



**TIBBİ GÖRÜNTÜ VERİ SETLERİNDE MATLAB İLE
BÜYÜK VERİ TEKNİKLERİ UYGULAMASI**

Eissa Faisal Ali AL-ZABIDI

**YÜKSEK LİSANS TEZİ
BİLGİSAYAR MÜHENDİSLİĞİ ANA BİLİM DALI**

**GAZİ ÜNİVERSİTESİ
FEN BİLİMLERİ ENSTİTÜSÜ**

EYLÜL 2019

Eissa Faisal Ali AL-ZABIDI tarafından hazırlanan “TIBBİ GÖRÜNTÜ VERİ SETLERİNDE MATLAB İLE BÜYÜK VERİ TEKNİKLERİ UYGULAMASI” adlı tez çalışması aşağıdaki jüri tarafından OY BİRLİĞİ ile Gazi Üniversitesi Bilgisayar Mühendisliği Ana Bilim Dalında YÜKSEK LİSANS TEZİ olarak kabul edilmiştir.

Danışman: Dr.Öğr.Üyesi Uraz YAVANOĞLU

Bilgisayar Mühendisliği Ana Bilim Dalı, Gazi Üniversitesi

Bu tezin, kapsam ve kalite olarak Yüksek Lisans Tezi olduğunu onaylıyorum.

.....

Başkan: Prof. Dr. M. Ali AKÇAYOL

Bilgisayar Mühendisliği Ana Bilim Dalı, Gazi Üniversitesi

Bu tezin, kapsam ve kalite olarak Yüksek Lisans Tezi olduğunu onaylıyorum.

.....

Üye: Dr.Öğr.Üyesi Adnan ÖZSOY

Bilgisayar Mühendisliği Ana Bilim Dalı, Hacettepe Üniversitesi

Bu tezin, kapsam ve kalite olarak Yüksek Lisans Tezi olduğunu onaylıyorum.

.....

Tez Savunma Tarihi: 19/09/2019

Jüri tarafından kabul edilen bu tezin Yüksek Lisans Tezi olması için gerekli şartları yerine getirdiğini onaylıyorum.

.....

Prof. Dr. Sena YAŞYERLİ

Fen Bilimleri Enstitüsü Müdürü

ETİK BEYAN

Gazi Üniversitesi Fen Bilimleri Enstitüsü Tez Yazım Kurallarına uygun olarak hazırladığım bu tez çalışmada;

- Tez içinde sunduğum verileri, bilgileri ve dokümanları akademik ve etik kurallar çerçevesinde elde ettiğimi,
- Tüm bilgi, belge, değerlendirme ve sonuçları bilimsel etik ve ahlak kurallarına uygun olarak sunduğumu,
- Tez çalışmada yararlandığım eserlerin tümüne uygun atıfta bulunarak kaynak gösterdiğimi,
- Kullanılan verilerde herhangi bir değişiklik yapmadığımı,
- Bu tezde sunduğum çalışmanın özgün olduğunu,

bildirir, aksi bir durumda aleyhime doğabilecek tüm hak kayıplarını kabullendiğimi beyan ederim.

Eissa Faisal Ali AL-ZABIDI

19/09/2019

TIBBİ GÖRÜNTÜ VERİ SETLERİNDE MATLAB İLE BÜYÜK VERİ TEKNİKLERİ UYGULAMASI

(Yüksek Lisans Tezi)

Eissa Faisal Ali AL-ZABIDI

GAZİ ÜNİVERSİTESİ
FEN BİLİMLERİ ENSTİTÜSÜ

Eylül 2019

ÖZET

Son yıllarda tıbbi görüntüler, hastalık teşhisi, tıbbi araştırma ve eğitimde önemli bir rol oynamaktadır. Günümüzde teknolojinin gelişmesiyle tıbbi görüntü çeşitleri de artmıştır. Bu artan görüntülerin geleneksel yöntemlerle etkin ve verimli bir şekilde depolanması, aranması, paylaşımı ve analiz edilmesi için de zorluklar ortaya çıkmıştır. Bu çalışmada, MATLAB ile Hadoop MapReduce kullanarak akciğer nodüllerinin iyi huylu ya da kötü huylu olup olmadığının sınıflandırılması için Bilgisayar Destekli Tanı (Computer Aided Diagnosis - CAD) sistemini önerdik. Sistem, radyologlar tarafından Bilgisayarlı Tomografi (Computed Tomography - CT) görüntü veri setinden benzer nodüllerin görüntülerini kullanıcıya aktarmaktadır. Tıbbi görüntü veri seti, Hadoop'un dağıtılmış dosya sistemi'nde (Hadoop Distributed File System - HDFS) saklanıp, görüntülere MATLAB aracılığıyla erişim sağlanmaktadır. MapReduce kullanılarak çoklu özniteliklerin paralel olarak elde edilmesi için İçerik-tabanlı Görüntü Erişimi (Content-based Image Retrieval - CBIR) yaklaşımı kullanılmıştır. Şekil, doku ve radyolog sayısal görüntü (Radiologist Quantified Image - RQI) öznitelikleri dâhil olmak üzere en uygun 26 öznitelik kümesi tanımlanmıştır. Bu öznitelikleri değerlendirmek için farklı SVM ve KNN sınıflandırıcıları kullandık. Deney sonuçlarımız, MATLAB'ta yerel MapReduce yönteminin, Hadoop küme MapReduce yönteminden ~ 5 kat daha hızlı olduğunu göstermektedir. SVM, sınıflandırıcılara göre %95.27 doğruluk (accuracy), %96.32 anma (recall), %94.76 kesinlik (precision) ve %95.53 F-skoru oranıyla daha başarılı sonuçlar elde edilmiştir.

Bilim Kodu : 92414

Anahtar Kelimeler : Tıbbi Görüntü, Öznitelik Çıkarma, CBIR, Hadoop, HDFS, MapReduce, MATLAB

Sayfa Adedi : 63

Danışman : Dr. Öğr. Üyesi Uraz YAVANOĞLU

APPLICATION OF BIG DATA TECHNIQUES FOR MEDICAL IMAGES DATASETS
WITH MATLAB
(M. Sc. Thesis)

Eissa Faisal Ali AL-ZABIDI

GAZİ UNIVERSITY
GRADUATE SCHOOL OF NATURAL AND APPLIED SCIENCES

September 2019

ABSTRACT

In the recent years, medical images played an important role in disease diagnosis, medical research and education. Today with the advances in technology in medical image capture devices and the huge increase in the number of medical images, challenges have emerged to retrieve and store relevant images to the user query efficiently and accurately. Furthermore, it is impossible to maintain and handle with this image using the conventional method. In this work, we presented a Computer-Aided Diagnosis (CAD) system for classifying lung nodules as either benign or malignant by using Hadoop MapReduce with MATLAB. The system retrieves images of similar nodules from a big publicly available Computed Tomography (CT) images dataset annotated by radiologists. They are stored in Hadoop Distributed File System (HDFS) and accessed via the MATLAB. A content-based image retrieval (CBIR) approach was used in order to extract multiple features parallelly by using the MapReduce. An optimal set of 26 features was defined, including shape, texture and Radiologist Quantified Image (RQI) features. We used different SVM and KNN classifiers to evaluate these features. Our experimental results show that the Local MapReduce method is ~5 times faster than Hadoop cluster MapReduce method in MATLAB. The proposed method improves the retrieval result to 95.27%, 96.32%, 94.76% and 95.53% in accuracy, recall, precision and F-score respectively using the SVM classifier.

Science Code : 92414

Key Words : Medical Image, Feature Extraction, CBIR, Hadoop, HDFS, MapReduce, MATLAB

Page Number : 63

Supervisor : Asst. Prof. Dr. Uraz YAVANOĞLU

TEŞEKKÜR

Bu çalışmanın gerçekleştirilmesinde, yüksek lisans eğitimim süresince bana cesaret veren saygıdeğer danışman hocam Dr. Öğr. Üyesi Uraz YAVANOĞLU'nun katkıları olmasaydı bu çalışma gerçekleştirilemezdi. O ve Gazi Üniversitesi'ndeki diğer değerli hocalarım, beni yönlendirdiler ve bana destek oldular, elimden gelenin en iyisini yapmam dışında başka bir alternatifini kabul etmediler. Onların katkıları olmasaydı, çalışma şimdi olduğu gibi tam olamazdı. Hepsine teşekkürlerimi sunuyorum.

Eşim Sabah SAAD'a da en samimi teşekkürlerimi sunuyorum. Nezaketi ve sabrı bu çalışmayı bitirmeme büyük katkı sağladı. Hayatımın en değerli varlıkları olan anne ve babama da hayatım boyunca verdikleri destek için hassaten teşekkür ederim. Tüm bunları mümkün kılan Tanrı'ya da şükürler olsun.

İÇİNDEKİLER

	Sayfa
ÖZET	iv
ABSTRACT.....	v
TEŞEKKÜR.....	vi
İÇİNDEKİLER	vii
ÇİZELGELERİN LİSTESİ.....	x
ŞEKİLLERİN LİSTESİ.....	xi
RESİMLERİN LİSTESİ	xii
SİMGELER VE KISALTMALAR.....	xiii
1. GİRİŞ.....	1
2. TANIMLAR VE KAVRAMLAR	5
2.1. Büyük Veri	5
2.2. Hadoop	6
2.2.1. Hadoop mimarisi.....	6
2.2.2. Hadoop'un dağıtılmış dosya sistemi (HDFS)	7
2.2.3. MapReduce	8
2.2.4. MATLAB ile Hadoop MapReduce	9
2.3. İçerik-tabanlı Görüntü Erişimi (CBIR)	10
2.4. Bilgisayar Destekli Tanı Sistemi (CAD).....	11
3. MEVCUT UYGULAMALAR (İLGİLİ ÇALIŞMALAR).....	13
3.1. Bilgisayar Destekli Tanı (CAD) Teknikleri ile Yapılan Çalışmalar	13
3.2. Görüntü Erişim Teknikleri ile Yapılan Çalışmalar	17
3.3. Hadoop Üzerinden Görüntü Erişim Teknikleri ile Yapılan Çalışmalar	17
4. ÖNERİLEN ÇÖZÜM.....	21

	Sayfa
4.1. Görüntüler Yükleme Aşaması.....	21
4.2. Görüntüler Erişim Aşaması.....	22
4.3. Akciğer CT Görüntü Veri Seti	23
4.4. Akciğer Nodül Görüntülerinin Ön-işlemesi	24
4.5. Öznitelik Çıkarımı	25
4.5.1. Radyolog sayısal görüntü öznitelikler (RQI)	26
4.5.2. 2D-şekil tabanlı öznitelik çıkarımı.....	27
4.5.3. GLCM tabanlı doku özniteliği çıkarımı.....	28
4.5.4. LBP tabanlı öznitelik çıkarımı	29
4.6. Sınıflandırma Algoritmaları	30
4.6.1. Destek vektör makinesi (SVM).....	31
4.6.2. K-en yakın komşu (KNN).....	31
4.6.3. Hedef oluşturma	32
4.6.4. K-katmanlı çapraz doğrulama	32
4.6.5. Karışıklık matrisi.....	32
4.6.6. Doğruluk değeri	33
4.6.7. Kesinlik değeri	34
4.6.8. Anma değeri	34
4.6.9. F-skoru değeri	34
4.6.10. ROC eğrisi.....	34
5. ÖNERİLEN SİSTEMİN UYGULANMASI.....	37
5.1. Sistem İhtiyaç Özellikleri.....	37
5.1.1. Donanım.....	37
5.1.2. Yazılım.....	38

	Sayfa
5.2. Sistemin Uygulanması	38
5.2.1. Veri seti oluřturma ve ykleme iřlemi.....	39
5.2.2. Grntler eriřim ve sınıflandırma iřlemi	41
6. NERİLEN SİSTEMİN DEĞERLENDİRİLMESİ	45
6.1. Ykleme ve Eriřim Srelerinin Sonuları.....	45
6.2. nerilen Sistemin Sınıflandırma Sonuları	46
6.3. Deney Sonularının Literatrdeki alıřmalarla Karřılařtırılması	49
7. SONU	53
KAYNAKLAR	55
ZGEMİř	63

ÇİZELGELERİN LİSTESİ

Çizelge	Sayfa
Çizelge 4.1. LIDC veri setinin dağılımı.....	23
Çizelge 4.2. LIDC veri setinde yer alan akciğer nodülleri üzerindeki durum ve dilimlerin dağılımı	25
Çizelge 4.3. LIDC veri setindeki radyolojistler ve ilişkili derecelendirme sistemi tarafından açıklanmalı özellikler.....	26
Çizelge 4.4. LIDC veri setinde bulunan akciğer nodül tipleri üzerindeki bulgu ve dilimlerin dağılımı	30
Çizelge 4.5. Karmaşıklık matrisi	33
Çizelge 5.1. Küme konfigürasyonu	37
Çizelge 5.2. Yazılım konfigürasyonu	38
Çizelge 6.1. Görüntüler yükleme ve erişim deneyleri çalışma süreleri	45
Çizelge 6.2. Tüm öznitelikler SVM ve KNN'nin sınıflandırmasının arasında en iyi özniteliklerin test edilmesi	46
Çizelge 6.3. Farklı literatürlerin performans karşılaştırması	50

ŞEKİLLERİN LİSTESİ

Şekil	Sayfa
Şekil 2.1. Hadoop mimarisi	6
Şekil 2.2. HDFS mimarisi	8
Şekil 2.3. MapReduce veri akışı	9
Şekil 2.4. MATLAB ile Hadoop mimarisi	10
Şekil 4.1. Önerilen model sistemin mimarisi	21
Şekil 4.2. Önerilen model sisteminin görüntüler yükleme aşaması	22
Şekil 4.3. Önerilen model sisteminin görüntüler erişim ve sınıflandırma aşaması	23
Şekil 4.4. LIDC veri setinden alınan nodül örneği ve dört radyoloji uzmanından verilen tanısal RQI öznitelik değerleri	24
Şekil 4.5. Akciğer nodülü görüntü ön-işleme ve öznitelik çıkarım diyagramı	26
Şekil 6.1. SVM algoritması test sonuçları	48
Şekil 6.2. Algoritma doğruluk oranı grafiği	48
Şekil 6.3. ROC eğrisi SVM ve KNN sınıflandırma grafiği	49

RESİMLERİN LİSTESİ

Resim	Sayfa
Resim 5.1. Grafiksel kullanıcı arayüzünün tasarımı (GUI)	38
Resim 5.2. Veri seti oluşturma ve yükleme işlemi	39
Resim 5.3 MATLAB'ta MapReduce aşamalı işlemi	40
Resim 5.4. HDFS’de sıralanmış (.mat) dosyaları formatı olarak veri seti görüntüleyin	40
Resim 5.5. HDFS’de sıralanmış (.dcm) görüntüleri formatı olarak veri seti görüntüleyin	41
Resim 5.6. Görüntüler seçme işlemi	41
Resim 5.7. Görüntüler sınıflandırma işlemi.....	42
Resim 5.8. Görüntüler SVM sınıflandırıcı değerlendirmesi	42
Resim 5.9. Görüntüler KNN sınıflandırıcı değerlendirmesi.....	43
Resim 5.10. Görüntüler erişim ve benzer olan görüntüler göstermesi	43

SİMGELER VE KISALTMALAR

Bu çalışmada kullanılmış simgeler ve kısaltmalar, açıklamaları ile birlikte aşağıda sunulmuştur.

Simgeler

Açıklamalar

2D

İki Boyutlu (Two Dimensional)

3D

Üç Boyutlu (Three Dimensional)

Kısaltmalar

Açıklamalar

AUC

Eğri Altındaki Alanı (Area Under the Curve)

BEMD

Bidimensional Ampirik Mod Ayrışımı

CAD

Bilgisayar Destekli Tanı (Computer Aided Diagnosis)

CBIR

İçerik-tabanlı Görüntü Erişimi

CBMIR

İçerik-tabanlı Tıbbi Görüntü Erişimi

CT

Bilgisayarlı Tomografi (Computed Tomography)

DICOM

Tıpta Dijital Görüntüleme ve İletişim

DFT

Ayrık Fourier Dönüşümü (Discrete Fourier Transform)

FDA

Amerikan Gıda ve İlaç Dairesi

FN

Yanlış Negatif (False Negative)

FNIH

Ulusal Sağlık Enstitüleri Vakfı

FP

Yanlış Pozitif (False Positive)

GDI

Grafik Aygıt Arabirimi (Graphics Device Interface)

GGD

Genelleştirilmiş Gauss Yoğunluğu

GLCM

Gri Seviye Eş-Oluşum Matrisi

GUI

Grafiksel Kullanıcı Arayüzü

HD

Sabit Disk (HardDisk)

HDFS

Hadoop'un dağıtılmış dosya sistemi

KNN

K-en yakın komşu (K-Nearest Neighbor)

LBP

Yerel İkili Örüntü (Local Binary Pattern)

LIDC

Akciğer Görüntü Veritabanı Konsorsiyumu

LGGP

Yerel Gri Gabor Deseni (Local Gray Gabor Pattern)

Kısaltmalar**Açıklamalar****LTrP**

Yerel Tetra Örüntü (Local Tetra Pattern)

MATLAB

MATris LABoratuvarı

MR

MapReduce

MRI

Manyetik Rezonans Görüntüleme

NCI

Ulusal Kanser Enstitüsü (National Cancer Institute)

PACS

Görüntü Arşivleme ve İletişim Sistemi

PC

Kişisel Bilgisayar (Personal Computer)

RAM

Rasgele Erişilebilir Hafıza (Random Access Memory)

ROC

Alıcı Operatör Özellikleri

RQI

Radyolog sayısal görüntü öznitelikler

SIFT

Ölçek Değişimsiz Öznitelik Dönüşümü

SVM

Destek vektör makinesi (Support Vector Machines)

TN

Doğru Negatif (True Negative)

TP

Doğru Pozitif (True Positive)

YARN

Yine Başka Bir Kaynak Müzakerecisi

XML

Genişletilebilir İşaretleme Dili

1. GİRİŞ

Tıbbi görüntüleme, hastalık teşhisinde, tıbbi araştırmalarda ve eğitimde özellikle son yıllarda önemli bir rol oynamıştır. Günümüzdeki, tıbbi görüntü alma cihazlarında sağlanan teknolojik ilerlemelerle birlikte, dünyada üretilen ve depolanan tıbbi görüntü miktarı katlanarak artmıştır [1] Bu büyük miktardaki görüntü gerçekten yararlı olsa da, ilgili görüntülerin saklanması ve kullanıcı sorgusuna alınması zor olabilmektedir. Dahası, bu miktarda bir görüntüyü geleneksel yöntemi kullanarak kolayca saklamak ve aramak mümkün değildir. Ancak, tıbbi görüntü erişimi, görüntülerin karmaşık yapısı nedeniyle tıbbi alanda daha zordur. Bunun için bu görüntülerin bir kullanıcı sorgusuna yüklemek ve erişimi için hızlı ve güvenli bir tekniğe ihtiyaç vardır [2].

Her bir çalışma için kullanılan tıbbi görüntü verisi birkaç megabayttan yüzlerce megabayta kadar değişebilmektedir [3]. ImageCLEF’de üretilen tıbbi görüntülerin hacim ve hızı ile ilgili ilginç bir örnek vermek gerekir ise, 2005 ve 2007 yılları arasında görüntülenen toplam görüntü sayısı 66.000 civarında iken; 2013 yılında her gün ortalama 300.000 görüntü kaydedilmiştir [4]. Bu görüntülerin depolanması ve analizi için büyük depolama kapasitesinin yanı sıra hızlı ve doğru algoritmalar gerekmektedir. Artan görüntü hacmine ek olarak, görüntülerin boyut ve kalitesi vb. bakımından da farklılık göstermekte olup, bu ise, özellikle veri entegrasyonu ve veri madenciliği gibi yeni zorlukları gündeme getirmektedir [5]. Bu çalışmada büyük hacimli tıbbi görüntü veri setlerine odaklanılmaktadır [6].

Modern tıpta farklı türlerin geliştirildiği tıbbi görüntüleme (Bilgisayarlı Tomografi (CT), Manyetik Rezonans Görüntüleme (MRI) vb.) önemli ilerlemeler kaydetmiştir. Günümüzde, bu tıbbi görüntüleme teknikleri tanı, tedavi, değerlendirme ve planlama amacıyla kullanılmaktadır [7]. Sadece bir CT muayenesi için alınan görüntülerin sayısının ve boyutunun bir kaç gigabayta kadar çıkabildiği bu durumda biz betimlenen içeriğin yorumlanabilirliğinin iyileştirilmesini teminen tıbbi görüntü analizlerinde büyük veri kavramlarının uygulanmasını önermekteyiz [2].

Tıbbi görüntü analizi, sağlık sektörü alanında olağanüstü bir üstünlüğe sahiptir [8]. Tıbbi görüntülemelerde, CT filmlerdeki görüntüleri yakalamak için çekici alanlar kullanan

filtreleme mekanizmalarından biridir. Akciğer kanseri, dünya çapında büyük kanser ölümlerine neden olan tehlikeli hastalıklardan biridir. Akciğer kanserinin erken teşhisi, bir hastanın hayatta kalma şansını arttırmanın tek yoludur. Kanser başlangıç evrelerinde teşhis edilirse sağkalım oranı artar. Akciğer kanserinin erken teşhisi basit bir iş değildir. Hastaların yaklaşık % 80'ine yalnızca kanser merkezinde ve ortaya çıkan evrelerinde etkili bir şekilde tanı konmaktadır [9]. Akciğer kanseri sayısı küresel olarak kadınlarda 2., erkeklerde ise 10. sırada yer almaktadır. Bu çalışmalarda verilen bilgiler, dört temel aşamayı içeren akciğer kanseri yerleşim çerçevesinin genel bir tasviridir. Akciğer kanseri, kadınlarda meme ve kolorektal kanserlerden sonra en sık görülen üçüncü kanserdir [10].

Öznitelik çıkarma işlemi, görüntü işlemede en basit ve etkili boyutluluk azaltma tekniklerinden biridir [11]. Çıkarılan öznitelikler, girilen verilerden ilgili bilgiyi azaltma yöntemiyle çeker. Akciğer kanseri sınıflandırması için mevcut araştırma çalışmaları ANN, SVM ve KNN gibi kullanılarak yapılmıştır. SVM, istatistiksel öğrenme hipotezine dayanan evrensel bir faydalı öğrenme yöntemidir [12]. Bununla birlikte, bu teknikler pahalıdır ve hayatta kalma şansının çok düşük olmasından dolayı akciğer kanserini ileri aşamalarında tespit eder. Kanserın erken teşhisi, hastalığın tamamen iyileştirilmesinde yardımcı olabilir. Bu nedenle erken evrede kanserli nodül oluşumunu saptamak için bir teknik geliştirme gerekliliği artmaktadır [13].

