri tabanından elde edilen nodül özelliği veri kümeleri dengesiz bir dağılıma sahiptir. Bu dengesiz dağılım baskın olan sınıfın yüksek oranda ve azınlıkta olan sınıfın daha düşük oranda tahmin edilmesine yol açmaktadır. Yüksek sınıflama başarımları ve düşük duyarlılık ölçümleri bu çıkarımı doğrulamaktadır. Veri kümeleri dengelendiğinde ise, sınıflama başarımının düşmesi ve duyarlılık değerinin yükselmesi, azınlıkta olan sınıfların daha iyi tahmin edildiğini ve baskın olan sınıfın görece tahmin oranının düştüğünü göstermektedir. Azınlıkta olan sınıfta gerçekleşen artış, sınıflama başarımına düşük oranda katkı sağlarken, duyarlılık oranına yüksek oranda katkı sağlamaktadır. Bunun nedeni, sınıflama başarımının tüm veri kümesindeki örneklerin toplamı üzerinden hesaplanması, duyarlılığın ise sınıf bazlı hesaplanarak ortalamalarının alınmasıdır. Belirlilik değeri ise dengeleme işleminden fazla

etkilenmemektedir. Bu durum, her iki koşulda da (dengelenmiş/dengelenmemiş veri) hatalı olumsuz oranının düşük olmasından kaynaklanmaktadır. Duyarlılık derecesinin düşük olmasının başka bir nedeni de, nodül özelliklerinin değerlendirmelerinin yüksek ölçekli (LIDC için 1-5 arası) olması sonucu belirsizliğin artması, duyarlılık ve sınıflama başarımları ölçümünün düşmesidir[10].

Sınıflama yöntemleri açısından bakıldığında, Rastgele Orman sınıflayıcı birçok nodül özelliğinde yüksek sınıflama başarımları sağlamıştır. Destek Vektör Makineleri ise duyarlılık değerinde yüksek oranda başarımları sahiptir. Belirlilik değerinde ise sınıflayıcılar arasında dikkate değer bir farklılık yoktur.

Sonuçlar incelendiğinde, farklı nodül özellikleri ve farklı başarımları ölçütleri bağlamında tahmin mekanizmaları oluşturulurken tek bir model yapısı kullanmak yerine farklı stratejilerin uygulanmasının daha doğru olacağı çıkarımı yapılabilir. Bu çalışma, çeşitli mekanizmaları ve bunların kombinasyonlarını inceleme açısından literatüre katkıda bulunacaktır. İleriki çalışmalarda, uzmanların anlaşma oranlarındaki farklılıkların da incelenmesi, derin sınıflama yaklaşımlarının doğrudan uyarlanması düşünülmektedir.

KAYNAKLAR (REFERENCES)

