

T.C.  
SAKARYA ÜNİVERSİTESİ  
FEN BİLİMLERİ ENSTİTÜSÜ

**GÖĞÜS HASTALIKLARININ TEŞHİS EDİLMESİNDE  
MAKİNE ÖĞRENME Sİ ALGORİTMALARININ  
KULLANILMASI**

**DOKTORA TEZİ**

**Amani YAHYAOUİ**

**Enstitü Anabilim Dalı : BİLGİSAYAR VE BİLİŞİM  
MÜHENDİSLİĞİ**  
**Tez Danışmanı : Prof. Dr. Nejat YUMUŞAK**

**Temmuz 2017**

T.C.  
SAKARYA ÜNİVERSİTESİ  
FEN BİLİMLERİ ENSTİTÜSÜ

**GÖĞÜS HASTALIKLARININ TEŞHİS EDİLMESİNDE  
MAKİNE ÖĞRENMESİ ALGORİTMALARININ  
KULLANILMASI**

**DOKTORA TEZİ**

**Amani YAHYAOUİ**

**Enstitü Anabilim Dalı : BİLGİSAYAR VE BİLİŞİM  
MÜHENDİSLİĞİ**

**Bu tez 05 / 07 /2017 tarihinde aşağıdaki jüri tarafından oybirliği/oyçokluğu ile kabul edilmiştir.**



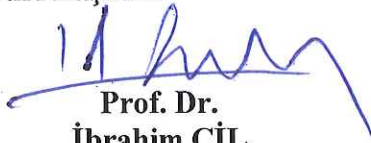
**Doç. Dr.  
Uğur GÜVENÇ  
Jüri Başkanı**



**Prof. Dr.  
Nejat YUMUŞAK  
Üye**



**Prof. Dr.  
Celal ÇEKEN  
Üye**



**Prof. Dr.  
İbrahim ÇİL  
Üye**



**Doç. Dr.  
Resul KARA  
Üye**

## **BEYAN**

Tez içindeki tüm verilerin akademik kurallar çerçevesinde tarafımdan elde edildiğini, görsel ve yazılı tüm bilgi ve sonuçların akademik ve etik kurallara uygun şekilde sunulduğunu, kullanılan verilerde herhangi bir tahrifat yapılmadığını, başkalarının eserlerinden yararlanılması durumunda bilimsel normlara uygun olarak atıfta bulunulduğunu, tezde yer alan verilerin bu üniversite veya başka bir üniversitede herhangi bir tez çalışmasında kullanılmadığını beyan ederim.

Amani YAHYAOUİ

24.05.2017



## TEŐEKKÜR

Bu alıőmanın tamamlanmasında katkılarında dolayđ aőađıdaki kiői ve kuruluőlara teőekkürü bir bor bilirim.

Türkiye gibi eđitim düzeyinin yüksek olduđu güzel bir ülkede eđitim fırsatı ve desteđi sundukları için YTB'ye (Yurtdıőı Türkler ve Akraba Topluluklar Baőkanlıđı) teőekkür ve őükranlarımı sunarım.

Doktora eđitimimin baőlangıcından beri benimle alıőmayı kabul ettiđi ve bu doktora tezi süresince yoğun iő temposuna rađmen bana ayırdıđı zaman ve yardımlarından ötürü Sayın Prof. Dr. Nejat Yumuőak hocama samimi teőekkürlerimi sunarım. Bu doktora alıőmasının baőarıyla tamamlanmasında kendisinin rehberliđi ve bana sađladıđı vizyon ile büyük katkıları olmuőtur. Sakarya Üniversitesi'ne, Bilgisayar ve Biliőim Bilimleri Fakültesi'ne ve Bilgisayar Mühendisliđi Bölümü'ne teőekkürlerimi sunarım.

Eđitim hayatım süresince verdikleri destekten ve fedakârlıktan ötürü sevgili aileme teőekkürlerimi sunarım.

## TABLE OF CONTENTS

TEŞEKKÜR .....	i
İÇİNDEKİLER .....	ii
SİMGELER VE KISALTMALAR LİSTESİ .....	v
ŞEKİLLER LİSTESİ .....	vi
TABLOLAR LİSTESİ .....	vii
ÖZET .....	viii
SUMMARY .....	ix
BÖLÜM 1.	
GİRİŞ .....	1
BÖLÜM 2.	
GÖĞÜS HASTALIKLARI .....	4
2.1. Giriş .....	4
2.2. Tüberküloz .....	5
2.2.1. Tüberküloz türleri .....	6
2.2.2. Tüberküloz belirtileri .....	6
2.2.3. Tüberküloz teşhisi .....	7
2.2.4. Tüberküloz tedavisi .....	7
2.3. COPD .....	8
2.3.1. COPD nedenleri .....	8
2.3.2. COPD belirtileri .....	9
2.3.3. COPD teşhisi .....	9
2.3.4. COPD tedavisi .....	9
2. 4. Zatürre .....	10
2.4.1. Zatürre nedenleri .....	10

2.4.1. Zatürre belirtileri.....	11
2.4.1. Zatürre teşhisi .....	12
2.4.1. Zatürre tedavisi .....	12
2.5. Akciğer Kanseri .....	12
2.5.1. Akciğer kanseri türleri .....	13
2.5.2. Akciğer kanseri nedenleri .....	14
2.5.3. Akciğer kanseri belirtileri .....	14
2.5.3. Akciğer kanseri teşhisi.....	15
2.5.3. Akciğer kanseri tedavisi .....	15
2.6. Astım .....	16
2.6.1. Astım nedenleri .....	16
2.6.2. Astım belirtileri .....	17
2.6.3. Astım teşhisi .....	17
2.6.4. Astım tedavisi .....	17
2.6. Sonuç .....	18

### BÖLÜM 3.

KULLANILAN MAKİNA ÖĞRENMESİ TEKNİKLERİ.....	19
3.1. Giriş .....	19
3.2. Destek Vektör Makinaları: SVM .....	19
3.2.1. SVM temelleri .....	19
3.2.1. SVM kullanarak göğüs hastalıklarının teşhisi .....	24
3.2.1. ASVM kullanarak göğüs hastalıklarının teşhisi .....	19
3.3. K En Yakın Komşular (KNN) .....	27
3.3.1. KNN temelleri .....	27
3.3.2. KNN kullanarak göğüs hastalıklarının teşhisi .....	28
3.4. Doğal Bayes .....	29
3.4.1. Doğal Bayes temelleri .....	29
3.4.2. NB kullanarak göğüs hastalıklarının teşhisi.....	31
3.5. Sonuç.....	32

## BÖLÜM 4.

UYGULAMA ÇALIŞMASI VE SONUÇLARI.....	34
4.1. Giriş.....	34
4.2. Çalışmanın İçeriği .....	34
4.2.1. Geliştirme ortamı .....	34
4.2.2. Kullanılan veritabanı.....	34
4.2.3. Performans ölçümü .....	37
4.3. Kullanılan Farklı Modeller.....	37
4.3.1. SVM Modeli.....	37
4.3.1.1 SVM sistem mimarisi .....	37
4.3.1.2 SVM sistem sonuçları.....	38
4.3.2. ASVM-Modeli.....	41
4.3.2.1 ASVM sistem mimarisi .....	41
4.3.2.2 ASVM sistem sonuçları.....	43
4.3.3. KNN-Modeli.....	43
4.3.3.1 KNN sistem mimarisi .....	44
4.3.3.2 KNN sistem sonuçları.....	45
4.3.4. NB-Modeli.....	46
4.3.4.1 NB sistem mimarisi .....	47
4.3.4.2 NB sistem sonuçları.....	48
4.4. Farklı sınıflandırma sistemlerinin karşılaştırılması.....	48

## BÖLÜM 5.

TARTIŞMA VE ÖNERİLER .....	54
KAYNAKLAR.....	56
ÖZGEÇMİŞ .....	62



## SİMGELER VE KISALTMALAR LİSTESİ

SVM	: Destek Vektör Makinaları
KNN	: K-En Yakın Komşu
NB	: Basit Bayes
ASVM	: Uyarlamalı Karar Destek Makinaları
QP	: Optimizasyon problemi
DT	: Karar Ağaçları
MLNN	: Çok katmanlı sinirsel ağlar
LASSO	: LASSO regresyonları
PCA	: Temel bileşen analizi

## ŞEKİLLER LİSTESİ

Şekil 2.1.	İnsan vücudundaki göğüs organları [12] .....	4
Şekil 2.2.	İnsan göğsündeki akciğer tüberkülozu [16].....	5
Şekil 2.3.	İnsan göğsündeki COPD [25].....	8
Şekil 2.4.	İnsan göğsündeki zatürre [29].....	10
Şekil 2.5.	İnsanda göğüs kanseri [35] .....	13
Şekil 2.6.	İnsan göğsündeki astım [43] .....	16
Şekil 3.1.	Doğrusal durum için temel SVM [53] .....	20
Şekil 3.2.	Doğrusal olmayan durumda SVM [54] .....	23
Şekil 3.3.	Eşleşme uzayı [54].....	23
Şekil 3.4.	ASVM algoritması .....	27
Şekil 3.5.	KNN temelleri [59] .....	29
Şekil 3.6.	NB temelleri.....	32
Şekil 4.1.	SVM kullanılarak göğüs hastalıkları teşhis sistemi.....	39
Şekil 4.2.	Her hastalık için sınıflandırma doğrulukları .....	42
Şekil 4.3.	ASVM kullanılarak göğüs hastalıkları teşhis sistemi .....	43
Şekil 4.4.	KNN kullanarak göğüs hastalıkları teşhis sistemi .....	45
Şekil 4.5.	NB kullanılarak göğüs hastalıkları teşhis sistemi .....	48
Şekil 4.6.	Tüberküloz için farklı sınıflandırma sistemlerinin karşılaştırılması .....	49
Şekil 4.7.	COPD için farklı sınıflandırma sistemlerinin karşılaştırılması.....	50
Şekil 4.8.	Zatürre için farklı sınıflandırma sistemlerinin karşılaştırılması .....	51
Şekil 4.9.	Astım için farklı sınıflandırma sistemlerinin karşılaştırılması .....	52
Şekil 4.10.	Akciğer kanseri için farklı sınıflandırma sistemlerinin karşılaştırılması .....	53

## TABLULAR LİSTESİ

Tablo 4.1.	Veri özellikleri özeti .....	36
Tablo 4.2.	SVM üzerindeki çekirdek etkisi .....	40
Tablo 4.3.	Polinom çekirdek derecesinin etkisi .....	40
Tablo 4.4.	SVM kullanılarak her hastalık için sınıflandırma doğrulukları .....	40
Tablo 4.5.	ASVM kullanarak her hastalık için sınıflandırma doğrulukları .....	41
Tablo 4.6.	KNN kullanarak her hastalık için sınıflandırma doğrulukları .....	44
Tablo 4.7.	En yakın komşu sayısının etkisi.....	46
Tablo 4.8.	Her hastalık için sınıflandırma doğrulukları .....	48

## ÖZET

Anahtar kelimeler: Yapay zekâ teknikleri, göğüs hastalıkları, destek vektör makinası, uyarlanabilir destek vektör makinası, K-en yakın komşular (K-NN), Basit Bayes sınıflandırması (NB).

Birçok hastalık insan yaşamını, sağlığını ve yaşam kalitesini birçok yönden etkileyerek tehdit etmektedir. Bunlar arasında tüberküloz (TB), kronik obstrüktif akciğer hastalığı (COPD), zatürre, astım ve akciğer kanseri gibi göğüs hastalıkları hem gelişmekte olan ülkelerde hem de gelişmiş ülkelerde önemli sağlık sorunları ve ölüm nedenlerinden başlıcaları olarak kabul edilmektedir. Bir hastalığın teşhisi doktorlar tarafından ne kadar erken konulursa, hastanın iyileşme olasılığı o kadar yükselmektedir. Bu anlamda, günümüzde uzman sistemler ve farklı yapay zeka teknikleri, tıbbi teşhis dahil olmak üzere çeşitli alanlardaki farklı problemleri çözmek için başarıyla kullanılmaktadır.

Bu doktora tezi kapsamında, göğüs hastalıklarını teşhis etmek için destek vektör makinaları (SVM), K-en yakın komşular (K-NN) ve Basit Bayes sınıflandırma (NB) metotları kullanılmıştır. Ayrıca, ilk kez uyarlamalı destek vektör makina metodunun (ASVM) göğüs hastalıkları teşhisindeki performansı da değerlendirilmiştir. Bu metot, SVM tekniğinin en uygun eşik değerini bulacak şekilde geliştirilmesine dayanmaktadır. Kullanılan bu yaklaşımlar Diyarbakır Göğüs Hastalıkları Hastanesi'nden alınan deneysel veriseti kullanılarak ve daha önceki çalışmalarda kullanılan sinir ağları yöntemleri ile karşılaştırılarak değerlendirilmiştir. Yapılan çalışmada elde edilen sonuçlar, kullanılan metotların, özellikle de ASVM metodunun başarımlarının yüksek olduğunu göstermiştir. ASVM metodu, öngörülebilir ve doğruluk oranı en yüksek sonuçları vererek göğüs hastalıklarında etkili bir şekilde kullanılabilceğini kanıtlamıştır.

# **CHEST DISEASES DIAGNOSIS BASED ON MACHINE LEARNING ALGORITHMS**

## **SUMMARY**

Keywords: Artificial intelligence techniques, chest diseases, support vector machine, adaptive support vector machine, K-Nearest Neighbors (K-NN), Naïve Bayes classification (NB).

Several diseases threaten human health by affecting longevity and its well-being in many ways. Among them, chest diseases as Tuberculosis (TB), Chronic Obstructive Pulmonary Disease (COPD), Pneumonia, Asthma, and Lung cancer are considered as serious health complications and one major cause of death in both developing and developed countries. Doctors confirm that the earlier a disease is diagnosed, the higher is the patient cure probability. In this context, expert systems and different artificial intelligence techniques have been successfully used to solve different problems in various domains including medical diagnosis.

In this thesis, we use the Support Vector Machines (SVM), the K-Nearest Neighbors (K-NN) and the Naive Bayes classification (NB) methods to diagnose chest diseases and for the first time, we examine the performance of the Adaptive Support Vector Machine (ASVM) method for chest disease diagnosis. This involves improving the SVM by finding its most appropriate Bias term value. These approaches are evaluated using an experimental dataset from Diyarbakir chest diseases hospital and comparing them with the Neural Network method used in previous studies. The experimental results showed the efficiency of these methods, especially ASVM, which could achieve promising results and confirmed that it can be efficiently used in chest diseases.

## **BÖLÜM 1. GİRİŞ**

Tıbbi teşhis (tanı), hastada bulunan rahatsızlığın doktorlar tarafından belirlenmesi ve ona uygun tedavinin verilmesi amacıyla gözlemlenen belirtilerin yorumlanmasına dayanan bir eylemdir. Tıbbi tanı, öncelikle hastanın tıbbi geçmişine ve doktor tarafından gerçekleştirilen fiziksel muayeneye dayanmaktadır [1].

Tıbbi tanının başarı derecesi, birden çok faktöre bağlı olup bunlardan en önemlileri doktorun bilimsel yeterliliği, deneyimi ve kullanılan teknik malzemelerdir.

Tıp alanında sürekli ve önemli gelişmeler kaydedilmesine rağmen, araştırmacılar ilaçla tedavinin kesin bir bilimsel sonuç olarak kabul edilmemesi görüşünde hem fikirdirler [2]. Doktorların yeterliliği veya kullanılan yüksek teknik ekipman kalitesi ne olursa olsun, tıbbi ve özellikle hastalığın teşhisindeki hatalar kaçınılmazdır [3].

Araştırmacılar, yanlış veya geç teşhis sayısının her geçen yıl arttığını ve bunların da dünyanın birçok yerinde yaşayan insanların ölümüne sebep olduğunu belirtmişlerdir [4-5]. Örneğin Amerika Birleşik Devletlerinde meydana gelen ölümlerde tıbbi hatalar ile ölümler, kardiyovasküler hastalıklar ve kanserden sonra en önde gelen üçüncü neden olarak tespit edilmiştir [6]. Erken teşhis konan bir hastanın tedavisinin başarı ile sonuçlanma olasılığı, erken teşhisin zamanına bağlı olarak genellikle yükselmektedir. Buna ek olarak, erken tanıdaki rehabilitasyon masrafları, geç tanıya kıyasla çok daha düşük olmaktadır.

Tıbbi hatalar, sürekli olarak yeni ve karmaşık hastalıkların ortaya çıkması, tıp alanında yeterli araştırmaların bulunmaması veya doktorların doğru karar vermesine yardımcı olan araçların bulunmamasından kaynaklanmaktadır [7].

Karmaşık hastalıkların bazılarında doğru teşhis kararı almak çok zor ve bazen de imkânsız olmaktadır. Bundan dolayı, araştırmacılar, özellikle iş, mühendislik, tıp vb. gibi önemli alanlarda karar destek sistemlerinin kullanılmasıyla giderek daha fazla ilgilenmektedirler [8-9].

Günümüzde karar destek sistemleri doktorların hastalıkları teşhis etmesinde ve karar vermelerinde yardımcı olan güçlü araçlar arasında yer almaktadır [10]. Karar destek sistemleri üzerine doktorların daha kolay tanı koymalarına yardımcı olacak birçok çalışma yapılmıştır. Bu anlamda, uzman sistemler ve yapay zeka teknikleri çeşitli tıp branşlarındaki farklı problemleri çözmek için başarıyla kullanılmaktadır [10].

Birçok hastalık insan sağlığını yaşam sürekliliği ve kalitesi gibi farklı açılardan etkileyerek tehdit etmektedir. Bunlar arasında, tüberküloz (TB), kronik obstrüktif akciğer hastalığı (COPD – KOAH), zatürre, astım ve akciğer kanseri gibi göğüs hastalıkları hem gelişmekte olan ülkelerde, hem de gelişmiş ülkelerde en ciddi sağlık komplikasyonu ve ölüm nedenlerinin başında gelmektedir [11].

Bu tez çalışmasında, doktorların göğüs hastalıklarını teşhis etmesine yardımcı olmak için bazı yapay zeka tekniklerini kullanan karar destek sistemlerinin kullanılması üzerinde çalışılmıştır. Bu teknikler arasından Destek Vektör Makinaları (SVM) ile sınıflandırma, K-En yakın komşu (K-NN) ve Doğal Bayes sınıflandırması (NB) kullanılmıştır. Ayrıca, iyi bilinen makina öğrenme tekniklerinden olan SVM; Uyarlamalı SVM (ASVM) olarak adlandırılıp başarı oranını arttırmak için en uygun eğilim (bias) değerini bulacak şekilde geliştirilmiştir. Bu çalışma ile ilk kez göğüs hastalıklarının tanısı için ASVM performansı test edilmektedir.

Kullanılan yöntemler Diyarbakır Göğüs Hastalıkları Hastanesi'nden alınan deneysel bir veri kümesi kullanılarak değerlendirilmiştir. Önerilen yöntemler önceki çalışmalarda kullanılan makine öğrenmesi yöntemleri ile karşılaştırılmıştır.

Bu tez çalışması dört ana bölümden oluşmaktadır:

Birinci bölümde, genel olarak göğüs hastalıkları ve topluma olan etkisi ile teşhis için kullanılan yöntemler hakkında genel bilgiler verilmiştir.

İkinci bölümde tüberküloz (TB), kronik obstrüktif akciğer hastalığı (KOAH), zatürre, astım ve akciğer kanseri hakkında detaylı bir literatür çalışması sunulmaktadır. Her hastalık türü için varsa tipi, semptomları, doktorlar tarafından yapılan teşhis çeşitleri, hastalığın nedenleri ve iyileştirme sürecinde doktorlar tarafından önerilen çeşitli tedaviler detaylandırılmaktadır.

