

153941

T.C

SELÇUK ÜNİVERSİTESİ

FEN BİLİMLERİ ENSTİTÜSÜ

**ÖZELLİK SEÇME (FS) İLE YAPAY BAĞIŞIKLIK
TANIMA SİSTEMİ (AIRS)
KULLANILARAK MEDİKAL TEŞHİSE GİDİŞ**

Kemal POLAT

YÜKSEK LİSANS TEZİ

ELEKTRİK-ELEKTRONİK MÜHENDİSLİĞİ A.B.D.

Konya, 2004

T.C
SELÇUK ÜNİVERSİTESİ
FEN BİLİMLERİ ENSTİTÜSÜ

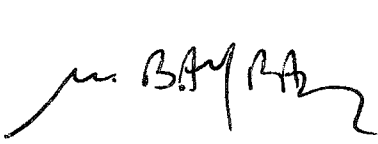
ÖZELLİK SEÇME (FS) İLE YAPAY BAĞIŞIKLIK TANIMA SİSTEMİ (AIRS)
KULLANILARAK MEDİKAL TEŞHİSE GİDİŞ

Kemal POLAT

YÜKSEK LİSANS TEZİ

ELEKTRİK-ELEKTRONİK MÜHENDİSLİĞİ ANABİLİM DALI

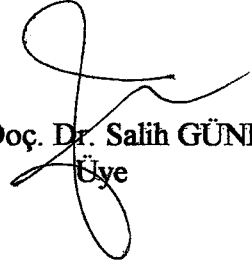
u tez 08/10/2004 tarihinde aşağıdaki jüri tarafından oybirliği/oyçokluğu ile kabul edilmiştir.



rof. Dr. Mehmet BAYRAK
Üye



Prof. Dr. Osman YILMAZ
Üye



Yrd. Doç. Dr. Salih GÜNEŞ
Üye

ÖZET

Yüksek Lisans Tezi

**ÖZELLİK SEÇME (FS) İLE YAPAY BAĞIŞIKLIK TANIMA SİSTEMİ
(AIRS) KULLANILARAK MEDİKAL TEŞHİSE GİDİŞ**

Kemal POLAT

Selçuk Üniversitesi

Fen Bilimleri Enstitüsü

Elektrik-Elektronik Mühendisliği Anabilim Dalı

Danışman: Yrd. Doç. Dr. Salih GÜNEŞ

Danışman: Prof. Dr. Mustafa ŞAHİN

2004, 90 sayfa

Jüri: Prof. Dr. Osman YILMAZ

Prof. Dr. Mehmet BAYRAK

Yrd. Doç. Dr. Salih GÜNEŞ

1900' lü yılların yarısından itibaren ön plana çıkan ve günümüzde karmaşık problemlerin ve analizlerin çoğunda başlıca çözüm yönelimini oluşturan Yapay Zeka Teknikleri, gün be gün yeni bir uygulama alanı kazanmakta ve gelişmektedir.

Yapay Zeka Teknikleri arasında henüz yeni diyebileceğimiz bir sistem Yapay Bağışıklık Sistemleri (YBS), problem çözme performansı bağlamında etkili ve umut vadeden bir sistem olarak karşımıza çıkmaktadır. Başlangıçta Bağışıklık Sistemindeki etkileşimleri modelleyerek Tıp alanındaki araştırmacıların Bağışıklık Sisteminin işleyişini daha iyi kavramasını sağlamak amacıyla ortaya çıkan Bağışıklık Sistemi modelleri YBS' nin ortaya çıkmasında öncü rolü oynamıştır. Bu

modellerde Baęışıklık Sistemindeki bazı özelliklerin (tanıma, hafıza oluřturma, yabancı organizmayı yok etme...vb.) arařtırmacıların dikkatini çekmesi üzerine, bu özelliklerden esinlenerek oluřturulmuř modeller geliřtirilerek klasik çözümlerinin başarılı olamadığı karmařık problemlerin çözümünde kullanılmaya başlanmıřtır.

Medikal teřhis, Tıp alanında oldukça önemlidir. Medikal analizde karar desteęi için modern, etkili ve verimli bir bilgisayar tabanlı bir yöntemin oluřturulması gereklidir. Bu tezde medikal teřhise yardımcı olabilecek bir yöntem sunulmuřtur.

Bu tezde, İrvine' daki California Üniversitesi Bilgisayar Bölümü makine öğrenmesi veritabanından elde edilen göęüs kanseri, kalp hastalığı, karacięer rahatsızlığı ve lenf hastalığı olmak üzere dört farklı hastalığın veritabanı kullanılarak ilgili hastalıklar, *Özellik Seçme (FS)* ile Yapay Baęışıklık Tanıma Sistemi (*AIRS*) algoritması ile sınıflandırılmıřtır. Ayrıca, *Matlab v6.5* programlama dili kullanılarak bu dört veri kümesi üzerinde bir sınıflandırma teknięi olarak *AIRS* ve *FS -AIRS* algoritmalarının etkinlięi gösterilmiřtir.