Problem tanımı

Günümüzde tıbbi cihazlar teknolojilerinin gelişmesiyle, üretilen görüntü verilerinin miktarı her geçen gün aşırı bir şekilde artıyor. Geleneksel yöntemi kullanarak bu görüntüyü kolayca saklamak ve aramak mümkün değildir. Ancak, tıbbi görüntü erişimi, görüntülerin karmaşık yapısı nedeniyle tıbbi alanda daha zordur. Doğru özniteliklerin seçimi, benzerlik ölçüm kriterleri, bir endeksleme mekanizması ve sorgu formülasyon tekniği CBIR sistemlerinin tasarımında göz önünde bulundurulması gereken ana faktörlerdir. Dahası, her öznitelik görüntülerden çıkarılan sadece görüntü içeriğinin belirli bir yönünü karakterize eder, erişim performansını artırmak için birden fazla öznitelik kullanılması zorunludur. Bu arada, özel bir öznitelik farklı görüntülerin içeriğini yansıtırken farklı bir öneme sahip olduğundan, farklı görüntü sorguları için özel bir öznitelik eşit derecede önemli değildir [14]. Bunun için bu görüntüleri bir kullanıcı sorgusuna yüklemek ve görüntülere erişim için hızlı ve güvenli bir tekniğe ihtiyaç vardır.

Araştırmanın amacı ve önemi

Bu tez çalışmasının amacı, ilgili geçmiş durumları göstererek hekimlerin veya Radyolojistlere öznitelik çıkarmak için teşhis desteği sağlamaktır. Bu nedenle, genellikle paralel olarak kullanılan MATLAB ile Hadoop MapReduce platformunu kullanarak büyük miktar görüntü veri setinden benzer nodüller görüntüleri elde edip iyi huylu veya kötü huylu akciğer nodülleri sınıflandırması için basitleştirilmiş bir yaklaşım önerilmektedir [15]. Bu önerilen sistemi kullanarak sonucu en kısa sürede ve daha yüksek performans bulmaya çalışılmaktadır.

Varsayımlar

Akciğer kanseri tanısında bir öğrenme aracı olarak hareket edebilir. Radyoloğa da yardımcı olabilecektir. CBIR tabanlı CAD sisteminin geliştirilmesine odaklanmaktadır [16]. Çoğu zaman, doktorlar ve stajyer doktorlar, bu nodülleri güvenli bir şekilde kötü huylu veya iyi huylu olarak etiketlemek için deneyimli radyologlara bağımlı olmak zorundadır. İçerik-tabanlı görüntü erişim (CBIR) sistemi, bu tür stajyer radyologlara / doktorlara, veri setindeki benzer nodüllerin örneklerini göstererek kendi kendine öğrenmede yardımcı olabilir. Akciğer nodülleri için CBIR sisteminde, sorgu ve öznitelik veri seti oluşturma sırasında nodüllerin sınırlarını el ile çizmek için radyologların kullanması gereklidir. Bu tezde, görüntülerin içeriğinin algoritmalarına odaklanılmaktadır. MATLAB çerçeve platformu altında uygulanacaktır. MATLAB, akademi ve sanayide yaygın olarak kullanılan bir görüntü işleme ve grafik arayüz programıdır. Alternatif ve uygun maliyetli de bir tanı aracıdır. MATLAB uygulamalarda, grafiksel arayüzler kullanımı sayesinde önemli avantaj sağlamaktadır.

Tezin katkıları

Bu çalışmanın katkısı, CT akciğer kanseri görüntülerini sınıflandırmak için birkaç uzman radyolog yardımıyla veri setinden benzer nodüllerin örneklerini göstererek bir öğrenme aracı geliştirilmiştir.

Tıbbi Görüntülerin, farklı hastanelerde depolanmasını ve paylaşılmasını sağlamak için bir sistem geliştirilmiştir. Ayrıca, ağı Hadoop Küme'deki bir düğüm eklemek basittir. Dahası, MATLAB'taki etkileşim kolaylığı için arayüzü sağlanmıştır.

Bu tez çalışmasında geliştirilmiş olan MATLAB ile Hadoop MapReduce kullanarak tıbbi büyük miktar görüntü veri seti analizi uygulaması getirileri şunlardır:

- Akciğer CT görüntüleri veri setinden, içerik-tabanlı Görüntüler Erişim (CBIR) sistemi tasarımıdır.
- Sistem seçilen akciğer görüntülerini, nodüllerini ve çıkarılan öznitelikleri HDFS'ye yüklemektedir.
- Sistem, belirli bir sorgu nodülü için benzer nodülleri HDFS'den arayıp göstermektedir.
- Sistem, akciğer kanserlerinin sınıflandırılmasında da kullanılmaktadır.
- Sistemi, junior radyolog için eğitim aracı olarak kullanılabilir.

Bu tezin 2. bölümünde büyük veri, Hadoop, CBIR ve CAD kavramların tanımı yapılmış, 3. bölümünde ise ilgili çalışmalar takdim edilmiştir. Tezin 4. bölümünde önerilen çözüm sunulmuş, 5. bölümünde ise önerilen sistemin uygulaması gösterilmiştir. Tezin 6. bölümünde önerilen çözümün değerlendirmesi yapılmıştır. Son bölüm olan 7. bölümde elde edilen sonuçlar incelenmiştir.

2. TANIMLAR VE KAVRAMLAR

Bu bölümde, bir sonraki bölümlerde kullanılan yöntem ve araçların doğru anlaşılması için önemli kavramların teorik olarak tanımı yapılacaktır.

2.1. Büyük Veri

Büyük veri, geleneksel bilgisayar teknikleri kullanılarak işlenemeyen büyük, hızlı, karmaşık ve değişken büyük veri miktarlarını tanımlamak için kullanılan çok popüler bir terimdir. Bilgilerin yakalanması, depolanması, dağıtılması, yönetimi ve analizini sağlamak için gelişmiş teknolojiler gerektirmektedir. [17].

Büyük veri, genellikle Büyüklük (Volume), Çeşitlilik (Variety), Hız (Velocity) ve Kalite(Veracity) 4V olarak tanımlanmaktadır [18].

1. Büyüklük (Volume): Kuruluşlar veya bireyler tarafından terabaytlardan zettabaytlara kadar üretilen büyük verinin miktarını ifade eder. Bu bilgi, önemli bilgiler elde etmek için daha fazla kullanılmalıdır.
2. Hız (Velocity): Büyük verinin işlenebileceği zamanı ifade eder. Bazı aktiviteler çok önemlidir ve acil cevaplara ihtiyaç duyar, bu nedenle hızlı işlemenin verimliliğini maksimuma çıkarmasıdır.
3. Çeşitlilik (Variety): Büyük veriyi oluşturan farklı kaynaklardan elde edilen veri türünü ifade eder. Bu veriler, metin, ses, video, sayısal, görüntüler, sosyal medya verileri gibi farklı biçimler olabilir. Bu veriler yapısal, yarı yapısal ve yapısal olmayan üç şekilde bulunmaktadır.
4. Kalite (Veracity): Verilerdeki önyargı, gürültü ve anormalliği ifade eder. Kalite, büyük verinin depolanan ve çıkartılan analiz edilen önemli özniteliklerinden biridir. Veri analizindeki Kalite, hacim ve hız gibi diğer özniteliklerle karşılaştırıldığında en büyük zorluktur.

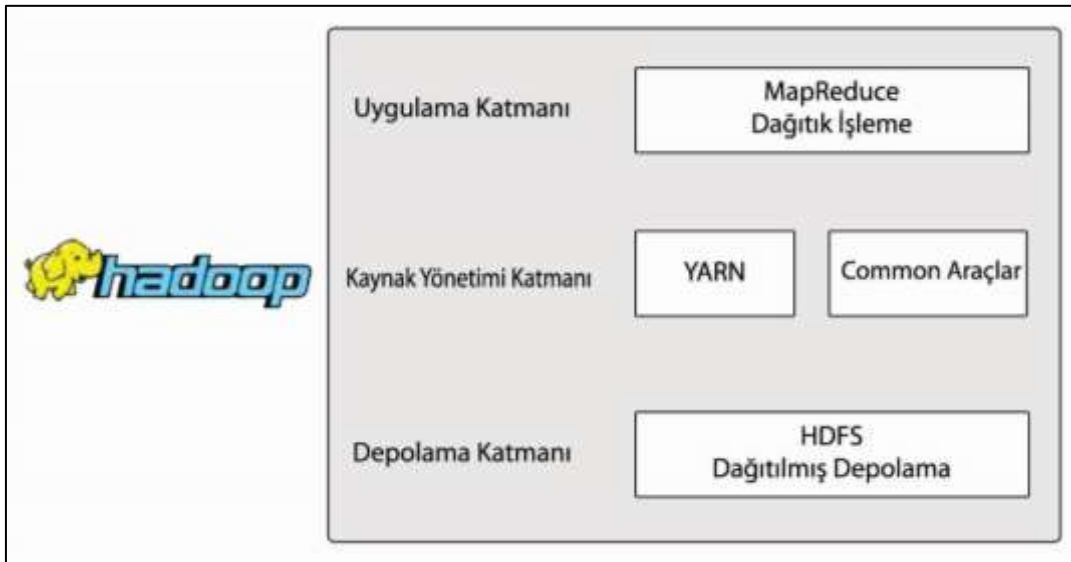
2.2. Hadoop

Apache Hadoop, Google tarafından geliştirilen MapReduce çerçevesinin açık kaynaklı bir uygulamasıdır. Apache Software Foundation ve Yahoo, 2004 yılında ilk sürümünü yayımladı ve yeni alt çalışmalarla çerçeveyi genişletmek için devam ediyor. Hadoop, basit programlama modelleri kullanarak büyük veri setlerinin bilgisayar kümeleri arasında dağıtılmasını sağlayan Java'da yazılmıştır. Hadoop çerçevesi uygulaması, bilgisayar kümeleri arasında dağıtılmış depolama ve hesaplama sağlayan bir ortamda çalışır. Hadoop, her biri yerel hesaplama ve depolama sunan tek bir sunucudan binlerce PC'ye ölçeklendirmek için tasarlanmıştır.

Kısacası, Hadoop, büyük miktarlarda veri üzerinde tam istatistiksel analiz gerçekleştirebileceği uygulamalar geliştirmek için kullanılmıştır [19].

2.2.1. Hadoop mimarisi

Hadoop'ta Hadoop MapReduce (uygulama katmanı) ve Hadoop'un dağıtılmış dosya sistemi_HDFS (Depolama katmanı) iki ana katman vardır.



Şekil 2.1. Hadoop mimarisi

Yukarıda belirtilen iki ana katmandan ayrı olarak, Hadoop çerçevesi kaynak yönetimi katmanında da YARN ve Hadoop Common içermektedir.

Yine Başka Bir Kaynak Müzakerecisi (YARN): Bu işi planlama ve küme kaynak yönetimi için bir çerçevedir.

Hadoop Common: Apache Hadoop çerçevesinin önemli bir modülüdür. Diğer Hadoop modüllerini destekleyen Common yardımcı programların, kütüphanelerin, gerekli Java dosyalarının ve komut dosyalarının koleksiyonunu ifade eder.

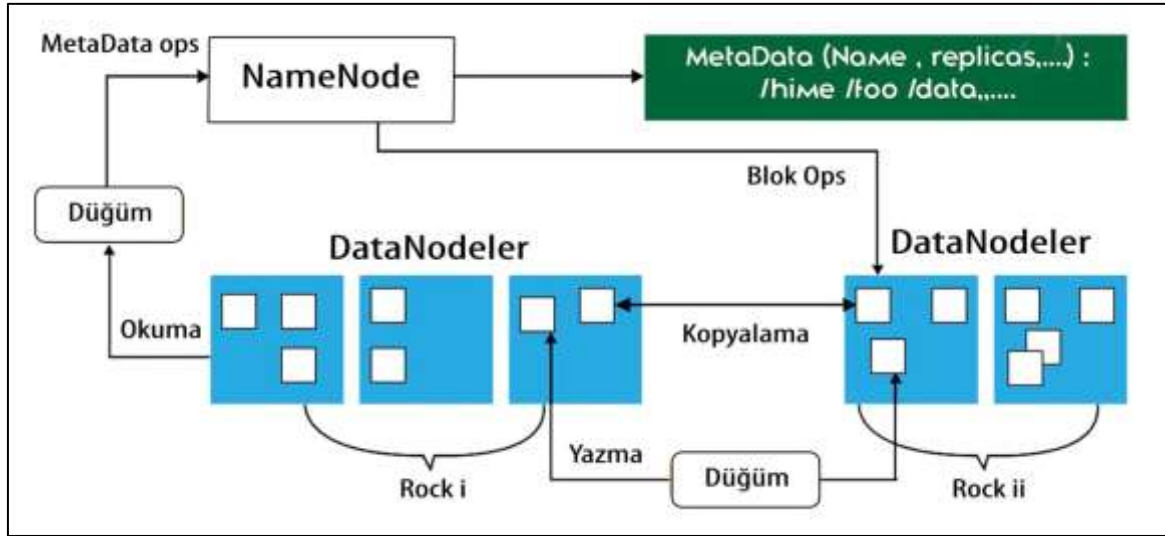
2.2.2. Hadoop'un dağıtılmış dosya sistemi (HDFS)

HDFS, verileri, küme üzerinde birden fazla sunucular arasında blok olarak dağıtan ve düğüm hatasını tolere etmek amacıyla farklı rafa farklı veri yedeklemesi oluşturan Hadoop sisteminin omurgasıdır. HDFS, uygulama verilerine yüksek verimli erişim sağlayan bir dağıtılmış dosya sistemidir. Bir HDFS kümesi, Namenode, verilerin bir master olarak nerede saklandığı bilgisini tutmaktadır. Datanode ise verilerin slave olarak birden fazla küme düğümü oluşturularak saklandığı makinalardır. Namenode, dosya sistemi ağacını, ağağa depolanan tüm dosya ve dizinlerin Metadata'lerini ve Datanode'lerde depolanan tüm blokların bulunduğu alanı yönetmekten sorumludur.

HDFS, verileri depolamak için Java programlarını destekler ve ölçeklenebilirlik özelliklerine sahiptir. Birden fazla donanım arasında dağıtılır. Şekil 2.2'de bir HDFS mimarisi gösterilmektedir. HDFS, master-slave mimarisini izler ve aşağıdaki unsurlara sahiptir. Düğüm, verilerine Namenode ve Datanode yardımıyla erişip saklayabilir.

Namenode, GNU / Linux işletim sistemini ve Namenode yazılımı içeren bir donanımdır. Namenode olan sistem master sunucu olarak davranır ve şu görevleri yapar: dosya sistemi namespace'ini yönetir, düğümlerin dosyalara erişimini düzenler ve aynı zamanda yeniden adlandırma, kapatma ve dosyaları açma gibi dosya sistemi işlemlerini yürütür.

Böyle bir durumda, tek bir Namenode başarısız olursa ikinci bir Namenode her zaman etkindir ve bu ikinci Namenode'den anlık görüntüler alır ve tüm bilgileri kendisi saklar. Namenode başarısız olursa, ikinci Namenode'den veri kurtarabilir.



Şekil 2.2. HDFS mimarisi

Datanode, GNU / Linux işletim sistemine ve Datanode yazılımına sahip bir donanımdır. Bir kümedeki her düğüm (donanımı / sistemi) için bir Datanode olacaktır. Bu düğümler, sistemlerinin veri depolamasını yönetir, istemci isteğine göre dosya sistemlerinde okuma-yazma işlemlerini gerçekleştirir ve Namenode'nin talimatlarına göre blok oluşturma, silme ve yedekleme gibi işlemleri gerçekleştirir. Genellikle kullanıcı verileri HDFS dosyalarında saklanır. Bir dosya sistemindeki dosya bir veya daha fazla bölüme ayrılır ve / veya ayrı veri düğümlerinde saklanır. Bu dosya bölümleri blok olarak adlandırılır. Yani, HDFS'nin okuyabileceği veya yazabileceği minimum veri miktarına Blok denir. Varsayılan blok boyutu 64 MB'dir, ancak HDFS yapılandırmasında değişiklik yapılması gerektiğinde artırılabilir.

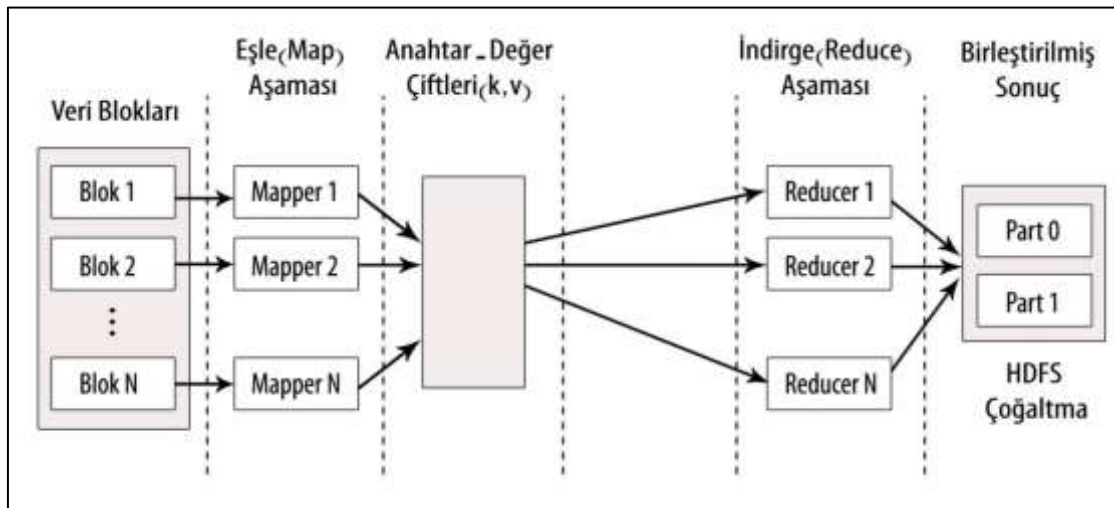
2.2.3. MapReduce

MapReduce, büyük veri setlerini birden fazla görüntünün paralel olarak işlenmesi için HDFS'nin üstündeki sonucu en iyi performansla veren bir programlama modelidir. MapReduce algoritması, Eşle(Map) aşaması, Karıştırma aşaması ve İndirge(Reduce) aşaması gibi üç aşamada çalışmaktadır. Map aşaması veya eşleyicisinin(mapper) giriş verilerinin işlemidir. Genellikle giriş verileri, HDFS sisteminde depolanır. Eşleyici, giriş dosyası satırına, satır geçirerek birkaç küçük blok halinde veri oluşturur.

Reduce aşaması, karıştırma aşaması ile birleşimdir. Reduce'nin işi, Eşleyiciden gelen verileri işlemektir. İşlemden sonra, HDFS'de depolanarak yeni bir çıktı veri seti üretir. Bir

MapReduce işi sırasında, Hadoop Map ve Reduce görevlerini kümedeki uygun sunuculara gönderir.

MapReduce çerçevesi, görev verme, görev tamamlama işlemini doğrulama ve düğümler arasındaki küme etrafındaki verileri kopyalama gibi tüm veri aktarma ayrıntılarını yönetir. Ağ trafiğini azaltan yerel disklerdeki verileri içeren düğümlerde gerçekleştirilir. Verilen görevlerin tamamlanmasından sonra, küme uygun bir sonuç elde etmek için verileri toplayıp, azaltıp Hadoop sunucusuna geri gönderir. Programcı sadece fonksiyonları çözüp her aşama için Hadoop, veri toplama, sıralama ve düğümler arasında geçen mesaj ile ilgilenir, arasındaki geçiş işler. Tek bir veri analiz programında, birden fazla Map ve Reduce aşaması olabilir. Bu işlevler büyük veriler üzerinde paralel olarak çalışmaktadır. MapReduce'nin veri akışı Şekil 2.3'te gösterilmektedir.



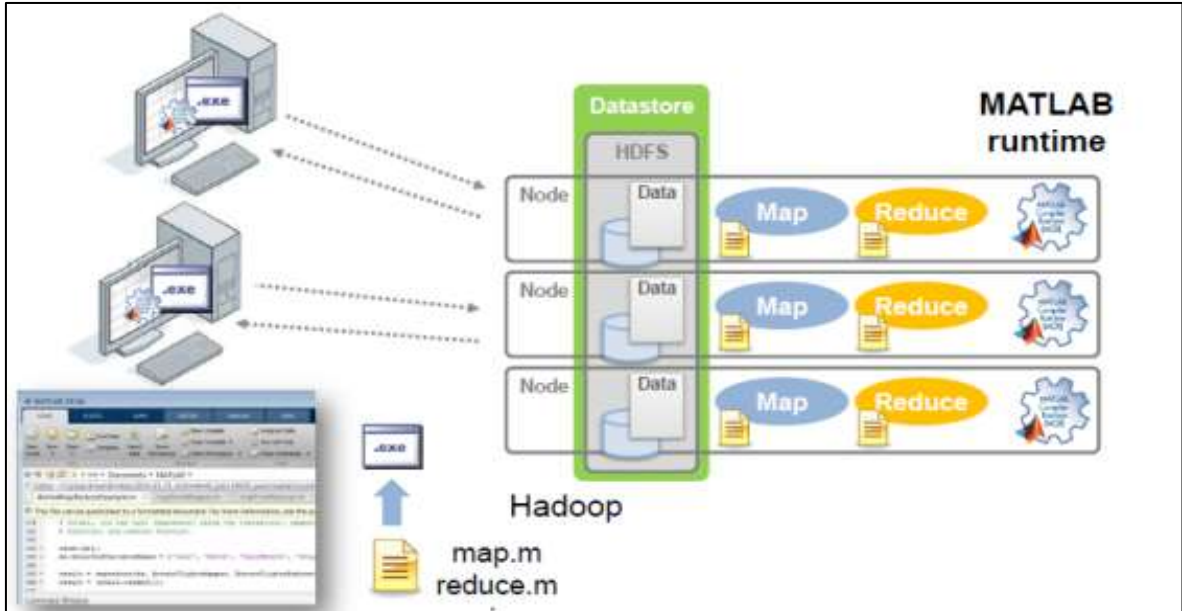
Şekil 2.3. MapReduce veri akışı

2.2.4. MATLAB ile Hadoop MapReduce

MATLAB, büyük veri setlerini analiz etmek için sayısız yeteneklere sahiptir. Bunlar arasında, MapReduce, büyük verileri filtreleme ve istatistik ile büyük veriye yönelik diğer genel analiz yöntemleri de bulunmaktadır. MATLAB'ta MapReduce veri setlerini analiz etmek için bir programlama tekniğidir. Verileri de "Datastore" kullanarak belleğe sığdıran küçük bloklar halinde işlemektedir. Her blok, işlenecek verileri biçimlendiren bir Map aşamasından geçer. Daha sonra ara veri bloklar, final bir sonuç elde etmek için ara sonuçları toplayan bir Reduce aşamasından geçer. MapReduce, çeşitli ortamlarda, tek bir

donanımda ya da Hadoop ile bir küme şeklinde genişletilebilir [20].

