

DEMEK: Çok Amaçlı Sayısal Mamografi Referans Veri Kümesi

Ozan AKÇAY^{a,1}, Adil ALPKOÇAK^b, Pınar BALCI^c, Oğuz DİCLE^c

^a Dokuz Eylül Üniversitesi Sağlık Bilimleri Enstitüsü Biyomekanik Anabilim Dalı, İzmir

^b Dokuz Eylül Üniversitesi Mühendislik Fakültesi Bilgisayar Mühendisliği Bölümü, İzmir

^c Dokuz Eylül Üniversitesi Tıp Fakültesi Dahili Tıp Bilimleri Bölümü Radyodiagnostik Anabilim Dalı, İzmir

MuDMaDs: A Multipurpose Digital Mammography Reference Dataset

Abstract. In this work, we present a reference dataset for human and machine learning in mammographic researches. The Dataset is called MuDMaDs, for “a Multipurpose Digital Mammography Reference Dataset”. To accomplish this work, an ontology that includes all the details of mammography is created at first. Then, the candidate cases to be in the dataset are automatically selected by a boolean information retrieval system retrospectively from the cases that have been examined from 2004 to 2008 in the department of Radiodiagnostics at Dokuz Eylül University Hospital. In the final step, the mammograms of these cases are examined by the domain experts to select the cases to be in the dataset and the necessary corrections and annotations over them are done by using the application developed.

Keywords. Mammography Dataset, Training set, Mammographic Ontology, OWL

Özet. Bu çalışmada, mamografi alanında insan ve makine eğitimi ile ilgili çalışmalarda yararlanılabilecek referans niteliğinde bir veri kümesinin hazırlanması amaçlandı. Oluşturulan ürüne kısaca DEMEK (Dokuz Eylül Mamografi Eğitim Kümesi) adı verildi. Bunun için öncelikle, mamografi ile ilgili tüm ayrıntıları kapsayacak bir ontoloji hazırlandı. Ardından geriye dönük olarak 2004–2008 yılları arasında Dokuz Eylül Üniversitesi Tıp Fakültesi Hastanesi Radyodiagnostik Anabilim Dalı’nda yapılan mamografi incelemeleri arasından veri kümesinde yer alması için aday olarak seçilen olgular, oluşturulan mantıksal bilgi erişim sistemi aracılığı ile otomatik bir yöntemle belirlendi. Daha sonra konunun uzmanları tarafından bu olguların mamogramları, oluşturulan yazılım kullanılarak gözden geçirilip gerekli düzeltme ve betimlemelerin yapılmasıyla veri kümesinde yer alacak olguların belirlenmesi amaçlandı.

Anahtar Sözcükler. Mamografi Veri kümesi, Eğitim kümesi, Mamografi ontolojisi

¹ Sorumlu Yazar: Dokuz Eylül Üniversitesi Sağlık Bilimleri Enstitüsü, Biyomekanik AD, İnciraltı, İzmir; E-posta: ozan.akcay@deu.edu.tr, ozan.akcay@deu-medical.info.

Giriş

Gelişen teknoloji ile birlikte artık röntgen filmleri veya film tarama cihazları, yerlerini hızla yüksek çözünürlüklü bilgisayar ekranlarına bırakmakta [1]. Sadece çekim teknikleri değil, görüntülerin incelenmesinde veya tanınmasında bilgisayar desteği sağlayan yazılımlar da gün geçtikçe gelişmekte. Ancak literatüre bakıldığında, bu konuda yapılmakta olan güncel yayınlarda, halen 1980lerin ortasında hazırlanmış veri kümelerinden faydalanılmakta olduğu görülmekte. Hem incelediği değişkenlerin az oluşu hem de içerdiği görüntülerin kalite sorunları yönünden günümüz teknolojisinin gerisinde kalmış bu veri kümelerinden bazılarının artık erişim bile bulunmamakta. Bugünün teknolojisiyle geliştirilmiş, daha çok sayıda ve güncel değişkenlerle anlamlandırılmış incelemeler barındıran, referans niteliğinde bir mamografi veri kümesinin, yazılım geliştiricilerin yararlanacağı iyi bir kaynak olmanın yanında, bu alanda yapılacak çalışmaların da daha nesnel bir şekilde ve karşılaştırmalı olarak değerlendirilmesini sağlayacağı bir gerçektir.

Bu çalışmanın amacı, mamografi alanındaki araştırmalarda gerek insan gerekse makine eğitimi için kullanılabilecek, mamografi incelemelerinde karşılaşılabilecek hemen her çeşit olguya örnek içeren, herkes tarafından erişilebilir ve evrensel ölçekte anlaşılabilir bir veri kümesi hazırlamaktır.