Literatürde konu ile ilgili yapılan çalıřmalarda, sınıflama doęruluęu bakımından karşılařtırılmıř ve geliřtirilen algoritmanın daha iyi performans sağladığı görölmüřtür.

Anahtar Kelimeler: Yapay Baęışıklık Sistemleri, Medikal teřhis, AIRS, Özellik seçme, FS-AIRS.

ABSTRACT

MS Thesis

**GOING TO MEDICAL DIAGNOSIS BY USING ARTIFICIAL IMMUNE RECOGNITION
SYSTEMS (AIRS) WITH FEATURE SELECTION (FS)****Kemal POLAT**

Selçuk University

Graduate School of Natural and Applied Sciences

Department of Electrical-Electronics Engineering

Supervisor: Asst. Prof. Dr. Salih GÜNEŞ**Supervisor: Prof. Dr. Mustafa ŞAHİN**

2004, 90 pages

Jury: Prof. Dr. Osman YILMAZ**Prof. Dr. Mehmet BAYRAK****Asst. Prof. Dr. Salih GÜNEŞ**

Artificial intelligence techniques which have come into existence especially after mid-1990s and which are the most important solution trend for complex problems are improving and having a new application field day by day.

Artificial Immune Systems (AIS), which we can say very new, is an effective and prosperous artificial intelligence area with respect to its problem solving performance. In the beginning, it was formed for helping medical experts to understand the working procedure of immune system in more detail by modelling interactions in immune system. But later, some propoerties of immune system (recognition, memory formation, elimination of non-self, ..etc) took the attention of researchers by this models and the techniques developed based on these properties

have begun to be used for complex problems for which traditional solutions were not successful.

Medical diagnosis is very important in medicine. It is necessary to form an efficient and effective computer-based method for decision support in medical analysis. In this thesis, a method that can help medical diagnosis was proposed.

The used medical data were taken from machine learning database of California University in Irvine and these data belong to breast cancer, heart disease, liver disorders and lymphoid disease. Each of data was classified with Artificial Immune Recognition System (AIRS) after a feature selection process (FS-AIRS). Comparisons were done with using pure AIRS and FS-AIRS for each dataset. These applications were done by using MATLAB 6.5 programming language.

The proposed method was compared with other studies in literature related to this concept and it was seen that proposed method had performed better than others.

Key Words: Artificial Immune Systems, Medical diagnosis, AIRS, Feature selection, FS-AIRS.

TEŞEKKÜR

Çalışmalarım sırasında özverili katkılarından dolayı; Selçuk Üniversitesi Mühendislik Mimarlık Fakültesi Elektrik-Elektronik Mühendisliği Bölümü öğretim üyesi danışmanım Yrd. Doç. Dr. Salih GÜNEŞ' e şükranlarımı sunarım.

İkinci danışmanım olarak her zaman yanımda olan yardımlarını hiç esirgemeyen, sık sık kendilerine giderek değerli bilgiler aldığım Selçuk Üniversitesi Meram Tıp Fakültesi Genel Cerrahi A.B.D. öğretim üyesi Prof. Dr. Mustafa ŞAHİN' e teşekkür ederim.

Yardımlarından dolayı Selçuk Üniversitesi Mühendislik Mimarlık Fakültesi Elektrik-Elektronik Mühendisliği Bölüm Başkanı Prof. Dr. Mehmet BAYRAK' a, yönlendirici katkılarından dolayı Arş. Gör. Seral ŞAHAN ve Arş. Gör. Halife KODAZ' a ve diğer bölüm hocalarıma en içten teşekkürlerimi sunarım.

Kemal POLAT
Ekim 2004, KONYA

İÇİNDEKİLER

	Sayfa no:
ÖZET	I
ABSTRACT	III
TEŞEKKÜR	V
İÇİNDEKİLER	VI
ŞEKİLLER LİSTESİ	IX
TABLolar LİSTESİ	XI
SİMGELER VE KISALTMALAR	XII
1. GİRİŞ	1
1.1. Tez Konusunun Tanıtılması	3
1.2. Tez Çalışmasının Amacı ve Önemi	4
1.3. Bu Konuda Yapılan Çalışmaların Tarihsel Gelişimi	5
2. BAĞIŞIKLIK SİSTEMİ	8
2.1. Genel Bilgiler	8
2.1.1. Bağışıklıkta rol alan temel birimler	10
2.2. Lenfositlerin Olgunlaşması, Aktivasyonu ve Regülasyonu	13
2.2.1. B Lenfositlerinin olgunlaşması	14
2.3. T Lenfositlerinin Aktivasyonu	15
2.3.1. Tanıma	16
2.3.2. Çoğalma	17
2.3.3. Farklılaşma: efektör hücrelere dönüşme	17
2.3.4. Efektör fonksiyonları	17
2.4. B Hücreleri Aktivasyonu ve Antikor (Ab) Üretimi	19
2.4.1. Antijen tanıma ve B hücreleri aktivasyonu	20
2.4.2. Protein antijenlerine karşı gösterilen T-bağımlı B hücreleri yanıtı	21
2.4.3. Yardımcı T hücrelerine bağımlı B hücreleri yanıtındaki son olaylar: germinal center reaksiyonları	22
2.4.4. T-Hücreleri bağımsız antijenlere B hücreleri yanıtları	23
3. YAPAY BAĞIŞIKLIK SİSTEMLERİ	24
3.1. Yapay Bağışıklık Sistemlerinin Tarihi	24
3.2. Bir Yapay Bağışıklık Sisteminde Kullanılan Gerekli Kavramlar	25