Üçüncü bölümde doktorların hastalıkları teşhis etmesine ve karar aşamasına yardımcı olacak farklı makine öğrenmesi algoritmaları, hazırlanmış olduğumuz sistem ile birlikte sunulmaktadır. Her tekniğin özellikleri detaylandırılmış, literatürde yer alan önemli ve farklı araştırmalardaki göğüs hastalıkları teşhisinde kullanımları ile birlikte sonuçları incelenerek sıralandırılmıştır.

Son bölüm, tez çalışmasında kullanılan farklı sistemlerin deneysel sunumu ve literatürde yer alan benzer araştırma sonuçlarının tez sonuçları ile karşılaştırılmasını kapsamaktadır.

Tez çalışmasının son bölümünde, tezin katkıları ve ana sonuçları özetleyen genel bir tartışma ile sonraki çalışmalar hakkında bilgiler verilmiştir.

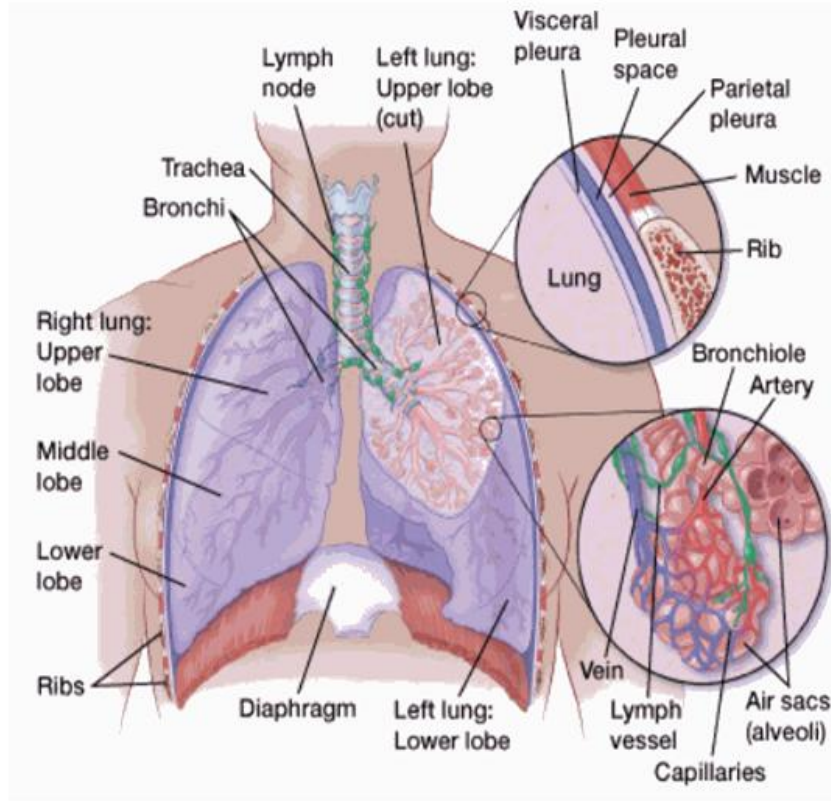


## BÖLÜM 2. GÖĞÜS HASTALIKLARI

### 2.1. Giriş

Göğüs, insan vücudunun boyun ve karın bölgeleri arasında yer alan, kalp, akciğerler, timüs bezi ve diğer çeşitli iç organlarını içeren bir bölgedir [12].

Göğüs, insan vücudu üzerinde önemli rollere sahip olup bunlardan en önemlileri akciğerlerin düzgün şekilde çalışması ve kalp, akciğerler, kan damarları gibi hayati organların korunmasını sağlamasıdır [11]. Şekil 2.1., insan vücudunda göğüs içerisinde yer alan farklı organları göstermektedir.



Şekil 2.1. İnsan vücudundaki göğüs organları[12].

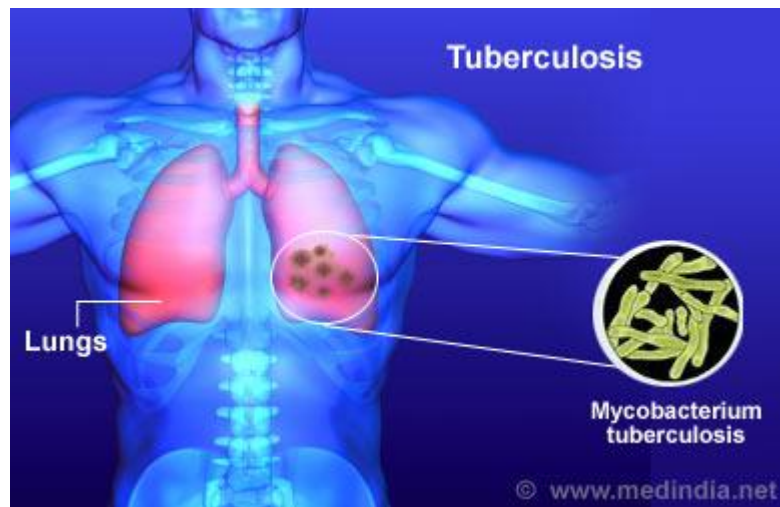
Göğüs hastalıkları ile ilgilenen tıp dalı, pulmonoloji [13], temel olarak göğsü tehdit eden hastalıkların tanısı ve tedavisi üzerinde çalışmaktadır.

Polmonologlar başta tüberküloz (TB), kronik obstrüktif akciğer hastalığı (KOAH - COPD), zatürre, astım ve akciğer kanseri gibi belirli tehlikeli hastalıkların teşhisi ve göğüs üzerindeki etkileri üzerinde çalışmaktadır [11].

## 2.2. Tüberküloz

Tüberküloz (TB), genellikle akciğerleri etkileyen tehlikeli, bulaşıcı, zehirli ve ölümcül sonuçlar doğuran bir hastalıktır. Sadece göğüs değil, aynı zamanda lenf ve beyni de etkileyen karmaşık bir hastalık olarak tanımlanmaktadır [14].

Akciğer tüberkülozu gözlemlenen insanlar, öksürme ve konuşma gibi basit eylemler ile hastalığı kolayca sağlıklı insanlara bulaştırdıkları için genellikle tehlikelilerdir [14]. Aslında, yetersiz beslenen veya zayıf bağışıklık sistemine sahip olan insanların savunma dirençleri düşüktür. Bu yüzden Afrika, Güneydoğu Asya ve benzeri koşullara sahip bölgelerde bu hastalık yoğun şekilde görülmektedir. TB, hastalık ve ölümlerin önemli bir nedeni olarak Dünya Sağlık Örgütü (WHO – DSÖ) tarafından 2014 yılında 1,5 milyon tüberküloz ölümü ile gösterilmiştir [15].



Şekil 2.2. İnsan göğsündeki akciğer tüberkülozu [16].

Tüberküloz, 0,3-0,6 mm. genişlik ve 1-4 mm. uzunluğa sahip çubuk şeklindeki mikro bakteri olarak isimlendirilen mikro organizmalardan kaynaklanmaktadır [17]. Tüberküloz bakterileri vücuda solunum yoluyla akciğerlere yerleşir ve kan yoluyla vücudun diğer bölgelerine taşınırlar [18, 19].

### **2.2.1. Tüberküloz türleri**

Tüberkülozun birçok türü bulunup bunlardan en önemli iki türü gizli TB ve aktif TB hastalığıdır [15]. İlk tür olan gizli TB, TB enfeksiyonu olarak da adlandırılmakta olup, bu hastalıkta bakteriler hiçbir belirti göstermeksizin inaktif bir durumda vücut içerisinde yaşarlar ve diğer kişilere bulaşmazlar. Bulaşıcı değildir ve tahlil sonuçlarında anormali gözlenmez, ancak tedavinin yokluğunda aktif duruma dönebilirler. DSÖ, gizli TB hastalığına sahip kişilerin %10'nunun aktif TB'ye dönüştüğünü belirtmiştir. İkinci tür TB hastalığında ise bakteriler aktif haldedir, bulaşan kişiler hastadır, belirtiler açıkça gözlemlenir ve diğer kişilere kolaylıkla bulaşabilir [15].

### **2.2.2. Tüberküloz belirtileri**

Gizli TB'ye sahip bir kişi belirti göstermezken, aktif TB bulaşmış bir kişi doktorlara göre kolaylıkla tanımlanmış belirtilere sahiptir. Bu belirtiler aşağıdaki gibi olabilmektedir [20]:

- Ciddi öksürük.
- Göğüs şiddetli ağrısı.
- Kan tükürme.
- Zayıflık.
- Önemli Kilo kaybı.
- İştahsızlık.
- Titreme,
- Ateş,
- Gece terlemesi.

Yukarıda yazılmış olan belirtilerin her hangi birinin ortaya çıkmasından sonra kişi ne kadar kısa sürede başvurursa tedavinin başarı oranını arttıracığından, hastaların derhal tıbbi yardım alması doktorlar tarafından önerilmektedir.

### **2.2.3. Tüberküloz teşhisi**

Tüberküloz teşhisi için bir çok teknik vardır, ancak görüntüleme tekniklerine dayanan testler en çok kullanılanlardır [18]. Bunlar arasından, göğüs tüberkülozu gibi hastalıkları teşhis etmek için X-ışınlarını kullanan tıbbi bir görüntüleme tekniği olan göğüs radyografisi (göğüs röntgeni) önem taşımaktadır.

Mikroskopi kullanılan diğer bir araç olmakla birlikte, TB bakterinin var olup olmadığını teyit etmek için hasta balgamını analiz eden bir yöntemdir. Tüberküloz teşhisi için bir başka teknik de mikroorganizma üretimidir. Balgam numunesinin mikro bakteri içerip içermediğini belirlemek için mikroorganizmaların uygun ortamda üretilmesine dayanır. Bu teknik, düşük sayıdaki bakterileri tespit edebildiğinden dolayı mikroskobiden etkili olduğu bilinmektedir [21].

### **2.2.4. Tüberküloz tedavisi**

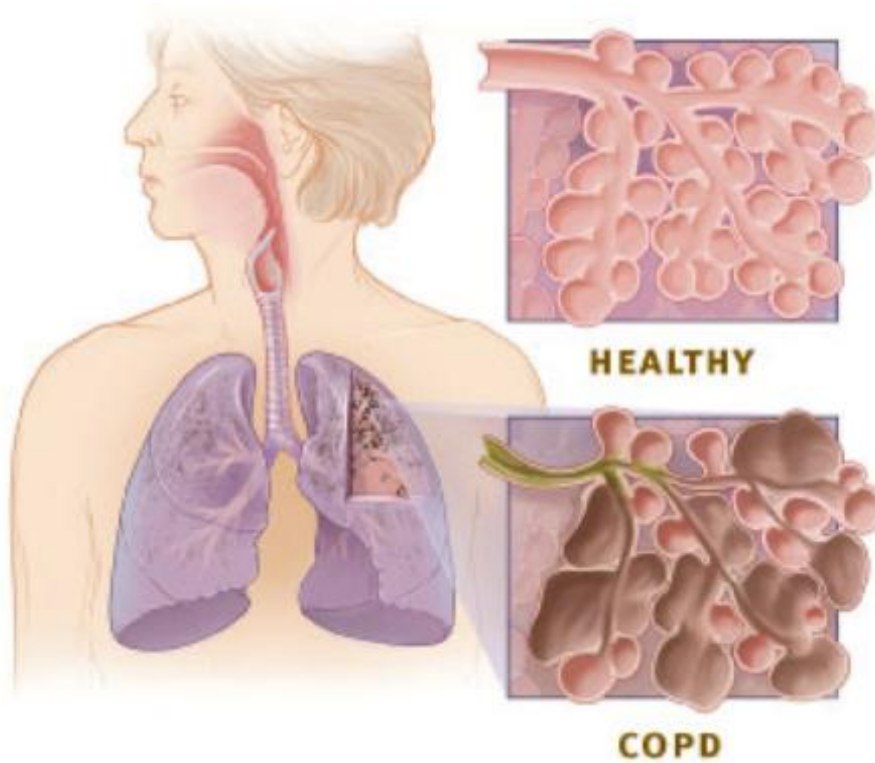
Çoğu insanda, TB uygun tedavi ile tamamen iyileştirilebilmektedir. TB tedavisi, TB'nin türüne bağlıdır ve diğer bakteri enfeksiyonlarının tedavisinden daha uzun sürebilmektedir. Tedavinin amacı enfekte olmuş organlardaki bakterileri yok etmektir. Gizli TB hastalığının tedavisinde tek tür ilaç kullanılır. Ancak aktif TB için doktorlar birkaç ilacın kombinasyonunu kullanmayı tercih ederler.

Standart TB tedavisi 6-9 ay arasında değişmektedir ve antibiyotikler baz alınarak yapılır. Dünya Sağlık Örgütüne göre 2000 ile 2014 yılları arasında TB tedavisi ile 43 milyon hayat kurtarılmıştır [15].

### 2.3. COPD

Kronik Obstrüktif Akciğer Hastalığı (KOAHA - Chronic Obstructive Pulmonary Disease (COPD)), akciğer hava yollarının kronik olarak iltihaplanması ile oluşan kompleks bir göğüs hastalığıdır [22]. COPD, son yıllarda ölümlerin hızla artmasına neden olan kronik hastalıklardandır [23]. Dünya Sağlık Örgütü ölüm nedenleri sıralamasında COPD hastalığının dördüncü sırada yer aldığını ve bu rakamın 2030 yılı itibarıyla üçüncü sıraya yükseleceğini açıklamıştır [24].

Bununla birlikte, bu hastalıkların erken teşhisi hayatta kalma oranını ve tedavilerin başarılı olma olasılığını önemli ölçüde arttırmaktadır.



Şekil 2.3. İnsan göğsündeki COPD [25].

#### 2.3.1. COPD nedenleri

COPD genellikle sigara, sigara dumanına maruz kalma veya hava kirliliği nedeniyle oluşmaktadır. COPD'nin diğer nedenleri arasında solunum yolu enfeksiyonları ve

toz, duman, kimyasal maddeler gibi çevresel faktörler bulunmaktadır [24]. Daha önce hiç sigara içmemiş veya sigara dumanına maruz kalmamış kişilerde de genetik yatkınlıktan ötürü bu hastalık görülebilmektedir.

### **2.3.2. COPD belirtileri**

COPD belirtileri erken evrede ortaya çıkmaz, hastalığın ileri evrelerinde görülür ve aşağıdakileri içerebilir [24]:

- Nefes alma sorunları
- Göğüste sıkışma
- Öksürük
- Solunum yolu enfeksiyonları
- Kilo kaybı

### **2.3.3. COPD teşhisi**

COPD teşhisi hastanın nefes akışının ölçülmesine dayanmaktadır. Bu anlamda spirometri yöntemi (EFR = Exploration of Respiratory Function – Solunum Fonksiyonunun Araştırılması) COPD’yi teşhis etmek ve şiddetini belirlemek için kullanılmaktadır [26].

Bu yöntem, akciğer fonksiyonlarını test etmek için solunan ve dışarı çıkan hava miktarı ile hızını ölçmeye dayalı sık kullanılan bir araçtır. COPD hastalığının teşhisinde akciğerlerde kalan havanın tespit edildiği akciğer röntgeni veya akciğer röntgeninden daha ayrıntılı veriler gösterebilen CT taraması yapılmaktadır. CT taraması, akciğerleri taramakta ve daha detaylı akciğer filmleri göstermektedir [26].

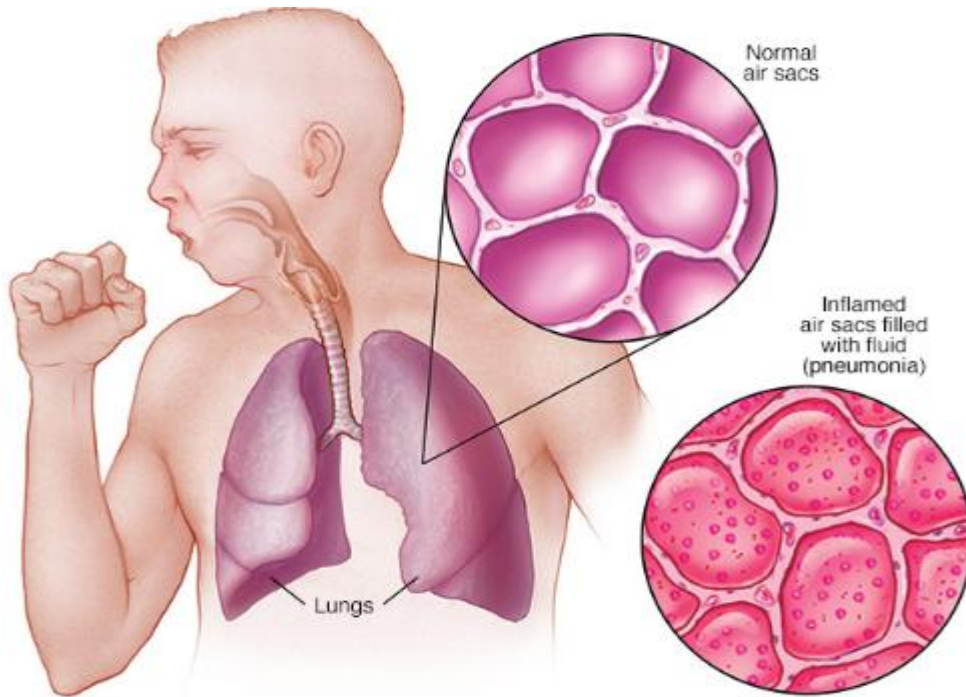
### **2.3.4. COPD tedavisi**

Doktorlar, COPD hastalığını tedavi etmek için farklı ilaçlar kullanmaktadır. Bazı doktorlar oksijen terapisi ve akciğer rehabilitasyon programı gibi akciğer terapileri

tercih ederken, bazı doktorlar ise antibiyotikleri tedavi esnasında tercih etmektedir. Ayrıca doktorlar, sigarayı bırakmanın tedaviyi büyük ölçüde etkileyen ilk ve en önemli adım olduğunu belirtmekte, tedavi sürecini hızlandıran ve hastalığın kötüye gitmesini önleyen bir etken olduğu fikrini paylaşmaktadır [26].

#### 2.4. Zatürre

Zatürre göğüs enfeksiyonu ile oluşan salgın bir hastalıktır [27]. Bu hastalık, özellikle iki ay ile yedi yaş arasındaki çocuklar başta olmak üzere her yaşta insanı etkileyebilmektedir. Hastalık, genellikle grip şeklinde ortaya çıkmaktadır. Tıbbi araştırmalardan önce zatürre birçok insanın ölümüne sebep olmuştur. WHO raporuna göre zatürre sebebiyle 2015 yılında yaklaşık olarak 922 bin çocuk kaybedilmiştir [28]. Ayrıca, çoğunluğu Güney Asya ve Afrika Sahra çevresinde olmak üzere, her yıl beş yaş altı yaklaşık olarak 1,2 milyon çocuk bu hastalıktan dolayı ölmektedir [28].



Şekil 2.4. İnsan göğsündeki zatürre [29].

### 2.4.1. Zatürre nedenleri

Zatürre mikroplardan kaynaklanmakta olup, bunlar arasında en çok bakteri, virüs veya mantar gibi diğer bulaşıcı etmenlerden meydana gelmektedir. Bu mikroplar, soğuk veya grip gibi bir solunum yolu enfeksiyonundan sonra veya viral bir hastalığın ortaya çıkmasından sonra, enfeksiyona maruz kalmış havadan solunum yoluyla vücuda girmektedir [30].

Sigara içen ve zatürre olmaya meyilli, zayıf bağışıklık sistemine sahip, yetersiz beslenen, kimyasal ürünlere maruz kalan kişiler veya hava kirliliği, astım veya COPD gibi diğer solunum yolları hastalıklarına sahip kişilerin zatürre olma olasılıkları diğer kişilere kıyasla daha fazladır.