MATLAB'ta fonksiyon olan MapReduce ve Datastore kullanarak, masaüstümüzde veya Hadoop'ta küme olarak algoritmalar geliştirebilmektedir. Daha sonra, HDFS'de depolanan tüm veri setine karşı Hadoop MapReduce çerçevesi içindeki bir kümede algoritmaları çalıştırmak için MATLAB Dağıtılmış Hesaplama Sunucusu kullanılabilir. Ek olarak MATLAB analitiği, MATLAB MapReduce tabanlı algoritmalar uygulamalar veya kütüphaneler oluşturmak için MATLAB Derleyici'sini kullanarak üretim Hadoop sistemleri ile entegre edilebilmektedir [21].



Şekil 2.4. MATLAB ile Hadoop MapReduce mimarisi

2.3. İçerik-tabanlı Görüntü Erişimi (CBIR)

İçerik-tabanlı görüntü erişimi (CBIR), görüntünün içeriğine göre görüntü veri setindeki belirli bir görüntüyü almak için kullanılan bir tekniktir [22]. İçerik, renk, doku ve şekil gibi görüntüyle ilişkilendirilmiş olan özniteliklerdir.

Bu öznitelikler kullanılarak, kullanıcının sorgusuna göre görüntü veri setinden alınmaktadır. CBIR görsel benzerlik olan görüntüleri elde etmeyi amaçlamaktadır.

Renk ile görüntü içeriği karşılaştırması, görüntüleri renk dağılımlarına göre eşleştirmeye

dayanır. Bu durumda görüntü özneliği, bir görüntüdeki belirli renk veya bu renkteki piksellerin oranını tanımlar. Bu nedenle, istenilen renk dağılımına sahip ve istenilen renkleri veya örnek bir görüntüyü belirterek renk aramaları yapıp benzer görüntüler elde edilebilir.

Renk histogramları, renk dağılım tanımlayıcılarını görüntüden çıkarmak için yaygın olarak kullanılır. Görüntüdeki piksellerin renginin bir istatistiğidir. İlk renk dağılımı uygun renk histogramı ile temsil edilir ve sonra bu histogramdan renk vektörü oluşturulur [22].

Doku da renk öznelik çıkarma gibi görüntü dokusuna göre elde edilir, ancak renkler yerine görüntülerde görsel desenler kullanılmaktadır. Dolayısıyla, tek bir renk mevcudiyeti veya bir piksel değerinin yoğunluğunun sonucu olmayan homojenliğe bakılmalıdır. Bazen de daha fazla uzamsal bilgi sağlamaktadır. Kontrast derecesi, yönlülük, düzenlilik, coarseness (kalitesizlik) ve rastgelelik gibi çeşitli doku öznelikleri vardır.

Görüntülerdeki dokusal öznelikleri çıkarma algoritmalarından Yerel İkili Örüntü ve GLCM güçlü bir öznelik çıkarma yöntemleridir. Bunlar sınıflandırıcıların performansının artıracak şekilde literatürde yaygın olarak araştırmacılar tarafından kullanılmaktadır [23].

Şekil tabanlı görüntü erişimi karşılaştırması, görüntü içindeki bölgelerin şekillerine bakarak sorgu görüntüsündeki gibi şekilleri aramaktadır. Kenar tespitleri, şekil özneliği çıkarımı için önemli parçalardır. Bu kenarlar, görüntüde çevrelerden daha parlak veya koyu olan noktaları veya bölgeleri göstermektedir. Görüntü filtreleme ve görüntü dönüşümleri farklı türler kullanılarak şekil-tabanlı görüntü erişimi için kullanılan çeşitli yöntemlerdir [24].

CBIR sisteminde dört ana zorluk şunlardır: (i) dizin oluşturma ve erişim amaçlı olarak geçen süreyi temsil eden indekslleme ve alma işleminin hesaplamalı karmaşıklığı, (ii) bellek ve disk alanı gereksinimleri, (iii) anlamını gösteren semantik alma performansı sistemin doğruluğunu gösterir ve (iv) kullanılabilirlik [25].

2.4. Bilgisayar Destekli Tanı Sistemi (CAD)

CAD tıbbi alanda görüntüleme analizi için çok önemlidir, çünkü görüntüleme en erken

dönemde herhangi bir hastalığı teşhis etmek için temel yöntemdir [26, 27]. MRG, X-Ray, Endoskopi, Ultrason gibi görüntüleme teknikleri vardır. CAD sistemleri görüntü kalitesini artırmak için kullanarak tıbbi görüntüleri doğru yorumlamaya yardımcı olur [28]. CAD sisteminin ana uygulaması insan vücudundaki anormalliği bulmaktır. Tüm bunlar arasında, tümörün tespiti tipik bir uygulamadır [29].

3. MEVCUT UYGULAMALAR (İLGİLİ ÇALIŞMALAR)

3.1. Bilgisayar Destekli Tanı (CAD) Teknikleri ile Yapılan Çalışmalar

Akciğer nodüllerinin otomatik sınıflandırılması için son yıllarda birçok CAD sistemi tasarlanmıştır [30]. Bu bölümde, otomatik sistemlerin çeşitli aşamalarında, ön işleme, akciğer segmentasyonu, akciğer bölgesinde nodül tespiti, nodül segmentasyonu ve hacim ve şekle dayalı analiz, büyüme, doku ve görünüm gibi ana adımlardan oluşan akciğer kanserinin teşhisi için birkaç çalışma kısaca anlatılmaktadır.

Lakshmanaprabu ve diğerleri (2019), bilgisayarlı tomografi (CT) akciğer görüntülerinde otomatik tanı sınıflandırma yöntemini sunmuşlardır. Modifiye Yerçekimi Arama Algoritması (MGSA), Optimal Derin Sinir Ağı (ODNN) ve Doğrusal Ayrımcılık Analizi (LDA) algoritmaları ile %94,56 doğruluk oranı elde etmişlerdir [10].

Nasrullah, Sang, Alam ve Xiang (2019), otomatik akciğer kanseri tespiti için Karışık Link Ağı (MixNet) ile Gradyan Artırma Makinesi (GBM) önermişlerdir. Yaptıkları sınıflama sonucunda %94 civarı anma oranı elde etmişlerdir [31].

Ardila ve diğerleri (2019), akciğer kanseri riskini tahmin etmek için LDCT ve NLST veri setlerinde hastanın mevcut ve önceki bilgisayarlı tomografi hacimlerini kullanarak Derin Sinir Ağları (CNN) algoritmasını önermişlerdir. Bu çalışma sonucunda ROC eğrisinin altında % 94,4'lük oranı başarıyı elde etmişlerdir [32].

Shen, Han, Aberle, Bui ve Hsu (2019), bilgisayarlı tomografi (CT) taramasında gözlenen belirli bir akciğer nodülün malign olup olmadığını tahmin etmek için Hiyerarşik Semantik Evrişimli Sinir Ağı (HSCNN) sunmuşlardır. Hem radyologlar tarafından bildirilen tanılama hem de yüksek düzeyde 3D CNN'den çıkarılan nodül özneteliklerinin üzerinde yaptıkları deneyde %91 doğruluk elde etmişlerdir [33].

Kasinathan ve diğerleri (2019), bilgisayar destekli tanı (CAD) sistemini geliştirmişlerdir. Bu çalışmada bilgisayarlı tomografi (CT) görüntüleri için aktif kontur modeli (ACM) ile önerilen model otomatik bir akciğer segmentasyon aracı kullanılmıştır. LIDC-IDRI veri

setinde akciğer tümörü tespit etmek için üç önemli öznitelik değerlendirilmiştir. Çok-ölçekli Gauss dağılımı ve derin Evrişimli Sinir Ağları (CNN) yöntemleri ile %97 doğruluk elde etmişlerdir [34].

Nasser ve Abu-Naser (2019), insan vücudunda akciğer kanserinin olduğunu tespit etmek için Yapay Sinir Ağı (ANN) geliştirmişlerdir. Model değerlendirmesiyle %96.67 doğruluk elde etmişlerdir [35].

Prabukumar, Agilandeewari ve Ganesan (2019), akciğer kanseri nodüllerinin tespiti için en uygun teşhis sistemini önermişlerdir. Çalışmada, Fuzzy C-Means (FCM) ve bölgesel büyüyen segmentasyon algoritmaları kullanarak geometrik ve doku öznitelikleri gibi sonuçlar elde edilmiştir. SVM sınıflandırıcısı, bu akciğer kanseri tipini iyi huylu veya kötü huylu olarak sınıflandırmamıza yardımcı olan bu optimal özellikleri kullanarak eğitilmiştir. Bu çalışma sonucunda ortalama olarak %98.51 oranında başarı elde edilmiştir [36].

Günay ve diğerleri (2019), akciğer kanseri teşhisi için Yapay Bağışıklık Sistemi (AIS) yöntemini uygulamışlardır. AIS özellikleri ağırlıkları ile değiştirilmiştir. Pearson Korelasyon Katsayıları (PCC) kullanılarak bu algoritmanın performansını artırmak için veri kümesi boyutu azaltılmıştır. AIS üzerinde yaptıkları deneyde %82 doğruluk oranı elde etmişlerdir [37].

Coudray ve diğerleri (2018), doğru ve otomatik olarak LUAD, LUSC veya normal akciğer nodüllerini sınıflandırmak için TCGA veri set üzerinde derin bir evrişimli sinir ağı (CNN) geliştirmişlerdir. Yaptıkları sınıflama sonucunda ROC eğrinin altındaki %97 oranı başarı elde etmişlerdir [38].

Nishio ve diğerleri (2018), benign nodül, primer akciğer kanseri ve metastatik akciğer kanseri arasında sınıflandırma için bilgisayar destekli bir tanı (CAD) yöntemi geliştirmişlerdir. Çalışmada, derin konvolüsyonel sinir ağının (DCNN) ve destek vektör makinesi (SVM) ile üç dik düzlemde rotasyon değişmeyen tek biçimli model yerel ikili desen kullanılarak gerçekleştirilmiştir. DCNN yöntemini kullanarak %68'in üzerinde doğruluğu elde etmişlerdir [39].

Zhu ve diğerleri (2018), radyomik imzaya dayalı skuamöz hücreli karsinom (SCC) ve

akciğer adenokarsinomu (ADC) nodüllerini ayırt etmek için bir sistem önermişlerdir. Histolojik alt tip algoritması ve GLCM kullanarak %83 anma oranıyla sınıflandırılmıştır [40].

Gong ve diğerleri (2018), kötü huylu ve iyi huylu akciğer nodülleri arasında sınıflandırma için bilgisayar destekli bir tanı (CAD) sestimi geliştirmişlerdir. Çalışmada, destek vektör makinesi, naif Bayes ve lineer ayırma analizi sınıflandırma yöntemlerini kullanarak ROC eğrileri (AUC) altında ortalama %94 başarı elde etmişlerdir [41].

Makaju ve diğerleri (2018), kanserli nodülün akciğer BT tarama görüntüsünden tespit edilmesi ve algılanması için havza segmentasyonu ve SVM kullanmışlardır. Nodülün Malign veya benign olarak sınıflandırılması için yeni sistem önerilmiştir. Önerilen modelde ortalama olarak %92 oranında başarı elde edilmiştir [42].

Wan, Hoogendoorn ve Haasdijk (2017), hesaplama maliyetini azaltan ve yüksek bir kesinliği korurken düşük sayıda yanlış pozitif elde eden 2D yöntemleri kullanılarak iki CAD sistemi geliştirilmiştir. Bu çalışmada, nodül adaylarını nodül veya nodül olmayan şeklinde sınıflandırmak için destek vektör makinesi (SVM) ve Çok Katmanlı Perceptron (MLP) sınıflandırma yöntemlerini kullanarak %87'nin üzerinde akış kesinliği elde etmişlerdir [43].

Nithila ve Kumar (2017), akciğer nodüllerinin tespiti için bir Swarm Yapay Sinir Ağı metodolojisi önermişlerdir. Çalışmada, nodül ve nodül olmayan, kısmen katı ve katı olmayan tiplere ayrılmıştır. Bu yöntem, akciğer tümörlerinin otomatik teşhisinde kullanılır. LIDC-IDRI ve SPIE-AAPM veri setleri kullanarak BPNN-VIW'nin sınıflandırma üzerinde yaptıkları deneyde katı nodüller üzerinde %92 doğruluk oranı elde etmişlerdir. Kısmen katı nodüller üzerinde ise %93 civarında başarı elde etmişlerdir [44].

Akram, Javed, Qamar, Khanum ve Hassan (2015), 2D ve 3D geometrik ve yoğunluk tabanlı istatistiksel özniteliklere dayalı Yapay Sinir Ağları (ANN) kullanarak akciğer nodül tespit sistemini önermişlerdir. Önerilen CAD sistemi, standart LIDC veri seti kullanılarak %96,68 doğruluk oranı elde edilmiştir [45].

Silva, Silva, Paiva ve Gattass (2016), morfoloji ve doku özniteliklerini akciğer nodüllerinin

malignite sınıflandırılması için derin bir öğrenme tekniği olan üç farklı Konvolüsyonel Sinir Ağı (CNN) mimarisi önermişlerdir. Metodoloji, LIDC-IDRI üzerinde yaptıkları deneyde %82.30 oranında doğruluk elde etmişlerdir [46].

Kumar, Wong ve Clausi (2015), akciğer nodüllerini kötü huylu veya iyi huylu olarak sınıflandırmak için bir otomatik kodlayıcıdan çıkarılan derin öznitelikleri kullanan bir CAD sistemi önermişlerdir. 10 kat çapraz doğrulama üzerinde NCI-LIDC veri setinden 4323 nodül içeren 4303 örnek kullanarak %75,01 ve %96,31 doğruluk oranıyla başarı elde edilmiştir [47].

Kuruvilla ve Gunavathi (2014), bilgisayarlı tomografi (CT) akciğer görüntüleri Yapay Sinir Ağı kullanılarak geliştirilen bir bilgisayar destekli sınıflandırma yöntemi sunmuşlardır. Yaptıkları sınıflama sonucunda %93,30 başarı elde etmişlerdir [48].

Chaudhary ve Singh (2012), CT taramalarında akciğer nodülü tespit etmek için bir yöntem sunmuşlardır. Çeşitli geliştirme ve bölümlleme teknikleri kullanarak daha doğru sonuçlar elde etmek için görüntü işlemede yapılan işlemler, görüntü ön işleme, bölümlleme ve öznitelik çıkarımı gibi işlemler yoluyla MATLAB'ı kullanmışlardır. İyi bir deneme ve daha verimli hale getirmek için birçok geliştirmeler gerekmektedir [49].

Krishnamurthy, Narasimhan ve Rengasamy (2016), CT taramasından malign akciğer nodüllerini sınıflandırmak için 3D morfoloji analizinden sonra otomatik segmentasyonu sunmuşlardır. AMRG ve EBFT algoritmaları juxtaplevral nodülleri olmadan sınırlandırmak için geliştirilmiştir. Sonuç olarak %88 kesinlik oranı elde edilmiştir [50].

Iva (2014), nodülün tespiti ve nodülün sınıflandırılması için bir algoritma sunmuştur. Önerilen algoritmada 3 bağımsız boyutluluk azaltma yöntemi (öznitelik seçme yöntemi, Fisher'in discriminant analizi ve temel bileşen analizi) kullanılarak yalnızca etkili öznitelikler seçilmiştir. Sonuçlar En Yakın Komşu Araması ve Torbalı Karar Ağaçları uygun sınıflandırma yöntemleriyle kullanılmış. KNN sınıflandırma yöntemini kullanarak ROC eğrileri (AUC) altındaki %91'in üzerinde akış ortalaması elde etmiştir [51].

Han ve diğerleri (2014), akciğer nodüllerin üzerinde iki boyutlu 2D doku öznitelikleri üç tanınmış türleri (Haralick, Gabor ve Yerel İkili Örüntü öznitelikleri) LIDC-IDR

oluşturulan akciğer görüntü veritabanı kullanarak karşılaştırmışlardır. Destek vektör makinesi (SVM) sınıflandırıcısı kullanarak üç öznitelik türü üzerinde yaptıkları deneyde %90 civarında doğruluk elde etmişlerdir. Haralick öznitelikleri, görüntü dilim kalınlığında ise %92,70 başarı elde etmişlerdir [52].

3.2. Görüntü Erişim Teknikleri ile Yapılan Çalışmalar

Bugünün görüntü erişim sistemlerinin tek ya da birden fazla öznitelik çıkartması ve kavramsal görüntü içerik açıklamaları arasında CBIR gibi çeşitli teknikler vardır. CBIR'da, görüntüler renk, doku ve şekiller gibi görsel içerikleri ile indekslenir [53].

Bu kısımda, örnek vermek amacıyla, önemli olan veya ilgi çeken birkaç çalışma kısaca anlatılmaktadır.

Khan, Hussain ve Alshaikhli (2012), benzerlik ölçümünün önemini belirtti. İki görüntü arasındaki benzerliği bulan üç benzerlik yöntemi önermişlerdir. Bu teknik daha sonra sorgu görüntülerini veri kümesinde depolanan görüntülerden karşılaştırmak için kullanılmıştır. CBIR, tek görüntü alma, benzer görüntü yöntemleri içindeki veri setine hızlı bir şekilde erişme ve sistemin doğru bir sonuç almasına yardımcı olmak için Euclidean Distance (Öklit Mesafesi) formülü ile karşılaştırmanın iki yöntemi açıklanmışlardır [54].

Kumar ve Senthilmurugan (2013), görüntülerden kullanıcılar için kolayca elde edilen düşük seviyeli öznitelikler ve önemli olan yüksek seviyeli öznitelikler arasındaki tutarsızlık ve semantik boşluk ile ilgili sorunlar CBIR sisteminde yoktur [55].

3.3. Hadoop Üzerinden Görüntü Erişim Teknikleri ile Yapılan Çalışmalar

Görüntü alma cihazları, teknolojik gelişmeler nedeniyle her gün çok fazla görüntü verileri üretmektedir. Görüntülerle ilgili bilgilerin çeşitlendirilmiş uygulamaları vardır. Tıbbi görüntüleme sistemleri, tüm tıbbi alanlarda giderek daha fazla dijitalleştirilmiş görüntüler üretir. CBIR sistemi, görüntü veritabanlarını etkin bir şekilde yönetmek için olası çözümlerden biridir. Ayrıca, bu kadar büyük bir veritabanına hızlı erişim, verimli bir hesaplama modelini gerektirir. Hadoop çerçevesi MapReduce Dağıtılmış hesaplama

modeline dayanan bulgulardan biridir. Bu kısımda, önemli olan veya ilgi çeken birkaç çalışma kısaca anlatılmaktadır.

Kusuma ve Megha (2015), Hadoop MapReduce işleme yönteminin uygulanmasıyla CBIR sisteminin paralel işlemesini gerçekleştirmek için bir çerçeve önermişlerdir. Önerilen çerçeve hesaplama süresi ve harita dağıtma / azaltma slot sayısının veri dağıtım ve tahsis açısından uygulanmıştır. Önerilen yöntem işleme süresini büyük ölçüde azaltmıştır ve görüntü işleme alanının diğer alanlarına da uygulanabilir [53].

Lohar ve diğerleri (2016), CBGIR simülasyon sistemini Hadoop ile sunmuşlardır. Simülasyon testinin sonuçları, Hadoop kullanarak tıbbi görüntü erişim sisteminin, görüntü saklama ve görüntü alma verimliliğini arttırdığını, büyük tıbbi görüntülerle ilgili daha iyi sonuç elde edildiğini göstermişlerdir. Önerilen sistemin görüntü alma verimliliğini daha da arttırması için Map ile Reduce görevi arasındaki verilerin iletim hızını azaltması gerekmektedir [56].

Markonis ve diğerleri (2012), tıbbi görüntü işleme sistemi kullanarak hızı arttırmak için MapReduce çerçevesini, (i) well-known makine öğrenme yöntemi ve SVM kullanarak akciğer dokusu sınıflandırması, (ii) içerik-tabanlı tıbbi görüntü indeksleme ve (iii) katı doku sınıflandırması için dalgacık analizi üç büyük ölçekli kullanmışlardır. Sonuçlarda, optimum SVM parametrelerini bulmak için toplam yürütme süresi yaklaşık 1000h'den yaklaşık 10h'ye düşmüştür [57].

Mahakulkar ve diğerleri (2016), tıbbi teşhis amacıyla ve büyük miktarda veri işlemek için Hadoop algoritmalarını kullanarak CBIR'yi önermişlerdir. Önerilen teknik anma oranını ortalama %44'den ortalama %72'ye çıkarmıştır. Bu önerilen algoritma, yeni algoritmalar gibi her zaman en iyi ve en kötü çalıştığı bölgelere sahiptir [58].

Trivedi (2018), MapReduce tekniği kullanarak görüntü verilerinin Hadoop üzerinden çok sayıda düğüme dağıtılması ve büyük veri tabanını yönetmek için Yerel Tetra Örüntü'yü (LTrP) kullanmıştır. Hadoop çerçevesinde sorgulanan görüntüyle eşleşen görüntülerin doğru şekilde alınmasını da kolaylaştırmak için CBIR sistemini sunmuştur. Önerilen

görüntü erişim sistemi Hadoop'ta uygulandığı için, bulut ortamına minimum ek yük ile uyumunun sağlanması çok kolaydır [59].