Bu makalenin bundan sonraki bölümleri şu şekilde düzenlenmiştir: İkinci bölümde geçmişte hazırlanmış belli başlı veri kümelerinin kısa bir incelemesi sunulmuştur. Üçüncü bölümde öncelikle farklı amaçlara hizmet edecek bir mamografi veri kümesinde olması arzu edilen temel özellikler sunularak, bu çalışmada hazırlanan veri kümesinin genel özellikleri, hazırlanma süreçleri gibi ayrıntılara yer verilmiştir. Dördüncü bölüm ise veri kümesi hakkında bir tartışma sunarak çalışmayı sonlandırmakta ve bu alanda gelecekte yapılabilecek çalışmalara yol göstermektedir.

Benzer Çalışmalar

Yapılan literatür taramasında, hem sayısallaştırılmış film hem de sayısal röntgen tabanlı veri kümelerinin oluşturulduğu, ancak çoğunluğunun güncelliğini yitirerek artık ulaşılabilirliğinin kalmadığı görülmüştür. Oluşturulan ilk veri kümelerinden olan **Nijmegen Sayısal Mamogram Veritabanı**, sadece 20 hastanın 40 adet sayısallaştırılmış mamografi görüntüsünü içerir. Resimler 12 bit CCD kamera ile etkin nokta çözünürlüğü 100 μm olacak şekilde sayısallaştırılmıştır. Her resim 2048x2048 nokta büyüklüğündedir. Tüm resimler en az bir mikrokalsifikasyon kümesi içerir ve kümede 7 iyi huylu, 13 kötü huylu lezyon örneği bulunur. Bu veri kümesi artık ulaşılabilir değildir [2]. Bir başka veri kümesi olan **Washington Üniversitesi Sayısal Mamografi Veritabanı**, 80 olgu için, biyopsi aracı tarafından memede lezyonun yerini belirlemek için çekilmiş tek bakış açılı sayısal mamografi resimlerinden oluşmaktadır. Veri kümesinde eşit sayıda iyi huylu ve kötü huylu lezyon bulunur ve içerdiği mikrokalsifikasyon kümesi ve kitle örneği sayıları da yine birbirine eşittir. Her resim 12 bit renk derinliğinde 512x512 nokta büyüklüğünde olup nokta çözünürlükleri 100 μm 'dir. Artık ulaşılamiyor olsa da bu veri kümesi, sayısal olarak çekilmiş ve zamanında herkes tarafından FTP aracılığıyla ulaşılabilen ilk örneklerdendir [3]. Bir başka örnek, **OWH (Office of Women's Health) Veritabanı**, 900 incelemede sağ ve sol memenin CC ve MLO bakış açısında çekilmiş ve 50 μm noktasal çözünürlükle 12 bit renk derinliğinde taranmış resimlerden oluşur. 540 normal olgu, 180 adet iyi huylu

ve 180 adet kötü huylu lezyon içeren incelemeden oluşan veri kümesinde ayrıca lezyonun yerini, özelliklerini ve patolojik sonucunu belirten bilgiler de yer almaktadır. **(Mini-)MIAS (Mammographic Image Analysis Society) Veritabanı**, 161 olguda sağ ve sol memenin MLO bakış açısıyla çekilmiş görüntüsünü içeren toplam 322 resim içerir. Orijinal veri kümesi artık ulaşılabilir olmadığından [2], bu veri kümesindeki resimlerin kenarları kesilip görüntüleri ortalanarak 1024×1024 nokta büyüklüğünde, 200 µm çözünürlüğe sahip resimlerden oluşan bu veritabanı oluşturulmuştur [4, 5]. Diğer bir örnek olan **LLNL/UCSF Veritabanı**, mikrokalsifikasyonlar üzerine çalışmak isteyenler için 50 hastanın sağ ve sol meme için CC ve ML bakış açılarında çekilip, 35 µm nokta çözünürlüğünde ve 12 bit renk derinliğinde sayısallaştırılmış toplam 197 mamografi görüntüsünü içerir [6]. Görüntülerin yanında, kalsifikasyon kümelerinin ve kümedeki belli başlı birkaç kalsifikasyonun sınırlarının işaretlendiği iki adet saydam dosyası ve olgu geçmişiyle incelemeyi yapan radyologun yorumları da yer alır [3]. Oluşturulmuş en geniş örneklerden olan **DDSM (Digital Database for Screening Mammography) Veritabanı**, 2620 inceleme için her hastanın sağ ve sol memesi için CC ve MLO bakış açısında çekilmiş resimlerini içerir [7]. Her inceleme için hastanın yaşı, mamogramın çekim tarihi, filmin sayısallaştırıldığı tarih ve uzman bir radyolog tarafından belirlenmiş ACR meme yoğunluk değeri bilgileri ve bunların yanında, normal olmayan her olgu için lezyonun işaretlendiği bir doğrulama dosyası, ve lezyon hakkında uzman bir radyolog tarafından yapılmış BI-RADS değerlendirmesi ve zorluk derecesi yer almaktadır. En güncel örneklerden olan **GPCALMA (Grid Platform for a Computer-Aided Library in MAMmography) Veritabanı** ise 1999'da geliştirilmeye başlanmış ve 967 hastanın toplam 3369 sayısallaştırılmış mamogramını içerir [8]. 1997–2002 arasında toplanan filmler tek bir CCD tarayıcı ile 2067×2657 nokta boyutlarında 85 µm çözünürlükte ve 12 bit renk derinliğinde sayısallaştırılmıştır. Bu aşamada filmlerin çekim değerlerine ulaşamadığından normalizasyon yapılamamıştır. Veri tabanında uzman radyologlar tarafından yapılmış meme dokusu, lezyon varlığı, yeri ve türü değerlendirmeleri ve hastanın yaşı, demografik bilgisi ve takip eden incelemeleri de yer almaktadır [9].