### 2.4.2. Zatürre belirtileri

Zatürre belirtileri hastalığın ciddiyetine, hastalığın türüne, hastanın yaşına ve genel sağlık durumuna bağlı olarak değişmektedir. Bu hastalığın belirtileri aşağıdaki şekilde tanımlanabilmektedir [28]:

- Nefes darlığı
- Göğüs ağrısı
- Öksürük
- Yüksek ateş
- Mide bulantısı
- Kusma
- İshal

Ek belirtiler de şu şekilde görülmektedir:

- Baş ağrısı
- İştahsızlık
- Halsizlik



- Yorgunluk
- Aşırı terleme

### **2.4.3. Zatürre teşhisi**

Zatürre belirtisi gözlemlenen kişilerde nefes darlığı veya göğüs ağrısı şiddetlenirse doktora başvurulmalıdır. Zatürre teşhisi zor hastalıklar arasında yer almaktadır. Bu nedenle, bu hastalığın teşhisi göğüs röntgen filmi, kan testleri, balgam testi, göğsün hatasız görüntüsünü oluşturmayı sağlayan CT taraması, akciğerlerin sorunlarını tanımlamak için solunum yollarına içerisinde kamera yer alan esnek bir tüp sokulması ile yapılan bronkoskopi gibi bir dizi teste dayanmaktadır [31].

### **2.4.4. Zatürre tedavisi**

Zatürre tedavisi, hastalığın şiddetine, yaşına ve sağlık durumuna göre değişmektedir. Evde ilaçlarla veya hastanede tedavi edilebilmektedir.

Zatürre tedavisi doktor tarafından belirlenen bir süre boyunca antibiyotik alımı veya aspirin gibi ateş düşürücü ilaçlar alınarak yapılabilmektedir. Zatürre olan kişi 65 yaşın üzerindeyse, solunumu hızlanırsa, kişi yer ve zaman kavramları karışırsa veya diğer hayati belirtileri zayıflarsa kesinlikle hastanede tedavi edilmelidir [32].

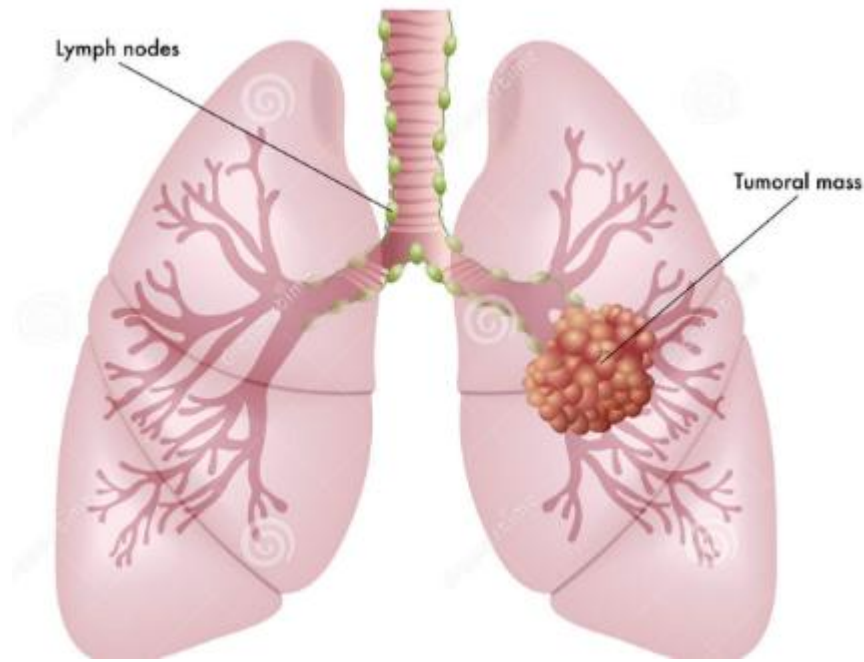
Ayrıca, zatürre tedavisi, tedavi sonrası çeşitli aşular kullanılarak devam edebilmektedir. Doktorlar zatürreden tamamen korunulamayacağını belirtmekte olup uygun beslenme yöntemleri, ağız diş hijyeninin korunması, sigara ve sigara içilen ortamlardan uzak durulması ile hastalık riskinin azaltılabileceğini belirtmişlerdir.

## **2.5. Akciğer Kanseri**

Kanser, insan yaşamını tehdit eden ve sonlandıran ölümcül hastalıkların başında gelmektedir. Beyin, göğüs, karaciğer, kemik ve kemik iliği gibi insan vücudunun

hayati organlarına zarar verebilen, hücrelerin hızlı ve kontrolsüz biçimde büyümesidir [33].

WHO istatistiklerine göre, akciğer kanseri kanser istatistikleri arasında en yaygın türdür [34]. Akciğer kanseri nedeniyle 1,59 milyon ölüm vakası kaydedilmiştir. Akciğer kanseri ile ölümlerin sayısı, meme ve prostat kanseri nedeniyle ölümlerin toplam sayısından daha fazladır [34].



Şekil 2.5. İnsanda göğüs kanseri [35].

### 2.5.1. Akciğer kanseri türleri

Kanser, etkilediği organla sınırlı olmayıp diğer organlara da bulaşabildiğinden tehlikeli bir hastalık olarak tanımlanmaktadır. İki tür akciğer kanseri bulunmaktadır. İlk tür akciğerlerde başlar, ikinci tür ise başka bir organda başlayarak akciğerleri etkiler.

İlk tür birincil akciğer kanseri olarak adlandırılır [36]. Birincil akciğer kanseri, küçük hücre akciğer kanseri (SCLC - small cell lung cancer) ve küçük olmayan hücre

akciğer kanseri (NSCLC - non-small cell lung cancer) olarak iki ana türü bulunmaktadır. SCLC kanserine, mikroskop altında hücrelerin çok küçük görülmesinden dolayı yulaf hücresi kanseri de denilmektedir. Sigara içenler arasında en sık gözlemlenen ve kötü huylu kanserler arasında en hızlı yayılım gösteren kanser bu türdür. Bu hastalık genellikle akciğerlerde görülse de nadiren pankreasta da ortaya çıkabilir [36]. İkinci tür olan NSCLC ise akciğer kanserleri arasında en yaygın olup, akciğer kanserlerinin yaklaşık %85'i bu türdür. SCLC türünden daha yavaş büyür, yayılır ve belirtiler sadece ileri evrelerde gözlemlenir.

İkinci tür akciğer kanseri başka bir organdan akciğerlere yayılmış olan, bu nedenle ikincil akciğer kanseri olarak adlandırılan kanser türüdür. Bu türün tedavisi hastalığın başladığı organın tespit edilmesi gerektiğinden çok daha karmaşıktır.

### **2.5.2. Akciğer kanserinin nedenleri**

Akciğer kanserinin nedenleri araştırmacılar tarafından günümüzde hala araştırma konusu olarak devam etmektedir. Doktorlar, akciğer kanseri ve sigarayı birbirleriyle ilişkilendirmiş, akciğer kanserinin başlıca nedenin sigara kullanımı olduğunu belirtmişlerdir. Akciğer kanseri riskini büyük ölçüde arttıran kirlilik, kimyasal ürünlere maruz kalma gibi diğer ikincil faktörler de bulunmaktadır [37].

### **2.5.3. Akciğer kanseri belirtileri**

Akciğer kanseri diğer akciğere bağlı hastalıklardan daha şiddetli belirtiler göstermektedir. Akciğer kanseri belirtileri, aşağıdaki gibi ortaya çıkabilir [38]:

- Öksürük
- Göğüste şiddetli ağrı
- Kilo kaybı
- Şiddetli halsizlik
- Baş ağrısı
- Zayıflık

- Nefes darlığı
- İştahsızlık
- Göz problemleri

#### **2.5.4. Akciğer kanseri teşhisi**

Hastalığın önemi ve şiddeti göz önüne alındığında, hastalığın tanımlanmasında kullanılan teknikler çok fazladır. Genellikle hastalığın varlığını kanıtlamak için doktorlar tüm bu teknikleri aynı anda kullanırlar.

Hastalığın tanımlanması doktorlar göğüs röntgeni ile başlamakta olup, röntgen ışınlarında tümör, beyaz-gri kitle halinde ortaya çıkar. Burada elde edilen sonuçlar kesin değildir. Bu nedenle hastalıklı vücudun iç organlarının ayrıntılı görüntülerini oluşturan bilgisayarlı tomografi (CT) taraması ile sonraki test gerçekleştirilir. CT taramasının pozitif çıkması durumunda A PET-CT taramasına geçilmektedir. Bu test daha ayrıntılı olup kanserli hücreler daha rahat tanımlanabilmektedir [39].

#### **2.5.5. Akciğer kanserinin tedavisi**

Akciğer kanserini tedavi etmek için doktorlar ilk olarak kanserin lokal veya başka organlara dağılıp dağılmadığını ve tümörün boyutunu ölçmek için hastalığın aşamasını belirlemektedirler. Bu aşama tedaviye başlamak ve uygun tedavinin seçimi için yardımcı olacak önemli bir adımdır [40].

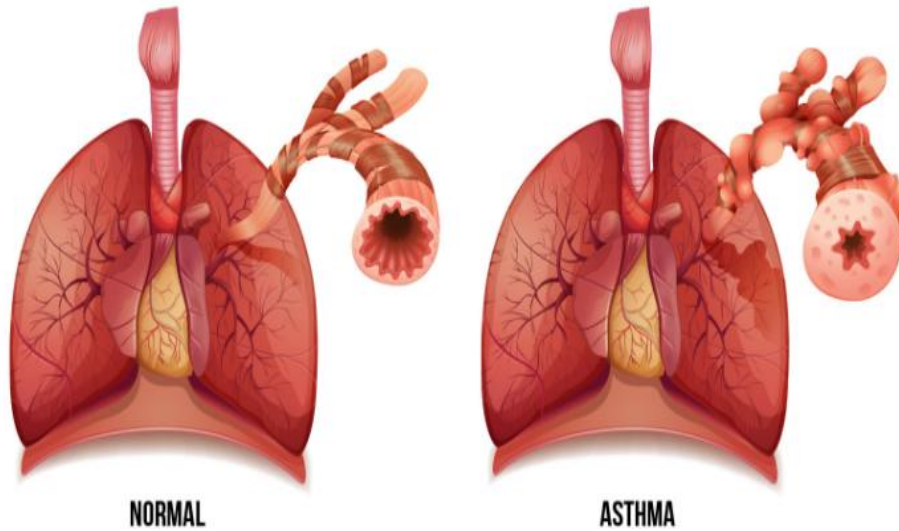
Eğer akciğer kanseri lokalse, bölgesel olarak radyoterapi veya cerrahi operasyonlar yoluyla tedavi yapılmaktadır. Eğer akciğer kanseri hastanın vücuduna yayılmış ise tüm kanserli hücreleri öldürmek için vücudun tamamında dolaşan ve sistematik tedavi adı verilen tedavi uygulanmaktadır.

Akciğer kanserinin tedavisi türüne bağlıdır. Birincil akciğer kanseri olan SCLC türünde kanser hücrelerini öldürmek için öncelikli olarak güçlü kimyasal maddeler içeren kemoterapi tedavisi uygulanmaktadır [40]. İkinci tür akciğer kanseri olan

NSCLC türünde ameliyat, kemoterapi, radyoterapi veya bu tekniklerin kombinasyonları ile tedavi gerçekleştirilebilir [40]. İkinci tür akciğer kanserini tedavi etmek için kanserli akciğer hücrelerinin ameliyat ile alınması, kanser hücrelerinin ısıtılarak radyo frekanslı ameliyat ile (Radiofrequency Ablation - RFA) öldürülmesi, biyolojik tedavi ile hücre bölünmelerinin durdurulması veya kanserli hücrelerin gelişimini durdurmaya yardımcı olmak için vücut içerisindeki bazı hormonların seviyesini düşürülmesi gibi yöntemler izlenmektedir [40].

## 2.6. Astım

Astım, akciğerlerin uzun süreli iltihaplanması ile oluşmakta olup bulaşıcı değildir, solunum yollarının enfekte olmasıdır [41]. Astım, insan vücudunda ilerlemesine bağlı olarak günde veya haftada birkaç kez çeşitli sürelerde gözlemlenen ataklarla her yaşta insanı etkileyen ve dünyada yaygın bir şekilde görülen bir hastalıktır. WHO'ya göre, yaklaşık 300 milyon kişi astım hastalığından şikâyetçidir [42]. Bu hastalık yaşamın tüm evrelerinde görülebilir. Genellikle çocukluk çağında görülmekle birlikte, ergenlik döneminde kaybolur ancak daha sonra tekrar ortaya çıkabilir.



Şekil 2.6. İnsan göğsündeki astım [43].

### **2.6.1. Astım nedenleri**

Doktorlar astım nedenlerinin yeterince iyi belirlenmediğini doğrularken, hastalığın çoklu genetik ve çevresel faktörlerin karmaşık bir etkileşiminin sonucu olduğunu belirtmektedir. Ayrıca, çevre kirliliği ve bazı alerjiler de astımın nedenleri arasındadır [44].

### **2.6.2. Astım belirtileri**

Astım belirtileri kişiden kişiye değişmekte olup, sıklıkla gözlemlenen belirtiler aşağıdaki gibidir [44]:

- Şiddetli nefes darlığı
- Hırıltı
- Göğüs daralması
- Göğüs ağrısı
- Öksürük
- Uyku problemleri
- Yorgunluk

Bu belirtiler astım atakları olarak bilinmekte olup gece ve sabahları şiddetlenmektedir. Astım belirtilerin şiddetine bağlı olarak, hafif astım, orta dereceli astım ve şiddetli astım olarak isimlendirilmektedir.

### **2.6.3. Astım teşhisi**

Astım, belirtilerin incelenmesi yoluyla kolay bir şekilde tespit edilmektedir. Bir hastanın astıma yakalanıp yakalanmadığını tespit etmek için öncelikle hastada gözlemlenen belirtilerin sorulduğu bir sağlık geçmişi anketi doktorlardan tarafından doldurtulmaktadır. Ardından, doktor bazı fiziksel muayene yöntemleri kullanarak gözler, kulaklar, burun ve deriyi kontrol etmektedir. Ayrıca, hasta solunumunu ölçmek için akciğer fonksiyon testi de doktorlar tarafından kullanılmaktadır [44].

#### 2.6.4. Astım tedavisi

Astım, tamamen yok edilemeyen fakat kontrol altına alınan bir hastalık olduğundan dolayı uzun süreli tedaviler uygulanmaktadır. Hastalığın tedavisi her zaman aynı değildir. Bir hastadan diğerine farklılık gösterebilir. Kullanılacak olan ilaç, hastanın yaşına ve sahip olduğu belirtilere bağlıdır. Doktorlar tarafından hastalığı kontrol altına almak için bronşları genişleten ve rahatlatan ilaçlar kullanılmaktadır [44].

#### 2.7. Sonuç

Bu bölümde en yaygın ve tehlikeli göğüs hastalıklarından olan tüberküloz (TB), kronik obstrüktif akciğer hastalığı (COPD), zatürre, akciğer kanseri ve astım tanıtılmıştır. Hastalıkların türleri, nedenleri, belirtileri, teşhisleri ve tedavileri açıklanmıştır.

Bu hastalıklarının nedenleri birbirlerine benzer olup genellikle çevre kirliliği ve sigara temellidir. Bu faktörlere bağlı olarak hastalarda göğüste ağrı, kanlı öksürük, zayıflık, yorgunluk, halsizlik, iştahsızlık, ateş, vb. gibi yaygın belirtiler gözlemlenmektedir.

Bu bölümde açıklanan hastalıkların ölümcül riskleri göz önüne alındığında, hastalıkların analiz ve tespit edilmesi için güvenilir yöntemlerin kullanılması önem taşımaktadır. Bir sonraki bölümde bu farklı göğüs hastalıklarını teşhis ve tespit etmek için kullanılan farklı teknikler ve yöntemler ayrıntılı bir şekilde tanıtılacaktır.

## **BÖLÜM 3. KULLANILAN MAKİNA ÖĞRENMESİ TEKNİKLERİ**

### **3.1. Giriş**

Bir önceki bölümde açıklandığı gibi tüberküloz (TB), kronik obstrüktif akciğer hastalığı (COPD), zatürre, astım ve akciğer kanseri gibi göğüs hastalıkları ciddi sağlık problemleri arasında yer almakta ve birincil ölüm nedeni olarak gösterilmektedir.

Doktorlar bir hastalığın teşhisini ne kadar erken evrede tespit ederlerse hastanın iyileşme olasılığını o denli artmaktadır. Bu anlamda, teşhis sırasında karşılaşılabilecek problemleri en aza indirebilmek için uzman sistemler ve çeşitli karar destek sistemleri başarılı bir şekilde kullanılmaya gelmektedir. Bu tez çalışmasında doktorların göğüs hastalıklarını teşhis etmesine yardımcı olmak için yapay zeka tekniklerine dayalı karar destek sistemleri kullanılmıştır. Bu teknikler arasında, göğüs hastalıklarının teşhisinde, Destek Vektör Makinaları (Support Vector Machines -SVM), k-En Yakın Komşu (k-Nearest Neighbor - KNN), Basit Bayes (Naive Bayes - NB) yer almakta olup tez kapsamında ilk defa Uyarlamalı Karar Destek Makinaları (Adaptive Support Vector Machine - ASVM) performansı incelenmiştir.

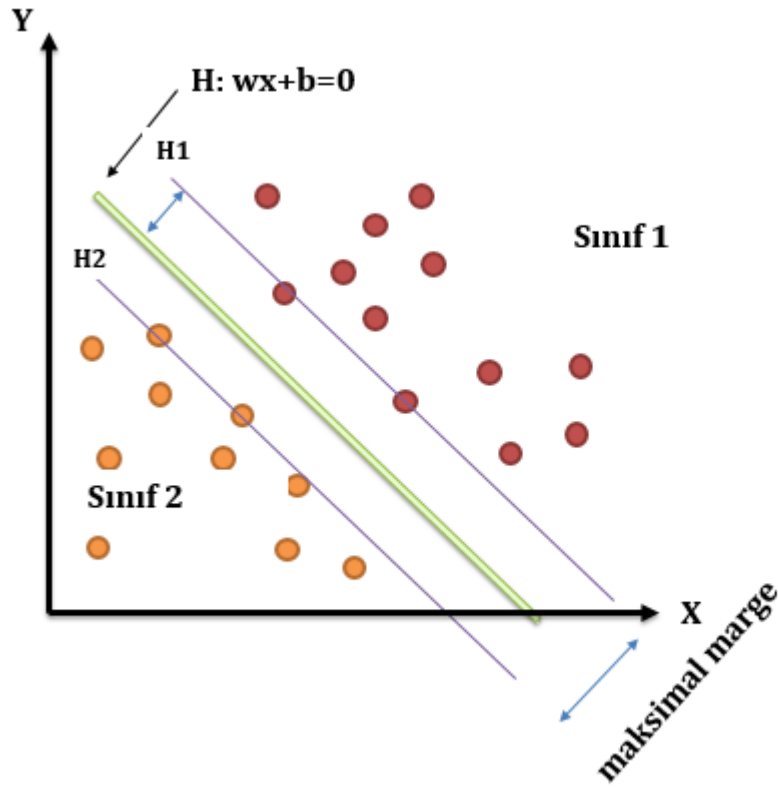
### **3.2. Destek Vektör Makinaları: SVM**

#### **3.2.1.SVM temelleri**

SVM, ikili veri sınıflandırması için ilk kez 1990'da Cortes ve Vapnik tanımlanan, ayrıklık ve regresyon problemlerini çözmek için kullanılan bir makina öğrenme tekniğidir [45, 50]. Günümüzde yüz tanıma [51], ses tanıma [52], tıbbi tanı [53], vb. Çalışma alanlarında kullanılmaktadır. SVM, Şekil 3.1.'de görüldüğü gibi D



$= \{(x, y), \dots, (x_m, y_m)\}$  bir veri setinin iki sınıfını optimal olarak ayırarak maksimum kenar hiperdüzlemi oluşturmaktadır. Burada  $x_i$  sınıf etiketi olmak üzere,  $x_i \in R_n$  ve  $y_i \in \{-1,1\}$  olmaktadır. Fakat hiperdüzlemlerin birçoğu iki sınıfı ayırabilmektedir. Böylece SVM optimal ayırıştırma hiperdüzlemi (Optimal Separating Hyperplane - OSH) olarak adlandırılan en uygun düzlemini bulmak için bir eğitim fazı kullanmaktadır. OSH,  $f(x) = w^T x - b$  fonksiyonunu uygulamakta olup, buradaki  $w$ ;  $m$  boyutlu vektörü,  $b$  (bias) ise iki sınıfı ayıran ve sınıflar arasındaki kenarı maksimize eden eşik değeri göstermektedir [52].