Jai-Andaloussi ve diğerleri (2013), MapReduce modeli ve HDFS depolama modeli uygulayarak CBIR sistemi önermişlerdir. Görüntülerin içeriğini karakterize etmek için BEMD-GGD ve BEMD-HHT iki yöntem kullanmışlardır. Görüntüler arasındaki benzerliği ölçmek için, görüntülerin imzaları arasındaki mesafeyi hesaplanmıştır. DDSM mamografi görüntü veri tabanında deneyerek sonuçlarda, CBIR'ın büyük tıbbi görüntü veritabanlarında uygulanmasının uygulanabilirliğini ve verimliliğini görmesini sağlamışlardır [60].

Venkatraman (2012), CBIR için bulut ortamında büyük veri koleksiyonundan oluşan bir MapReduce Sınır Ağı Yöntemi önermiştir. Renkli görüntüleri içeriklerine göre sınıflandırmak için doğal dil sorgularını kullanmıştır. Gerçek zamanlı olarak doğru sonuçlara varmak için bulut kümelerinde Map ve Reduce işlevlerini uygulamıştır [61].

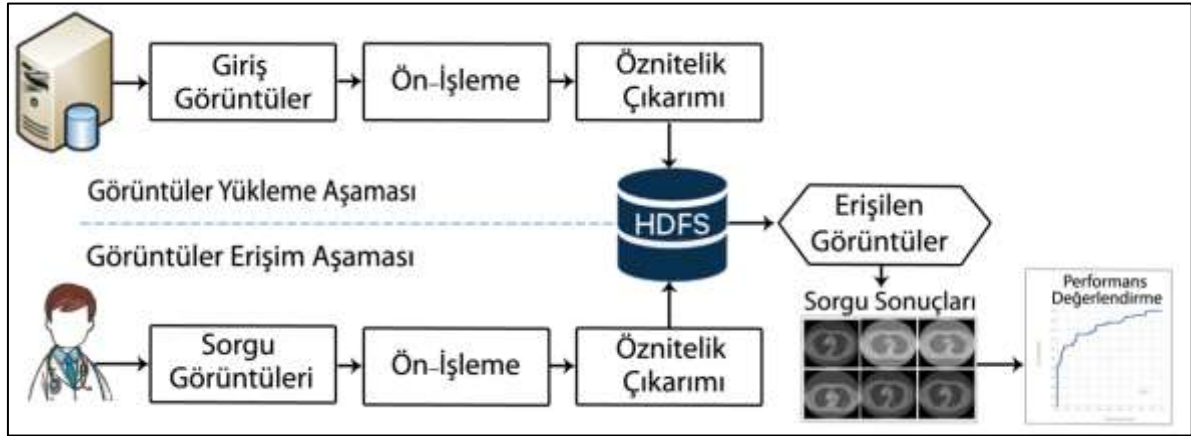
Patil ve Pradeepani (2016), Hadoop'un MapReduce Çerçevesini kullanarak tıbbi görüntüleri analiz etmek için bir yöntem önermişlerdir. Mevcut tıbbi görüntülerin öznitelik kütüphanesini paralellik olarak depolamak için HDFS kullanmışlardır. Eşleştirme ve erişim işlemleri MapReduce tarafından gerçekleştirilmiştir. Bir master ve üç slave'ler kümesi oluşturulmuştur. Düğüm makinesi, istemci tarafından sorgu düğümlerini ana düğüme yüklemek için kullanılmıştır. İstemci görüntüyü istemci makine üzerinden yüklediğinde, ana düğüm onu kabul eder ve bir MapReduce işi oluşturur. İşleri yürütmek için slaver'ler vererek görevlere ayrılmıştır. Paralel uygulama sonucunda, Hadoop tabanlı tıbbi görüntü erişim sistemi, geleneksel görüntü erişim sistemlerine göre daha az zaman almıştır [62].

Dravyakar (2014), görüntünün içeriğini analiz etmek için bir CBIR sistemi önermiştir. Öneren sistem kullanıcının belirli bir görüntü, veritabanından renk, şekil ve dokudaki birleştirilmiştir değerlere bağlı olarak yükleyip benzer görüntüleri almıştır. Önerilen sistem, görüntülerin içeriği ve HDFS kullanılarak büyük görüntü veri setleri üzerinde görüntülerin yükleme ve arama tekniklerini birleştirmiştir [63].

Önceki literatürde, birçok araştırma, görüntü erişim sistemlerinin dağıtılmış hesaplama araçları ve paralel işlem paradigmasında daha verimli çalıştığını gösterildi. Hadoop çerçeveleri, MapReduce algoritmalarını kullanarak güvenilir ve yüksek veri depolama sağlar. Dolayısıyla, büyük veri setlerindeki CBIR sistemleri, Hadoop çerçevesi kullanılarak geleneksel yaklaşımlardan daha etkilidir [64].

4. ÖNERİLEN ÇÖZÜM

Önerilen sistem, kullanıcılar için, radyoloji uzmanlarının nodülleri daha erken tanımlama ve tespit etmelerine yardımcı olan bir uygulama arayüzü önermektedir. Çalışmada, akciğer görüntülerinin HDFS'ye yüklenmesi ve kullanıcı bir uzmanın nodülü iyi huylu veya kötü huylu olarak sınıflandırmak istediğinde, belirli bir görüntüye benzerlikler gösteren diğer görüntülerin HDSF'den alınmasını iki aşama olarak sağlamaktadır. Bu görüntüler ile radyolog uzmanın XML dosyalarında açıklamalara bağlı olarak nodülleri tespit edebilmektedir. Şekil 4.1'de Hadoop kümesi üzerine konuşlandırılan bu model sistemin mimarisini göstermektedir. Bu aşamalardan her birinin detayları takip eden alt-bölümlerde açıklanacaktır.

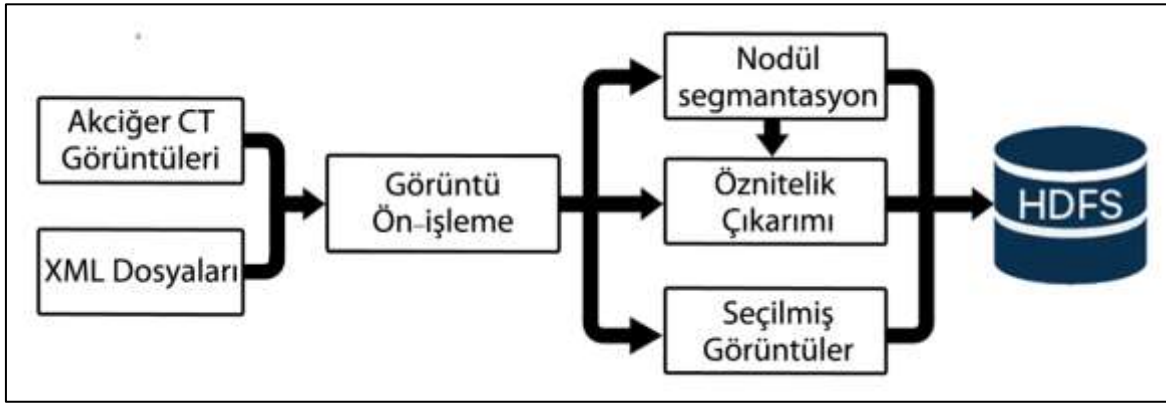


Şekil 4.1. Önerilen model sistemin mimarisini

4.1. Görüntüler Yükleme Aşaması

Sistemde, grafik kullanıcı arayüzü (GUI) üzerindeki tek bir noktada görüntü klasör yolunu vererek kullanıcının HDSF olarak adlandırılan Hadoop veri seti üzerine bir veya daha fazla akciğer tomografisi görüntülerini yüklemesine imkân vermektedir. Her bir klasör, klinik göğüs tomografisinden alınan DICOM görüntülerini ve ilgili XML dosyalarını barındırmaktadır. Kullanıcının yüklediği görüntülerin sayısı, XML dosyalarında uzman radyolog tarafından tanımlanan nodül sayısına bağlı olarak değişebilir. Her bir tomografi çok sayıda görüntü içerebilir ve bunlar bütün görüntülere işlem yapılırsa uzun zaman almaktadır. Bu senaryoda, sadece açıklamaların XML dosyalarındaki seçilen görüntüleri okumayı gerçekleştirip spesifik nodül sonuçlarını ikili görüntülere çevirmektedir. Görüntü

dönüştürüldükten sonra RQI öznitelikleri her bir radyolog için bağımsız olarak çıkarılır. İkili görüntüler, şekil ve doku öznitelikleri çıkarılır. Çıkarılan farklı öznitelikler bir araya getirilerek bir öznitelik vektörü oluşturulur. Çıkarılan özniteliklerle dönüştürülen nodül ikili görüntüleri (.mat) formatında saklanıp seçilen DICOM görüntüleri ile birlikte HDSF'ye yüklenmektedir. Şekil 4.1 önerilen model sistemin yükleme aşamasını göstermektedir.

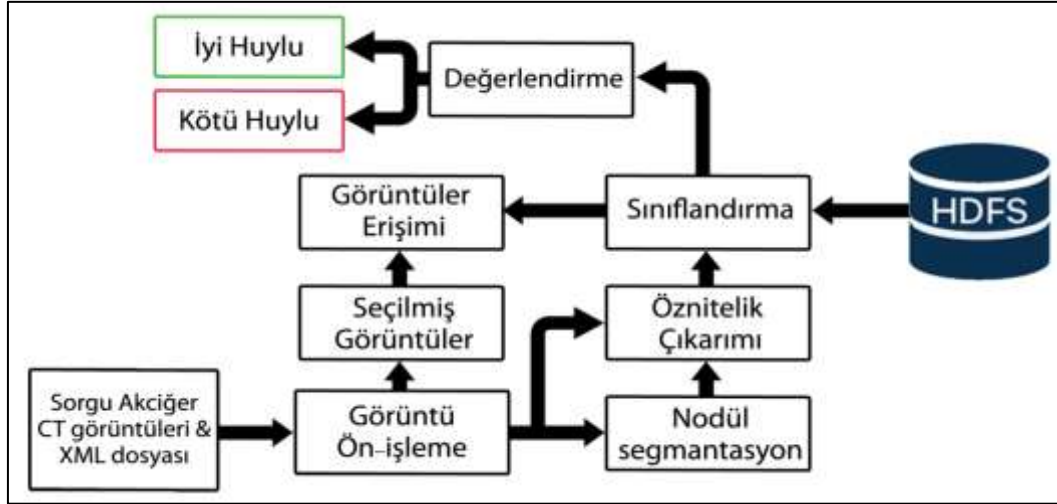


Şekil 4.2. Önerilen model sisteminin görüntüler yükleme aşaması

4.2. Görüntüler Erişim Aşaması

Yükleme aşamasına benzer bir şekilde, sistem kullanıcıya, sorgu görüntüsü ile HDSF veri setinde görüntü bir GUI kullanarak arama ve görüntü almayı sağlamaktadır. Bir radyolog uzman belirli öznitelikteki görüntüleri aramak istediğinde, sorgu tomografi görüntü seçerek XML dosyasında depolanan nodul özniteliklerini vermek zorundadır. Aranılan ikili nodül görüntü öznitelikleri ile veri setinde depolanan değerlerle eşleştirmektedir. Şekil 4.2 önerilen model sistemin erişim ve sınıflandırma aşamasını göstermektedir. Nodül görüntülerinin erişim algoritması aşağıdaki gibi özetlenmiştir:

1. Sorgu görüntüsünün özniteliklerinin çıkarımı.
2. HDFS veri setindeki görüntülerin özniteliklerinin çıkarımı.
3. Akciğer nodülünün iyi huylu veya kötü huylu olarak sınıflandırması.
4. Sorgu görüntüsüne benzerlik olan görüntüleri erişimi.



Şekil 4.3. Önerilen model sisteminin görüntüler erişim ve sınıflandırma aşaması

4.3. Akciğer CT Görüntü Veri Set

Deneylelerimizde kullanan görüntü verileri, akciğer görüntü veri tabanı konsorsiyumu (Lung Image Dataset Consortium - LIDC) adı verilen çevrimiçi bir kaynaktan yüklenmiştir [65], Bahse konu kaynak, ulusal kanser enstitüsü (National Cancer Institute - NCI) tarafından kurulmuş ve ulusal sağlık enstitüleri vakfı (Foundation for The National Institutes of Health - FNIH) tarafından geliştirilmiştir. Bu çalışmalar (Amerikan) gıda ve ilaç idaresi (Food and Drug Administration - FDA) tarafından da desteklenmiştir. Akciğer veri setinin detaylı içeriği Çizelge 4.1’de yer almaktadır.






Çizelge 4.1. LIDC veri setinin dağılımı

Koleksiyon İstatistikleri	Değer
Modaliteler	CT (Bilgisayarlı Tomografi)
Görüntülerin sayısı	244,527
Görüntülerin formatı	DICOM
CT tarama sayısı	1,018
Görüntüler boyutu	124 GB
Sınıf sayısı	2 (“iyi huylu” veya “kötü huylu”)

LIDC veri setinde akciğer kanseri taraması için yapılan 1018 göğüs tomografisi yer almaktadır. Bunlardan her biri bir dizi CT görüntüsü ile bu görüntülere ilişkin XML dosyalarını içermektedir. XML dosyaları nodüllere dair açıklamaları ve habislik

değerlendirmelerini vermektedir. Habisliğe ilişkin olarak en fazla 4 radyoloji uzmanı tarafından verilen değerlendirmelere göre, 1 düşük habisliği ifade ederken, 5 yüksek derece habislik anlamına gelmektedir. Örnek bir nodül, tanı öznetelik değerleri ile birlikte Şekil 4.4'te sunulmaktadır.

İncelik	İncelik	İç Yapı	Kireçlenme	Küresellik	Marjın	Lobulasyon	Spikülasyon	Doku	Habitslik
1	5	1	6	3	3	3	4	5	5
2	5	1	6	4	4	5	5	5	5
3	5	1	6	3	2	3	3	5	5
4	5	1	6	5	4	1	5	4	5



Habitslik Derecesi	1	2	5	3	4
--------------------	---	---	---	---	---

Şekil 4.4. LIDC veri setinden alınan nodül örneği ve dört radyoloji uzmanından verilen tanısal RQI öznetelik değerler

Bu açıklamalarda, tomografi dilimleri üzerinde manuel olarak işaretlenen nodül sınırları üzerindeki piksellerin lokasyon konturları da dahil edilmiştir [66]. Her bir akciğer nodülü, çapı ≥ 3 mm olan büyük nodül ve çapı < 3 mm olan non-nodül olarak sınıflanmıştır [67].

Bu çalışmada, çapı 3mm'ye eşit veya daha büyük olan akciğer nodülleri analiz edilmektedir. Habislik değeri ise 3'den az olan nodüller seçim dışı bırakılmıştır [68]. Bu deneyde kullanılan nihai veri setinde 4179 akciğer nodülü yer almakta olup, bunların 2555 tanesi iyi huylu, 1624 tanesi ise kötü huyludur.

4.4. Akciğer Nodül Görüntülerinin Ön-işlemesi

CAD sisteminin geliştirilmesi için kullanılan uygun görüntülerin oluşturulması için görüntüler ön-işlemeye tabi tutulmuştur. İlk olarak, dilim-dilim çalışarak XML dosyalarındaki nodül sınır noktalarını ikili görüntülere dönüştürerek akciğer nodüllerini çıkarmaktayız [43]. Her bir nodül için, her bir radyolog tarafından gerçekleştirilen işaretlemeler, üzerlerinde hacim ölçümlerinin yapılabileceği 3D ikili görüntülere dönüştürüldü [69]. Her bir açıklamayı veri setimiz için benzersiz bir örnek olarak ele almaktayız. Örneğin 4 radyologdan 3'ü tarafından işaretlemeleri yapılan bir nodülün 3 adet 3D ikili görüntüsü olacaktır [70]. İkinci işlem olarak, her bir nodül için, RQI görüntü özneteliklerini çıkarıyoruz. Bu öznetelikler, dokuz adet öznetelikten oluşmakta (İncelik, İç Yapı, Kireçlenme, Küresellik, Marjin, Lobulasyon, Spikülasyon, Doku ve Habislik) ve

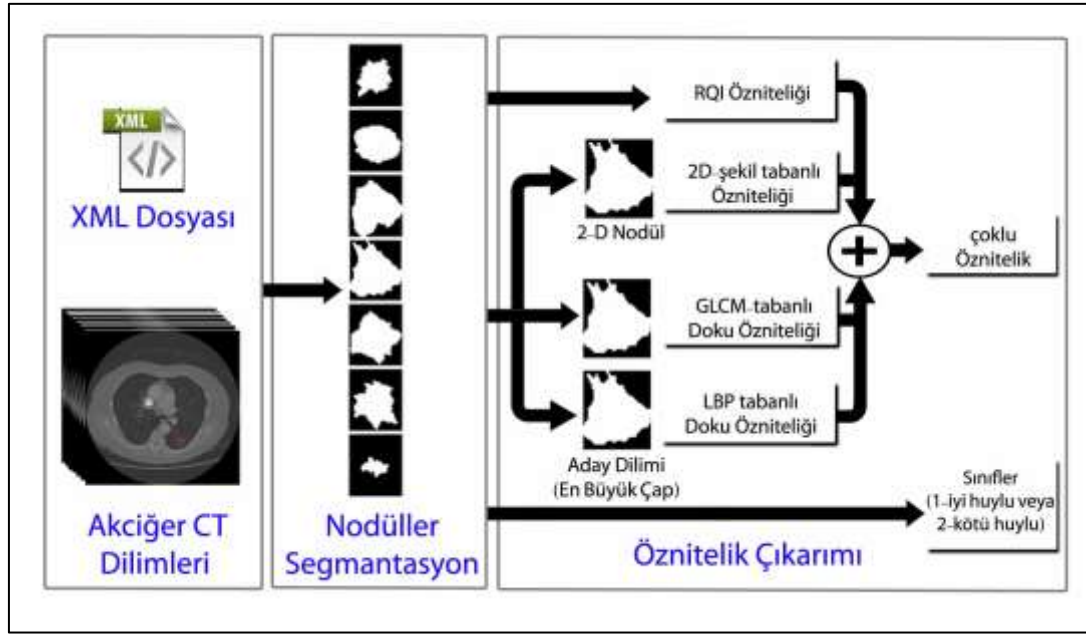
1'den 5'e kadar tam sayılarla ifade edilmektedir. LIDC veri setinin inşa edilmesinde izlenen kurallara göre, nodüller, her bir radyolog tarafından, kompozit habislik derecesi olarak 5'lik bir skala ile derecelendirilmiştir [71]. Nodülleri sınıflandırmadan önce, modellerin eğitilmesi gereklidir ve eğitim de bir hedef çıktı belirlenmesini gerektirmektedir. Bu nedenle, habislik değeri 1 ve 2 olan nodüller iyi huylu; 4 ve 5 olan nodüller ise kötü huylu olarak kabul edilmiş; 3 değerine sahip nodüllerin habisliği ise belirsiz olarak kabul edilmiştir [72]. Çalışmada, radyologlar tarafından belirsiz habisliğe sahip olarak gösterilen nodülleri kapsam dışı tutmaktayız. Daha sonra hedef değişkenler belirlenmiştir. Bu kapsamda, iyi huylu olan aday nodüllere 1, kötü huylu olan aday nodüllere ise 2 değeri verilmiştir. Her bir nodul 8 adet RQI özniteliği ve habislik kategorisi (nodüllere karşılık gelen hedef etiket) içermektedir. Veri setimiz 4.179 nodülden oluşmakta olup, bunların 2555 tanesi iyi huylu, 1624 tanesi ise kötü huylu nodül olup, sınıflandırılmak üzere gerçekleştirilmektedir [52]. Bu veri tabanının içeriğine ilişkin detaylar Çizelge 4.2'de sunulmaktadır.

Çizelge 4.2. LIDC veri setinde yer alan akciğer nodülleri üzerindeki durum ve dilimlerin dağılımı

Akciğer nodülü tipi	CT tarama sayısı	Dilimler Sayısı	Nodül sayısı
İyi huylu	359	12593	2555
Kötü huylu	488	14046	1624
Toplam	847	26639	4179

4.5. Öznitelik Çıkarımı

Öznitelik çıkarımı, girdi görüntülerinin öznitelik setlerine dönüştürülmesi işlemidir. Bu aşamada, kısımlara ayrılmış tomografi görüntülerinde bağlantılı bileşen nodül adayı için öznitelikler çıkarılmaktadır. Bu öznitelikler nodül adayı hakkında bilgi vermektedirler ve üç kategoriye ayrılmaktadırlar: (Radiologist Quantified Image - RQI) öznitelikler, 2D-şekil tabanlı öznitelikler ve doku-tabanlı öznitelikler (Gri Seviye Eş-Oluşum Matrisi (Gray-Level Co-occurrence Matrix - GLCM) ve Yerel İkili Örüntü (Local Binary Pattern - LBP)). RQI öznitelikleri XML dosyasındaki manuel açıklamalardan çıkarılmaktadır. Doku öznitelik ve 2D-şekil tabanlı öznitelikler tomografideki tek bir dilimden çıkarılmaktadır. Akciğer görüntülerinde ön-işleme ve öznitelik çıkarım işlemleri Şekil 4.5'te gösterilmektedir.



Şekil. 4.5. Akciğer nodülü görüntü ön-işleme ve öznitelik çıkarım diyagramı

4.5.1. Radyolog sayısal görüntü öznitelikler (RQI)

Sekiz nicelik nodül öznitelikleri, her Radyoloğun bağımsız olarak değerlendirildiği nodülün subjektif karakteristikleridir (İncelik, İç Yapı, Kireçlenme, Küresellik, Marjin, Lobulasyon, Spikülasyon, Doku ve Habislik).

Sekiz nicelik nodül öznitelikleri ve ilgili malignite derecelendirmesi, özniteliklerin ilgili derecelendirme sistemleri ile birlikte, Çizelge 4.3'te listelenip açıklanmaktadır.

Veri setinden örnek bir nodül, atanan tanısal öznitelik değerleri ile birlikte, Şekil 4.4'te gösterilmiştir.