1. Gereç ve Yöntem

Mamografi veri kümesinin hazırlanma aşamasına geçmeden önce, ideal bir veri kümesinde hangi özelliklerin olması gerektiğini belirlerken yardımcı olması amacıyla, bundan önce yapılmış benzer çalışmalarda oluşturulan veri kümelerinin sahip olduğu özellikler incelendi. Buna göre, bir sayısal mamografi veri kümesi oluştururken dikkate alınacak başlıca özellikler aşağıdaki gibi sıralanabilir:

Uygun olgu seçimi: Klinik uygulama sırasında karşılaşılabilecek her türden olguyu içermelidir. Normal (lezyonsuz) olguların yanında, meme kanserinin göstergesi olan tüm lezyon tiplerine örnek içermelidir. Normal olgular arasında lezyonlarla güçlükle ayırt edilebilecek meme yapısına sahip örnekler bulunabilir. Örnekler mamografide deneyimli radyologlar tarafından toplanmalı, her olgu standart bakış açılarında çekilmiş 4 adet resim içermelidir (her iki meme için CC ve MLO) [10, 11].

Her olgu için doğru değerlendirme: Her olgu için biyopsi ile doğrulama yapılmış olmalı, biyopsi gerektirmeyen durumlarda (normal olgular) kişinin en az üç yıllık izlemesinin yapılmış ve durumu değişmemiş olması gerekmektedir [12]. Lezyonlu olgular için de aynı şekilde en az 3 yıllık izleme yapılmış ve durumlarının kesin kanıtlanmış olması gerekmektedir. Zaman ve maliyet açısından düşünüldüğünde hiç

pratik bir yöntem olmasa da, biyopsi yapılmış olan durumlarda, patoloğların her zaman tanı koymada fikir birliğine sahip olmadıkları göz önünde bulundurularak, iki ayrı hastaneden iki patoloğın görüşleri istenmeli, uyuşmazlık durumunda bağımsız üçüncü bir patoloğın yardımına başvurulmalıdır. Lezyonun yeri ve sınırları her zaman nesnel olarak belirlenemez. Bu demektir ki sayısal görüntü üzerinde yer alan her pikselin bir lezyonu gösteriyor olma olasılığı vardır. En doğru kararın verilmesi için izlenecek yöntem, bir grup radyolog lezyonların yer ve sınırlarını belirledikten sonra yapılan işaretlemelerin ortalamasının alınmasıdır.

Her olguya eşlik eden bilgiler: Hastanın klinik öz geçmişi ve soy geçmişi, inceleme sırasındaki yaşı, varsa önceki biyopsi sonuçları ve kanserin ilerleme aşamasının [13] da yer alması faydalı olacaktır. İdeal durumda, mamografi uzmanlarından oluşan bir grup radyoloğın, lezyonların zorluk derecelerini ve meme yoğunluğunu saptamaları gerekir. Meme yoğunluğu saptanırken BI-RADS ölçeği [14] veya memenin % kaç glandüler doku içerdiğinin belirlenmesi yeterli olacaktır. Lezyonun zorluk derecesinin belirlenmesinde ise, 0–5 arası bir ölçek kullanılabilir. Bu ölçeğe göre, 0, görünür lezyon içermeyen; 1, sadece mamografi konusunda uzman radyologlarca belirlenebilen lezyonlar içeren; 2, uzman bir radyoloğın belirleyebildiği lezyonlar içeren; 3, iyi mamografi eğitimi almış bir gözlemci tarafından belirlenebilen lezyonlar içeren; 4, kısıtlı mamografi eğitimi almış gözlemci tarafından belirlenebilen lezyonlar içeren; 5 ise, eğitim almamış bir gözlemci tarafından bile belirlenebilecek kadar bariz lezyonlar içeren görüntüleri belirtir [15].

Sayısal görüntünün özellikleri: Teknik olarak kusursuz bir mamografi görüntüsü için görüntünün çekim özelliklerini belirleyen çekim süresi ve x-ışını enerji miktarının yanında, sayısallaştırılmış dosyanın da uygun renk derinliğine ve nokta çözünürlüğüne sahip olması ve kayıpsız bir şekilde saklanması gereklidir. Sayısal mamografide, kalsifikasyon gibi ince detayların yakalanabilmesi için, nokta büyüklüğünün 40–60 µm olması uygundur [16].