Şekil 3.1. Doğrusal durum için temel SVM [53].

Şekil 3.1.'de gösterilen en uygun hiperdüzlem H değeri; H1 ile H2 düzlemlerinin kendisine paralel olduğunu ve H'ye en yakın noktalardan geçtiğini doğrular. Bu H değerleri destek vektörleri (SVs) olarak adlandırılmaktadır [52]. Böylece, SVM'ler H1 ile H2 arasındaki uzaklık da olan iki sınıf arasındaki maksimum kenarı en uygun hiperdüzlem olarak seçmektedir. En geniş kenara sahip olan hiperdüzlem  $\frac{2}{\|w\|}$

şeklinde gösterilmektedir. Matematiksel olarak, maksimum kenara sahip olan en uygun hiperdüzlem, Denklem 3.2. sabit olmak koşuluyla aşağıda açıklandığı gibi ikinci dereceden programlama (Quadratic Programming – QP) optimizasyon problemi kullanılarak Denklem 3.1.'i minimize ederek bulunabilmektedir [54]:

$$\min_{w,b} \frac{1}{2} \|w\|^2 \quad (3.1)$$

$$y_i(w^T x_i + b) > 1, i = a, \dots, m \quad (3.2)$$

Yukarıdaki problem, Denklem 3.3.'de ifade edildiği gibi Lagrangian çarpanları metodu kullanılarak çözülebilmektedir.

$$Q(w, b, \alpha) = \frac{1}{2} w^T w - \sum_{i=1}^m \alpha_i \{y_i (w^T x_i + b) - 1\} \quad (3.3)$$

Burada  $\alpha_i$  Lagrange çarpanıdır. En uygun hedef Q fonksiyonu w ile b'ye göre en düşük hale getirilirken  $\alpha_i$  kullanılarak maksimize edilmektedir (Denklem 3.4).

$$\begin{cases} \frac{\partial Q}{\partial w} = 0 \\ \frac{\partial Q}{\partial b} = 0 \end{cases} \quad (3.4)$$

Denklem 3.4. kullanılarak elde edilen sonuç:

$$\begin{cases} w = \sum_{i=1}^m \alpha_i y_i x_i \\ \sum_{i=1}^m \alpha_i y_i = 0 \end{cases} \quad (3.5)$$

Dolayısıyla, Denklem 3.3.'de, Denklem 3.4. ve Denklem 3.5.'in değiştirilmesi ile aşağıdaki ikili maksimize problemi elde edilmektedir.

$$\left\{ \begin{array}{l} \text{Maksimize } Q(\alpha) = \sum_{i=1}^m \alpha_i - \sum_{i=1}^m \sum_{j=1}^m \alpha_i \alpha_j y_i y_j x_i x_j \\ \text{Buan göre } \sum_{i=1}^m \alpha_i y_i = 0 \\ \alpha_i \geq 0 \end{array} \right. \quad (3.6)$$

Burada,  $\alpha_i \neq 0$  Lagrange katsayılarına sahip  $x^i$  örnekleri destek vektörleri tanımlanmaktadır.

Yeni bir  $x$  verisinin sınıflandırma problemi aşağıdaki karar fonksiyonu ile çözülmektedir (Denklemler 3.7.).

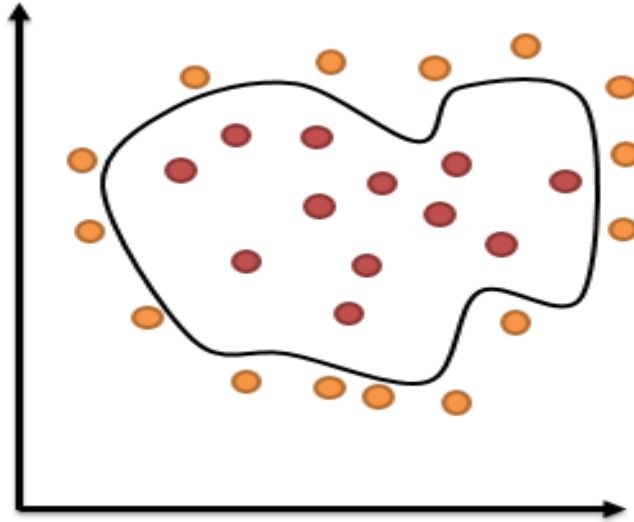
$$\left\{ \begin{array}{l} H(x) = \sum_S \alpha_i y_i x^T x_i + b \\ b = y_i - w^T x_i \end{array} \right. \quad (3.7)$$

Böylece, sınıf karar aşağıdaki şekilde yapılabilir:

- Eğer  $H(x) > 0$  ise  $x \in +1$
- Eğer  $H(x) < 0$  ise  $x \in -1$
- Eğer  $H(x) = 0$  ise  $x$  sınıflandırılmaz.

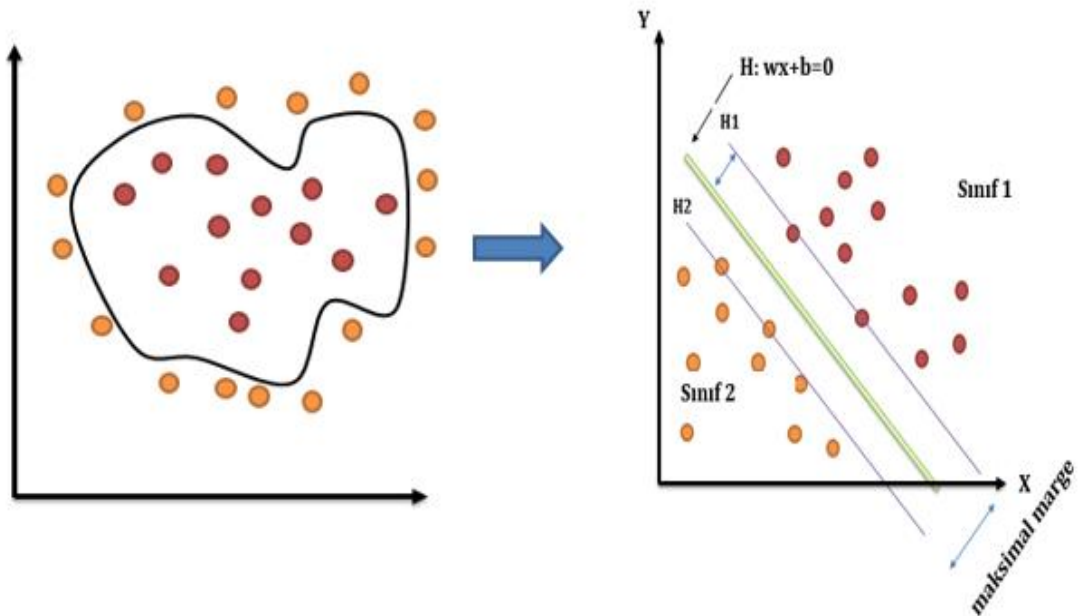
Veriler, Şekil 3.2.'deki durumda görüldüğü gibi doğrusal olarak ayrılabilir değildir. Bu nedenle, doğrusal olmayan uzayı yeni doğrusal yüksek boyutlu uzayla eşleştirerek eğitim setini doğrusal olarak ayrıştırılabilir şekilde Denklem 3.8'de görüldüğü gibi çekirdek fonksiyonuyla ifade ederek eşleştirilmektedir.

$$k(x_i, x_j) = \phi(x_i)^T \phi(x_j) \quad (3.8)$$



Şekil 3.2. Doğrusal olmayan durumda SVM [54].

Dolayısıyla, bu dönüşüm uzayı, SVM çekirdeği seçilirken dolaylı olarak tanımlanan bir  $F = \{\phi\{x\} | x \in X\}$  “eşleştirme fonksiyonu” kullanılarak gerçekleştirilir. Yeni uzaya “Özellikler uzayı” adı verilmektedir (Şekil 3.3.) [54].



Şekil 3.3. Eşleşme uzayı [54].

SVM uygulanırken birden çok çekirdek fonksiyonu kullanılmaktadır. Bunlardan en sık kullanılan fonksiyonlar aşağıdaki gibidir [54]:

Doğrusal çekirdek:  $K(x_i, x_j) = x_i^T x_j$

Polinom çekirdeği:  $K(x_i, x_j) = (\gamma x_i^T x_j + r)^d, \gamma > 0$

RBF çekirdeği:  $K(x_i, x_j) = \exp(-\gamma \|x_i - x_j\|^2), \gamma > 0$

Sigmoid çekirdeği:  $K(x_i, x_j) = \tanh(\gamma x_i^T x_j + r)$

SVM çekirdeğinin seçimi henüz çözülmemiş bir problemdir. SVM seçimi, performansı sınırlayan bir faktördür. Doğru çekirdeği seçmek için araştırmacılar genellikle testler uygularlar. Buradaki tez çalışmasında öncelikle standart doğrusal çekirdek kullanılmış, diğer SVM çekirdekleri üzerinde de bazı testler yapılmıştır.

### 3.2.2. SVM kullanılarak göğüs hastalıklarının teşhisi

Araştırmacıların birçoğu hastalıkların teşhisinde SVM kullanma yoluna gitmişlerdir. Bunun yanı sıra yapılan birçok çalışmada SVM'nin göğüs hastalıklarını teşhis etmede güçlü bir araç olduğu kanıtlanmıştır.

Kaynak [55]'de, akciğer kanserinin teşhisinde SVM ile Boosting, Karar Ağaçları, k-en yakın komşu, LASSO regresyonları, Sinir Ağları ve Karar Ormanı gibi diğer yöntemlerin performansını içeren karşılaştırmalı bir çalışma sunulmaktadır. Boosting metodu %88 teşhis doğruluğu elde ederken, karar ağaçları %75 oranında teşhis doğruluğu elde etmiştir. Buna ek olarak, k-en yakın komşu %73, LASSO regresyonları ise %90 teşhis doğruluğu elde etmiştir. Ayrıca çalışmada sinir ağları %89, karar ormanları metodu %90 oranında doğru teşhis elde etmiştir. Bununla birlikte en yüksek teşhis doğruluğu %93 oran ile SVM tarafından sağlanmıştır [55].

SVM, çocuklukta görülen zatürre hastalığının teşhisinde başarılı sonuçlar elde etmiştir [56]. Bu çalışmada, klinikten elde edilen veriseti üzerinde iki farklı makina

öğrenmesi sınıflandırıcısı olan k-en yakın komşu (KNN) ve Doğal Bayes, SVM ile karşılaştırılmıştır. SVM %77 oranında doğru teşhis ile en iyi performansı sergilerken, bunu %70 ile KNN ve %68 ile Doğal Bayes takip etmiştir [56].

Temel bileşen analizi (Principal Component Analysis - PCA) olarak adlandırılan teknik ile SVM'nin birlikte kullanılmasının iyi sonuçlar verdiği belirtilmiştir [57]. Bu çalışmada yazarlar SVM ve PCA'yı astım hastalığını teşhis etmede kullanmışlardır. Elde edilen sonuçlar %95,54 oranında hastalıklı kişilerin doğru tahmin edildiğini göstermiştir.

### 3.2.3. ASVM temelleri

Bir önceki bölümde açıklandığı gibi, SVM, b eşik değeri adı verilmiş düzeltilmiş bir terimi en uygun hiperdüzlem  $f(x)=w^T x+b$  denklemi içerisinde barındırmaktadır. Bu çalışmada, hastalıkların teşhis edilmesinde en yüksek başarı oranlarının elde edilmesi için ASVM uygulanmıştır. Gerçekte ASVM, SVM içerisindeki en uygun b eşik değerini bulmayı sağlayan algoritmanın uygulanma fikrine dayanmaktadır [58]. Algoritma, araştırmacı tarafından verilen eşik değeri uzayındaki minimum ile maksimum  $S_{\min}$  ve  $S_{\max}$  parça sayıları ve  $b_{\min}$  ve  $b_{\max}$  minimum ile maksimum eşik değerleri kullanılarak eşik değeri uzaylarının eşit parçalara ayrılması, her parçanın farklı eşik değerlerinin test edilmesi ve en yüksek hastalık teşhis başarı oranına sahip en uygun parçalarının elde edilmesine dayanmaktadır. ASVM algoritması aşağıdaki gibi özetlenebilmektedir [58]:

Başla

Ata  $D = \{(x_1, y_1), \dots, (x_n, y_n)\}$ ,  $x \in R_n$  ve  $y_i \in \{-1, 1\}$ .

Ata  $b_{\min}, b_{\max}$  (minimum ve maksimum eşik değeri parametreleri),  $S_{\min}, S_{\max}$  (minimum ve maksimum parça sayıları),  $\Delta_{\min}$  ( $b_{\min}$  ve  $b_{\max}$  arasındaki minimum fark).

$b_{\max} - b_{\min} > \Delta_{\min}$  ve  $S_{\min} < S_{\max}$  olduğu sürece Tekrarla

$(b_{\min}, b_{\max})$  aralığında,  $S_{\min} - 2$  noktasına eşit uzaklıkları tespit et.

SVM model eğitim verisiyle eşit:  $\text{model} = \text{svmtrain}(D(x), y, \dots)$

Sistemi farklı  $S_{\min} - 2$  eşik değerleri ile test et.

En yüksek iki doğru sınıflandırma oranına sahip  $R_1$  ve  $R_2$ 'yi seç.

Eğer  $R_1$  ve  $R_2$ ,  $b_{\min}$  veya  $b_{\max}$  ise

$$S_{\min} = S_{\min} + 1$$

Değilse

$b_{\min}$  değerini en yüksek ikinci doğru sınıflandırma oranına sahip  $R_2$  eşik değeri olarak ata.

$b_{\max}$  değerini en yüksek ikinci doğru sınıflandırma oranına sahip  $R_1$  eşik değeri olarak ata.

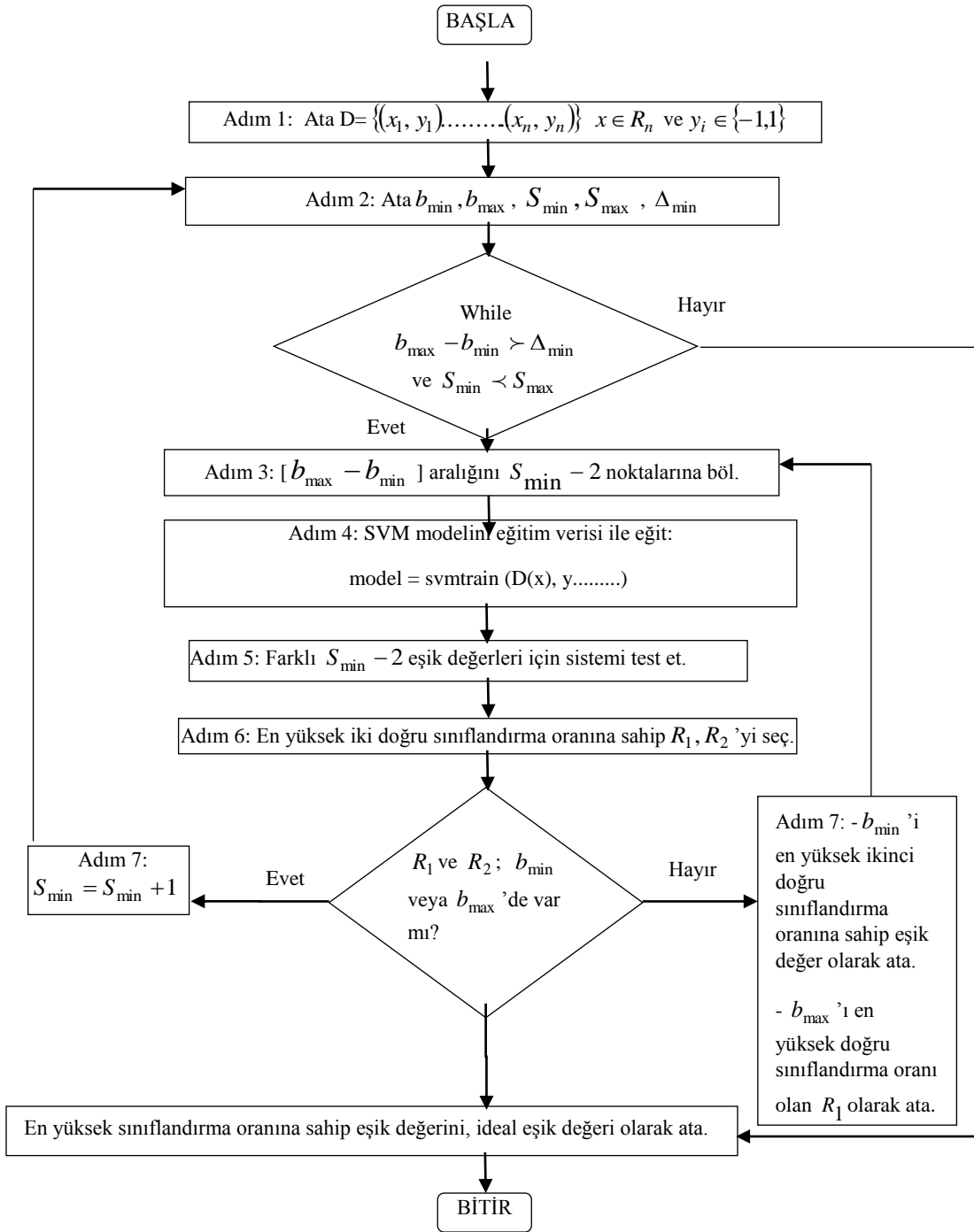
Koşul Sonu

Döngü Sonu

En yüksek sınıflandırma oranına sahip eşik değerini, ideal eşik değeri olarak ata.

Bitir

ASVM algoritması şekil 3.4.'de gösterildiği gibi özetlenebilir.



Şekil 3.4. ASVM algoritması.