1. Ferlay J., Soerjomataram I., Dikshit R., Eser S., Mathers C., Rebelo M., Parkin D.M., Forman D., Bray F., Cancer incidence and mortality worldwide: Sources, methods and major patterns in GLOBOCAN 2012, *International Journal of Cancer*, 136 (5), E359-E386, 2015.
2. Austin J.H.M., Muller N.L., Friedman P.J., Hansell D.M., Naidich D.P., RemyJardin,M., Webb,W.R., Zerhouni,E.A., Glossary of terms for CT of the lungs: Recommendations of the Nomenclature Committee of the Fleischner Society, *Radiology*, 200 (2), 327-331, 1996.
3. Armato S.G., McLennan G., Bidaut L., McNitt-Gray M.F., Meyer C.R., Reeves A.P., Zhao, B.S., Aberle, D.R., Henschke C.I., Hoffman E.A., Kazerooni E.A., MacMahon H., van Beek E.J.R., Yankelevitz D., Biancardi A.M., Bland P.H., Brown M.S., Engelmann R.M., Laderach G.E., Max D., Pais R.C., Qing D.P.Y., Roberts R.Y., Smith A.R., Starkey A., Batra P., Caligiuri, P., Farooqi, A., Gladish, G.W., Jude, C.M., Munden, R.F., Petkovska, I., Quint, L.E., Schwartz, L.H., Sundaram B., Dodd, L.E., Fenimore, C., Gur, D., Petrick N., Freymann J., Kirby, J., Hughes, B., Castele, A.V., Gupta, S., Sallam, M., Heath, M.D., Kuhn, M.H., Dharaiya, E., Burns, R., Fryd, D.S., Salganicoff, M., Anand, V., Shreter, U., Vastagh, S., Croft, B.Y., Clarke, L.P., The Lung Image Database Consortium, (LIDC) and Image Database Resource Initiative (IDRI): A Completed Reference Database of Lung Nodules on CT Scans, *Medical Physics*, 38 (2), 915-931, 2011.
4. van Klaveren R.J., Oudkerk M., Prokop M., Scholten, E.T., Nackaerts, K., Vernhout, R., van Iersel, C.A., van den Bergh K.A.M., van't Westeinde S., van der Aalst C., Thunnissen, E., Xu, D.M., Wang, Y., Zhao, Y.R., Gietema, H.A., de Hoop, B.J., Groen, H.J.M., de Bock, G.H., van Ooijen, P., Weenink, C., Verschakelen, J., Lammers, J.W.J., Timens, W., Willebrand D., Vink, A., Mali, W., de Koning, H.J., Management of Lung Nodules Detected by Volume CT Scanning, *New England Journal of Medicine*, 361 (23), 2221-2229, 2009.
5. Jabon S.A., Raicu D.S., Furst J.D., Content-based versus semantic-based retrieval: an LIDC case study, *Medical Imaging 2009: Image Perception, Observer Performance, and Technology Assessment*. International Society for Optics and Photonics, 72631L, 2009.
6. Junior J.R.F., Oliveira, M.C., de Azevedo-Marques, P.M., Cloud-based NoSQL open database of pulmonary nodules for computer-aided lung cancer diagnosis and reproducible research, *Journal of digital imaging*, 29 (6), 716-729, 2016.
7. Vinay K., Rao, A., Kumar, G.H., Predication of Lung Nodule Characteristic Rating using Best Classifier Model, *International Journal of Computer Applications*, 56 (18), 2012.
8. Zinovev D., Furst J., Raicu D., Building an ensemble of probabilistic classifiers for lung nodule interpretation, *IEEE 10th International Conference on Machine Learning and Applications and Workshops*, Honolulu, ABD, 155-161, 18-21 Aralık, 2011.
9. Zinovev D., Raicu D., Furst J., Armato III S.G., Predicting radiological panel opinions using a panel of machine learning classifiers, *Algorithms*, 2 (4), 1473-1502, 2009.
10. Kaya A., Can A.B., A weighted rule based method for predicting malignancy of pulmonary nodules by nodule characteristics, *Journal of Biomedical Informatics*, 56, 69-79, 2015.
11. Kaya,A., Can A.B., eFis: A Fuzzy Inference Method for Predicting Malignancy of Small Pulmonary Nodules Image Analysis and Recognition. *ICIAR 2014. Lecture Notes in Computer Science*, Springer, 8815, 2014.
12. Li G., Kim H., Tan, J.K., Ishikawa, S., Hirano, Y., Kido, S., Tachibana, R., Semantic characteristics prediction of pulmonary nodule using Artificial Neural Networks, *35th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, Osaka, Japonya, 5465-5468, 3-7 Haziran, 2013.
13. Lee M.C., Boroczky L., Sungur-Stasik, K., Cann, A.D., Borczuk, A.C., Kawut, S.M., Powell, C.A., Computer-aided diagnosis of pulmonary nodules using a two-step approach for feature selection and classifier ensemble construction, *Artificial Intelligence in Medicine*, 50(1) 43-53, 2010.
14. Zhao L., Lee M.C., Boroczky L., Vloemans V., Opfer, R., Comparison of computer-aided diagnosis performance and radiologist readings on the LIDC pulmonary nodule dataset, *International Society for Optics and Photonics Medical Imaging*, 6915, 69151I 2008.