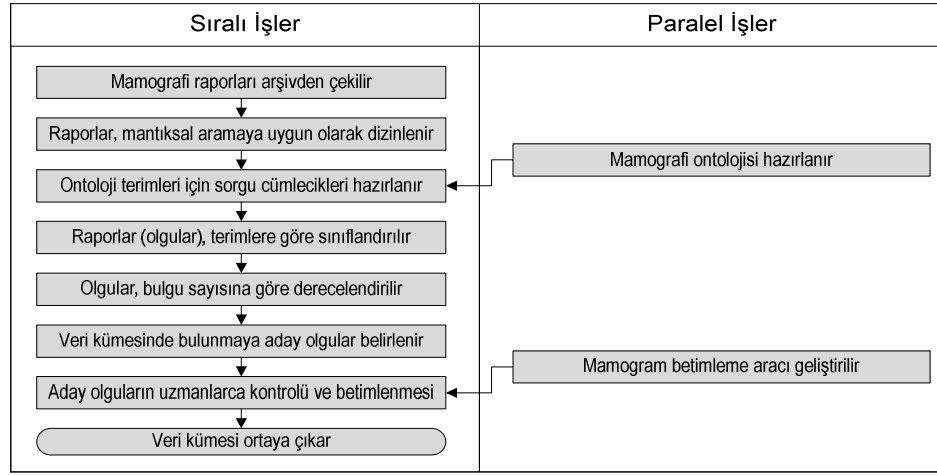
Veri kümesinin yapısı: Veri kümesinin alt kümelere ayrılmasında olası bir yöntem biri eğitim, diğeri sınama amaçlı, iki küme oluşturulmasıdır. Eğitim kümesi ne kadar büyük olursa, BDT de o kadar tutarlı olur. Ancak bu, eğitilen sistemi sınamak için geriye az miktarda verinin kalmasına neden olacağından yapılan testlerin güvenilirliği azalacaktır.

Veri kümesinin dağıtılması: Oluşturulan veri kümesi, internet üzerinden herkesin erişimine açık olmalıdır. Böylece, denemesi yapılan BDT sisteminin veya önerilecek sınıflandırma algoritmalarının verimliliklerini ölçmek ve başka çalışmalarda aynı amaçla oluşturulmuş benzerleri ile kıyaslama yapabilmek için kullanılacak bir veri kümesi bulunmuş olur. Veri kümesinin dağıtılması aynı zamanda, bu çalışmada ortaya konulacak sonuçların başka araştırmacılar tarafından da denenip doğrulanmasını sağlayacaktır.

Hazırlanan veri kümesindeki olgular, geriye dönük olarak Dokuz Eylül Üniversitesi Tıp Fakültesi Hastanesi Radyodiyagnostik Anabilim Dalı'nda, 2004 yılından 2008 yılı Kasım ayına kadar yapılmış incelemeler arasından seçildi. Her olgu için standart bakış açıları olan CC ve MLO ile çekilmiş DICOM formatındaki dörder adet görüntünün (bazı olgular için daha az veya daha çok) yanında, görüntülerde yer alan lezyonların yerlerini ve çeşitli özelliklerini gösteren belli bir biçime sahip ek işaretleme bilgileri ve yapılan incelemeye ait radyolojik değerlendirme raporu yer alacaktır. Bu verilerin tamamı, hastaların ve hekimlerin kimlik bilgilerinden arındırılmıştır.

Veri kümesi hazırlanırken, içeriğini anlamsal olarak yapılandırmak için öncelikle bir ontolojiye ihtiyaç duyuldu. Bunun için, konuyla ilgili önceden hazırlanmış olan ontolojiler incelendi ancak ihtiyacı tam olarak karşılamadıkları görüldüğü için yeni bir tane hazırlandı. Ontoloji kullanılarak yapılacak betimleme ile böylece, olgulardaki bulguların neyi ifade ettiği hem insan hem de makine tarafından anlamlandırılabilir hale getirildi. Hazırlanan ontolojinin bu konuyla ilgili mümkün olan en geniş kitle tarafından anlaşılır ve karşılaştırılabilir olması amacıyla, temeli, American College of Radiology (ACR)'nin yayınladığı BI-RADS değerlendirme ölçeğine dayandırıldı.

Veri kümesinde hangi olguların yer alacağıının belirlenmesi sırasında izlenen yol Şekil 1'deki akış diyagramında gösterilmiştir.