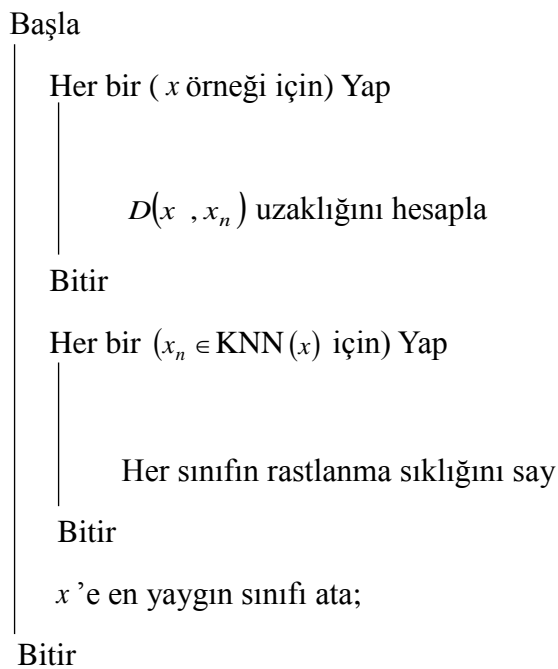
### 3.3. K En Yakın Komşular (KNN)

#### 3.3.1. KNN temelleri

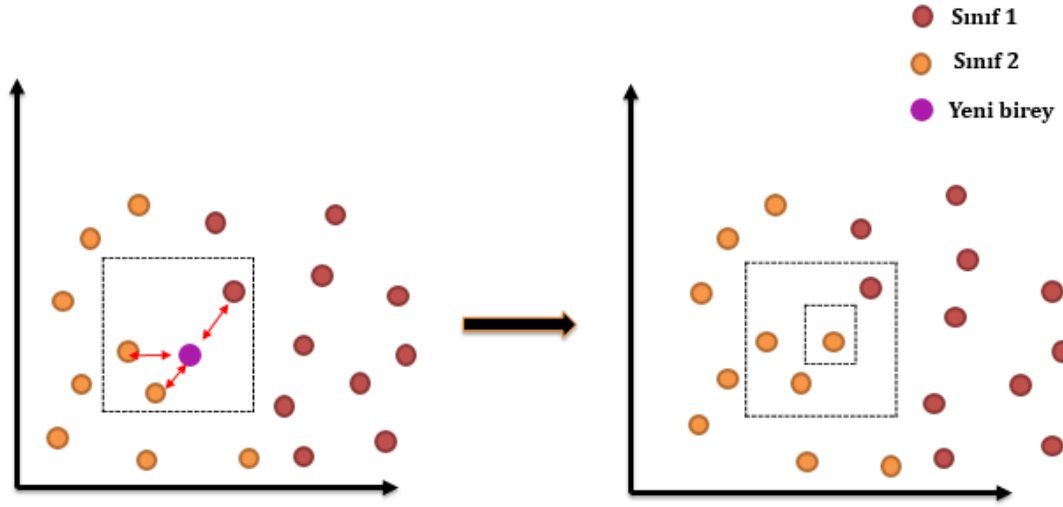
K-en yakın komşular algoritması (K-Nearest Neighbours - K-NN algorithm) ilk olarak Fix ve Hodges tarafından 1951 yılında ortaya atılmış olup en basit yapay zeka tekniklerindedir. Tüm durumları depolayan ve benzerlik ölçütlerine dayanarak yeni durumları sınıflandıran bir bellek fikrine dayanan oldukça basit bir algoritmadır [59]. Bu sınıflandırma algoritmasının temelleri aşağıdaki gibidir:

Verilen  $D(x_1, x_2, \dots, x_n)$  eğitim veriseti için uzaklık fonksiyonu  $d$  ve bir  $k$  tamsayısı tanımlanır.  $x \in R^n$  için yeni bir nokta test edilir ve bir karar verilir. Algoritma,  $D$  veriseti içerisinde  $x$ 'e en yakın  $k$  noktalarını  $d$  uzaklık fonksiyonu içerisinde aramaktadır. Bu uzaklık mesafesi Euclid uzaklığı olarak adlandırılmaktadır. Bu işlemden sonra kendi komşuları arasındaki en yaygın olan sınıfı  $x$ 'e atar [59].

En temel K-NN algoritması aşağıdaki şekilde özetlenmektedir:



Şekil 3.5.  $k=3$  için K-NN'nin çalışma prensibini göstermektedir.



Şekil 3.5. KNN temelleri [59].

$k$  parametresi kullanıcı tarafından belirlenmelidir:  $k \in \mathbb{N}$ . İkili sınıflandırmada, eşit oylamanın önüne geçmek için tek sayıda  $k$  seçmek avantajlıdır.

### 3.3.2. KNN kullanarak göğüs hastalıklarının teşhisi

Daha önce belirtildiği gib, KNN en basit yöntemler arasında yer almakta olup göğüs hastalıklarının teşhisinde araştırmacılar tarafından sıklıkla kullanılmıştır.

Asha ve arkadaşları  $k$ -ortalama ile K-NN'yi birlikte kullanarak tüberküloz tahmini için hibrit bir model önermişlerdir [19]. Kullanılan veriler, tüberküloz hastalarının 700 kayıtlarını içermektedir. Teşhis için 12 öznitelik kullanılmıştır. Uygulamada Bilgi Analizi için Waikato Ortamı (Waikato Environment for Knowledge Analysis - WEKA) kullanılmış olup performans analizi hassasiyet, yeniden çağırma, kappa istatistikleri ve doğrulama gibi farklı istatistiksel ölçütler kullanılarak gerçekleştirilmiştir. KNN %95,4 ile iyi bir performans göstermiştir [19].

Bir başka çalışmada, KNN ve genetik algoritma (GA) birleşiminden oluşan KNN'nin optimize edilmiş yeni bir versiyonu akciğer kanserinin erken safha teşhisinde kullanılmıştır [59]. Bu yöntemde, geliştirilmiş görüntüler KNN algoritmasına sınıflandırılmak üzere girdi olarak verilmiş, ardından sınıflandırma görevini optimize etmek için genetik algoritma kullanılmıştır. Geliştirilen yöntem %90 doğru sınıflandırma ile çok iyi sonuçlara ulaşmıştır.

Kaynak [60], KNN algoritmasını astım tahmini için geliştirmiştir. Çalışma 169 astım hastası ve 85 astım olmayan hastanın Tahran hastanesinden alınan veritabanı bilgileri ile gerçekleştirilmiştir. Elde edilen sonuçlar KNN algoritmasının diğer yöntemlere göre daha etkili olduğunu göstermiştir.

Ayrıca, İran'daki bir hastaneden alınan ve 600 astım hastasına ait veritabanı bilgileri kullanılan bir başka çalışmada KNN ile çok iyi sonuçlar elde edilerek %98 doğrulukla hastalık belirlenmiştir [61].

### **3.4. Doğal Bayes**

#### **3.4.1. Doğal Bayes Temelleri**

Basit Bayes (Naive Bayes – NB) ağları gözetimli öğrenme ve yapay zeka ile veri madenciliği içinde yer alan en basit metottur [62]. 1950'de geliştirilen graf ve olasılık teorilerini birleştiren Bayes teoremine dayanan, olasılık ve istatistiksel bir sınıflandırma yöntemidir [63].

NB, karmaşık sorunları çözmek için doğal araçlara dayanan, anlaşılır ve özel bir yöntemdir. Özellikle birbirinden bağımsız öznitelikler mevcut ve veritabanının önemli olduğu durumlarda kullanılır [64].

Şekil 3.6.'da Basit Bayes ağının genel yapısı gösterilmiştir. NB, [62] [65]'de belirtilen Bayes teoremi ile temelde ana bir düğüm ve azalan düğümlerden oluşmaktadır.

$$P(H | D) = \frac{p(H)p(D | H)}{p(D)} \quad (3.9)$$

Bu denklem,  $P(H | D)$ ,  $D$  olasılığı bilindiğinde  $H$ 'nin gerçekleşme olasılığını bulmaktadır.

- $p(H)$ :  $H$  hipotezinin öncelikli olasılığı,
- $p(D)$ :  $D$  verisinin olasılığı,
- $P(H | D)$ :  $H$  hipotezindeki  $D$  verisinin olma ihtimali.

Bir sınıflandırmada,  $D$  sınıflandırılacak olan veriyi ve  $H$  sınıftaki hipotezi tanımlar. Başka bir deyişle, verilen bir  $x_i$  için,  $x_i$ 'nin  $C_j$  sınıfına ait olma sonraki olasılığı şu şekilde hesaplanır:

$$P(H = C_j | D = x_i) = \frac{p(H = C_j)p(D = x_i | H = C_j)}{p(D = x_i)} \quad (3.10)$$

Bu durumda  $x_i$ 'nin ait olduğu sınıf belirlenmeye çalışılmaktadır.  $P(H = C_j | D = x_i)$  olasılığını maksimize eden değer korunmalıdır. Bu işlem aşağıdaki şekilde formüle edilir:

$$\hat{C}_j = \arg \max_{C_j} \frac{p(H = C_j)p(D = x_i | H = C_j)}{p(D = x_i)} \quad (3.11)$$

$P(D = x_i)$ ,  $C_j$ 'ye bağlı olmadığından yukarıdaki denklem aşağıdaki gibi sadeleştirilebilir:

$$\hat{C}_j = \arg \max_{C_j} p(H = C_j)p(D = x_i | H = C_j) \quad (3.12)$$

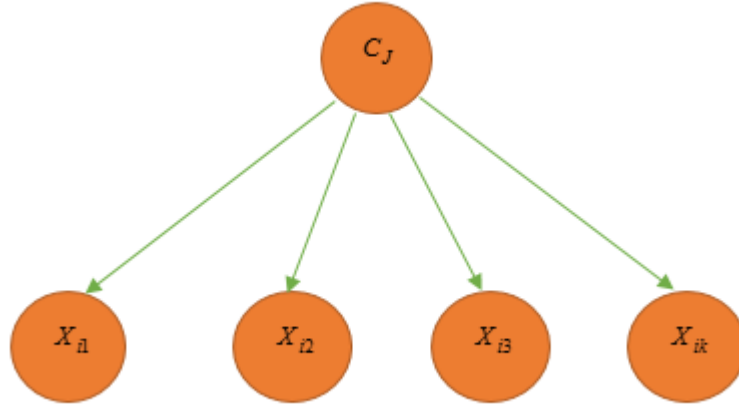
$x_i$  verisi genellikle elemanlar vektörü olarak temsil edilmektedir. Bu vektörün her özniteliği  $x_i$ 'nin belirgin bir değerine karşılık gelmektedir.  $x_i$ 'nin bir sınıfa atanması, sadece kendi değerlerine bağlıdır. Böylece  $x_i$ ,  $x_i = (x_{i1}, x_{i2}, \dots, x_{ik})$  şeklinde ifade edilebilir ve aşağıdaki şekilde gösterilebilir:

$$\hat{C}_j = \arg \max_{C_j} p(H = C_j) p(D = x_{i1}, x_{i2}, \dots, x_{ik} / H = C_j) \quad (3.13)$$

Basit Bayes'de  $x_i$  vektör niteliklerinin birbirinden bağımsız olduğu varsayılmaktadır. Bu varsayım her zaman doğru olmadığından bu metod "basit" olarak adlandırılmaktadır. Fakat bu kıstasa rağmen, Basit Bayes etkili ve verimli bir sınıflandırma metodudur. Bu varsayım göz önüne alındığında aşağıdaki gibi bir denklem oluşturulabilir:

$$p(x_{i1}, x_{i2}, \dots, x_{ik} / C_j) = \prod_{k=1}^k p(x_{ik} / C_j) \quad (3.14)$$

Böylece, maksimize edilecek  $P(C_j)$  miktarı Bayes ağına eklenen olasılığa gelmekte olup, aşağıdaki şekilde yapısal olarak sunulmaktadır (Şekil 3.6).



Şekil 3.6. NB temelleri.

Sınıflandırmada, öğrenme aşaması  $P(C_j)$  ve  $p(x_{ik} | C_j)$  farklı olasılıklarının etiketlenmiş bir kökten eğitilmesidir. Test aşaması ise çarpımı maksimize eden  $\hat{C}_j$  değerinin incelenmesidir.

### 3.4.2. NB kullanarak göğüs hastalıklarının teşhisi

NB günümüzde birçok araştırmannın konusu olmuştur. Bu ağlar, metin kategorisinde çıkarım ve öğrenme ile otomatik tıbbi teşhis gibi birçok alanda kendini kanıtlamış yöntemlerdir [66]. Bu nedenle birçok araştırmacı göğüs hastalıkları teşhisinde NB yöntemini kullanma yoluna gitmişlerdir.

Kaynak [67], farklı veri madenciliği teknikleri arasından, astım tahmini için Doğal Bayes ve sinir ağını birleştirmiştir. Göğüs hastanesinden elde edilen 1024 ham veri, veri tabanı olarak kullanılmış, astım teşhisine NB yönteminin etkinliği yazarlar tarafından doğrulanmıştır.

Diğer bir çalışmada, COPD hastalığının tahmininde NB yönteminin kullanılması önerilmiştir [68]. Kullanılan veri tabanı 992 veriden oluşmaktadır. NB modelinde yaş, cinsiyet, ırk, sigara kullanım geçmişi ve 8 farklı ek tanı değişkeni kullanılmıştır. Elde edilen sonuçlar kabul edilebilir sınırlar içindedir ve %83,3 doğruluk oranı bulunmuştur.

NB, yüzyılın hastalığı olan kanserin tahmininde de bazı araştırmacılar tarafından kullanılmıştır [66] [69]. Göğüs kanserinin tahmininde NB kullanılmış ve araştırmalar sonunda elde edilen tüm sonuçların NB'nin doktorlara göğüs kanseri tahmininde yardımcı olabilecek, uygulanabilir bir çözüm olabileceği doğrulanmıştır.

Bu tez çalışmasının da ilgilendiği göğüs hastalıklarından göğüs kanseri teşhisinde NP kullanımını konu alan çalışmalar da bulunmaktadır [69]. Araştırmacılar, göğüs kanseri teşhisinde 100 adet kanserli ve kansersiz hasta verisini kullanarak Basit Bayes algoritması ile akciğer kanseri tespiti yapmışlardır. Göğüs kanseri tahmininde NB tarafından kullanılan öz nitelikler yaş, cinsiyet, hava kirliliği, alkol kullanımı, toz alerjisi, mesleki risk unsurları, genetik risk unsurları, kronik akciğer hastalığı, dengeli diyet, obezite, sigara, pasif sigara içiciliği, göğüs ağrısı, kanlı öksürük, halsizlik, kilo kaybı, nefes darlığı, hırıltılı solunum, yutma güçlüğü, sıklıkla soğuğa maruz kalma, kuru öksürük ve horlamadır. Sınıflandırma sonuçları 100 hastanın

dođru sınıflandırılması ve böylece hastalık tahmininin NB ile etkili bir şekilde yapılabileceđini göstermiştir.

### **3.5. Sonu**

Bu bölümde Bölüm 1’de de belirtilen farklı göđüs hastalıklarını teşhis etmek ve belirlemek için kullanılan farklı ayrıştırıcı sınıflandırma yöntemlerini kapsamaktadır. SVM, ASVM, KNN ve NB olmak üzere dört farklı sınıflandırma yöntemi sunulmuştur. Her yöntemin özellikleri ve göđüs hastalıklarının teşhisinde kullanım farklılıkları tanımlanmıştır.

## **BÖLÜM 4. MAKİNE ÖĞRENMESİ ALGORİTMALARI İLE GÖĞÜS HASTALIKLARI TEŞHİSİ**

### **4.1. Giriş**

Bu bölümde, göğüs hastalıkları teşhisinde tez kapsamında önerilen yaklaşımların tanımlanması amaçlanmaktadır. Öncelikle kullanılan veritabanı, geliştirme ortamı, uygulama koşulları ve performans ölçüm yöntemleri gibi çalışmanın temelleri anlatılmaktadır. Daha sonra, göğüs hastalıkları teşhisinde kullanılan SVM, ASVM, KNN ve NB sınıflandırma yöntemlerinin farklı modelleri tanımlanmaktadır.

### **4.2. Çalışmanın İçeriği**

#### **4.2.1. Geliştirme ortamı**

Uygulama Matlab R2008a üzerinde tamamlanmıştır. Matlab 1970 yılında Cleve Moler tarafından geliştirilen ve bilimsel problemleri çözmek için vektör, matris ve mantıksal işlemlerin geliştirilmesini sağlayan dördüncü nesil bir programlama dilidir.

#### **4.2.2. Kullanılan veritabanı**

Hazırlanan modeller, Türkiye'nin güneyinde yer alan Diyarbakır'da bir hastanenin göğüs hastalıkları departmanından alınan veriseti ile test edilmiştir. Veri tabanında 6 hastalık sınıfı, bu sınıfların belirlenmesinde kullanılacak olan 32 bağımsız değişken ve toplamda 357 adet örneklemden aşağıdaki gibi oluşmaktadır.

- Sınıf 1: Tüberküloz (50)
- Sınıf 2: COPD (71)
- Sınıf 3: Zatürre (60)



- Sınıf 4: Astım (44)
- Sınıf 5: Akciğer kanseri (32)
- Sınıf 6: Sağlıklı (100)

Veriseti; öksürük, yüksek vücut ısısı, göğüs ağrısı, zayıflık, tempolu çalışmada solunum güçlüğü, göğüs hırıltısı, göğüste baskı, balgam, solunum yollarındaki ses, sigara alışkanlığı, akyuvar sayısı (WBC), alyuvar sayısı (RBC), trombosit (PLT), hematokrit (HCT), hemoglobin (HGB), albumin2, alkalın fosfat 2 L, alanin aminotransferaz (ALT), amilaz, aspartat aminotransferaz (AST), bilirubin (toplam + direkt), CK/keratin kinaz toplamı, CK-MB, demir (SERUM), gama-glutamil transferaz (GGT), glukoz, HDL kolektrol, kalsiyum (CA), kandaki üre azotu (BUN), klor (CL), kolektrol, kreatinin, laktik dehidrojenaz (LDH), potasyum (K), sodyum (NA), toplam protein, triglesid, ürik asit olmak üzere toplam 38 farklı özellikten oluşan laboratuvar verilerini kapsamaktadır.

Her özellik için Tablo 4.1.'de gösterildiği gibi minimum ve maksimum veya standart değer olmak üzere iki gösterge verilmektedir.

Tablo 4.1. Veri özellikleri özeti.

Özellikler	Minimum	Standart değer	Maksimum
Öksürük	-	0 or 1	-
Vücut sıcaklığı	-	0 or 1	-
Göğüs ağrısı	-	0 or 1	-
Zayıflık	-	0 or 1	-
Tempolu çalışmada solunum güçlüğü	-	0 or 1	-
Göğüs hırıltısı	-	0 or 1	-
Göğüste baskı	-	0 or 1	-
Balgam	-	0 or 1	-
Solunumda ses	-	0 or 1	-

Tablo 4.1. (Devamı).

Özellikler	Minimum	Standart değer	Maksimum
Sigara		0 or 1	
WBC	4	-	11
RBC	4.2	-	6.2
HCT	35	-	55
PLT	12	-	18
PLT	150	-	400
ALBUMIN 2	3,5	-	5,2
Alkalin Fosfat	38	-	155
ALT	5	-	41
Amilaz	28	-	100
AST	5	-	37
Bilirubin	0	-	1
Keratin Kinaz	15	-	170
CK-MB	6	-	24
Demir	37	-	145
GGT	8	-	61
Glikoz	70	-	110
HDL Kolesterol	35	-	65
Kalsiyum (Ca)	8,4	-	10,8
BUN	13	-	50
Klor (Cl)	96	-	110
Kolesterol	40	-	200
Keratinin	0,4	-	1,5
LDH	240	-	480
Potasyum	3,5	-	5,5
Sodyum (Na)	140	-	157
Protein	6,4	-	8,3
Trigliserid	40	-	200
Ürik asit	0	-	7

### 4.2.3. Performans ölçümü

Sistem performansını değerlendirmede sınıflandırma doğruluğu, değerlendirme ölçütü olarak kullanılmıştır. Sınıflandırma doğruluğu, doğru sınıflandırılmış ögeleri toplam örnek sayısına bölerek elde edilmektedir [19].

## 4.3. Kullanılan Farklı Modeller

Bu bölüm içerisinde SVM, ASVM, KNN ve NB'yi kapsayan farklı sistem mimarileri sunulmaktadır.

### 4.3.1. SVM modeli

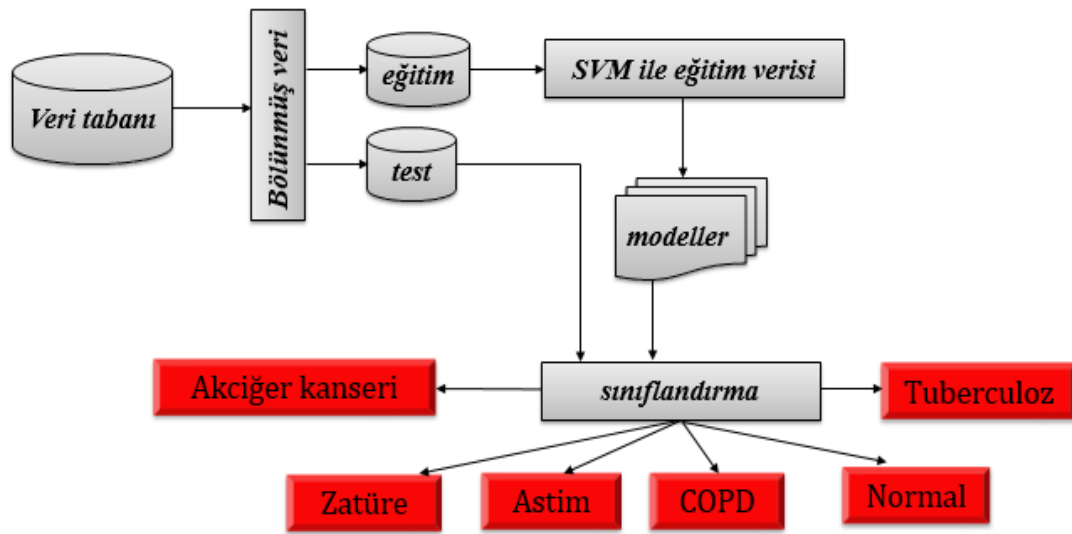
Bölümde kullanılan SVM sistem mimarisinin yanında, bu mimari üzerinde gerçekleştirilen çeşitli testlerin sonuçları gösterilmektedir. Ayrıca, bazı teknik seçimlerin sınıflandırma oranları üzerindeki etkisi de sunulmaktadır.