15. El-Baz A., Beache, G.M., Gimel'farb, G., Suzuki, K., Okada, K., Elnakib, A., Soliman, A., Abdollahi, B., Computer-aided diagnosis systems for lung cancer: challenges and methodologies, *International Journal of Biomedical Imaging*, 2013, 942353, 2013.
16. Hua K.L., Hsu, C.H., Hidayati, H.C., Cheng, W.H., Chen, Y.J., Computer-aided classification of lung nodules on computed tomography images via deep learning technique, *Oncotargets and Therapy*, 8, 2015-2022, 2015.
17. Roth H.R., Lu, L., Liu, J.M., Yao, J.H., Seff, A., Cherry, K., Kim, L., Summers, R.M., Improving Computer-Aided Detection Using Convolutional Neural Networks and Random View Aggregation, *IEEE Transactions on Medical Imaging*, 35 (5), 1170-1181, 2016.
18. Dou Q., Chen, H., Yu, L.Q., Zhao, L., Qin, J., Wang, D.F., Mok, V.C.T., Shi, L., Heng, P.A., Automatic Detection of Cerebral Microbleeds From MR Images via 3D Convolutional Neural Networks, *IEEE Transactions on Medical Imaging*, 35 (5), 1182-1195, 2016.
19. Cheng J.Z., Ni, D., Chou, Y.H., Qin, J., Tiu, C.M., Chang, Y.C., Huang, C.S., Shen, D., Chen, C.M., Computer-Aided Diagnosis with Deep Learning Architecture: Applications to Breast Lesions in US Images and Pulmonary Nodules in CT Scans, *Scientific Reports*, 6, 2016.
20. Pan W.K., A survey of transfer learning for collaborative recommendation with auxiliary data, *Neurocomputing*, 177, 447-453, 2016.
21. Shin H.C., Roth, H.R., Gao, M.C., Lu, L., Xu, Z.Y., Nogues, I., Yao, J.H., Mollura, D., Summers, R.M., Deep Convolutional Neural Networks for Computer-Aided Detection: CNN Architectures, Dataset Characteristics and Transfer Learning, *IEEE Transactions on Medical Imaging*, 35 (5), 1285-1298, 2016.
22. Deng J., Dong, W., Socher, R., Li, L.J., Li, K., Li, F.F., ImageNet: A Large-Scale Hierarchical Image Database, *IEEE Conference on Computer Vision and Pattern Recognition*, 248-255, Miami, ABD, 20-25 Haziran, 2009.
23. Krizhevsky A., Sutskever, I., Hinton, G.E., Imagenet classification with deep convolutional neural networks, *25th International Conference on Neural Information Processing Systems*, Siem Reap, Kamboçya, 1097-1105, 14-16 Aralık, 2012.
24. Cortes C., Vapnik, V., Support-Vector Networks, *Machine Learning*, 20 (3), 273-297, 1995.
25. Breiman L., Random forests, *Machine Learning*, 45 (1), 5-32, 2001.
26. Fix E., Hodges, J.L., Discriminatory Analysis - Nonparametric Discrimination - Consistency Properties, *International Statistical Review*, 57 (3), 238-247, 1989.
27. Freund Y., Schapire, R.E., A decision-theoretic generalization of on-line learning and an application to boosting, *Journal of Computer and System Sciences*, 55 (1), 119-139, 1997.
28. He H., Bai, Y., Garcia, E.A., Li, S., ADASYN: Adaptive synthetic sampling approach for imbalanced learning, *IEEE International Joint Conference on Neural Networks*, (IEEE World Congress on Computational Intelligence), Hong Kong, Çin, 1-8 Haziran, 2008.
29. Haralick R.M., Shapiro, L.G., *Computer and Robot Vision*, 1st Edition, Addison-Wesley, 1992.
30. Shen W., Zhou, M., Yang, F., Dong, D., Yang, C., Zang, Y., Tian, J., Learning from Experts: Developing Transferable Deep Features for Patient-Level Lung Cancer Prediction, *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2016: 19th International Conference*, Atina, Yunanistan, 124-131, 17-21 Kasım, 2016.
31. Raykar V.C., Yu, S.P., Zhao, L.H., Valadez, G.H., Florin, C., Bogoni, L., Moy, L., Learning From Crowds, *Journal of Machine Learning Research*, 11, 1297-1322, 2010.
32. Kononenko I., Simec, E., RobnikSikonja, M., Overcoming the myopia of inductive learning algorithms with RELIEFF, *Applied Intelligence*, 7 (1), 39-55, 1997.
33. Celik C., Bilge, H.S., Feature Selection with Weighted Conditional Mutual Information, *Journal of the Faculty of Engineering and Architecture of Gazi University*, 30 (4), 585-596, 2015.
34. Akben S.B., Alkan A., Density-Based Feature Extraction to Improve the Classification Performance in the Datasets Having Low Correlation between Attributes, *Journal of the Faculty of Engineering and Architecture of Gazi University*, 30 (4), 597-603, 2015.
35. Ochs R., Kim H.J., Angel, E., Panknin, C., McNitt-Graya, M., Brown, M., Forming a reference standard from LIDC data: Impact of reader agreement on reported CAD performance - art. no. 65142A, *Medical Imaging 2007: Computer-Aided Diagnosis*, Pts 1 and 2, 6514, A5142-A5142, 2007.
36. Xie Y., Zhang, J., Xia, Y., Fulham, M., Zhang, Y., Fusing texture, shape and deep model-learned information at decision level for automated classification of lung nodules on chest CT, *Information Fusion*, 42, 102-110, 2018.