Çizelge 4.3. LIDC veri setindeki radyolojistler ve ilişkili derecelendirme sistemi tarafından açıklanmalı öznitelikler

Öznitelik	(1) İncelik	(2) İç Yapı	(3) Kireçlenme
oylama sistemi	1 - Son derece İnce 2 - Orta Derecede İnce 3 - Oldukça İnce 4 - Orta Açıkça 5 - Açık	1-Yumuşak Doku 2 - Akışkan 3 - Yağ 4 – Hava	1 - Patlamış Mısır 2 - Lamine 3 - Katı 4 - Merkez Olmayan 5 - Merkez 6 – Yok

Çizelge 4.3. (devam) LIDC veri setindeki radyolojistler ve ilişkili derecelendirme sistemi tarafından açıklanmalı öznitelikler

Öznitelik	(4) Küresellik	(5) Marjin	(6) Lobulasyon
oylama sistemi	1 - Doğrusal 2 - 3 - Oval 4 - 5 - Yuvarlak	1 - Kötü Tanımlanmış 2 - 3 - 4 - 5 - Keskin	1 - Yok 2 - 3 - 4 - 5 - İşaretli
Öznitelik	(7) Spikülasyon	(8) Doku	(9) Habislik
oylama sistemi	1 - Yok 2 - 3 - 4 - 5 - İşaretli	1 - Katı Olmayan 2 - 3 - Parça Katı / Karışık 4 - 5 - Katı	1 - Çok Muhtemel Değil 2 - Orta Olasılıkla 3 - Belirsiz 4 - Orta Şüpheli 5 - Çok Şüpheli

4.5. 2D-şekil tabanlı öznitelik çıkarımı

Çalışmada, 2D şekil tabanlı öznitelikleri, her bir nodül için bütün görüntülere ait alan en büyük olan dilimden hesaplanır. Bu öznitelikleri işlem prosedürleri MATLAB tabanlı yazılım aracılığıyla aşağıda tarif edilmektedir.

1. Alan (A): Akciğer dilimindeki gerçek piksel sayısını, bir büyüklük olarak hesaplanmaktadır [73].
2. Nodül çapı (D): Akciğer diliminde maksimum sınırlayıcı kutu uzunluğu olarak hesaplanmaktadır. Matematiksel formda aşağıdaki şekilde göstermektedir.

$$\text{Çap } (D) = \frac{2}{\pi} \sqrt{A} \quad (4.1)$$

Buradaki A (Alan) nesnenin akciğer dilimindeki piksel sayısı göstermektedir.

3. Yan eksen uzunluğu: Akciğer dilimindeki nodül bölgesi ile aynı normalize ikinci merkezi anlar vardır elips küçük eksen piksel uzunluğu bir sayısal olarak hesaplanmaktadır [74].

4. Çevre: Akciğer nodül Bölgenin sınır etrafında Mesafe bir sayısal olarak dönmektedir. Bölgenin kenarları boyunca piksellerin her bitişik çifti arasındaki mesafenin hesaplanması ile çevre hesaplanmaktadır. Bu rakam, bu nesne için çevre hesaplamasına dahil pikselleri göstermektedir.

MATLAB'ta nodüllerin bölgelerinin ölçü özellikleri hesaplamak için regionprops fonksiyonu kullanılmaktadır [75].

4.5.3. GLCM tabanlı doku özniteliği çıkarımı

Gri Seviye Eş-Oluşum Matrisi (GLCM), iki farklı pikseli arasındaki belirli gri seviyelerin diğer gri seviyelerle birlikte nasıl gerçekleştiğinin yollarını tanımlamak için öznitelik çıkartma yöntemidir [76]. Bu çalışmada akciğer nodüllerinin Haralick doku öznitelikleri GLCM kullanılarak hesaplanmıştır. GLCM'de görüntü yoğunluğuna göre matris boytusu gri seviyeye ölçeklendirilir. Bu matris $P(i, j | d, \theta)$ olarak pikseller arasındaki uzaklık d ve açı θ fonksiyonundan oluşur. Örüntü matrisinin 0° , 45° , 90° ve 135° 'lik dört farklı açıdan hesaplanabilir. Bu çalışmada görüntü tanımda $d=1$ ve $\theta=0$ parametreleri için her bir GLCM'de kontrast, korelasyon, enerji ve homojenlik olmak üzere 4 doku özniteliği çıkarılmıştır. MATLAB'ta graycoprops fonksiyonu vasıtasıyla hesaplanmaktadır [77].

1. Kontrast (K_t), bir piksel (p) ile komşusu arasındaki yoğunluk veya gri seviye varyasyonlarının ölçümüdür.

$$K_t = \sum_{i,j} |i - j|^2 p_{i,j} \quad (4.2)$$

Buradaki i ve j satır ve sütun indislerini göstermektedir.

2. Korelasyon (K_n), bir pikselin tüm görüntünün üzerindeki komşusuyla ne kadar ilişkili olduğuna dair bir ölçümüdür.

$$K_n = \sum_{i,j} \frac{p_{i,j}(i - \mu_i)(j - \mu_j)}{\sigma_i \sigma_j} \quad (4.3)$$

Buradaki μ_i , μ_j , σ_i ve σ_j değerleri, ortalaması ve standart sapmasıdır.

3. Enerji (E), görüntünün homojenliğinin ve kare elemanların toplamını ölçüsüdür. Görüntü ne kadar homojen olursa enerji değeri o kadar büyük olur.

$$E = \sum_{i,j} (p_{i,j})^2 \quad (4.4)$$

4. Homojenlik (H), elementlerin dağılımının diyagonaline yakınlığını ölçen bir değer döndürmüştür.

$$Kn = \sum_{i,j} \frac{p_{i,j}}{1+|i-j|} \quad (4.5)$$

4.5.4. LBP tabanlı öznelik çıkarımı

Haralick texture, GLCM'ye dayalı olarak global bir doku temsili hesaplayan özneliklerin aksine Yerel İkili Örüntü (LBP) dokunun yerel bir temsili hesaplanmıştır. Bu yerel gösterim, her pikseli, çevresindeki pikseller mahallesiyile karşılaştırarak oluşturulmuştur. LBP, çeşitli bilgisayar görüntüsü uygulamaları için başarılı bir şekilde kullanılmıştır [77]. Merkezi pikselin etrafındaki farklı yarıçaplara göre değişik sayıda komşu piksele dayalı, birçok elverişli örüntü teklif edilmiştir. Ancak, standart örnek olarak getirilen ve uygulamalarda en çok kullanılan temel LBP operatörü sekiz komşu piksele dayanmaktadır [78]. Merkez pikselin yoğunluğu, komşusuna eşit veya ondan büyükse, değeri 1 olarak aksi takdirde ise 0 olarak ayarlanmıştır. 8 çevreleyen piksel ile, toplam $2^8 = 256$ olası LBP kodu kombinasyonuna sahiptir. Herhangi bir komşu pikselden saat yönünde veya saatin ters yönünde başlayabilmektedir. Bu ikili testin sonuçları 8-bitlik bir dizide saklanır, sonra ondalık basamağa dönüştürülmüştür. Eşitlik (4.8)'de matematiksel olarak ifade edilmiştir.

$$LBP(x) = \sum_{i=0}^7 S(P_i - P_c) 2^i \quad (4.6)$$

$$S = \begin{cases} 1, & t \geq 0 \\ 0, & t < 0 \end{cases}$$

Burada P_i ve P_c sırasıyla ele alınan komşular ve karşılaştırılan merkez değeri göstermektedir.

Bu değer LBP 2D dizisinde saklanmıştır. 8 komşu pikseleden dolayı LBP 2D dizimiz minimum 0 ve maksimum 255 değerine sahiptir.

Görüntü bloklara bölünerek LBP hesaplaması yapılır ve daha sonra her bloğun histogramı çıkarılır. Histogram oluşturulduktan sonra tüm histogramlar tek bir histogramda birleştirilmiştir.

Seçilmiş nodül özniteliklerinin dağılımı Çizelge 4.4'te verilmektedir.

Çizelge 4.4. LIDC veri setinde bulunan akciğer nodül tipleri üzerindeki bulgu ve dilimlerin dağılımı

Öznitelikler	Açıklama	Boyutlar
RQI	Radyolog sayısal görüntü öznitelikleri, XML dosyalarından çıkarılır. 8 nicelikli nodül özneliğinden oluşur: İncelik, İç Yapı, Kireçlenme, Küresellik, Marjin, Lobulasyon, Spikülasyon ve Doku. 1-5 arası tamsayı değerleri alanmaktadır.	8
2D-şekil tabanlı	2D şekil tabanlı öznitelikler çıkarılır. Özniteliklerde, nodül alan, nodül çapı, çevre ve yan eksen uzunluğu hesaplanır.	4
GLCM	Gri Seviye Eş-Oluşum Matrisi	4
LBP	Yerel İkili Örüntü	10
Toplam		26

4.6. Sınıflandırma Algoritmaları

Sınıflandırıcıların amacı, öznitelik alanını bir kategoriden örnekleri temsil eden bölgelere bölmektir [79]. Bu çalışmada, sınıflandırıcılar akciğer nodüllerinin görüntülerinden elde edilen öznitelikleri analiz etmiş ve daha sonra nodülün malignite düzeyinin kötü huylu veya iyi huylu olduğunu tespit etmiştir. Sınıflandırma yapmak için birçok algoritma mevcuttur. Bu algoritmaların kendi içerisinde özellikle başarılı oldukları alanlar bulunmaktadır. Bu farklılıklar algoritmaların yapıları veya üzerinde çalışacağı verinin

yapısından kaynaklı olabilmektedir. Eğitim ve test işlemlerinde belli olan sınıf bilgileriyle eğitim ve test işlemleri yapılır.

Nodül görüntüleri sınıflandırmak için literatürde çokça çalışma yapılmış ve başarılı yöntemler geliştirilmiştir. Bu yöntemlerden öne çıkanlar Destekçi Vektör Makineleri (SVM), k-en yakın komşu (KNN) gibi yöntemlerdir. Özellikle SVM çok sınıflı çalışma alanlarında başarılı ve uygulaması kolay bir sınıflandırma yöntemi olarak belirtilmektedir.

Bu çalışmada her dilimdeki çıkarılan nodül adaylarından malignite kategorisini (“iyi huylu” veya “kötü huylu”) sınıflandırma amaçlandığı için KNN ve SVM algoritmaları kullanılmaktadır [78].

4.6.1. Destek vektör makinesi (SVM)

Destek vektör makinesi yöntemi sınıflandırma işlemlerinde kullanılan etkin ve basit sınıflandırma yapabilen bir yöntemdir. Bu yöntem sınıflandırma için bir düzlemde sıralanan gruplar arasında her iki gruba en uzak sınırı çizerek grupları sınıflara böler. SVM temel amacı, bir sınıftaki tüm örneklerin diğer sınıf örneklerinden ayrı tutulduğu optimum bir sınır çizgi bulmaktır [80]. SVM, öznelilik çıkarımlı literatürde genel olarak biyoinformatik verilerinin sınıflandırılmasında yaygın olarak kullanılır [81].

4.6.2. K-en yakın komşu (KNN)

K-en yakın komşu (KNN) yöntemi, eğitim verilerinden bir fonksiyonu öğrenmek için denetimli bir makine öğrenme tekniğidir. Nesneleri, öznelilik alanındaki en yakın eğitim verilerine göre sınıflandırır. KNN, her bir eğitim seti ögesini sınıfını belirten ilgili etikete sahip bir demet olarak görür. Her demet n-boyutlu bir uzayda bir nokta olarak verilmiştir, böylece yeni bir X demet verildiğinde, KNN, eğitim setinde X’den tüm demetlere kadar mesafeleri hesaplar ve daha sonra X’teki k en yakın demet içinde en sık sınıfa atar [82].

Temelde, KNN yöntemi, k parametresini önceden belirlemelidir ve en iyi K seçimi verilere bağlıdır. Genel olarak, daha büyük K değerleri, gürültünün sınıflandırma üzerindeki etkisini azaltacaktır. Ancak sınıflar arasındaki sınırların daha az belirgin olmasına neden olur [82].

4.6.3. Hedef oluřturma

Modeller nodül adaylarını sınıflandırmadan önce nodül malignite kategorisinin istatistiksel yaklaşımını, radyolog tarafından atanan malignite değerlerini eşleřtirerek “kötü huylu” veya “iyi huylu” olmak üzere ikili sınıflandırma problemi olarak ele alıyoruz. Bu nedenle, 3’ün altındaki malignite değerleri (yani 1 ve 2) iyi huylu olarak sınıflandırılır ve 3’ün üzerindeki değerler (yani 4 ve 5) kötü huylu olarak sınıflandırılır. Daha sonra, hedef deęişken, iyi huylu olan nodül adaylarına 1 deęerle, kötü huylu olan nodül adaylarına ise 2 deęerle atandığında oluřturulur.

4.6.4. K-Katmanlı apraz doęrulama

Veri setimizdeki akcięer nodülü görüntülerini “kötü huylu” veya “iyi huylu” olarak sınıflandırmak için SVM ve KNN’i kullandık. Sınıflandırma alıřmalarında sınıflandırıcının güvenilirliğini ölçmek için veri seti belirli oranlarda bölünerek bir kısmı eęitim setine bir kısmı ise test setine ayrıştırılmaktadır. Buradaki temel ama eęitim seti ve test veri setlerinin aynı olmaması ve sınıflandırıcının ezber yapmasını önlemektir.

K-katmanlı apraz doęrulama yönteminde veri seti k için belirlenen sayı kadar paraya bölünür ve bu paraların her biriyle tek tek eęitilir ve test edilir [40]. Yani $k=10$ seilince veri seti 10 paraya bölünür ve 9 parasıyla eęitim yapıp kalan bir parayla da test yapılır. Daha sonra test için seilen para dönüřtürülerek 10 defa bu iřlem tekrarlanır. Böylece bütün veriler için eęitilen sınıf aynı zamanda bütün veriler için de test edilmiř olur.

4.6.5. Karıřıklık matrisi

Karıřıklık matrisi bir deęerlendiricinin ürettięi sonuçların performansının gösterildięi özel bir tablo yapısıdır [83]. Bu tabloyu oluřturmak için sütunlara deęerlendiricinin tahminleri ve satırlara da gerek durumlar yazılır. Satır ve sütunlar doldurulup karıřıklık matrisi oluřturulduktan sonra sistemin tahminleri gerekle ne kadar uyuyor, sistemin a sınıfı dediklerinin gerekte kaı a sınıfında řeklinde okumalar yapılabilmektedir. Karıřıklık matrisi uyarlanabilir olabileceęi gibi sıklıkla kullanılan versiyonu iki sınıf için izelge 4.5’te gösterilmiřtir. Tüm aday nodüller doęru pozitif, doęru negatif, yanlış pozitif ve yanlış negatif olarak sınıflandırılmıřtır.

Çizelge 4.5. Karmaşıklık matrisi

		Tahmin edilen	
		kötü huylu	iyi huylu
Gerçek	kötü huylu	TP	FN
	iyi huylu	FP	TN

Karmaşıklık matrisinde hesaplanan değerlerin açıklaması şu şekildedir:

TP (doğru pozitif): Veri setinde kötü huylu nodül sınıfında olup, kötü huylu nodül sınıfında tahmin edilen örnek sayısı.

FN (yanlış negatif): Veri setinde kötü huylu nodül sınıfında olup, iyi huylu nodül olarak tahmin edilen örnek sayısı.

FP (yanlış pozitif): Veri setinde iyi huylu nodül sınıfında olup, kötü huylu nodül olarak tahmin edilen örnek sayısı.

TN (doğru negatif): Veri setinde iyi huylu nodül sınıfında olup, iyi huylu nodül olarak tahmin edilen örnek sayısı.

Karışıklık matrisi elde edildikten sonra, performans ölçümü için en yaygın yöntemler doğruluk (accuracy), kesinlik (precision), anma (recall) ve F-skoru değerlerine ait ölçütlerdir.

4.6.6. Doğruluk değeri

Bir algoritmanın doğru olarak sınıflandırma yaptığı örnek sayısının, toplam örnek sayısına oranı bu değeri vermektedir. Bu hesaplama aşağıdaki şekilde Eşitlik. 4.7'de gösterilmiştir.

$$\text{Doğruluk} = \frac{(TN + TP)}{(TN + TP + FN + FP)} \quad (4.7)$$

4.6.7. Kesinlik değeri

Kesinlik değeri, kötü huylu nodül olarak sınıflandırılan örneklerin sayısının tüm kötü huylu nodül olarak tahmin edilen örneklerle oranıdır. Bu hesaplama aşağıdaki şekilde Eşitlik. 4.8’de gösterilmiştir.

$$Kesinlik = \frac{TP}{(TP + FP)} \quad (4.8)$$

4.6.8. Anma değeri

Anma değeri, kötü huylu nodül olarak sınıflandırılan örnek sayısı ile gerçek değeri kötü huylu nodül olan örneklerin sayısına olan oran şeklindedir. Bu hesaplama aşağıdaki şekilde Eşitlik. 4.9’da gösterilmiştir.

$$Anma = \frac{TP}{(TP + FN)} \quad (4.9)$$

4.6.9. F-skoru değeri

F-skoru, kesinlik ve anma değerlerinin harmonik ortalamasının alınmasıyla elde edilmektedir. Bu ölçüt iki sınıflandırıcının tek ölçüt ile değerlendirilebilmesi için kullanılmaktadır. Bu hesaplama aşağıdaki şekilde Eşitlik. 4.10’da gösterilmiştir.

$$F - Skoru = 2 * \frac{anma * kesinlik}{anma + kesinlik} \quad (4.10)$$

4.6.10. ROC eğrisi

ROC ölçütü, veri seti üzerinde gerçekleştirilen sınıflama işleminin performans ölçümünde kullanılan bir diğer önemli ölçüttür. Bu ölçütte sınıflandırma işleminin hangi oranda doğru tahminde bulunduğu farklı bir yaklaşım kullanılarak incelenmektedir. ROC eğrisinin elde edilmesinde anma ve seçicilik ile ifade edilen iki kavram kullanılmaktadır. Y ekseninde hesaplanan anma değeri X ekseninde ise hesaplanan (1 - seçicilik) değeri yer

almaktadır. Bu iki değerin kesişimleri ile elde edilen noktaların birleşimi ile ROC eğrisi elde edilmektedir.

$$Se\check{c}icilik = \frac{TN}{(TN + FP)} \quad (4.11)$$

ROC analizinden elde edilen değeri, ROC eğrisinin altında kalan alanın büyüklüğünü göstermektedir. ROC testinde 1 en iyi değeri, 0,5 ise başarısız bir sınıflandırma yapıldığını göstermektedir. ROC değeri 1 olduğunda veri seti üzerinde hiçbir yanlış tahmin yapılmadığı anlaşılmaktadır ve bu ROC eğrisi koordinat düzleminde (0,0), (1,0) ve (1,1) noktalarını birleştiren bir doğru çizmektedir. ROC eğrisi $y=x$ fonksiyonuna eğilim gösterdiği kadarıyla başarısız bir işlem gerçekleştirildiğini ortaya koymaktadır.

5. ÖNERİLEN SİSTEMİN OLUŞTURMASI

Bu çalışmada kullanılan deney veri seti, en az bir radyolog tarafından açıklama düşülen 4.179 akciğer nodülünü kapsamaktadır. Diğer taraftan, bu kapsamda, belirgin olmayan nodüllerin yer aldığı dilimler atılmış ve sonuç olarak veri setinde 26,639 dilim yer almıştır. Bu veri setinin detaylı içeriği Çizelge 4.2’de listelenmiştir. Veri setinden model oluşturulduktan sonra doğruluk oranını test etmek için $k=10$ alınarak, k -katlamalı çapraz doğrulama yöntemi kullanılmıştır.

5.1. Sistem İhtiyaç Özellikleri

5.1.1. Donanım

Deneyler Hadoop kümesi üzerinde bir master düğüm ve iki slave düğüm olmak üzere üç düğüm üzerinde yapılmıştır: Her bir düğümün konfigürasyonu ve küme içindeki oynadığı rol Çizelge 5.1’de gösterilmektedir.

Çizelge 5.1. Küme konfigürasyonu

Düğüm adı	İşlemci	RAM	HD	Ağ
Master	Core i7-3610QM CPU@2.30GHZx8	16GB	1TB	Wifi/DSL(16 Mbps)
Slave1	Core i7-4200M CPU@2.50GHZx4	8GB	500GB	Wifi/DSL(16 Mbps)
Slave2	Core i5-4200M CPU@2.50GHZx4	8GB	500GB	Wifi/DSL(16 Mbps)

Master düğüm, kullanıcı arayüzü olarak çalışırken iki ana Hadoop işlemine, Namenode ve JobTracker, ev sahipliği yapmaktadır. Namenode, ilgili Datanode işlemleri aracılığı ile slave makineleri koordine ederek, dosyaların ve her bir HDFS bloğunun kaydını tutarak HDFS’yi yönetmektedir. JobTracker, her bir hesaplama düğümünün TaskTracker’ını yöneten MapReduce çerçeve ana işlemidir. Yük-dengesi ve hata toleransının sağlanması; tüm düğümlerin kendi paylarına düşen girdi veri yığını almalarını sağlamak; çalıştırılmayan bölümleri yeniden dağıtmak, görevleri arasında yer alan hususlardandır.

5.1.2. Yazılım

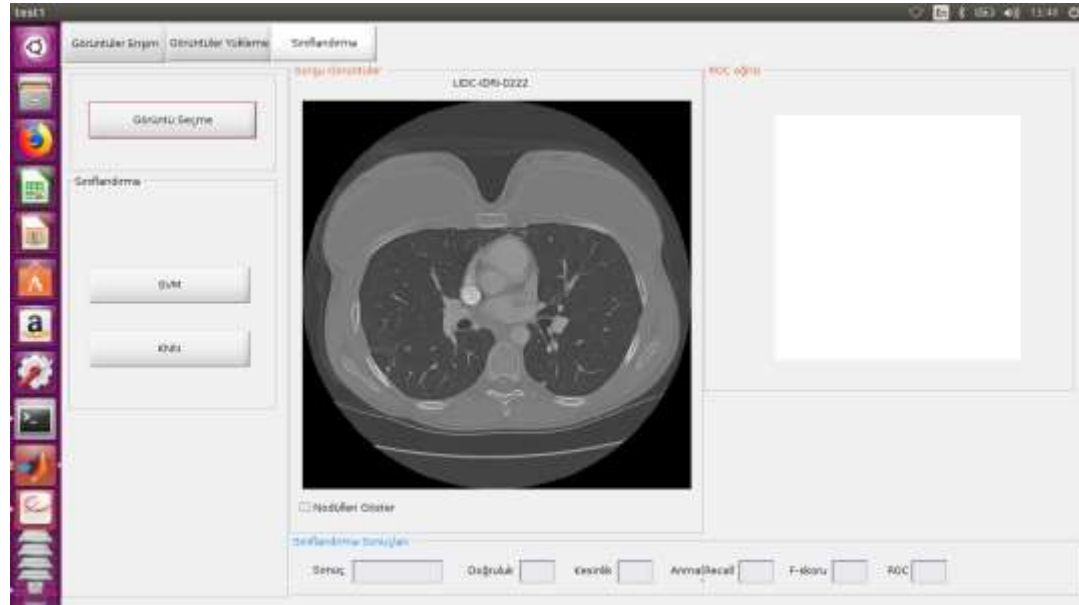
Önerdiğimiz sistemde kullanılan yazılıma ait detaylar Çizelge 5.2’de yer almaktadır.