#### 4.3.1.1. SVM sistem mimarisi

SVM için hangi çekirdeğin seçilmesi gerektiği henüz çözülememiş bir problem olup bu da SVM'nin bir sınırlamasıdır. Genellikle, doğru çekirdeği seçmek için araştırmacılar tarafından bir takım testler yapılmalıdır. Bu test çalışmasında öncelikle doğrusal standart çekirdek kullanılmış olup, sonrasında sistem üzerindeki diğer çekirdeklerin performansı da test edilmiştir.

Tez kapsamında kullanılan SVM yapısı Şekil 4.1.'de sunulmaktadır. Şekilde görüldüğü üzere, veriler öncelikle veri bölme metodu olarak çapraz doğrulama kullanılarak, eğitim ve test olmak üzere iki gruba ayrılmıştır. Çapraz doğrulama metodu Şekil 4.1.'de görüldüğü gibi bir dışsallaştırma (hold-out) yöntemidir. Verilerin iki örnekleme bölünmesiyle oluşturulur. İlk örnekleme eğitim için ayrılırken, ikinci örnekleme test için ayrılmaktadır. Model eğitim örnekleri üzerinde oluşturulur ve test örnekleri ile doğrulanır. Böylece, sınıflandırma işlemi sırasında

SVM, sınıf deęişkenleri ile verilen eğitim verisini öğrenerek referans modeli oluşturmaya başlar. Ardından, test verileri daha önce elde edilen modele dayalı olarak sınıflandırma tahmini yapar. Sonuç olarak, sınıf deęerleri “1” olan hastalar TB’den etkilenen kişiler iken, sınıf deęeri “2” olan kişiler COPD’den etkilenen kişiler olarak belirlenir. Bunun dışında, sınıf deęeri “3” olanlar zatürre, sınıf deęeri “4” olanlar astım, sınıf deęeri “5” olanlar akcięer kanseri ve sınıf deęeri “6” olanlar saęlıklı kişileri tanımlamaktadır.



Şekil 4.1. SVM kullanılarak göęüs hastalıkları teşhis sistemi.

#### 4.3.1.2. SVM sistem sonuçları

Bu tez çalışmasında, SVM çekirdeęinin sınıflandırma hızı ve sistemimiz üzerindeki performansı incelenmiştir. Test edilen farklı çekirdekler; doğrusal çekirdek, RBF çekirdeęi, ikinci dereceden polinom çekirdeęi ve kuadratik çekirdektir. Tablo 4.2.’de, SVM yönteminin tüm göęüs hastalıklarında farklı çekirdekleri kullanarak iyi sonuçlar verdięi açıkça görülebilmektedir.

Çekirdeklerle ilgili olarak ise en iyi sonuçların çoęunluęu çoklu çekirdek tarafından belirlenmiştir. Örneęin çoklu çekirdek %91,87 doğrulukla TB teşhisi yaparken, doğrusal çekirdek %87,08 doğruluęa sahiptir. Astım için çoklu çekirdek %97,23

doğruluğa sahipten, kuadratik çekirdek %93,82 doğruluğa sahiptir. Ayrıca, çoklu çekirdek sağlıklı kişilerin teşhisinde en iyi sonuçları vermiştir.

Araştırmacılar, çekirdek fonksiyonunun seçimlerinin problemlere bağlı olarak değişebileceğini belirtmişlerdir. Dolayısıyla, tıbbi teşhis probleminde gereken çekirdeğin polinom olduğu ortaya çıkmıştır.

Tablo 4.2. SVM üzerindeki çekirdek etkisi.

	Doğrusal	RBF	Çoklu	Kuadratik
Tüberküloz	87,08	90,94	<b>91,87</b>	83,71
COPD	78,65	81,46	<b>90,59</b>	86,52
Zatürre	75,84	85,96	89,58	<b>89,93</b>
Astım	81,46	94,38	<b>97,23</b>	93,82
Akciğer kanseri	<b>96,63</b>	93,26	96,70	94,94
Sağlıklı	90,57	71,91	<b>92,01</b>	90,45

En iyi çekirdeğin polinom olduğunun belirlenmesinden sonra, polinom çekirdeğinin en iyi derecesinin bulunması gerekmektedir. Polinom derecesindeki etkinin sonucu Tablo 4.3.'de verilmiştir.

Tablo 4.3. Polinom çekirdek derecesinin etkisi.

	2	3	4	5	6	7	8	9	10
Tüberküloz	91,87	<b>92,87</b>	87,83	87,78	83,71	88,09	87,73	87,24	87,08
COPD	90,59	<b>92,59</b>	83,71	83,29	82,70	76,03	67,98	61,94	57,24
Zatürre	89,58	<b>91,68</b>	82,96	83,57	83,03	82,96	82,99	83,01	83,02
Astım	97,23	<b>98,23</b>	93,82	93,68	93,37	91,85	90,69	80,90	73,28
Akciğer kanseri	96,70	<b>96,80</b>	96,25	96,79	95,17	94,38	82,18	73,03	65,92
Sağlıklı	92,01	<b>93,01</b>	82,77	90,34	77,87	76,78	76,08	75,56	75,16

Tablo 4.3.'de polinom çekirdeğinin en iyi derecesinin 3 olduğu açıkça gözlemlenmektedir. Göğüs hastalıkları teşhisinde SVM performansının test edilmesi

ve daha önceki çalışmalarla karşılaştırılmasıyla Tablo 4.4. ve Şekil 4.2.'deki sonuçlar elde edilmiştir.

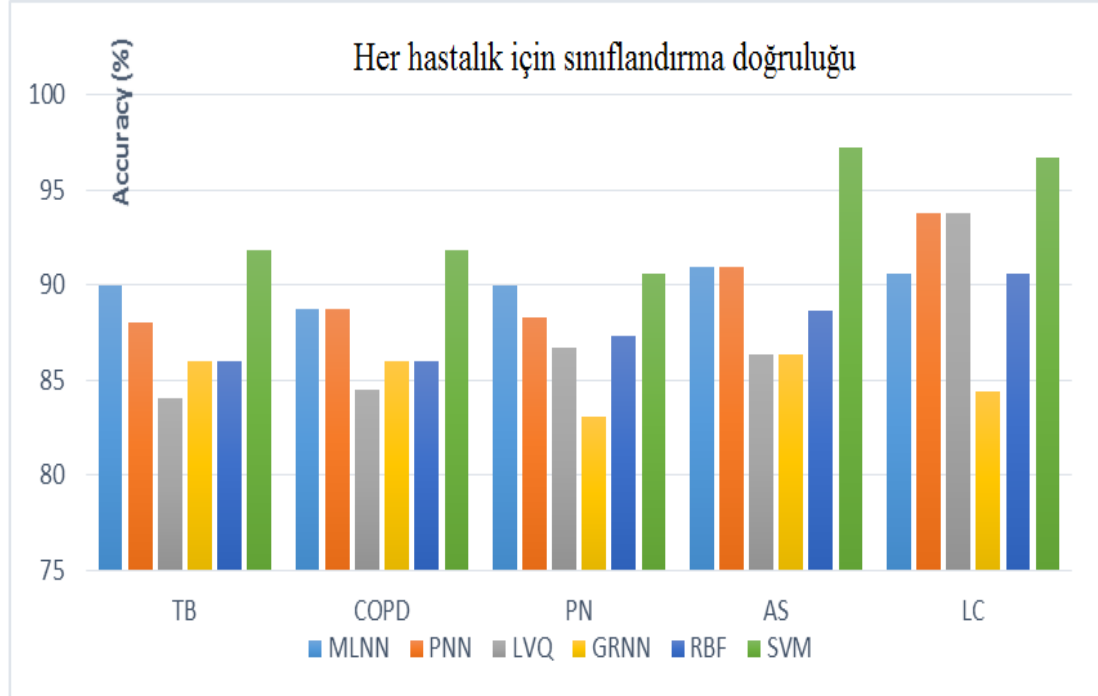
Elde edilen sonuçların analiz edilmesiyle, bu tez çalışmasıyla daha önceki çalışmaların sonuçlarının belirgin şekilde farklı olduğu tespit edilmiştir. Daha önceki sınıflandırma sonuçları; %90 doğrulukla TB, %88,73 doğrulukla COPD teşhisini, iki gizli katmanlı LM ile MLNN metodu kullanarak elde edilmiştir [14]. Bunun yanında zatürre teşhisinde MLNN metodu %90 doğruluk elde etmiştir [14]. Astım teşhisinde PNN metodu kullanılarak %90,91 doğruluk, akciğer kanserinde LVQ kullanılarak %93,75 doğruluk elde edilmiştir. Son olarak yine MLNN kullanılarak sağlıklı insanların %96'sı doğru teşhis edilmiştir. Dolayısıyla, kaynak [14]'de yazarların açıklanabilir sonuçlar ve kabul edilebilir çözümlerle göğüs hastalıklarının teşhisini yapabildikleri fakat daha yüksek doğruluk oranları veremedikleri gözlemlenmiştir.

Tablo 4.4. SVM kullanılarak her hastalık için sınıflandırma doğrulukları.

	MLNN	PNN	LVQ	GRNN	RBF	SVM
Tüberküloz	90,00	88,00	84,00	86,00	86,00	<b>92,87</b>
COPD	88,73	88,73	84,51	83,10	87,32	<b>92,59</b>
Zatürre	90,00	88,33	86,67	88,33	90,00	<b>91,68</b>
Astım	90,91	90,91	86,37	86,37	88,64	<b>98,23</b>
Akciğer kanseri	96,63	93,75	93,75	84,38	90,63	<b>96,80</b>
Sağlıklı	<b>96,00</b>	93,00	94,00	95,00	95,00	93,01

Kullanılan yöntemlere bakıldığında SVM metodunun tüm göğüs hastalıklarında iyi sonuçlar verdiği Tablo 4.4. ve Şekil 4.2. incelendiğinde görülebilmektedir. Yine açıkça görüldüğü üzere SVM yönteminin kaynak [14]'de sunulan sonuçlara kıyasla daha iyi sınıflandırma yaptığı tespit edilmiştir. Örneğin, SVM, TB teşhisinde %92,87 doğruluk, COPD teşhisinde %92,59 doğruluk, astım teşhisinde %98,23 doğruluk ve akciğer teşhisinde %96,80 doğruluk elde etmiştir. Dolayısıyla, Şekil 4.2'de verilen grafikte görüldüğü gibi SVM kullanılarak incelenen beş göğüs hastalığının sınıflandırma oranları, genel olarak önceki yöntemlerle kıyaslandığında daha

yüksektir [14]. Sonuç olarak, SVM'nin farklı sinir ağı türlerini kullanan önceki çalışmalara göre daha uygun olduğu görülmüştür.



Şekil 4.2. Her hastalık için sınıflandırma doğrulukları.

### 4.3.2. ASVM-Modeli

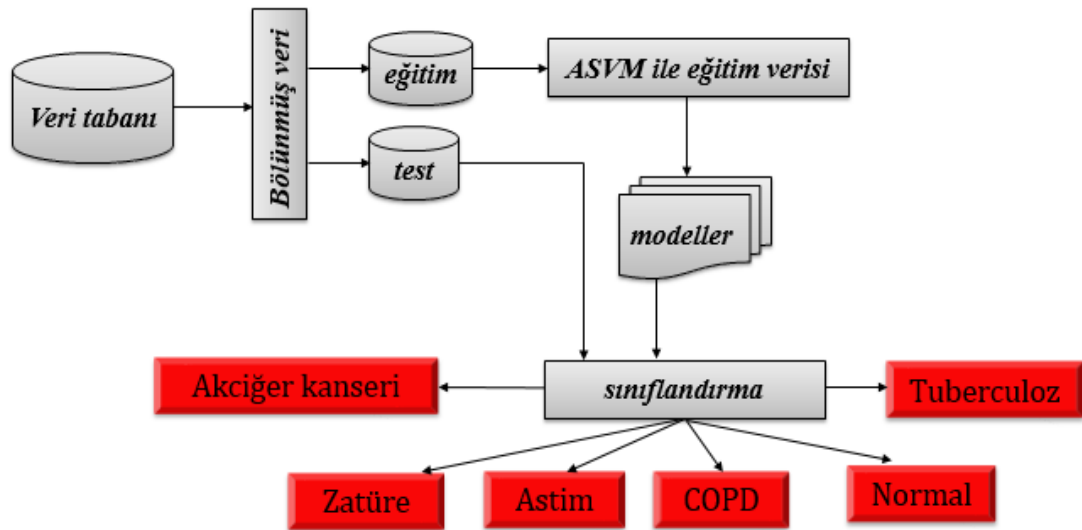
Bu bölümde kullanılan ASVM sistem mimarisi ve bu mimari üzerinde gerçekleştirilen test sonuçları anlatılmaktadır. Ayrıca, SVM ve ASVM karşılaştırması da sunulmaktadır.

#### 4.3.2.1. ASVM Sistem Mimarisi

Tez çalışması kapsamında kullanılan ASVM mimarisi, Şekil 4.3.'de gösterilmektedir. Bir önceki bölümde açıklanan SVM modeli gibi ASVM modeli de Şekil 4.3.'de sunulan prensiple çalışmaktadır. Bu model eğitim ve test aşaması olmak üzere iki kısımda gerçekleştirilmiştir. Eğitim aşamasında, test aşamasında kullanılan modeli oluşturmak için eğitim verileri sınıf değişkenleri ile sisteme verilir.

Bir önceki bölümde açıklandığı gibi, ASVM'nin ana fikri, eğitim aşamasından sonra sınıflandırma işlemini gerçekleştirmek için kullanılacak ideal eşik değerini (bias) bulmaktır. ASVM algoritması kullanılarak en iyi eşik değeri bulunacak ve test aşamasında kullanılacaktır. Test aşamasında test verileri sınıf değişkenleri olmadan verilmekte olup, bu değişkenler tahmin edilmektedir.

Sınıf değerleri "1" olan hastalar TB'den etkilenen kişiler iken, sınıf değeri "2" olan kişiler COPD'den etkilenen kişiler olarak belirlenir. Bunun dışında, sınıf değeri "3" olanlar zatürre, sınıf değeri "4" olanlar astım, sınıf değeri "5" olanlar akciğer kanseri ve sınıf değeri "6" olanlar sağlıklı kişileri tanımlamaktadır.



Şekil 4.3. ASVM kullanılarak göğüs hastalıkları teşhis sistemi.

#### 4.3.2.2. ASVM Sistem Sonuçları

Tüm göğüs hastalıklarında ASVM metodunun en iyi sınıflandırma doğruluğu elde ettiği Tablo 4.5.'de görülebilmektedir. Aslında ASVM, SVM sınıflandırmasının doğruluğunu arttırmaktadır. Örneğin, ASVM kullanılarak TB teşhisinde %93,56 sınıflandırma doğruluğu elde edilirken, SVM kullanılarak bu doğruluk %92,87 olmaktadır. Bunun yanında ASVM ile COPD teşhisinde %92,63 başarı elde edilirken, SVM ile bu başarı %92,59'dur. ASVM, astım teşhisinde de %98,45 başarı ile %98,23 orana sahip



SVM'ye kıyasla daha iyi performans sergilemiştir. Bu sonuçlar, göğüs hastalıklarının ilk teşhisinde ASVM kullanmanın diğer yöntemler ile karşılaştırıldığında daha iyi bir metot olduğunu kanıtlamaktadır.

Tablo 4.5. ASVM kullanarak her hastalık için sınıflandırma doğrulukları.

Metot	Doğruluk					
	Tüberküloz	COPD	Zatürre	Astım	A.kanseri	Sağlıklı
SVM	92,87	92,59	91,68	98,23	96,80	93,01
ASVM	<b>93,56</b>	<b>92,63</b>	<b>91,70</b>	<b>98,45</b>	<b>97,80</b>	<b>98,11</b>

ASVM'nin başarısı, SVM'de kullanılan standart eşik değerinin uyarlanan bir değer ile değiştirilmesinden kaynaklanmaktadır. Böylece, ASVM kullanılarak elde edilen sınıflandırma doğruluğu hem standart SVM kullanılarak elde edilen başarıdan, hem de daha önceki metotların (MLNN, PNN, LVQ, GRNN ve RBF) başarısından belirgin bir şekilde daha yüksek olmaktadır. Bu sonuçlar, eşik değeri eklenmesiyle hazırlanan ASVM metodunun, göğüs hastalıkları teşhisinde sınıflandırma doğruluğunu arttırması nedeniyle avantajını göstermiştir.

### 4.3.3. KNN-Modeli

Bu bölümde, kullanılan KNN sistem mimarisi ve üzerinde gerçekleştirilen test sonuçları anlatılmaktadır. Ayrıca, sınıflandırma oranında K-en yakın komşu sayısının etkisi de incelenmiştir.

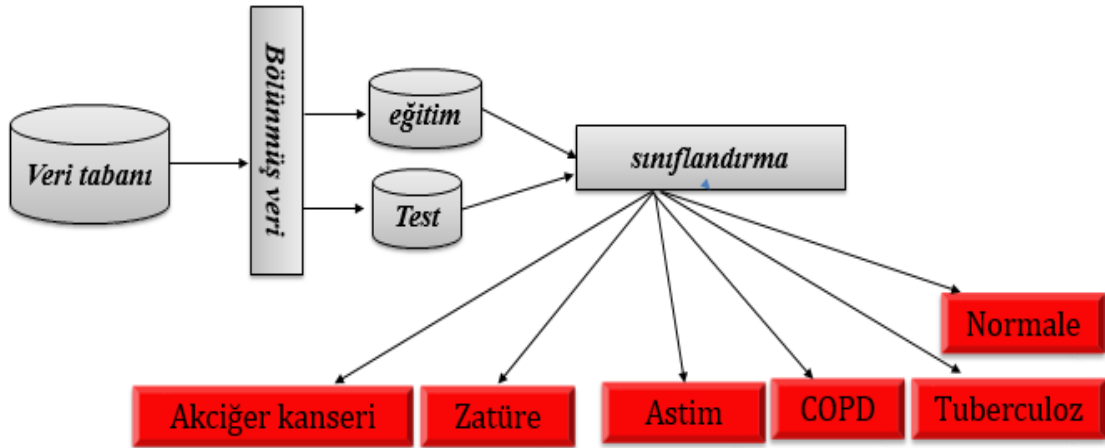
#### 4.3.3.1. KNN Sistem Mimarisi

Tez kapsamında kullanılan KNN mimarisi Şekil 4.4.'de sunulmaktadır. En yakın komşu sınıflandırmasının kullanımı ayırık analiz kapsamında değerlendirilmektedir. Tıbbi teşhis alanında KNN algoritması başarılı bir şekilde uygulanmasına rağmen en yakın komşu sayısının (K) belirlenmesi hala başlı başına bir problem olarak görülmektedir. Bu sayının (K) seçimi bir takım deneysel çalışmalar ile tespit

edilebilmektedir. Literatürde K'nın en iyi değerinin genellikle sınıfların sayısının kareköküne yakın değerler olduğu belirtilmiştir [70].

K-NN metodu çok basittir. Tüm yeni girdilerin K-en yakın komşularına Öklid uzaklıkları hesaplanarak elde edilen, doğrudan bir sınıflandırma yöntemidir. Tez kapsamında en iyi K sayısının belirlenmesine çalışılmıştır.