Şekil 1. Olgu seçimi akış diyagramı

Buna göre, Dokuz Eylül Üniversitesi Tıp Fakültesi Hastanesi Radyodiyagnostik Anabilim Dalı'nda, 2004 yılından 2008 yılı Kasım ayına kadar geçen sürede yazılmış olan 50 bine yakın (tam olarak 49874 adet) mamografi raporu, içeriklerinde metin bazlı arama olanağı sunmayan RADOS sisteminden kopyalanarak üzerlerinde mantıksal arama yapılabilecek şekilde dizinlendikleri ve ayrıntıları ilgili makalede [17] belirtilmiş olan bir bilgi erişim sistemine (kısaca "*BES*" veya literatürdeki İngilizce karşılığıyla "*Information Retrieval System*") aktarıldı. Böylece, geliştirilen bu sistem ve arayüzler kullanılarak, rapor metinleri üzerinde *VE*, *VEYA*, *DEĞİL* gibi mantıksal bağlaçlar ve -olası yazım yanlışlarına takılmamak için- *joker ifadeler* yoluyla *ve/veya* sözcüklerin birbirlerine *görece konumlarına* göre çeşitli sorgular yapılabilmektedir. Örnek olması açısından

Tablo 1'de bazı sınıflar için yazılan sorgu cümlecikleri görülmektedir.

Tablo 1. Bazı sınıflar için örnek MBES sorgu cümlecikleri

Sınıf	MBES Sorgusu
kitle şekli	Yuvarlak
	lob?l*
	şekilsiz / düzensiz
	şekils* OR bel?rs* OR d?zens*
kitle kenar özelliği	Oval
	ovo* OR oval*
	düzgün / keskin
	((d?zg?n /2 s?n?rl*) OR (s?n?rl* /2 d?zg?n) OR (bel?rg?n /2 s?n?rl*) OR (kesk?n /2 s?n?rl*) OR (s?n?rl* /2 bel?r*) OR (s?n?rl* /2 kesk*)) NOT (olmayan OR belirsiz)
	Mikrolobüler
	mikrolob* OR mikrolob*
	silik / parankimle örtülü
kitle yoğunluğu	(parank?m* /5 s?n?r*) OR (s?n?r* /5 parank?m*)
	sınırı tanımlanamayan / düzensiz / belirsiz
	(d?zens?z /3 s?n?rl*) OR (s?n?rl* /3 d?zens?z) OR (bel?rs?z /3 s?n?rl*) OR (s?n?rl* /3 bel?rs?z)
	ışınsal / spiküle
	ışınsal OR sp?k?l*
	düzgün konturlu
	(d?zg?n* /2 kont?r*) OR (kont?r* /2 d?zg?n*)
kitle yoğunluğu	yüksek yoğunluklu
	hiperdens OR (y?kse* /4 d?ns*) OR (d?ns* /4 y?kse*) OR (y?kse* /4 yo?un*) OR (yo?un* /2 y?kse*)
	eş yoğunluklu / izodens
	iz*d?ns* OR (e? /2 yo?un*)
kitle yoğunluğu	düşük yoğunluklu
	hipodens OR (d???k* /4 d?ns*) OR (d?ns* /4 d???k*) OR (d???k* /4 yo?un*) OR (yo?un* /2 d???k*)
kitle yoğunluğu	yağ içerikli / radyolüsent
	(ya? /1 içeri*) OR rad*ol*n* NOT radyod?ns*

BES kullanılarak, önceden oluşturulmuş olan ontolojide yer alan terimlerin ifade ettiği yapısal veya şekilsel özellik, lezyon veya bulguları barındıran raporların belirlenmesi amacıyla, ontolojideki her terim için, konunun uzmanlarıyla ortak bir çalışma yürütülerek, uygun sorgu cümlecikleri hazırlandı. Bu sorgu cümlecikleri hazırlanırken ilgili terim için mümkün olan en fazla sayıda sonuca ulaşılması hedeflendi. Bu aşamadan sonra, geliştirilen bu yöntemler yardımıyla, büyük çoğunluğunun BI-RADS'a uygun bir biçimde yazıldığı bilinen raporların ait olduğu olguların ontolojideki her terimle doğrudan ilgili olup olmadığına göre sınıflandırmaları yapıldı. Tablo 2'de her sınıfın ilişkilendirildiği olgu sayıları görülmektedir.