Çizelge 5.2. Yazılım konfigürasyonu

Yazılım adı	Versiyon
İşletim sistemi	Ubuntu 14.06, 64-bit
Apache Hadoop	2.8.1
Java	Java(TM) SE Runtime Environment (build 1.8.0_66-b17)
MATLAB	R2017b

5.2. Sistemin Uygulanması

Kullanıcı ve sistem arasındaki etkileşim kolaylığı için sistemde GUI sağlanmıştır. GUI MATLAB geliştirilmiştir.

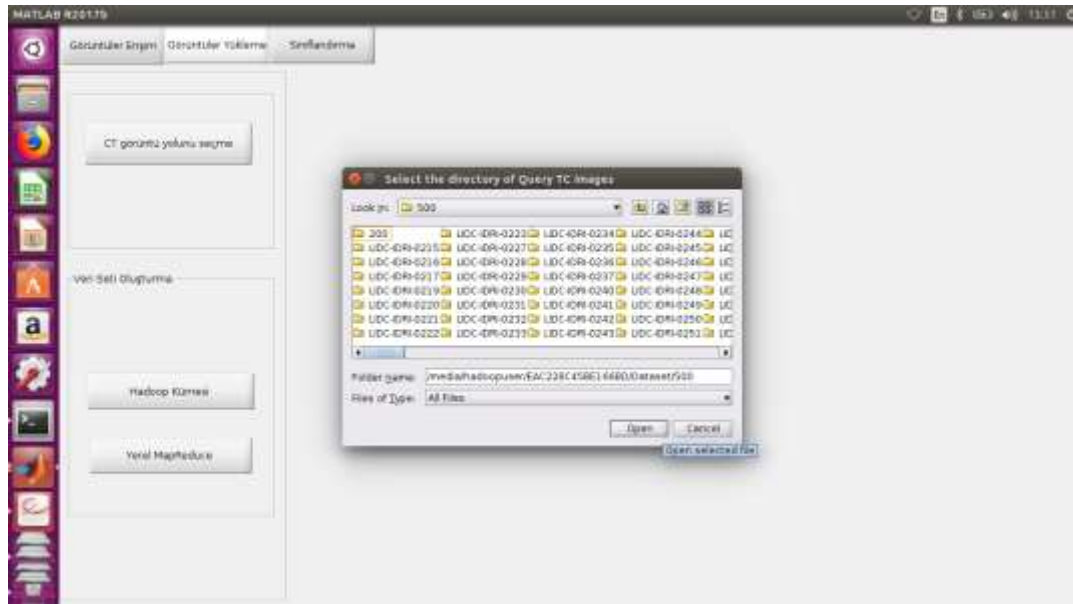


Resim 5.1. Grafiksel kullanıcı arayüzünün tasarımı (GUI)

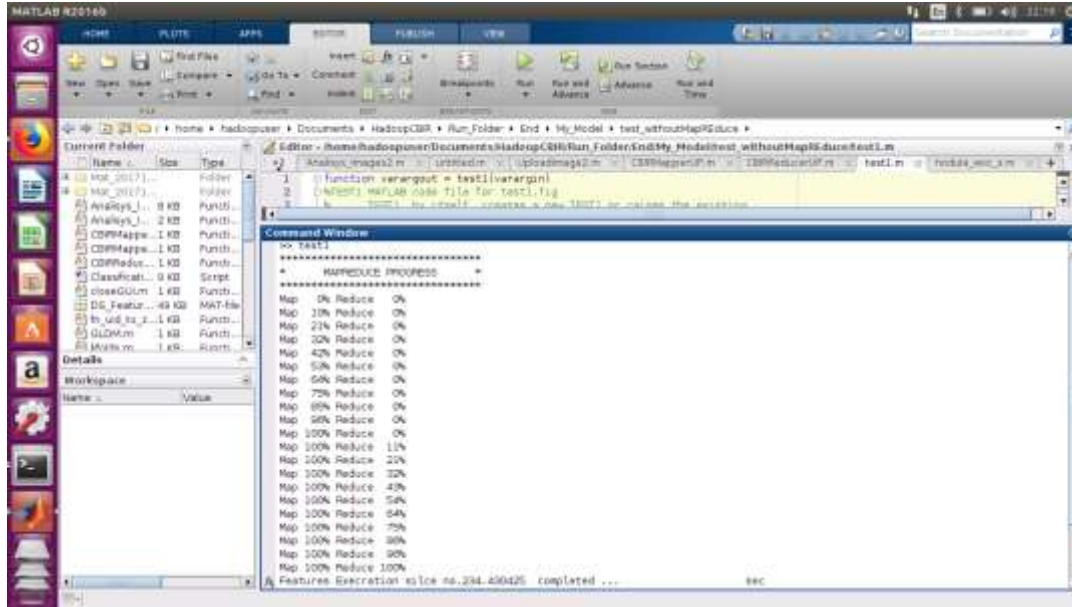
5.2.1. Veri seti oluşturma ve yükleme işlemi

Önerdiğimiz sistemde, Hadoop kümesinde kodu çalıştırmak için MATLAB'taki GUI geliştirilmiştir. Dışa aktarılan m dosyası, küme ve düğüm makine düzgün bir şekilde yapılandırıldıktan sonra MATLAB Hadoop kümesinde çalıştırılır. Bu GUI'nin iki ana işlemi vardır. Veri seti oluşturma ve yükleme işlemi ve görüntüler erişim ve sınıflandırma işlemidir. Kullanıcı yükle düğmesini tıkladığında, görüntüyü Hadoop'un dağıtılmış dosya sistemi'ne (HDFS) yükleme yoluyla oluşturulur, radyologlar tarafından belirlenen görüntülerden öznitelik ve nodülleri çıkararak HDFS'ye yüklenir.

Resim 5.2 ve 5.3'te verilerin HDFS veri kümesinde gerçekte nasıl depolandıklarını açıklar. Önerdiğimiz sistemde, HDFS görüntüleri depolamak için kullanılır. HDFS depolamada, veri kümesi mat biçiminde saklanır. Veriler daha küçük parçalara ayrılmıştır (bloklar adı verilir) ve bu parçalar Hadoop kümesi boyunca dağılmıştır. Hadoop bize ikili dosyaları okuma / yazma imkânı sunar. Sonuç olarak, baytlara dönüştürülebilen her şey HDFS'ye kaydedilebilir. Bu, büyük miktar görüntü veri seti işleme için istenen ölçeklenebilirliği sağlar.

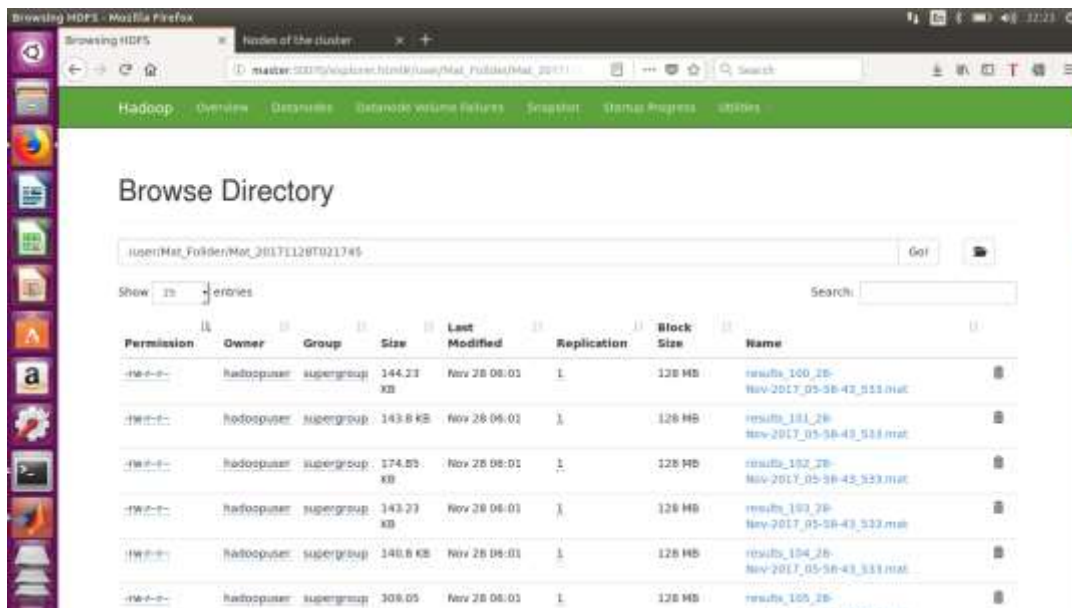


Resim 5.2. Veri seti oluşturma ve yükleme işlemi

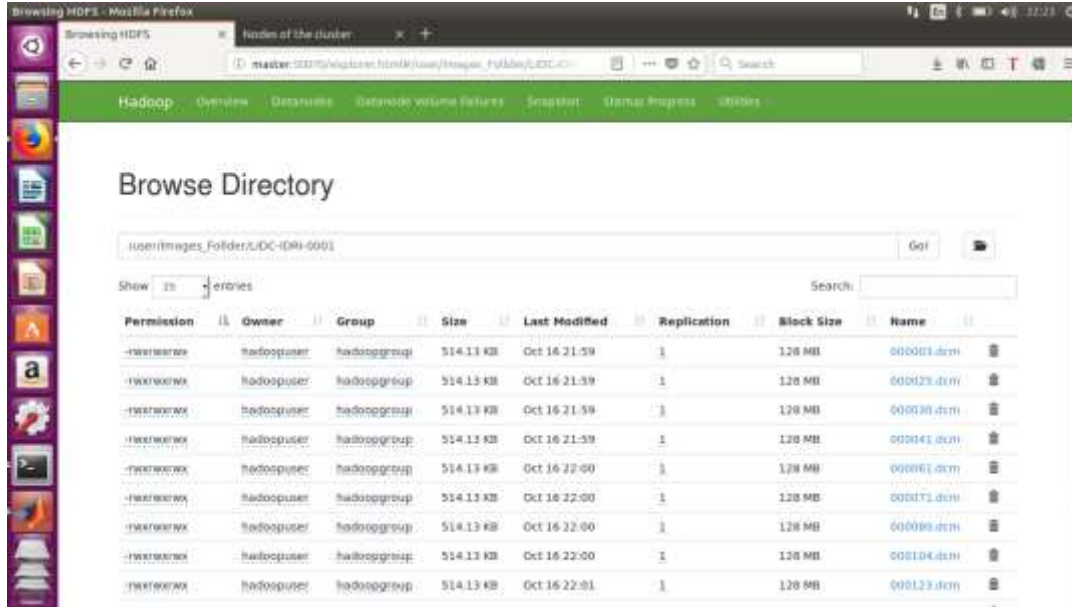


Resim 5.3 MATLAB'ta MapReduce aşamalı işlemi

HDFS'de depolanan görüntü verileri, metin dosyası biçimindedir ve bu metin dosyaları, görüntüler hakkında hiçbir bilgiyi Hadoop veri kümesi yöneticisine bile göstermez. Görüntüler, boyut olarak birbirlerinden ayırt edilemez, yani tüm görüntüler sunucuda aynı boyutta görünür. Ayrıca, görüntülerin verileri ikili dosyadadır. Bu nedenle, görüntüler görülemez. Bir yöneticinin bazı görüntülere erişmek istediği bir senaryo düşünün. Tüm verileri arayabilir, ancak HDFS veri kümelerinde resimlerle ilgili bilgi bulunamayacaktır. Resim 5.4 ve 5.5'te gösterilir.



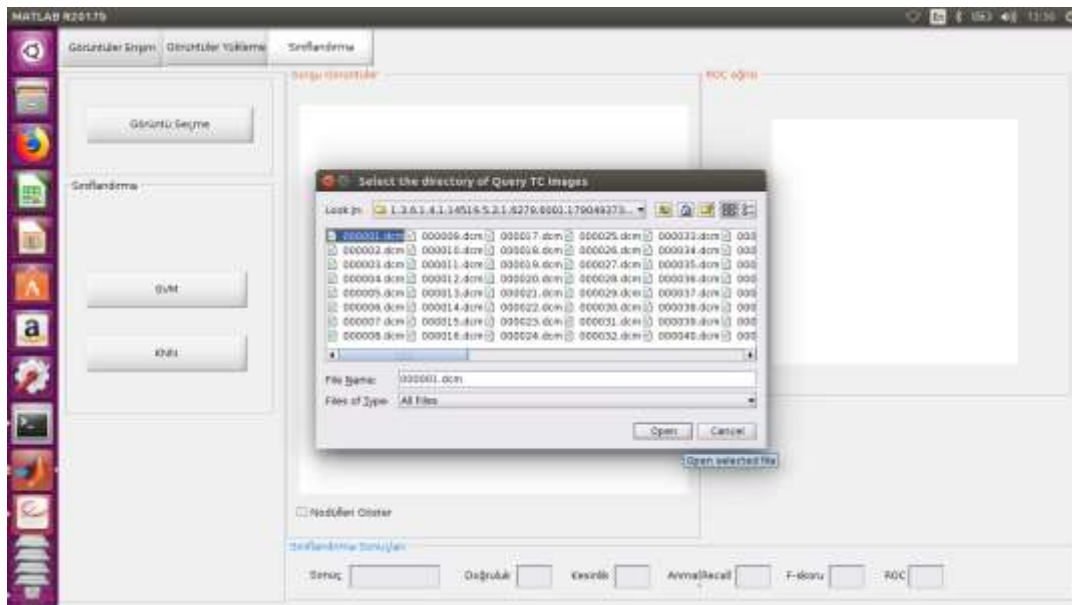
Resim 5.4. HDFS'de sıralanmış mat dosyaları formatı olarak veri seti görüntüleyin



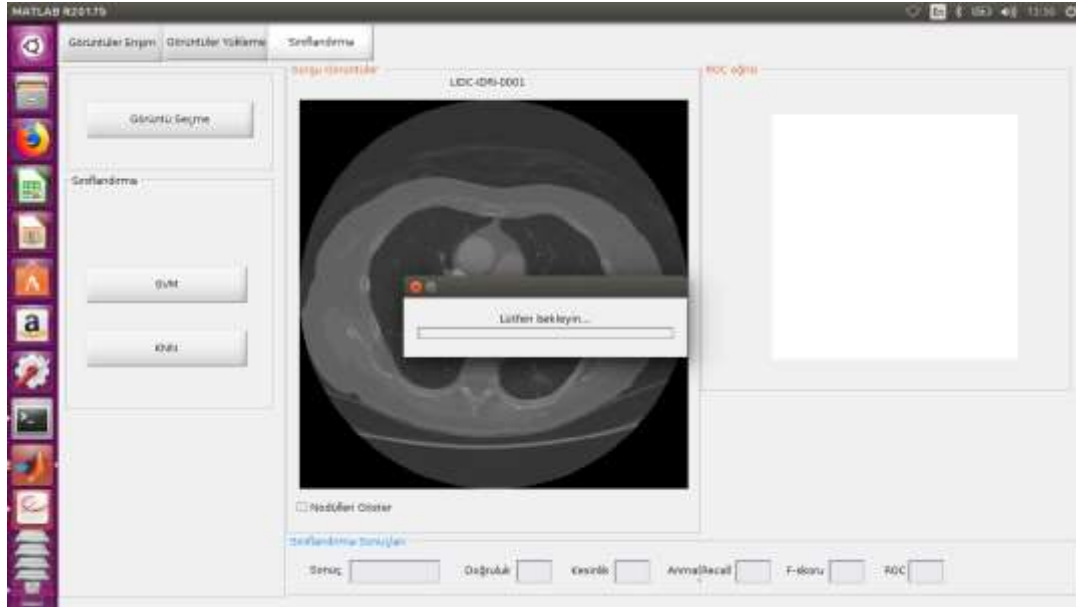
Resim 5.5. HDFS’de sıralanmış dcm görüntüleri formatı olarak veri seti görüntüleyin

5.2.2. Görüntüler erişim ve sınıflandırma işlemi

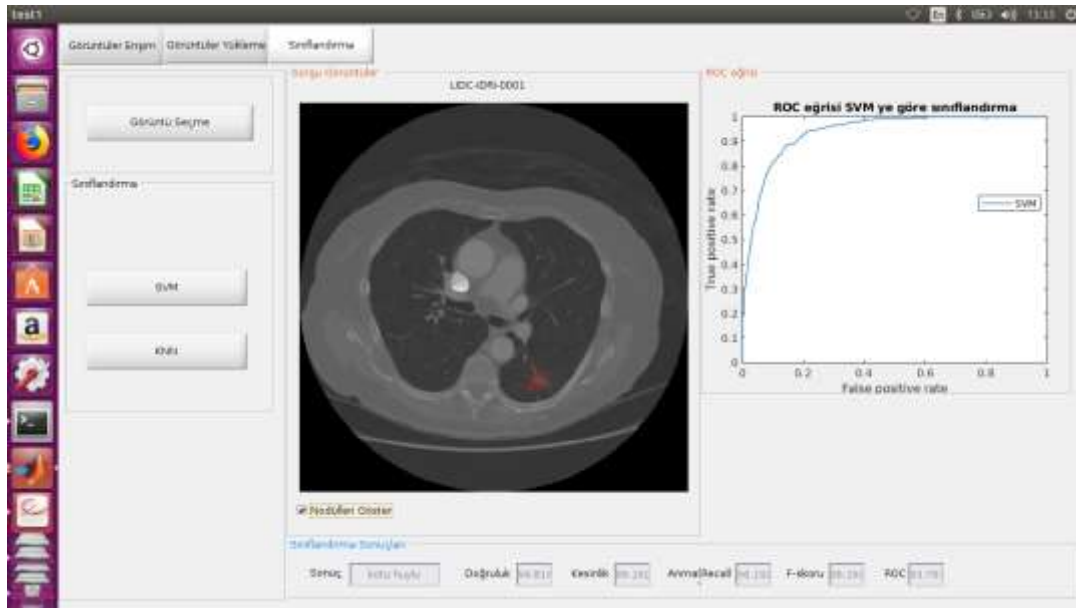
Görüntüler erişim ve sınıflandırma işleminin Resim 5.6, 5.7 ve 5.8’de gösterilen yükleme işlemi gibidir. Ondan GUI, CT tarama görüntüsünü sorgulamak için bir yol sağlanmış ve bunun nodülü çıkarılmıştır. Sorgu nodülü görüntüsü daha sonra Hadoop’un dağıtılmış dosya sistemi (HDFS) veri kümesinde karşılaştırılır. Bu aşamada, sorgu nodülü iyi huylu veya kötü huylu olarak SVM ve KNN ile sınıflandırılmaktadır.



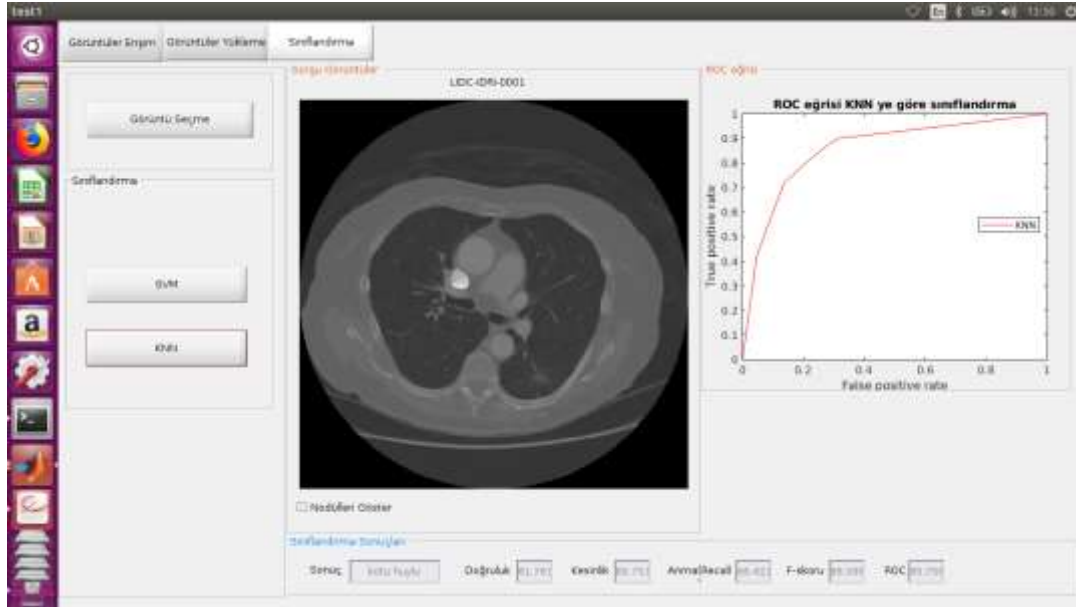
Resim 5.6. Görüntüler seçme işlemi



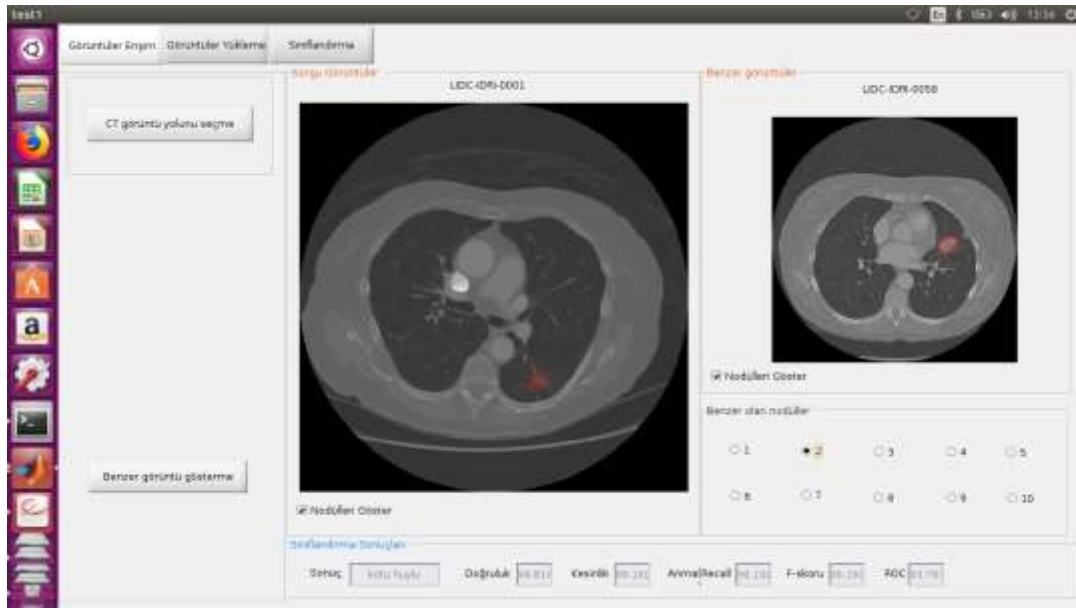
Resim 5.7. Görüntüler sınıflandırma işlemi



Resim 5.8. Görüntüler SVM sınıflandırıcı değerlendirilmesi



Resim 5.9. Görüntüler KNN sınıflandırıcı değerlendirilmesi



Resim 5.10. Görüntüler erişim ve benzer olan görüntüler göstermesi

Resim 5.10'da benzer olan görüntüler gösterilmektedir. Daha önce tartışıldığı gibi, Yükleme için aynı işlemler yapılır. Test için, görüntüler veri setinde aynı dosya biçiminde, tüm veri seti görüntüleri ile karşılaştırılır. Sonuç birkaç benzer görüntüleri göstermeyi sağlamaktadır.