Şekil 4.4.'de sunulduğu üzere, eğitim ve test olmak üzere veritabanını ikiye ayırdıktan ve K sayısını seçtikten sonra, model olmadan doğrudan bir sınıflandırma yapılmıştır. Buna göre, sınıf değerleri "1" olan hastalar TB'den etkilenen kişiler iken, sınıf değeri "2" olan kişiler COPD'den etkilenen kişiler olarak belirlenir. Bunun dışında, sınıf değeri "3" olanlar zatürre, sınıf değeri "4" olanlar astım, sınıf değeri "5" olanlar akciğer kanseri ve sınıf değeri "6" olanlar sağlıklı kişileri tanımlamaktadır.



Şekil 4.4. KNN kullanarak göğüs hastalıkları teşhis sistemi.

#### 4.3.3.2. KNN sistem sonuçları

Göğüs hastalıkları teşhisinde KNN'nin K=10 ile testinden elde edilen sonuçlar Tablo 4.6.'da sunulmuştur.

Sonuçlar elde edildikten sonra analiz edilmiş ve SVM ile ASVM tarafından elde edilen önceki sonuçlarla karşılaştırılmıştır. Bu testte elde edilen sonuçların çok düşük olduğu anlaşılmıştır. KNN ile elde edilen en iyi sonuç %61 doğrulukla tüberküloz teşhisidir.

Tablo 4.6. KNN kullanarak her hastalık için sınıflandırma doğrulukları.

		Doğruluk				
Metot	Tüberküloz	COPD	Zatürre	Astım	A.kanseri	Sağlıklı
KNN	61,00	57,00	<b>57,03</b>	52,08	59,80	48,01

Elde edilen sonuçları iyileştirmek için en yakın komşu sayısının sınıflandırma hızı ve sistem performansı üzerindeki etkisi incelenmiştir. İnceleme sonuçları Tablo 4.7.'de sunulmaktadır. Bir önceki bölümde açıklandığı gibi K'nın en iyi değeri, sınıf sayısının kareköküne yaklaşan değeridir. Bu tez kapsamında 6 sınıf mevcuttur. Bu yüzden KNN, üçe yaklaşan değerleri ile test edilmiştir (6'nın karekökü). Test edilen en yakın komşu sayısı sırasıyla 1, 2, 3, 4, 5, 6'dır.

Tablo 4.7. En yakın komşu sayısının etkisi.

	1	2	3	4	5	6
Tüberküloz	60,00	61,00	<b>75,00</b>	74,58	53,00	53,01
COPD	45,00	46,58	<b>56,74</b>	55,88	53,28	53,00
Zatürre	58,78	59,25	<b>59,30</b>	59,10	55,86	54,12
Astım	44,56	44,65	<b>44,50</b>	44,51	42,15	41,89
Akciğer kanseri	60,00	60,01	<b>61,25</b>	59,12	59,00	45,62
Sağlıklı	58,20	59,36	<b>60,48</b>	61,00	48,55	48,50

Tablo 4.7.'de görüldüğü üzere, sonuçlar hala düşüktür ve daha önceki metotlar ile karşılaştırıldığında kabul edilebilir değildir. Ayrıca, en iyi K-en yakın komşu görüldüğü üzere üçtür. Teşhis doğruluğu K=1'den, K=3 değerine kadar artmakta, ardından K=4, 5, 6 için azalmaktadır. Örneğin K=2 için tüberküloz teşhisinde %61 doğruluk elde edilmiş olup, K=3 için bu değer %75'dir. Fakat K=4 için aynı hastalık %74,58 doğruluk oranı ile azalmış ve bu azalma K=6 için %58,01'e düşmüştür. Elde edilen sonuçlar KNN'nin göğüs hastalıkları teşhisinde iyi bir çözüm olamayacağını ve performansının diğer yöntemlerden daha düşük olduğunu göstermektedir.

#### **4.3.4.NB-Modeli**

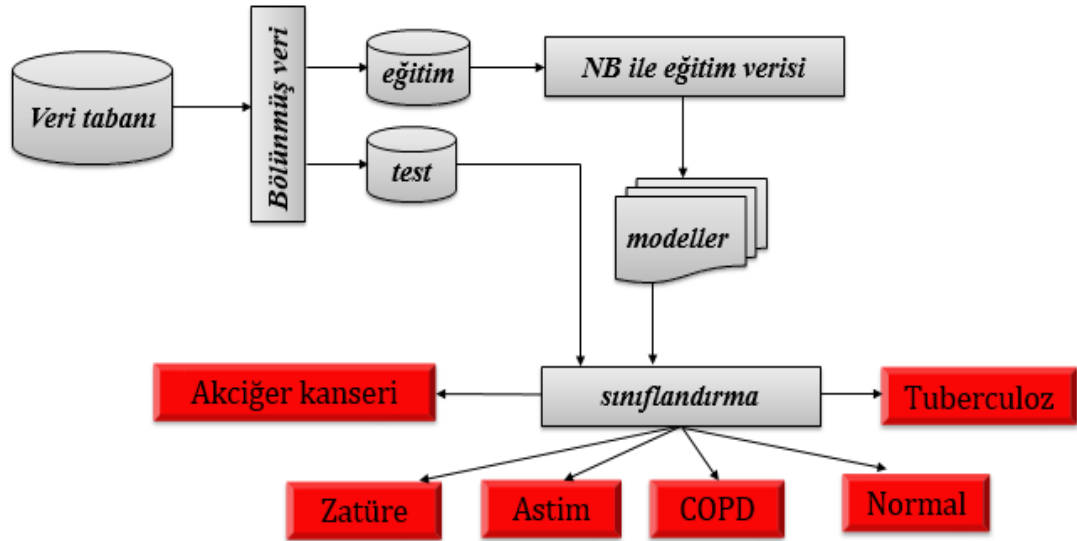
Bu bölümde kullanılan NB sistem mimarisi ve diğer sınıflandırıcılar gibi NB sınıflandırıcısı üzerinde gerçekleştirilen testler ve test sonuçları sunulmaktadır.

##### **4.3.4.1. NB Sistem Mimarisi**

Diğer sınıflandırıcılar gibi NB sınıflandırıcısının da mimarisi sunulmaktadır. Şekil 4.5, NB göğüs hastalıkları teşhis sisteminin yapısını göstermektedir.

SVM sisteminde olduğu gibi, NB sistemi için de eğitim aşamasında bir model oluşturulmuştur. Test aşaması sırasında sınıflandırma yapılmış olup, sınıflar tahmin edilmeye çalışılmıştır.

Sınıf değerleri "1" olan hastalar TB'den etkilenen kişiler iken, sınıf değeri "2" olan kişiler COPD'den etkilenen kişiler olarak belirlenir. Bunun dışında, sınıf değeri "3" olanlar zatürre, sınıf değeri "4" olanlar astım, sınıf değeri "5" olanlar akciğer kanseri ve sınıf değeri "6" olanlar sağlıklı kişileri tanımlamaktadır.



Şekil 4.5. NB kullanılarak göğüs hastalıkları teşhis sistemi.

#### 4.3.4.2. NB Sistem Sonuçları

NB sistemi kullanılarak elde edilen sonuçlar, Tablo 4.8.'de sunulmaktadır. Sonuçların SVM veya ASVM kullanılarak elde edilen sonuçlardan daha düşük olduğu, ancak KNN kullanılarak elde edilen sonuçlardan daha yüksek olduğu görülmüştür. Örneğin, NB, TB hastalığı için %73,06 doğruluk, COPD için %77,00, astım için %73,06 ve akciğer kanseri için %71,80 doğruluk elde etmiştir.

Tablo 4.8. Her hastalık için sınıflandırma doğrulukları.

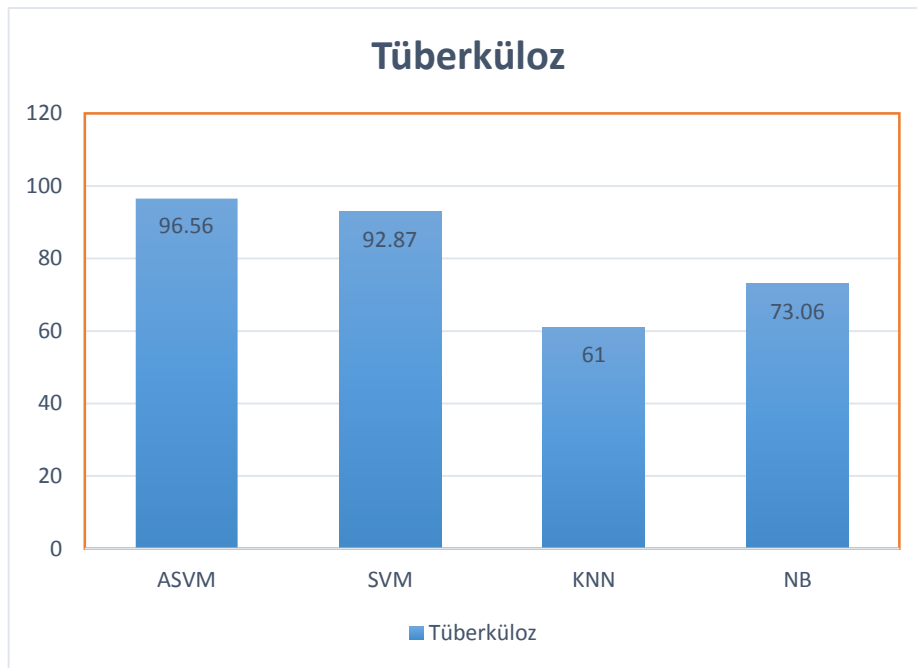
Metot	Doğruluk					
	Tüberküloz	COPD	Zatürre	Astım	A.kanseri	Sağlıklı
NB	73,06	77,00	65,00	73,06	71,80	68,01

Genel olarak özetlenecek olursa, NB kullanılarak elde edilen sınıflandırma oranları %65,00 ve %77 arasında olup, düşüktür.

#### 4.4. Farklı sınıflandırma sistemlerinin karşılaştırması

Bu bölümde, her hastalık için farklı sınıflandırıcılar test edilmiş ve elde edilen sonuçlar karşılaştırılmıştır.

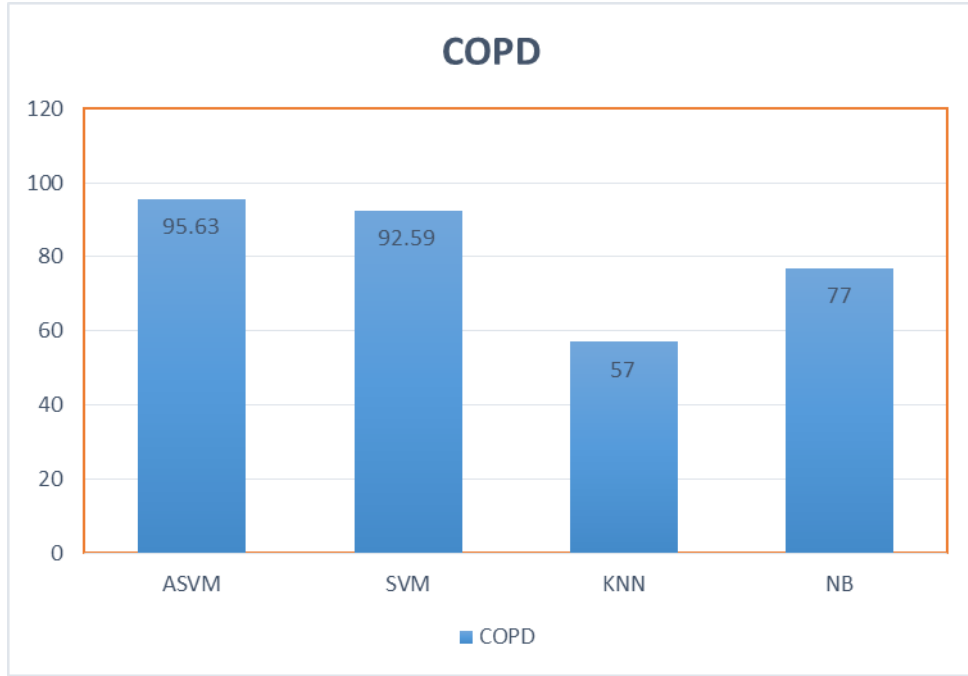
İlk önce, tüberküloz hastalığının farklı sınıflandırıcılar ile teşhisi karşılaştırmalı olarak sunulacaktır. Şekil 4.6.'da görüldüğü üzere tüberküloz teşhisinde KNN ve NB sırasıyla %61 ve %73,6 doğruluk oranları ile en düşük sonuçları vermiştir. En iyi teşhis doğruluk oranı ASVM ile %96,56 olarak elde edilmiş olup, bunu SVM %92,87 ile takip etmiştir. Bu sonuçlarda görüldüğü üzere, uyarlamalı terim olan eşik değeri (Bias – B) kullanılması ASVM'nin SVM'den daha iyi sonuçlar vermesini sağlamıştır.



Şekil 4.6. Tüberküloz için farklı sınıflandırma sistemlerinin karşılaştırılması.

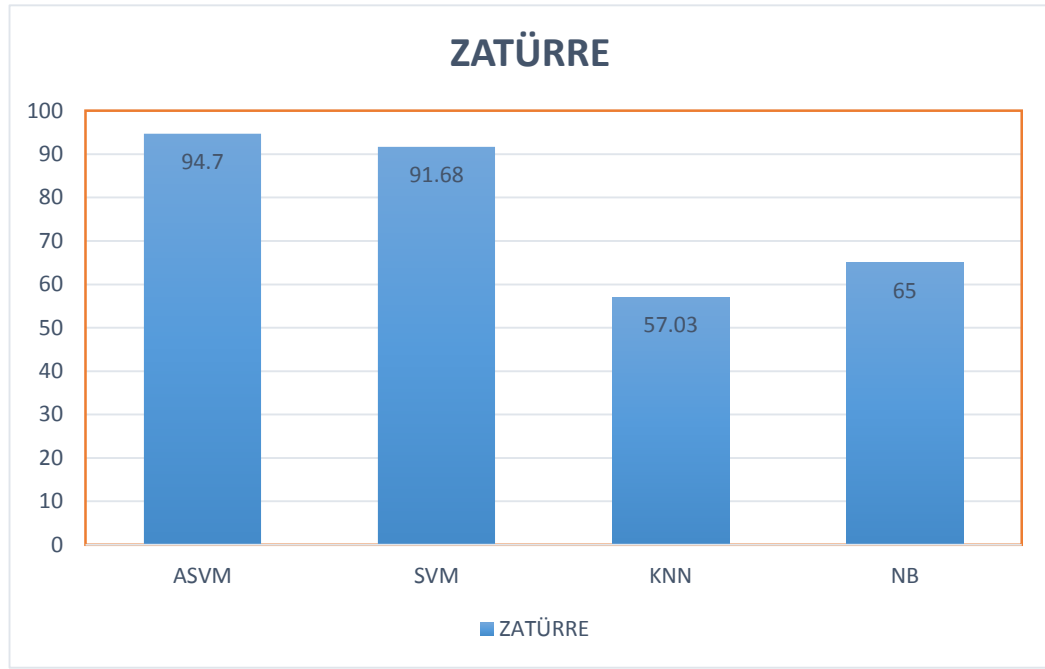
Tüberküloz için elde edilen sonuçların karşılaştırılmasından sonra COPD hastalığının farklı sınıflandırıcılarda elde edilen sonuçları anlatılacaktır. Şekil 4.7.'de görüldüğü üzere ASVM, %92,59 teşhis doğruluk oranına sahip SVM'yi, %95,63 teşhis doğruluk oranıyla geliştirmiştir. Ayrıca NB, %77 teşhis doğruluk

oranıyla ortalama bir sonuca sahip olup KNN ise %57 oranla en düşük teşhis doğruluğuna sahiptir.



Şekil 4.7. COPD için farklı sınıflandırma sistemlerinin karşılaştırılması.

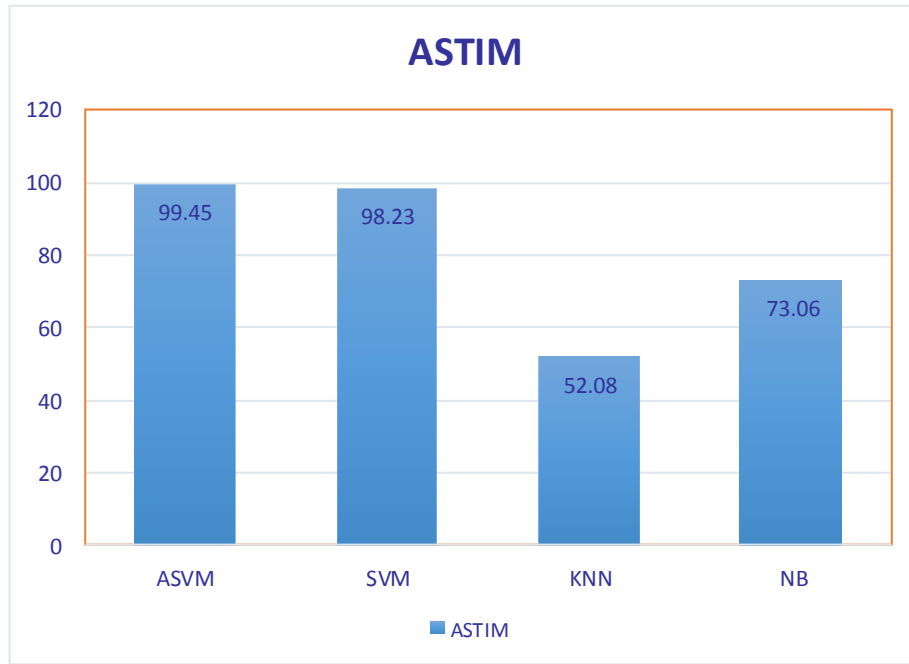
Tüberküloz ve COPD için elde edilen sonuçların karşılaştırılmasının ardından, Şekil 4.8.'de zatürre için elde edilen sonuçlar sunulmaktadır. Daha önceki sonuçlarda da görüldüğü gibi, ASVM, %91,68 doğruluk oranına sahip SVM'yi, %94,7 doğruluk değeri ile geliştirmiş ve en iyi sonucu vermiştir. ASVM ve SVM sonuçları çok iyi ve kabul edilebilirdir. Bu bakımdan hastalıkların tahmininde doktorların karar almalarında yardımcı olarak kullanılabilirler. KNN yöntemi, diğer hastalıklarda olduğu gibi zatürre teşhisinde düşük sonuçlar vermektedir. Bu, bir eğitim modeli kullanmadan, girdiği doğrudan sınıflandıran bir sınıflandırma stratejisi kullanılmasıyla açıklanabilir.



Şekil 4.8. Zatürre için farklı sınıflandırma sistemlerinin karşılaştırılması.