Tablo 2. Ontolojiyle uyumlu ortaya çıkan sınıflar ve ilişkili olgu sayıları

sıra no	sınıf adı	olgu sayısı
1	meme dokusu - yağ dokusundan zengin parankim	13051
2	meme dokusu - yağ ve fibroglandüler doku içeren	20421
3	meme dokusu – heterojen dens	(*)
4	meme dokusu - çok dens	5093
5	kitle şekli - yuvarlak	329
6	kitle şekli - lobüler	2220
7	kitle şekli - şekilsiz/düzensiz	952
8	kitle şekli - oval	1104
9	kitle kenar - düzgün / keskin	3973
10	kitle kenar - mikrolobüler	57
11	kitle kenar - silik / parankimle örtülü	16
12	kitle kenar - sınırı tanımlanamayan / düzensiz/belirsiz	668
13	kitle kenar - ırsal / spiküle	497
14	kitle kenar - düzgün konturlu	1772
15	kitle yoğunluğu - yüksek yoğunluklu	193
16	kitle yoğunluğu - eş yoğunluklu / izodens	14
17	kitle yoğunluğu - düşük yoğunluklu	770
18	kitle yoğunluğu - yağ içerikli / radyolüsent	323
19	benign - cilt/deride / dermal	29
20	benign - damar / vasküler	97
21	benign - kaba /patlamış mısır	1021
22	benign - çubuk şeklinde	11
23	benign - yuvarlak / round	57
24	benign - merkezi lucent-centered	9
25	benign - eggshell or ring	13
26	benign - milk of calcium	2
27	benign - sutur	41
28	benign - distrofik	137
29	benign - noktasal / punktat	153
30	orta derece şüpheli - amorf / belirsiz	128
31	orta derece şüpheli - kaba heterojen	(²)
32	malign - ince pleomorfik	83
33	malign - ince lineer / dallanan / çizgisel	16
34	kalsifikasyon dağılımı - kümeli / grup halinde	494
35	kalsifikasyon dağılımı - segmental	79
36	kalsifikasyon dağılımı - bölgesel	42
37	kalsifikasyon dağılımı - yaygın / dağınmık / saçılmış	1734
38	özel durumlar - tübüler dansite	25

² Bu sayılar çalışmanın ilgili aşaması sonuçlandıktan sonra ortaya çıkacaktır.

39	özel durumlar - intramammar lenf bezi	22
40	özel durumlar - asimetrik meme dokusu	133
41	özel durumlar - fokal asimetrik dansite	27
42	eşlik eden bulgu - deride çekintiler	22
43	eşlik eden bulgu - meme başı çekintisi	41
44	eşlik eden bulgu - deri kalınlaşması	23
45	eşlik eden bulgu - trabekular kalınlaşma	26
46	eşlik eden bulgu - cilt lezyonu	38
47	eşlik eden bulgu - aksiller lenf bezi	27
48	eşlik eden bulgu - yapısal bozulma	12

Gelinen bu noktada, uzman bir gözün süzgecinden geçmeden, otomatik bir yöntemle yapılan sınıflandırmanın tüm bulgu türlerini %100 başarıyla ayırt etmiş olduğunu beklemek doğru olmaz. Bu nedenle, uzmanların olabildiğince çok ontolojik terimle ilişkilendirilmiş raporların arasından en uygunlarını veri kümesi için seçmelerini sağlamak amacıyla, her rapor için, ilgili olduğu ontoloji terimlerinin çokluğuna göre bir derecelendirme yapıldı. Bu derecelendirmeye “*raporun içerdiği bulgu*” sayısı daha fazla olanlar öncelikli olacak şekilde her sınıf için en fazla 150 adet olgu belirlenerek uzmanların bu önerilen olgular arasından seçim yapmaları amaçlandı.

Veri kümesinde yer alması için toplamda, birden çok sınıf için önerilmiş olanları ayrı ayrı sayarsak 4077 olgu önerildi (2058 eşsiz olgu). Önerilen bu olgulara ait mamogramların uzman radyologlar tarafından bir daha incelenmesi yoluyla, bu aşamaya kadar bilgisayar tarafından yapılan sınıflandırmanın doğruluğunun ölçülmesinin yanında, geliştirilmiş olan bir mamografi gösterim ve betimleme aracını kullanarak radyologların sınıflandırmalarıyla ilgili olası düzeltmelerde bulunmaları, mamogramlar üzerinde gereken işaretleme ve betimlemeleri yapmaları ve veri kümesinin içinde yer alacak olguların hangileri olacağına karar vererek kümeye son şeklini vermeleri amaçlandı.

Geliştirilen betimleme uygulaması sayısal mamografi alanında piyasada kullanılan herhangi bir gösterim aracının taşıması olası temel özelliklere ve bu çalışmaya özgü özel yeteneklere sahiptir. Bunlar arasında dosya sisteminden veya bir DICOM sunucudan çağırarak resim göstermek, gösterim sırasında resmin daha rahat incelenmesi için parlaklık ve zıtlık değerleri üzerinde kolayca anlık değişiklik yapmak, resmin belli bir bölgesini daire, dörtgen veya çokgen kullanarak seçmek gibi özelliklerin yanında, oluşturulan ontoloji yardımı ile (ontolojiden bağımsız bir şekilde) seçilen bölgenin özelliklerinin betimlemesini yapabilmek, resmin önceden yapılmış olan sınıflandırmasını değiştirmek veya resmi incelenen olgunun veri kümesinde yer alıp almayacağını işaretlemek gibi bu çalışmaya özgü yetenekler barındırır. Uygulama kullanılarak yapılan bu işaretleme ve betimlemeler veri tabanına kaydedilir ve istendiğinde bir XML dosyası olarak da alınabilir.