6. ÖNERİLEN SİSTEMİN DEĞERLENDİRİLMESİ

Başlangıçta, deneyler, değerlendirme ve test amacıyla daha ufak bir görüntü seti üzerinde yapılmıştır. Bu set üzerinde çalıştırılan kod ve gerçekleştirilen tüm operasyonlar (.mat) formatında dosyalarda depolanmıştır. Bu sistem, farklı sayılarda dilim üzerinde tomografi görüntüsüne imkan vermektedir. Veri setindeki tomografi görüntüleri (.dcm) formatında olup, 512x512 piksel boyutundadır. Deneyler, özneteliklerin çıkarımı, yükleme, arama ve erişim gibi HDFS'deki görüntülerin depolanmasına benzer nodul aşamaları içermiştir.

6.1. Yükleme ve Erişim Sürelerinin Sonuçları

Bu görüntüler bir çok dilim içeren tomografi kayıtlarında depolanmıştır. Tomografi boyutları 1, 10, 20, 30, 50, 100, 200, 500, 847'dir. Deneyler MATLAB yüklü bulunan bir master ve iki slave bilgisayardan oluşan Lokal MapReduce ve Hadoop kümesi ile gerçekleştirilmiştir. Her bir bilgisayar, farklı işlemci, RAM ve HD özelliklerine sahiptir. HDFS'de saklanacak olan veri GUI yardımıyla girilmiştir. Bir tomografi görüntüsünden 847 CT tomografi görüntüsüne kadar farklı sayıda görüntü için tek bir noktada yükleme, veri tabanı üzerinde belirli bir nodül arama ve bu nodüle iyi huylu veya kötü huylu olarak sınıflandırma için geçen süreyi test ettik. Yerel MapReduce ve Hadoop küme yöntemleri arasında yükleme ve erişim süreleri bakımından görülen farklılıklara dair sonuçlar çizelge 6.1'de yer almaktadır.

Çizelge 6.1. Görüntüler yükleme ve erişim deneyleri çalışma süreleri

CT sayısı	Nodül sayısı	Yükleme süresi (sn)		Erişim süresi (sn)	
		Yerel MR	Hadoop küme	Yerel MR	Hadoop küme
1	4	12.11	59.02	1.16	1.06
10	42	29.65	190.55	3.52	4.21
20	119	70.02	318.46	7.99	10.11
30	162	71.33	419.68	10.83	15.47
50	305	113.52	545.95	17.77	21.92
100	554	221.45	1621.86	37.37	48.76
200	1129	430.61	2275.28	73.48	84.67
500	2496	967.82	5326.34	190.75	227.46
847	4273	1749.01	8981.14	227.46	830.23

6.2. Önerilen Sistemin Sınıflandırma Sonuçları

Bu bölümde, farklı öznitelik çıkarım KNN ve SVM sınıflandırmaları ile testler gerçekleştirilmektedir. Test seçeneği olarak tüm sınıflandırıcılar için $k=10$ alınarak, k -katlamalı çapraz doğrulama (15 fold cross-validation) yöntemi kullanılmıştır. Bu yöntemde veri kümesi 10 parçaya bölünür ve 10 parçadan 1 tanesi test için geriye kalan 9 tanesi eğitim için kullanılır.

Önerilen sistemin veri seti üzerindeki karmaşıklık matrisi kullanılarak tüm aday nodüller sınıflandırılmıştır. Karışıklık matrisi elde edildikten sonra, performans ölçümü için doğruluk (accuracy), anma (recall), kesinlik (precision), F-skoru ve ROC eğrisi altındaki alan oranları hesaplanmaktadır.

SVM ve KNN sınıflandırmaları ve öznitelik çıkarım algoritmaları kullanılarak yapılan test sonuçları Çizelge 6.2'deki verilmiştir.

Çizelge 6.2. Öznitelikler SVM ve KNN'nin sınıflandırmasında performans karşılaştırma tablosu

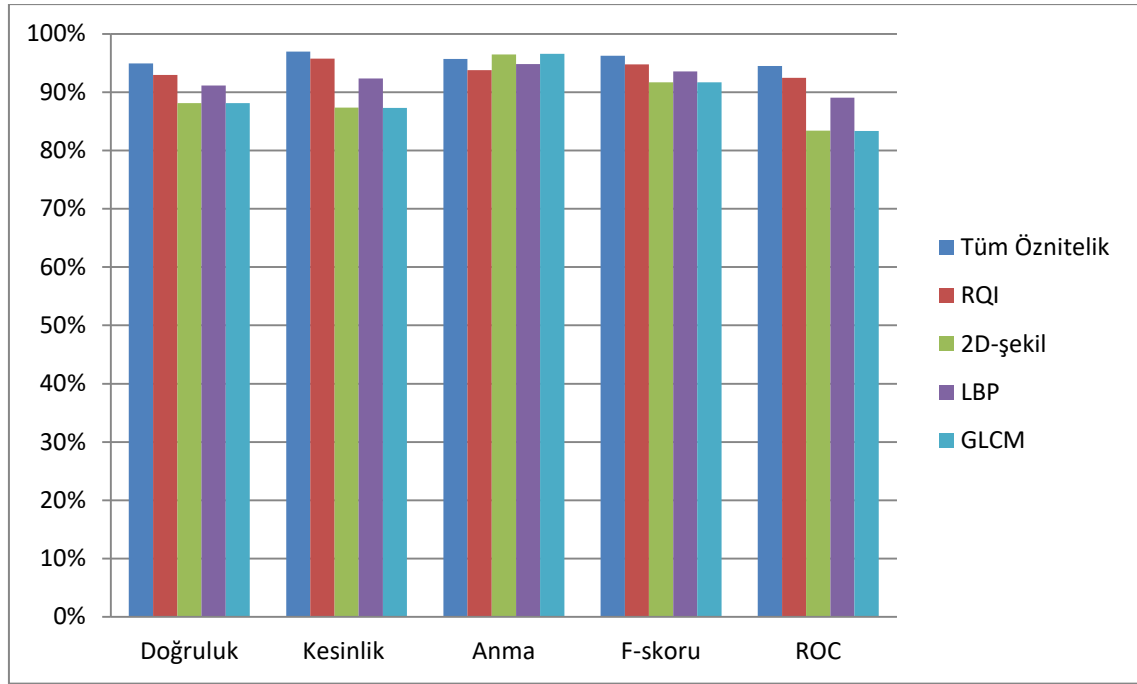
Algoritmalar		Karmaşıklık matrisi		Doğruluk (%)	Kesinlik (%)	Anma (%)	F-skoru (%)	ROC (%)
Tüm (26 öznitelik)	SVM	755	34	94.91	96.97	95.69	96.24	94.48
		25	346					
	KNN	699	90	85.26	89.62	88.59	89.10	83.38
		81	290					
RQI (8 öznitelik)	SVM	740	49	92.93	95.73	93.79	94.75	92.45
		33	338					
	KNN	774	65	83.53	85.18	91.76	88.35	78.90
		126	245					

Çizelge 6.2. (devam) Öznitelikler SVM ve KNN'nin sınıflandırmasında performans karşılaştırma tablosu

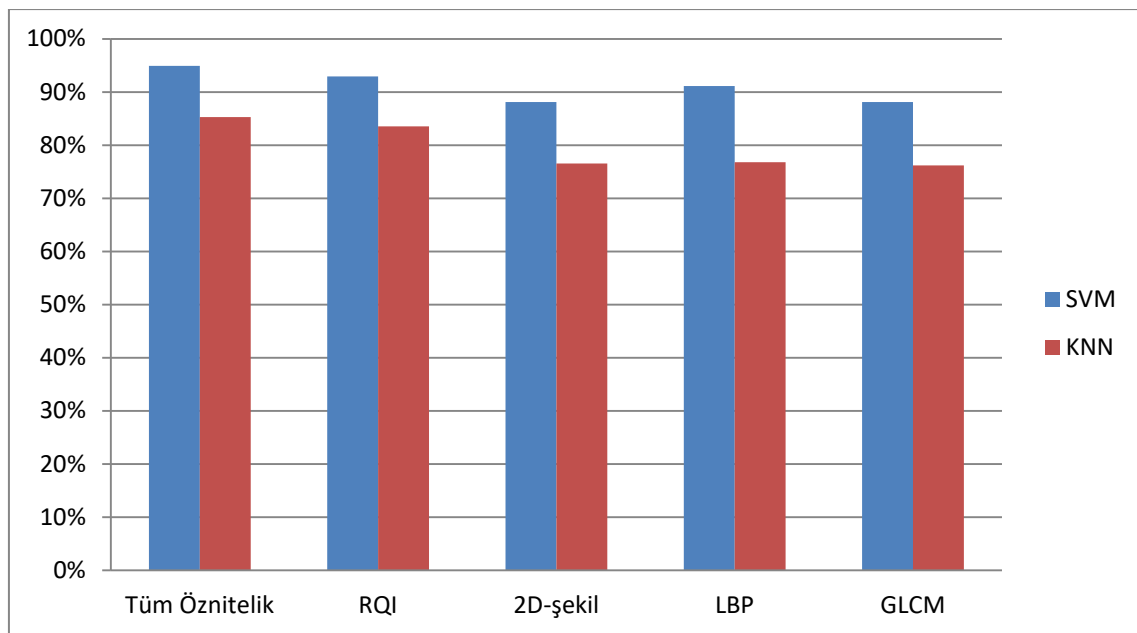
Algoritmalar		Karmaşıklık matrisi		Doğruluk (%)	Kesinlik (%)	Anma (%)	F-skoru (%)	ROC (%)
2D-şekil (4 öznitelik)	SVM	761	28	88.10	87.37	96.45	91.69	83.40
		110	261					
	KNN	668	121	76.55	81.56	84.66	83.08	71.98
		151	220					
LBP (10 öznitelik)	SVM	748	41	91.12	92.34	94.80	93.55	89.04
		62	309					
	KNN	671	118	76.81	81.63	85.04	83.30	72.17
		151	220					
GLCM (4 öznitelik)	SVM	762	27	88.10	87.28	96.58	91.70	83.33
		111	260					
	KNN	663	126	76.21	81.55	84.03	82.77	71.80
		150	221					

Çizelge 6.2’de ve Şekil 6.3’te tüm nodül öznitelik grupları için doğruluk, kesinlik, anma, F-skoru ve ROC kriterlerine göre SVM ve KNN sınıflama sonuçları gösterilmiştir. Burada amaç, hangi sınıflayıcının, hangi öznitelik grupların üzerinde daha yüksek performans sağladığını incelemektir. Test sonuçları veri setlerin üzerinde yapılan deneylerde, RQI öznitelikler ile birleştirilen değer öznitelikler (2D şekel, LBP ve GLCM), kesinlik ölçümlerinde en yüksek başarıyı elde etmiştir. Anma ölçümünde, birleştirilmeyen 2D şekel ve GLCM öznitelikleri daha yüksek başarıya ulaşmıştır. Genel olarak başarımlar incelendiğinde, tüm özniteliklerin birlikte kullanımı ile SVM sınıflama yöntemini kullanılarak elde edilen sonuçların diğerlerinden iyi olduğu tespit edilmiştir. SVM

algoritması ise diğer algoritmaya kıyasla daha yüksek performans sağlamıştır. Şekil 6.3'te de farklı öznitelik grupları üzerinde elde edilen sonuçlar ve farklı SVM ve KNN algoritmaları ile gerçekleştirilen testlere ait test sonuçları görülmektedir. SVM algoritması kullanılarak doğruluk, anma, kesinlik ve F-skoru %94.91, %96.97, %95.69 ve %96.24'e yükselttiğini göstermektedir.

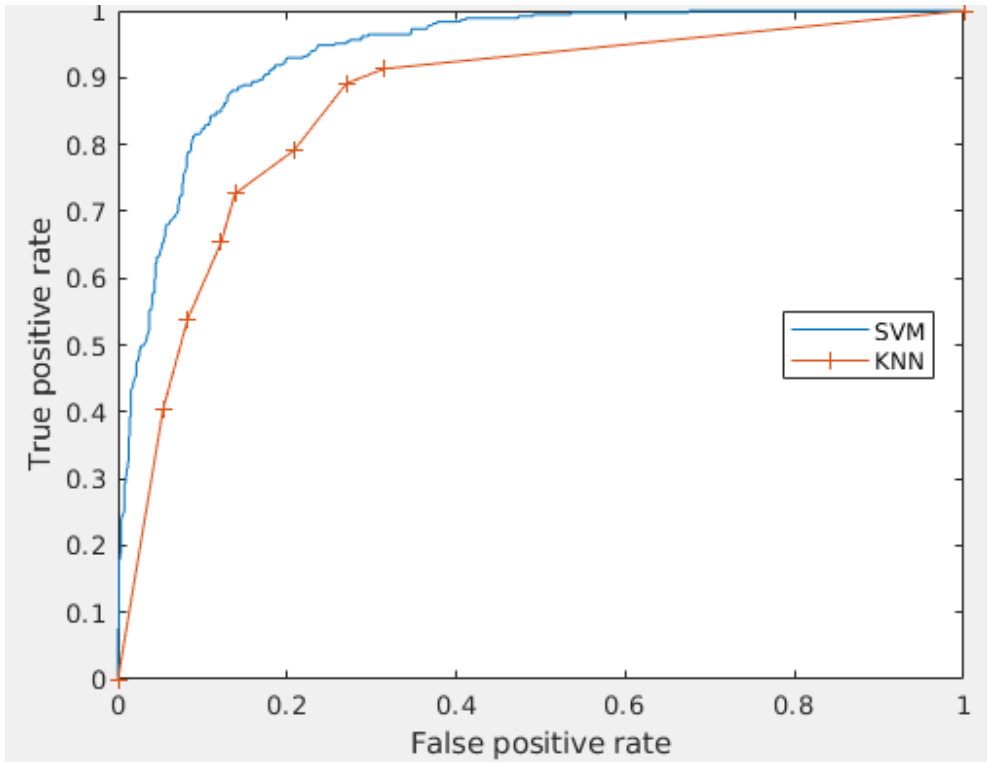


Şekil 6.3. SVM algoritması test sonuçları



Şekil 6.4. Algoritma doğruluk oranı grafiği

Yapılan deneyler sonucunda, Şekil 6.4'te doğruluk performans SVM sınıflayıcı, tüm öznitelikler ile en yüksek başarımlar ölçümlerine ulaşmıştır. SVM sınıflayıcıda tüm özniteliklerde başarımlar çok daha yüksek olmasına rağmen, 2D şekel özniteliklerle yapılan deneylerde %7 civarında doğruluk düşüşü gerçekleşmiştir. KNN sınıflayıcıda ise SVM sınıflayıcının aksine tüm özniteliklerde başarımlar çok daha düşük ve %10 civarında doğruluk düşüşü gerçekleşmiştir.



Şekil 6.5. ROC eğrisi SVM ve KNN sınıflandırma grafiği

Şekil 6.5'teki SVM üzerinde yapılan deneyde ROC eğrinin altındaki %93 oranı elde edilirken, KNN'de %85 oranı elde edilmektedir. Bu değerleri dikkate alındığında, en yüksek performansa sahip algoritmanın SVM olduğu görülmektedir.

6.3. Deney Sonuçlarının Literatürdeki Çalışmalarla Karşılaştırılması

LIDC veri seti, açık erişimli bir kaynak oldan dolayı birçok farklı çalışmada kullanılmıştır. Ancak böyle veri seti kullanılmasına rağmen, farklı Kullanılan yöntemleri ve farklı varsayımlar yüzünden literatürdeki çalışmaları doğrudan karşılaştırmak zordur [84].

Çizelge 6.3'te, çalışmalarında nodül özelliklerini sınıflama performans sonuçları ve kullanılan yöntemleri verilmiştir. Bu çalışmada en yüksek başarıyı gösteren SVM, sınıflayıcının tüm öznitelik grupları için anma oranı diğer çalışmalarla karşılaştırılabilir bir değere ulaşmıştır. Genel olarak yapılan nodül sınıflandırma deneylerinin sonuçları diğer literatürlerin sonuçları ile karşılaştırıldığında nodül sınıflandırmasının başarısının benzer başarıya ulaşmıştır.

Çizelge 6.3. Farklı literatürlerin performans karşılaştırması

Litratür	Kullanılan yöntem	Veri seti	Performans oranı
Lakshmanaprabu, S. K. (2019) [10]	ODNN, MGSA ve LDA	100 Ak.Gö.	recall=96.2% spec=94.2% accu=94.56%
Nasrullah, N. (2019) [31]	GBM ve MixNet	LIDC-IDRI	recall=94% spec=90% AUC=99%
Ardila, D. (2019) [32]	CNN	NLCST	AUC=94.4%
Kavitha, M. S. (2019) [33]	HSCNN	LIDC	accu=91% spec=76% recall=93% AUC=93%
Kasinathan, G. (2019) [34]	CNN ve MSGD	LIDC-IDRI	accu=97% recall=89% spec=91%
Nasser, I. M. (2019) [35]	ANN	anket	accu=96.67%
Prabukumar, M. (2019) [36]	FCM ve SVM	ELCAP	recall=98.13 spec=98.79% accu=98.51% AUC=99%
Günay, M. (2019) [37]	AIS ve PCC	-	accu=82%
Coudray, N. (2018) [38]	CNN	TCGA	AUC=97%
Nishio, M. (2018) [39]	DCNN, LBP-TOP ve SVM	Aquilion ONE	accu=68.0%
Zhu, X. (2018) [40]	Histolojik ve GLCM	NSCLC	recall=83% spec=93% AUC=90%
Gong, J. (2018) [41]	SVM, NBC ve LDA	LIDC	AUC=94%

Çizelge 6.3. (devam) Farklı literatürlerin performans karşılaştırması

Litratur	Kullanılan Teknik/Yöntem	Veri seti	Performans doğruluğu
Makaju, S. (2018) [42]	SVM	LIDC	accu=92%
Wan, C. C. (2017) [26]	SVM ve MLP	LIDC	recall=93% spec = 86 % prec = 87%
Akram, S. (2015) [28]	ANN	LIDC	accu= 96.68% recall=96.95% spec=96.39% AUC=99%
Silva, G. L. (2016) [29]	CNN	LIDC-IDRI	accu = 82.3% recall = 79.4% spec = 83.8%
Kumar, D. (2015) [30]	otomatik kodlayıcı ve derin öznitelikler	LIDC	accu=75.01% recall=83.35%
Krishnamurthy S. (2016) [33]	üç boyutlu segmentasyon ve Şekil öznitelikleri	LIDC	recall=88% spec=84.05%
Iva, M. B. (2014) [34]	KNN ve Karar ağaçları yöntemi	PND0	KNN- recall=91.4% AUC=94%
Han, F. (2014) [35]	SVM, Haralick, Gabor ve LBP	LIDC	AUC=92.70
Kuruvilla, J. (2014) [31]	ANN	LIDC	accu=93.3% recall=91.4% spec=100%
Önerilen sistem	SVM, çoklu öznitelikler	LIDC	accu=94.91 recall=95.69 prec=96.97

Yukarıdaki Çizelge 6.3'te doğruluk oranı (accu-accuracy), anma oranı (recall), kesinlik oranı (prec-precision), seçicilik oranı (spec-specificity) ve (AUC) ROC eğrinin altındaki oranını göstermektedir.

7. SONUÇ

Bu tezde, geçmiş akciğer tomografi görüntülerinden oluşan büyük miktarda görüntü veri setlerinde akciğer nodüllerinin sınıflandırılması ve teşhisi amacıyla kullanılmıştır. Birkaç uzman radyolog yardımıyla stajyer doktorlara teşhis desteği sağlamaya çalışılmıştır. Farklı algoritmalarla doğru ve daha kısa sürede iyi huylu veya kötü huylu akciğer nodülleri sınıflandırılmaya çalışılmıştır. Bu amaçla, bu görüntü işlemenin Hadoop MapReduce çerçevesinde paralel olarak kullanılması ve HDFS’de depolanması önerilmiştir.

Deneyleri gerçekleştirmek için öncelikle önerilen sistemi hesaplama süresi bakımından iki farklı yön oluşturulmuştur. Birincisi tomografi görüntülerinin HDFS’ye yüklenmesi, ikincisi ise HDFS’de saklanan veri setinden benzer nodüllerin alınmasıdır. İyi huylu nodüllerin kötü huylu nodüllerden ayırt edilebilmesi için doku, şekil ve RQI öznitelikleri çıkarılmıştır. Elde edilen veri setleri ile oluşturulan öznitelikler denenmiş ve başarılı sonuçlar elde edilmiştir.