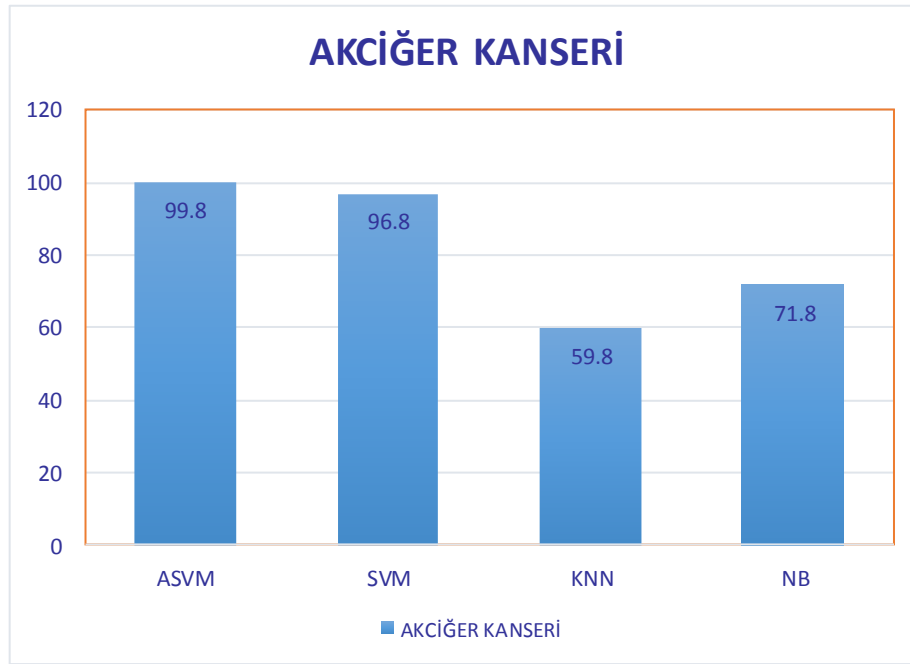
Şekil 4.9.'da elde edilen sonuçlar incelendiğinde astım hastalığının teşhisinde en iyi doğruluk oranı yaklaşık %100 ile ASVM tarafından ulaşılmıştır. Bu sonuç ASVM'nin tıbbi teşhislerde kullanılabilir bir yöntem olduğunu ve sadece göğüs hastalıkları değil, diğer hastalık türleri için de kullanılabilirliğini göstermektedir. Benzer şekilde ASVM tarafından elde edilen sonuçlara yakın sonuçlar veren SVM için de bu durum geçerlidir. Astım hastalığı için KNN ve NB metotları sırasıyla %52,8 ve %73,06 en düşük sonuçları verdiklerinden kullanılabilir çözüm olmadıkları belirlenmiştir.





Şekil 4.9. Astım için farklı sınıflandırma sistemlerinin karşılaştırılması.

Son olarak, kullanılan farklı sınıflandırma metotları için akciğer kanseri hastalığı teşhisinde elde edilen sonuçlar incelenmektedir. Şekil 4.10.'da ASVM yöntemi rekabet gücünü bu hastalık teşhisinde de kanıtlamış olup %99,8'lik bir doğruluk oranı ile en iyi sonuçları vermiştir. Bu yöntemi %96,8'lik teşhis doğruluğu ile SVM takip etmiştir.



Şekil 4.10. Akciğer kanseri için farklı sınıflandırma sistemlerinin karşılaştırılması.

## **BÖLÜM 5. TARTIŞMA VE ÖNERİLER**

Bu çalışmada, göğüs hastalıklarını teşhis etmek için destek vektör makinaları (SVM), K-en yakın komşular (K-NN) ve Basit Bayes sınıflandırma (NB) metotları kullanılmıştır. Ayrıca, ilk kez uyarlamalı destek vektör makina metodunun (ASVM) göğüs hastalıkları teşhisindeki performansı da değerlendirilmiştir. Bu bölümde, tez kapsamında oluşturulan farklı sistemlerin performanslarını değerlendirmek için yapılan deneysel içerikler sunulmaktadır. Gerçekleştirilen deneyler çeşitlilik göstermektedir. Öncelikle, SVM sistemi, ardından ASVM, KNN ve NB sistemleri değerlendirilmiş olup bu farklı metotların performanslarının karşılaştırılması bu bölüm içerisinde gerçekleştirilmiştir.

TB, COPD, zütürre, astım ve akciğer kanseri gibi göğüs hastalıkları solunum yollarını tehdit eden hastalıklar arasındadır. Günümüzde bir çok araştırmacı, bu tür hastalıkları saptamak ve teşhis etmek için kullanılabilir etkili yöntemleri araştırmaktadır. Yapılan çalışmalar ve araştırmalar doktorların daha kolay ve doğru teşhis koymalarını yardımcı olabilecek etkili karar destek sistemleri üzerine yoğunlaşmaktadır. Bu kapsamda, uzman sistemler ve farklı yapay zeka teknikleri tıp alanı dahil olmak üzere farklı alanlar üzerindeki etkinliklerini ve doğruluklarını kanıtlamaya devam etmektedir. Bu tez çalışmasında, göğüs hastalıklarının teşhisinde SVM, ASVM, KNN ve NB olmak üzere bazı yapay zeka teknikleri kullanılmıştır. Kullanılan bu teknikler, tanıtılmış, açıklanmış, test edilmiş ve daha önceki diğer teknikler ile karşılaştırılmıştır. Bu çalışma, ASVM'yi kullanarak göğüs hastalıklarının teşhisi üzerine yoğunlaşan ilk çalışmadır. Diyarbakır Göğüs Hastalıkları Hastanesi'nden (Türkiye'nin güneydoğusu) toplanan deneysel verileri kullanarak SVM ve ASVM sınıflandırıcılarının etkinliği doğrulanmıştır. Burada SVM, göğüs hastalıklarının teşhis probleminde çok iyi ve kabul edilebilir sonuçlar ortaya koymuştur. ASVM ise daha önceki çalışmalarla karşılaştırıldığında elde ettiği

sonularla en iyi yntem olduėunu kanıtlamıřtır. Genel olarak zetlemek gerekirse, gės hastalıkları teřhisinde ASVM yntemi bařarılı bir řekilde kullanılabilir olup, doktorlar tarafından hastalıkla ilgili kararlar almak iin tercih edilebilir.

Bu tezin devamı olarak gelecekte ařaėıdaki konularda arařtırma alıřmaları yapılabilir:

1. SVM, ASVM, KNN ve NB yntemlerinin farklı olası kararlarının paralel deėerlendirmesi.
2. Veri tabanında kullanılan zelliklerin elde edilen sonulara etkisinin arařtırılması
3. nerilen yntemin farklı hastanelerdeki, farklı ve daha byk veri tabanında test edilmesi.

## KAYNAKLAR

- [1] [https://en.wikipedia.org/wiki/Medical\\_diagnosis.](https://en.wikipedia.org/wiki/Medical_diagnosis.), Erişim Tarihi: 11.09.2016.
- [2] <https://fr.wikipedia.org/wiki/Médecine> ., Erişim Tarihi: 11.08.2016.
- [3] [www.who.int/patientsafety/topics/primary-care/technical\\_series/en/.](http://www.who.int/patientsafety/topics/primary-care/technical_series/en/), Erişim Tarihi: 11.09.2016.
- [4] Woolever, D. R; The impact of a patient safety program on medical error reporting.2005.
- [5] [www.sciencedaily.com/releases/2016/05/160504085309.htm](http://www.sciencedaily.com/releases/2016/05/160504085309.htm) ./, Erişim Tarihi: 11.09.2016.
- [6] [www.usnews.com/news/articles/2016-05-03/medical-errors-are-third-leading-cause-of-death-in-the-us](http://www.usnews.com/news/articles/2016-05-03/medical-errors-are-third-leading-cause-of-death-in-the-us) ., Erişim Tarihi: 11.10.2016.
- [7] [https://en.wikipedia.org/wiki/Medical\\_error.](https://en.wikipedia.org/wiki/Medical_error.), Erişim Tarihi: 12.10.2016.
- [8] Castaneda, C., Nalley, K., Mannion, C., Bhattacharyya, P., Blake, P., Pecora, A, Suh, K. S., Clinical decision support systems for improving diagnostic accuracy and achieving precision medicine. *Journal of clinical bioinformatics*, 5(1), 4, 2015.
- [9] Pannu, A., Artificial intelligence and its application in different areas. *Artificial Intelligence*, 4(10), 2015.
- [10] Angeli, C., Diagnostic expert systems: From expert's knowledge to real-time systems. *Advanced knowledge based systems: Model, applications & research*, 1: 50-73, 2010.
- [11] Er, O., Yumuşak, N., Temurtas, F., Chest disease diagnosis using artificial neural networks. *Expert Syst Appl.*, 37(12): 7648-7655, 2010.
- [12] [//en.wikipedia.org/wiki/Chest\\_\(disambiguation\).](https://en.wikipedia.org/wiki/Chest_(disambiguation).), Erişim Tarihi: 11.10.2016.
- [13] <https://en.wikipedia.org/wiki/Pulmonology.>, Erişim Tarihi: 11.10.2016.

- [14] Er, O., Temurtas, F., Tanrikulu, A. Ç., Tuberculosis disease diagnosis using artificial neural networks. *Journal of Medical Systems.*, 34(3): 299-302, 2010.
- [15] [www.who.int/tb/publications/global\\_report/en/](http://www.who.int/tb/publications/global_report/en/), Erişim Tarihi: 11.10.2016.
- [16] <http://www.medindia.net/patients/patientinfo/tuberculosis.htm>., Erişim Tarihi: 11.10.2016.
- [17] Ducati, D.G., Netto, A.R., Basso, L.A, Santos, S., The resumption of consumption: a review on tuberculosis. *Memórias do Instituto Oswaldo Cruz.*, 101(7): 697-714, 2006.
- [18] Panicker, R.O., Soman, B., Saini, G., Rajan, J., A review of automatic methods based on image processing techniques for tuberculosis detection from microscopic sputum smear images. *Journal of Medical Systems.*, 40(1): 17, 2016.
- [19] Asha, T., Natarajan, s., Murthy, K.N., A data mining approach to the diagnosis of tuberculosis by cascading clustering and classification. *Journal of Computing.*, 3(4): 1-8, 2011.
- [20] Sivaranjani, S., Analysis of Tuberculosis in chest using SVM classifier. *National Conference On Research Advance in Communication, Computation, Electrical Science and Structures, India*, 18-21, 2015.
- [21] [www.msf.fr/](http://www.msf.fr/)., Erişim Tarihi: 11.10.2016.
- [22] Cheplygina, V., Sørensen, L., Tax, D.M.J., Pedersen, J.H., Loog, M., Bruijne, M.D., Classification of COPD with multiple instance learning. *Pattern Recognition (ICPR), 2014 22nd International Conference on IEEE, Sweden*, 1508-1513, 2014.
- [23] Afonso, A.SM., Verhamme, K.MC., Sturkenboom, M. CJM., Brusselle. G.G.O., COPD in the general population: prevalence, incidence and survival. *Respiratory medicine.*, 105(12): 1872-1884, 2011.
- [24] [www.who.int/respiratory/copd/en/](http://www.who.int/respiratory/copd/en/)., Erişim Tarihi: 15.11.2016.
- [25] <https://healdove.com/disease-illness/How-Do-You-Know-When-You-Are-In-COPD-Crisis>., Erişim Tarihi: 11.10.2016.
- [26] [www.atousante.com](http://www.atousante.com)., Erişim Tarihi: 16.11.2016.

- [27] Er, O., Temurtas, F., Tanrikulu, A. Ç., A comparative study on chronic obstructive pulmonary and pneumonia disease diagnosis using neural networks and artificial immune system. *Journal of Medical System.*, 33(6): 33-36, 2009.
- [28] [www.who.int/mediacentre/factsheets/fs331/en/](http://www.who.int/mediacentre/factsheets/fs331/en/), Erişim Tarihi: 16.11.2016.
- [29] <http://www.mayoclinic.org/diseases-conditions/pneumonia/symptoms-causes/dxc-20204678> , Erişim Tarihi: 11.10.2016.
- [30] Oliveira, L., Nouer, S., Ribeiro, L., Andrade, A., Computer-aided diagnosis in chest radiography for detection of childhood pneumonia. *Int J med Inform.*, 77(8): 555-564, 2008.
- [31] [www.nhlbi.nih.gov/health/health-topics/topics/pnu/diagnosis](http://www.nhlbi.nih.gov/health/health-topics/topics/pnu/diagnosis)., Erişim Tarihi: 16.11.2016.
- [32] [www.mayoclinic.org/diseasesconditions/pneumonia/basics/treatment/](http://www.mayoclinic.org/diseasesconditions/pneumonia/basics/treatment/), Erişim Tarihi: 16.11.2016.
- [33] [www.who.int/mediacentre/factsheets/fs297/en/](http://www.who.int/mediacentre/factsheets/fs297/en/), Erişim Tarihi: 16.11.2016.
- [34] Bhuvanewari, P., Therese, A. B., Detection of Cancer in Lung with K-NN Classification Using Genetic Algorithm. *Procedia Materials Science.*, 10: 433-440, 2010.
- [35] <https://www.mountelizabeth.com.sg/facilities-services/cent-excellence/cancer/lung-cancer>., Erişim Tarihi: 11.10.2016.
- [36] [www.cancerresearchuk.org/about-cancer/lung-cancer/stages-typesgrades/m-types](http://www.cancerresearchuk.org/about-cancer/lung-cancer/stages-typesgrades/m-types) ., Erişim Tarihi: 23.11.2016.
- [37] Swati, P.T., Vrishali A.C., Classification of lung tumor using SVM. *International Journal of Computational Engineering Research.*, 2(5): 1254-1257, 2012.
- [38] [www.lungcancer.org/](http://www.lungcancer.org/), Erişim Tarihi: 26.11.2016.
- [39] [www.cancer.org](http://www.cancer.org) ., Erişim Tarihi: 26.11.2016.
- [40] [www.cancerresearchuk.org/about-cancer/lung-cancer/treatment](http://www.cancerresearchuk.org/about-cancer/lung-cancer/treatment)., Erişim Tarihi: 26.11.2016.

- [41] Prasad, B.D.C.N., Prasad, P. E. S. N., Sagar Y., An approach to develop expert systems in medical diagnosis using machine learning algorithms (asthma) and a performance study. *International Journal on Soft Computing.*, 2(1): 26-33, 2011.
- [42] [www.who.int/topics/asthma/en/](http://www.who.int/topics/asthma/en/), Eriřim Tarihi: 26.11.2016.
- [43] <https://sensiseeds.com/en/blog/top-6-benefits-of-cannabis-for-asthma/>, Eriřim Tarihi: 11.10.2016.
- [44] [www.who.int/mediacentre/factsheets/fs307/en/](http://www.who.int/mediacentre/factsheets/fs307/en/), Eriřim Tarihi: 27.11.2016.
- [45] Cortes, C., Vapnik, V., Support-vector networks. *Machine Learning*, 273-297, 1995.
- [46] Veropoulos, K., Campbell, C., Cristianini, N., Controlling the sensitivity of support vector machines. *Proceeding of the International Joint Conference on AI, Sweden*, 55-60, 1999.
- [47] Osowski, S., Siwek, K., Markiewicz, T., MLP and SVM networks-a comparative study. *Proceedings of the 6th Nordic Signal Processing Symposium–NORSIG, Finland*, 9-11, 2004.
- [48] Osowski, S., Tran, L., Markiewicz, T., Support vector machine-based expert system for reliable heartbeat recognition. *IEEE Transactions on Biomedical Engineering.*, 51(4): 582-589, 2004.
- [49] Comak, E., Arslan, A., Turkoglu, I., A decision support system based on support vector machines for diagnosis of the heart valve diseases. *Computers in Biology and Medicine.*, 37(1): 21-27, 2007.
- [50] Meier, T. B., Desphande, A. S., Vergun, S., Prabhakaran, V., Support vector machine classification and characterization of age-related reorganization of functional brain networks. *NeuroImage.*, 60(1): 601-613, 2012.
- [51] Osuna, E., Freund, R. M., Girosi, F., Training support vector machines: an application to face detection. *Computer vision and pattern recognition, 1997. Proceedings., 1997 IEEE computer society conference on, Puerto Rico*, 130-136, 1997.
- [52] Trabelsi, I., Ben Ayed, D., On the use of different feature extraction methods for linear and nonlinear kernel. *2012 6th International Conference on Sciences of Electronics, Technologies of Information and Telecommunications (SETIT), Tunisia*, 2012.



- [53] Bhatia, S., Prakash, P., Pillai, G. N., Training support vector machines: SVM based decision support system for heart disease classification with integer-coded genetic algorithm to select critical features. Proceedings of the World Congress on Engineering and Computer Science, USA, 34-38, 2008.
- [54] Djeflal, A., Utilisation de methode support vector machines dans l'analyse des bases de donnees. Mohamed Khider-Biskra University, Biskra, Algeria, PhD thesis, 1999.
- [55] Sun, T., Wang, J., Li, X., Guo, X., Comparative evaluation of support vector machines for computer aided diagnosis of lung cancer in CT based on a multi-dimensional data set. Computer Methods and Programs in Biomedicine., 111(2): 519-524, 2013.
- [56] Sousa, R. T., Marques, O., Soares, F., Spoto, E. S., Comparative performance analysis of machine learning classifiers in detection of childhood pneumonia using chest radiographs. Procedia Computer Science., 18: 2579-2582, 2013.
- [57] Chatzimichail, E., Paraskakis, E., Sitzimi, M., Rigas, A., An intelligent system approach for asthma prediction in symptomatic preschool children. Computational and Mathematical Methods in Medicine., 2013 (2013): 1-6, 2013.
- [58] Gürbüz, E., Kılıç, E., A new adaptive support vector machine for diagnosis of diseases. Expert Systems., 31(5): 389-397, 2013.
- [59] Soltani, T., Langarizadeh, M., Zolnoori, M., Intelligent diagnosis of asthma using machine learning algorithms. International Research Journal of Applied and Basic Sciences., 5(1): 140-145, 2015.
- [60] Peyman, R.H., taha, S.S., Ruhollah, k., Gheibi, M., Moftian, N., Prediction of Asthma control levels using data mining methods: an evidence-based approach. Bmj open., 7(1), 192, 2017.
- [61] Langarizadeh, M., Moghbeli, F., Applying naive bayesian networks to disease prediction: a systematic review. Acta informatica medica., 24(5): 364, 2016.
- [62] Medhekar, D.S., Bote, M. P., Deshmukh, S. D., Heart disease prediction system using naive bayes. International journal of enhanced research in science technology & engineering., 2(3), 2013.
- [63] Subbalakshmi, G., Ramesh, K., Rao, M.C., Decision support in heart disease prediction system using naive bayes. Indian journal of computer science and engineering., 2(2): 170-176, 2011.

- [64] Jamoussi, S., Une nouvelle représentation vectorielle pour la classification sémantique. *Apprentissage automatique par le TAL.*, 50: 23-57, 2009.
- [65] Rashmi, M., Usha, K. P., Cancer diagnosis using naive bayes algorithm. *International journal of recent trends in engineering & research.*, 2(5): 376-380, 2016.
- [66] Aneja, S., Lal, S., Effective asthma disease prediction using naive Bayes-Neural network fusion technique. *Parallel, Distributed and Grid Computing (PDGC), 2014 International Conference on IEEE*, 137-140, 2014.
- [67] Himes, B.E., DAI, Y., Kohane, I. S., Scott T. W., Marco F. R., Prediction of chronic obstructive pulmonary disease (COPD) in asthma patients using electronic medical records. *Journal of the American Medical Informatics Association.*, 16(3): 371-379, 2009.
- [68] Kharya, S., Soni, S., Weighted Naive Bayes Classifier: A Predictive Model for Breast Cancer Detection. *International Journal of Computer Applications.*, 133(9): 32-37, 2016.
- [69] Christopher, T., P, Banu, J.J., Study of Classification Algorithm for Lung Cancer Prediction. *International Journal of Innovative Science, Engineering & Technology.*, 3(2): 42-49, 2016.
- [70] Maleki, M., Eroğlu., Kübra, Aydemir,, Önder,. A new method for selection optimum k value in k-NN classification algorithm. *Signal Processing and Communications Applications Conference IEEE*, 2013.

## **ÖZGEÇMİŞ**

Amani YAHYAOUİ, 1987 yılında Tunus'ta doğdu. İlk öğretim ve liseyi Tunus'ta BEJA şehrinde tamamladı. Tunus'da gerçekleştirilen ulusal üniversite giriş sınavından sonra, 2006 yılında Tunus'da bulunan Chartage Üniversitesi'nde Uygulamalı Hesap, Bilgisayar Bilimleri ve Sistemleri bölümünde üniversite eğitime başladı. 2010 yılında lisans eğitimini tamamladı. 2012'de Tunis Üniversitesi'nde yüksek lisansını tamamladı. 2012 yılında, Sakarya Üniversitesi'nde burslu olarak Bilgisayar ve Bilişim Bilimleri Fakültesi, Bilgisayar Mühendisliği bölümünde doktora eğitimine başladı.