Tablo 3. Veri kümelerinin genel karşılaştırması (FT: Film Tarama; SR: Sayısal Röntgen; Resim Sayısı hanesindeki $a \times b = c$ denkleminde a: olgu başına çekim sayısı, b: olgu sayısı, c: veri kümesindeki bildirilen resim sayısı)

Veri Kümesi	Elde Tipi	Nokta Çözün (µm)	Renk Derinliği (bit)	Resim Boyutu (nokta)	Patolojisine göre olgu sayısı				Sınıf Sayısı	Resim Sayısı	Ulaşırlılık
					Normal	İyi Huylu	Kötü Huylu	Toplam			
Nijmegen	FT	100	12	2048×2048	0	7	13	20	?	2×20=40	×
WÜ	SR	100	12	512×512	0	40	40	80	?	1×80=80	×

OWH	FT	50	12	?	540	180	180	900	?	4×900	?
Mini-MIAS	FT	200	8	1024×1024	?	?	?	161	12	2×161 =322	Web
LLNL/UCSF	FT	35	12	?	10	32	8	50	5	4×50 ≈397	Posta
DDSM	FT	42-50	12-16	?	695	1011	914	2620	4	4×2620	Web
GPCAL MA	FT	85	12	2067×2657	306	661		967	19	3369	?
DEMEK	SR	70	12	3328×4096	?	?		?	48	?	Web

Çalışmanın gelinen aşamasında, veri kümesi için önerilen olguların resimlerinin uzmanlar tarafından incelenmesi sırasında hastanede kurulu olan radyoloji işletim sistemine bağlı olarak çalışılmadı. Bunun yerine, önerilen olgu kümesinin resimleri, önceden, radyoloji işletim sisteminin ilgili arayüzleri kullanılarak topluca kopyalandı ve betimleme uygulamasının bu resimler üzerinde çalışması sağlandı. Böylece, büyük boyutlu resim verilerine yerel disk üzerinden erişerek, hem yüksek iş yoğunluğu olan bir sunucuya ve yerel ağa bağlanarak yapılması muhtemel olumsuz etkilerin veya istemsiz veri kaybının, hem de kullanıcıya zaman kaybettirilmesinin önüne geçildi.

Hazırlamakta olan veri kümesinde, gelinen aşamaya kadar ağırlıklı olarak lezyonlu olguların seçilmesi üzerinde duruldu. Veri kümesinin hazırlanması devam etmekte ve henüz giriş bölümünde belirtilen “normal” nitelikli örneklerin toplanması aşamasına geçilmedi. Normal örneklerin belirlenmesinde öncelikle üzerinde durulacak nokta, lezyonsuz olmasına rağmen, lezyonlu örneklerden güçlükle ayırt edilebilir görünüm özelliklerine sahip olguların bulunup veri kümesine eklenmesi olacaktır.

Hazırlanmakta olan veri kümesinin daha önce yapılmış benzerlerinin olmasına rağmen, hem öncüllerinin taşımadığı yeni niteliklere sahip olması için, hem de araştırma amaçlı bir mamografi veri kümesi oluşturulurken göz önünde bulundurulması gereken noktaları eksiksiz kapsayacak şekilde hazırlanması için ileri seviye çaba harcandı. Tablo 3’de, literatürde yer alan belli başlı mamografi veri kümelerinin yapısal özellikleri ile DEMEK karşılaştırmalı olarak verilmiştir.

2. Tartışma ve Sonuç

Gelinen aşamada veri kümesine girmeye aday olarak, uzman radyologlar tarafından mamogramları kontrol edilerek gerekli düzeltmelerden ve işaretlemelerden geçirilmesi için 2058 olgu belirlendi. Bu aşama tamamlandıktan sonra toplamda yer alan 48 farklı sınıf için uygun örnekler belirlenmiş ve veri kümesine son şekli verilmiş olacaktır.

Mamografi veri kümesi oluşturulurken, içinde hangi olguların yer alması gerektiğine karar vermek kadar, içereceği resimlerin de yüksek kaliteli ve hatasız olarak toplanmış olması önemlidir. Şimdiye kadar yapılmış çalışmalar incelendiğinde, çalışmaların büyük çoğunluğundaki resimlerin film tarama yöntemiyle toplanmış olduğu, bu nedenle tarama hatalarına neden olacak dış etkilerden tamamıyla kaçınılmadığı görülmektedir. Bu çalışmada hazırlanmakta olan veri kümesinde ise resimler sayısal röntgen aracılığıyla elde edilmiş olduğu için, hem daha yüksek bir zıtlık (“contrast”) değerine sahiptir, hem de öncekilerde karşılaşılan bilgi kaybı sorunlarını içermez [18].

Şu an söz konusu alanda erişilebilir olan veri kümelerinin büyük çoğunluğu geçen yüzyılda hazırlanmıştır. Dolayısıyla, günümüz teknolojisinden yoksun yeteneklerle

elde edilmiş, nokta boyutu, nokta çokluğu, renk derinliği gibi özellikleri bakımından görece düşük standartlı resimler içermektedir. Günümüzde sayısal mamografi kullanımı sayıca film tabanlı kurulumlara göre hâlen az olmasına rağmen, özellikle son 5 yılda büyük bir hızla yaygınlaşmıştır [19]. Bu durum, bu yöntemle elde edilmiş referans alınabilecek bir veri kümesine olan ihtiyacı ortaya koyarak bu çalışma için kuvvetli bir motivasyon oluşturmaktadır.