Ele alınan herhangi bir akciğer nodül görüntüsü için hem ona benzer nodül görüntüleri elde edilebildiğinin hem de iyi huylu veya kötü huylu akciğer nodüllerinin sınıflandırılabilirdiğinin doğruluğunu kontrol etmek için bir takım deneyler yapılmıştır.

Deneyler esnasında bir çok dilim içeren tomografi görüntüleri 1, 10, 20, 30, 50, 100, 200, 500 ve 847’ye bölünmüştür. Deneyler MATLAB yüklü bulunan bir master ve iki slave bilgisayardan oluşan Lokal MapReduce ve Hadoop kümesi ile gerçekleştirilmiştir.

Oluşturulan veri seti içerisindeki 4.179 civarındaki nodülden, 2555 tanesi iyi huylu, 1624 tanesi ise kötü huylu nodül olup, sınıflandırılmak üzere incelenmiştir. Veri setimizdeki akciğer nodülü görüntülerini “kötü huylu” veya “iyi huylu” olarak sınıflandırmak için SVM ve KNN’i kullandık. Sınıflandırma, doğruluk oranını test etmek için k=10 alınarak, k-katlamalı çapraz doğrulama yöntemi kullanılmıştır.

Çalışmamızda denenen modelde, SVM yöntemiyle elde edilen sonuçlar ile KNN yöntemiyle elde edilen sonuçlar karşılaştırılmıştır. Sonuçlara göre SVM yöntemi, başarı oranını yaklaşık %10 doğruluk oranında daha artırmaktadır.

Hem literatürlerin sonuçları ile hem de tez kapsamında yapılmış olan nodül sınıflandırması deneylerinin sonuçları ile karşılaştırıldığında nodül sınıflandırmasının başarısının yaklaşık olarak aynı düzeyde olduğu ortaya konulmuştur.

Tıbbi Görüntülerin, farklı hastanelerde depolanmasını ve paylaşılmasını sağlamak için bir sistem geliştirilmiştir. Ayrıca, ağa Hadoop Küme'deki bir düğüm eklemek basittir. Dahası, MATLAB'taki etkileşim kolaylığı için arayüzü sağlanmıştır.

Gelecek çalışmalarda tez kapsamına Hadoop kümesindeki performansı artmak için düğüm sayısı artması ve HDFS'de güvenli bir şekilde görüntü depolanma ve erişimi için farklı şifreleme teknikleri kullanması planlanmaktadır. Görüntülerin içerdikleri nodülleri otomatik olarak tespit etme işlevi deney başarısını arttırmak için farklı oluşturulacak öznitelikler ve farklı makine öğrenmesi yöntemleri kullanılarak geliştirilmesi planlanmaktadır.

KAYNAKLAR

1. Branstetter, B. F. (2010). *Practical imaging informatics: foundations and applications for PACS professionals*. New York: Springer, 49–66.
2. Belle, A., Thiagarajan, R., Soroushmehr, S. M., Navidi, F., Beard, D. A., Najarian, K. (2015). Big data analytics in healthcare. *BioMed research international*, 2015, 1-16.
3. Suganya, R., Rajaram, S., Abdullah, A. S. (2018). *Big Data in Medical Image Processing*. ABD: CRC Press, 1-45.
4. Widmer, A., Schaer, R., Markonis, D., Müller, H. (2014). *Gesture Interaction for Content-based Medical Image Retrieval*. International Conference on Multimedia Retrieval (ICMR), 503.
5. Frost, S. (2012). Drowning in Big Data? Reducing Information Technology Complexities and Costs For Healthcare Organizations. *Frost and Sullivan*, 20(3), 1-14.
6. Dimitrov, D. V. (2016). Medical internet of things and big data in healthcare. *Healthcare informatics research*, 22(3), 156–163.
7. Ritter, F., Boskamp, T., Homeyer, A., Laue, H., Schwier, M., Link, F., Peitgen, H. O. (2011). Medical image analysis. *IEEE pulse*, 2(6), 60-70.
8. Rattan, S., Kaur, S., Kansal, N. and Kaur, J. (2017). *An optimized lung cancer classification system for computed tomography images*. International Conference on Image Information Processing (ICIIP), 1-6.
9. Wutsqa, D. U. and Mandadara, H. L. R. (2017). *Lung cancer classification using radial basis function neural network model with point operation*. International Congress on Image and Signal Processing, BioMedical Engineering and Informatics (CISP-BMEI), 1-6.
10. Lakshmanaprabu, S. K., Mohanty, S. N., Shankar, K., Arunkumar, N. and Ramirez, G. (2019). Optimal deep learning model for classification of lung cancer on CT images. *Future Generation Computer Systems*, 92, 374-382.
11. El-Sherbiny, B., Nabil, N., El-Naby, S. H., Emad, Y., Ayman, N., Mohiy, T. and AbdelRaouf, A. (2018). *Blb (brain/lung cancer detection and segmentation and breast dense calculation)*. First International Workshop on Deep and Representation Learning (IWDRL), 41-47.
12. Shankar, K., Lakshmanaprabu, S. K., Gupta, D., Maseleno, A. and Albuquerque, V. H. C. (2018). Optimal feature-based multi-kernel SVM approach for thyroid disease classification. *The Journal of Supercomputing*, 74, 1-16.
13. Sarker, P., Shuvo, M. M. H., Hossain, Z. and Hasan, S. (2017). *Segmentation and classification of lung tumor from 3D CT image using K-means clustering algorithm*. International Conference on Advances in Electrical Engineering (ICAEE), 731-736.

14. Vahid, A. H. (2012). *Fusion and combination methods for multimodal content based medical image retrieval*, Doktora Tezi, Dokuz Eylül Üniversitesi Fen Bilimleri Enstitüsü, İzmir.
15. Willingham, D. (2015). *Big Data Analysis, Analytics With MATLAB*. 15th International Conference on Accelerator and Large Experimental Control Systems, Melbourne, 9–12.
16. Zinovev, D., Raicu, D., Furst, J., Armato III, S. (2009). Predicting radiological panel opinions using a panel of machine learning classifiers. *Algorithms*, 2(4), 1473-1502.
17. Nagorny, K., Lima-Monteiro, P., Barata, J. and Colombo, A. W. (2017). Big data analysis in smart manufacturing: a review. *International Journal of Communications, Network and System Sciences*, 10(3), 31-58.
18. McAfee, A., Brynjolfsson, E., Davenport, T. H., Patil, D. J., Barton, D. (2012). Big data: The management revolution. *Harvard Business Review*, 1(90), 61–67.
19. Shvachko, K., Kuang, H., Radia, S., Chansler, R. (2010). *The Hadoop distributed file system*. 2010 IEEE 26th Symposium on Mass Storage Systems and Technologies (MSST), United States, 1–10.
20. Willingham, D. (2015). *Big Data Analysis and Analytics with MATLAB*. 15th International Conference on Accelerator and Large Experimental Physics Control Systems (ICALEPCS'15), Melbourne, 656–659.
21. İnternet: Getting Started with MapReduce. (2019). URL: www.mathworks.com/help/MATLAB/import_export/getting-started-with-MapReduce.html, Son Erişim Tarihi: 30.06.2019.
22. Rao, A., Srihari, R. K., Zhang, Z. (1999). *Spatial color histograms for content-based image retrieval*. 11th International Conference on Tools with Artificial Intelligence, Chicago, 183–186.
23. Nixon, M., Aguado, A. S. (2012). *Feature extraction and image processing for computer vision* (Third Edition). United Kingdom: Academic Press, 137-212.
24. Zhou, W., Montoya, J., Gutjahr, R., Ferrero, A., Halawish, A., Kappler, S. and Leng, S. (2017). Lung nodule volume quantification and shape differentiation with an ultra-high resolution technique on a photon-counting detector computed tomography system. *Journal of Medical Imaging*, 4(4), 043502.
25. Kumar, R. S., Senthilmurugan, M. (2013). Content-based image retrieval system in medical applications. *International Journal of Engineering Research and Technology*, 2(3), 1-5.
26. Doi, K. (2007). Computer-aided diagnosis in medical imaging: historical review, current status and future potential. *Computerized Medical Imaging and Graphics*, 31, 198–211.

27. Li, Q., Nishikawa, R. M. (2015). *Computer-aided detection and diagnosis in medical imaging* (First Edition). ABD: CRC Press, 1-12.
28. Chen, C.-M., Chou, Y.-H., Tagawa, N., Do, Y. (2013). Computer-aided detection and diagnosis in medical imaging. *Computational and Mathematical Methods in Medicine*, 2013, 1-2.
29. Beutel, J., Fitzpatrick, J. M., Horii, S. C., Kundel, H. L., Metter, R. L. Van. (2000). Computer-aided diagnosis in mammography. *Handbook of medical imaging*, 2, 1-70.
30. Sahu, S. P., Londhe, N. D., Verma, S. (2017). An Automated System for the Detection of Lung Cancer in CT data at Early Stages : Review. *International Journal of Control Theory and Application*, 18(10), 133-144.
31. Nasrullah, N., Sang, J., Alam, M. S. and Xiang, H. (2019). Automated detection and classification for early stage lung cancer on CT images using deep learning. In Pattern Recognition and Tracking. *International Society for Optics and Photonics*, 10995, 109-117.
32. Ardila, D., Kiraly, A. P., Bharadwaj, S., Choi, B., Reicher, J. J., Peng, L. and Naidich, D. P. (2019). End-to-end lung cancer screening with three-dimensional deep learning on low-dose chest computed tomography. *Nature medicine*, 25(6), 954.
33. Shen, S., Han, S. X., Aberle, D. R., Bui, A. A. and Hsu, W. (2019). An interpretable deep hierarchical semantic convolutional neural network for lung nodule malignancy classification. *Expert Systems with Applications*, 128, 84-95.
34. Kasinathan, G., Jayakumar, S., Gandomi, A. H., Ramachandran, M., Fong, S. J. and Patan, R. (2019). Automated 3-D lung tumor detection and classification by an active contour model and CNN classifier. *Expert Systems with Applications*, 134, 112-119.
35. Nasser, I. M. and Abu-Naser, S. S. (2019). Lung Cancer Detection Using Artificial Neural Network. *International Journal of Engineering and Information Systems (IJEAIS)*, 3(3), 17-23.
36. Prabukumar, M., Agilandeewari, L. and Ganesan, K. (2019). An intelligent lung cancer diagnosis system using cuckoo search optimization and support vector machine classifier. *Journal of Ambient Intelligence and Humanized Computing*, 10(1), 267-293.
37. Günay, M., Orman, Z., Ensari, T., Oukid, S. and Benblidia, N. (2019). *Diagnosis of Lung Cancer Using Artificial Immune System*. Scientific Meeting on Electrical-Electronics and Biomedical Engineering and Computer Science (EBBT), 1-4.
38. Coudray, N., Ocampo, P. S., Sakellaropoulos, T., Narula, N., Snuderl, M., Fenyö, D. and Tsirigos, A. (2018). Classification and mutation prediction from non-small cell lung cancer histopathology images using deep learning. *Nature medicine*, 24(10), 1559.

39. Nishio, M., Sugiyama, O., Yakami, M., Ueno, S., Kubo, T., Kuroda, T. and Togashi, K. (2018). Computer-aided diagnosis of lung nodule classification between benign nodule, primary lung cancer and metastatic lung cancer at different image size using deep convolutional neural network with transfer learning. *PloS one*, 13(7), 1-12.
40. Zhu, X., Dong, D., Chen, Z., Fang, M., Zhang, L., Song, J. and Tian, J. (2018). Radiomic signature as a diagnostic factor for histologic subtype classification of non-small cell lung cancer. *European radiology*, 28(7), 2772-2778.
41. Gong, J., Liu, J. Y., Sun, X. W., Zheng, B. and Nie, S. D. (2018). Computer-aided diagnosis of lung cancer: the effect of training data sets on classification accuracy of lung nodules. *Physics in Medicine and Biology*, 63(3), 1-11.
42. Makaju, S., Prasad, P. W. C., Alsadoon, A., Singh, A. K. and Elchouemi, A. (2018). Lung cancer detection using CT scan images. *Procedia Computer Science*, 125, 107-114.
43. Wan, C. C., Hoogendoorn, M., Haasdijk, E. (2017). *Lung nodules detection in CT scans*, M. Sc. Thesis, Vrije University, Amsterdam.
44. Nithila, E. E., Kumar, S. S. (2017). Automatic detection of solitary pulmonary nodules using swarm intelligence optimized neural networks on CT images. *Engineering Science and Technology, an International Journal (JESTECH)*, 20, 1192–1202.
45. Akram, S., Javed, M. Y., Qamar, U., Khanum, A., Hassan, A. (2015). Artificial neural network based classification of lungs nodule using hybrid features from computerized tomographic images. *Applied Mathematics and Information Sciences (AMIS)*, 9, 183–195.
46. Silva, G. L. F. da, Silva, A. C., Paiva, A. C. De, Gattass, M. (2016). Classification of malignancy of lung nodules in CT images using Convolutional Neural Network. *OncoTargets and therapy*, 3(9), 2481–2489.
47. Kumar, D., Wong, A., Clausi, D. A. (2015). *Lung Nodule Classification Using Deep Features in CT Images*. 12th Conference on Computer and Robot Vision, Nova Scotia, 133–138.
48. Kuruvilla, J., Gunavathi, K. (2014). Lung cancer classification using neural networks for CT images. *Computer Methods and Programs in Biomedicine*, 113, 202–209.
49. Chaudhary, A., Singh, S. S. (2012). *Lung cancer detection on CT images by using image processing*. International Conference on Computational Science (ICCS), 142–146.
50. Krishnamurthy, S., Narasimhan, G., Rengasamy, U. (2016). Three-dimensional lung nodule segmentation and shape variance analysis to detect lung cancer with reduced false positives. *Proceedings of the Institution of Mechanical Engineers, Part H: Journal of Engineering in Medicine*, 230, 58–70.

51. Iva, M. B. L. (2014). *Reduction of False Positives in Lung Nodule Detection Algorithm*, M. Sc. Thesis, Czech Teknik University, Prag.
52. Han, F. (2014). Texture Feature Analysis for Computer-Aided Diagnosis on Pulmonary Nodules. *Journal of Digital Imaging*, 28, 99–115.
53. Kusuma, B., Megha, P. A. (2015). Survey on Content Based Image Retrieval Using Mapreduce Over Hadoop. *International Journal of Advanced Research in Computer Science and Electronics Engineering (IJARCSEE)*, 2(10), 5–7.
54. Khan, S. M. H., Hussain, A., Alshaikhli, I. F. T. (2012). *Comparative study on content-based image retrieval (CBIR)*. International Conference on Advanced Computer Science Applications and Technologies (ACSAT), Kuala Lumpur, 61–66.
55. Kumar, R. S., Senthilmurugan, M. (2013). Content-based image retrieval system in medical applications. *International Journal of Engineering Research and Technology*, 2(3), 1-5.
56. Lohar, N., Chavan, D., Arade, S., Jadhav, A., Chikmurge, D. (2016). Content Based Image Retrieval System over Hadoop Using MapReduce. *International Journal of Scientific Research in Science, Engineering and Technology*, 1(2), 1-3
57. Markonis, D., Schaer, R., Eggel, I., Müller, H., Depeursinge, A. (2012). *Using MapReduce for large-scale medical image analysis*. 2nd International on Healthcare Informatics, Imaging and Systems Biology (HISB), Washington, 1-10.
58. Mahakulkar, C., Jagtap, M., Harale, N., Babbur, S., Avinash, P. (2016). Content Based Image Retrieval System for Medical Diagnosis using Hadoop. *International Journal for Scientific Research and Development*, 8(4), 83–85.
59. Trivedi, U. (2018). Local Tetra Pattern for Image Retrieval System Using Hadoop. *International Journal of Advance Engineering and Research Development*, 3, 2–5.
60. Jai-Andaloussi, S., Elabdouli, A., Chaffai, A., Madrane, N., Sekkaki, A. (2013). *Medical content based image retrieval by using the Hadoop framework*. T20th International Conference on Telecommunications (ICT), Casablanca, 1–5.
61. Venkatraman, S., Kulkarni, S. (2012). *MapReduce neural network framework for efficient content based image retrieval from large datasets in the cloud*. 12th International Conference on Hybrid Intelligent Systems (HIS), Pune, 63–68.
62. Patil, J. S., Pradeepani, G. (2016). Two dimensional medical images diagnosis using MapReduce. *Indian Journal of Science and Technology*, 9, 2-5.
63. Dravyakar, S. P. (2014). *Private content based image retrieval using Hadoop*, M. Sc. Thesis, Pune University Engineering Faculty, Pune.
64. Rajakumar, K., Sudheer, D. (2016). A review of visual information retrieval on massive image data using Hadoop. *International Journal of Control Theory and Applications*, 9, 425–430.

65. Kalpathy-Cramer, J., Freymann, J. B., Kirby, J. S., Kinahan, P. E. and Prior, F. W. (2014). Quantitative imaging network: data sharing and competitive AlgorithmValidation leveraging the cancer imaging archive. *Translational oncology*, 7(1), 147-152.
66. Armato, S. G. (2011). 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, 915–931.
67. Armato, S. G. (2015). Special Section Guest Editorial: LUNGx Challenge for computerized lung nodule classification: reflections and lessons learned. *Journal of Medical Imaging*, 2(2), 1-5.
68. Jacobs, C.(2016). Computer-aided detection of pulmonary nodules: a comparative study using the public LIDC/IDRI database. *European radiology*, 26(7), 2139–2147.
69. Macmahon, H. (2008). Different Size Metrics for Pulmonary Nodule Measurements. *World Health*, 14, 1475–1485.
70. Hancock, M. C., Magnan, J. F. (2016). Lung nodule malignancy classification using only radiologist-quantified image features as inputs to statistical learning algorithms: probing the Lung Image Database Consortium dataset with two statistical learning methods. *Journal of Medical Imaging*, 3(4), 1-15.
71. Dasovich, G., Kim, R., Raicu, D. S., Furst, J. D. (2010). A Model for the Relationship between Semantic and Content Based Similarity using LIDC. *International Society for Optics and Photonics*, 7624,1–10.
72. Shen, W., Zhou, M., Yang, F., Yang, C., Tian, J. (2015). *Multi-scale convolutional neural networks for lung nodule classification*. 24th International Conference on Information Processing in Medical Imaging, Isle of Skye, 588–599.
73. Hossen Bhuiyan, M., Azad, I., Uddin, M. (2013). Image Processing for Skin Cancer Features Extraction. *International Journal of Scientific and Engineering Research*, 4(2), 1–6.
74. Lehmann, G. and Legland, D. (2012). Efficient N-dimensional surface estimation using Crofton formula and run-length encoding. *Kitware INC*, 1-11.
75. İnternt: R2017b, I. (2017). Measure properties of 3-D volumetric image regions - MATLAB regionprops3 - MathWorks Deutschland. URL: https://de.mathworks.com/help/images/ref/regionprops3.html#mw_9b2ab6ef-9787-48aa-8c87-4cff7a71b23f, Son Erişim Tarihi: 30.06.2019.
76. Therasse, P., Arbuck, S. G., Eisenhauer, E. A., Wanders, J., Kaplan, R. S., Rubinstein, L., Gwyther, S. G. (2005). New guidelines to evaluate the response to treatment in solid tumors. *Breast Cancer*, 12, 16-27.

77. Manivannan, K., Aggarwal, P., Devabhaktuni, V., Kumar, A., Nims, D., Bhattacharya, P. (2012). Particulate matter characterization by gray level co-occurrence matrix based support vector machines. *Journal of hazardous materials*, 223, 94-103.
78. Ramamurthy, B. and Chandran, K. R. (2015). CBMIR: Content Based Medical Image Retrieval Using Multilevel Hybrid Approach. *International Journal of Computers Communications and Control*, 10(3), 382-389.
79. Duda, R. O., Hart, P. E. and Stork, D. G. (2001). Pattern classification. *International Journal of Computational Intelligence and Applications*, 1, 335-339.
80. Vapnik, V. (2013). *The nature of statistical learning theory*. Springer science and business media (Second Edition). New York: Springer Verlag, 35-97.
81. Bahadure, N. B., Ray, A. K. and Thethi, H. P. (2017). Image analysis for MRI based brain tumor detection and feature extraction using biologically inspired BWT and SVM. *International journal of biomedical imaging*, 2017, 1-13.
82. Duda, R. O., Hart, P. E. and Stork, D. G. (2000). *Pattern classification* (Second Edition). New York: John Wiley and Sons, 75-90.
83. Harrison Jr, J. H. (2008). Introduction to the mining of clinical data. *Clinics in laboratory medicine*, 28(1), 1-7.
84. Kaya, A., Keçeli, A. S. and Can, A. B. (2018). Akciğer nodül özelliklerinin tahmininde çeşitli sınıflama stratejilerinin incelenmesi. *Gazi Üniversitesi Mühendislik-Mimarlık Fakültesi Dergisi*, 2(34), 709-726.

ÖZGEÇMİŞ

Kişisel Bilgiler

Soyadı, adı : AL-ZABIDI, Eissa Faisal Ali
 Uyuğu : YEMEN
 Doğum tarihi ve yeri : 01.01.1985, Yemen
 Medeni hali : Evli
 Telefon : 0 (507) 028 08 55
 e-mail : eissa.al.zabidi@gazi.edu.tr



Eğitim

Derece	Eğitim Birimi	Mezuniyet Tarihi
Yüksek lisans	Gazi Üniversitesi/Bilgisayar Mühendisliği	2019
Lisans	Hodidah Üniversitesi/Bilgisayar Mühendisliği	2008
Lise	Althawrah Koleji	2003

İş Deneyimi

Yıl	Yer	Görev
2008 - 2013	Hodidah Üniversitesi	Bilgisayar Öğretmenliği
2009 – 2012	Ulusal Kanser Kontrol Vakfı (NCCF) - Hodeidah	Medya Görevlisi

Yabancı Dil

İngilizce, Türkçe, Arapça

Yayınlar

1. Alzabidi, E. ve Yavanoğlu, U. (2018). *Application of Big Data Analysis for Medical Image Datasets with MATLAB*. 7th International Conference on Advanced Technologies ICAT'18, Antalya, 669-675.

Hobiler

Yüzme, Futbol, Kayak.



GAZİ GELECEKTİR..