Bu çalışmada yeni bir mamografi ontolojisi oluşturulup, bir mamografi veri kümesinde ilk defa bir ontoloji ile uyumlu 48 farklı özellik belirlenerek, içerdiği olgular ve bulguları bu özelliklere göre sınıflandırılmıştır. Önceki hiçbir çalışmada, sunulan veri bu kadar çok sayıda özelliğe göre sınıflandırılmamış ve bu çalışmada hedeflenen betimleme ayrıntısına ulaşılamamıştır. Bu durum, yapılan bu çalışmanın en kuvvetli yönlerinden biridir.

Teşekkür

Bu çalışma, Türkiye Bilimsel ve Teknik Araştırma Kurumu, TÜBİTAK, tarafından EEEAG/107E217 numaralı araştırma projesi kapsamında desteklenmiştir.

Kaynakça

- [1] Rao VM, Levin DC, Parker L, Frangos AJ, Recent trends in mammography utilization in the Medicare population: is there a cause for concern? J Am Coll Radiol. 2008 May;5(5):652-6
- [2] <http://marathon.csee.usf.edu/Mammography/OtherResources.html>
- [3] Nishikawa RM, Mammographic databases. Breast Dis. 1998 Aug;10(3-4):137-50.
- [4] <http://peipa.essex.ac.uk/info/mias.html>
- [5] Suckling J, et al, The Mammographic Image Analysis Society Digital Mammogram Database Exerpta Medica. International Congress Series 1994;1069:375-378.
- [6] Ashby, A.E.; Hernandez, J.M.; Logan, C.M.; Mascio, L.N.; Frankel, S.; Kegelmeyer, W.P., Jr., "UCSF/LLNL high resolution digital mammogram library," Engineering in Medicine and Biology Society, 1995., IEEE 17th Annual Conference, pp.539-540 vol.1, 20-25 Sep 1995
- [7] The Digital Database for Screening Mammography, Michael Heath, Kevin Bowyer, Daniel Kopans, Richard Moore and W. Philip Kegelmeyer, in Proceedings of the Fifth International Workshop on Digital Mammography, M.J. Yaffe, ed., 212-218, Medical Physics Publishing, 2001. ISBN 1-930524-00-5.
- [8] Lauria A, GPCALMA: Implementation in Italian hospitals of a computer aided detection system for breast lesions by mammography examination, Physica Medica (2008), doi:10.1016/j.ejmp.2008.05.002
- [9] Lauria A, et al. GPCALMA: An Italian Mammographic Database of Digitized Images for Research. In: Susan M. Astley et al. (Eds.): IWDM 2006, LNCS 4046, pp. 384-391, 2006.
- [10] Bassett LW, Gold RH. Breast radiography using the oblique projection. Radiology. 1983 Nov;149(2):585-7.
- [11] Eklund GW, Cardenosa G. The art of mammographic positioning. Radiol Clin North Am. 1992 Jan;30(1):21-53.
- [12] Breast Screening Frequency Trial Group. The frequency of breast cancer screening: results from the UKCCCR Randomised Trial. United Kingdom Co-ordinating Committee on Cancer Research. Eur J Cancer. 2002 Jul;38(11):1458-64.
- [13] http://en.wikipedia.org/wiki/TNM_staging
- [14] The American College of Radiology Breast Imaging Reporting and Data System (BI-RADS™), Fourth Edition, http://www.acr.org/SecondaryMainMenuCategories/quality_safety/BIRADSAtlas.aspx
- [15] Schmidt RA, Nishikawa RM, Osnis R, Screibman KL, Giger ML, Doi K. Computerized detection of lesions missed by mammography. In: Doi K, Giger ML, Nishikawa RM, Schmidt RA, eds. Digital Mammography '96. Amsterdam, Elsevier Science; 1996:105-110.

- [16] Wirth M, Lyon J, Fraschini M, Nikitenko D. "The Effect of Mammogram Databases on Algorithm Performance." in: Proceedings of 17th IEEE Symposium on Computer-Based Medical Systems (CBMS '04). p. 15, 2004.
- [17] Berber T, Alpkoçak A, Mamografi Raporları için Mantıksal Bilgi Erişim Sistemi, 14. Biyomedikal Mühendisliği Ulusal Toplantısı (BIYOMUT'09), İzmir, 2009
- [18] Obenauer S, Luftner-Nagel S, von Heyden D, Munzel U, Baum F, Grabbe E. Screen film vs full-field digital mammography: image quality, detectability and characterization of lesions. *Eur Radiol.* 2002 Jul;12(7):1697–702.
- [19] <http://www.fda.gov/cdrh/mammography/archives/scorecard-statistics-archive.html>.