3.3. Şekil Uzayı Modeli	27
3.4. Antijen-Antikor Gösterimleri ve Duyarlılıklar	29
3.5. Yapay Tanıma Toplari ve Kaynak Rekabeti	32
3.6. Bellek Hücreleri, Mutasyon ve Klonlama Seçimi	32
3.6. Bağışıklık Ağları ve Hücreler Arası Etkileşimler	34
3.7. Sınıflandırma	34
3.8. Klonlama Seçimi Algoritması	35
3.9. Takviye ile Öğrenme (Reinforcement Learning) ve Bağışıklık Belleği	38
3.10.Örüntü Tanıma (Pattern Recognition).....	39
4. AIRS ALGORİTMASI VE ÖZELLİK SEÇME (FS) İŞLEMİ.....	40
4.1. AIRS Algoritmasının Tanımı ve Algoritmada Kullanılan Tanımlamalar.....	40
4.1.1. Algoritmanın rutinleri.....	43
4.1.2. AIRS algoritmasının ayrıntıları.....	53
4.2. Özellik Seçme (<i>Feature Selection</i>) Prosedürü.....	56
4.2.1. Özellik seçme işleminin tanımı ve kullanılan yöntemler	56
5. MEDİKAL VERİTABANI.....	59
5.1 Göğüs Kanseri Veritabanı	59
5.1.1. Göğüs kanseri hastalığının tanımı	59
5.1.2. Göğüs kanseri veritabanı bilgisi.....	60
5.1.3. Göğüs kanserinin geçmiş tahmin performansı.....	60
5.2. Kalp Hastalığı Veritabanı.....	61
5.2.1. Kalp hastalığının tanımı.....	61
5.2.2. Kalp hastalığı veritabanı bilgisi	62
5.2.3. Kalp hastalığının geçmiş tahmin performansı	63
5.3. Karaciğer Rahatsızlığı Veritabanı.....	64
5.3.1. Karaciğer rahatsızlığı tanımı.....	64
5.3.2. Karaciğer rahatsızlığı veritabanı bilgisi.....	64
5.3.3. Karaciğer rahatsızlığı veritabanının geçmiş tahmin performansı	65
5.4. Lenf hastalığı Veritabanı	66
5.4.1. Lenf hastalığı bilgisi	66
5.4.2. Lenf hastalığı veritabanı bilgisi.....	66
5.4.3. Lenf hastalığı Veritabanının Geçmiş Tahmin Performansı	67

6. AIRS VE FS-AIRS' NİN MEDİKAL VERİ SETİ ÜZERİNDEKİ SINIFLAMA PERFORMANSLARI	69
6.2. Kalp Hastalığı Veritabanının AIRS ve FS-AIRS Sınıflama Performansı	72
6.3. Karaciğer Rahatsızlığı Veritabanının AIRS ve FS-AIRS Sınıflama Performansı	76
6.4. Lenf hastalığı Veritabanının AIRS ve FS-AIRS Sınıflama Performansı	78
7. SONUÇLAR ve ÖNERİLER	82
8. KAYNAKLAR	84



ŞEKİLLER LİSTESİ

Sayfa no:

Şekil 2.1 Bağışıklık sisteminin hücreleri ve salgılarının yapısal bölümü.....	11
Şekil 2.2 Antikor.....	13
Şekil 2.3 Fagositik hücreler.....	14
Şekil 2.4 B lenfositinin olgunlaşmasında kemik iliğindeki seçme işlemi.....	16
Şekil 2.5 Antijen sunan hücreler ile T hücresi etkileşimleri.....	20
Şekil 2.6 Birinci ve ikincil humoral yanıtın kinetikleri.....	21
Şekil 3.1 Yapay Bağışıklık Sisteminin katmanlı yapısı.....	28
Şekil 3.2 Tanıma topu kavramı ve repertuar gösterimi.....	30
Şekil 3.3 Hamming uzaklığı ve duyarlılık hesabı.....	33
Şekil 3.4 Klonlama seçimi algoritmasının biyolojik olarak gösterimi.....	38
Şekil 3.5 Klonlama seçimi algoritmasının blok diyagramı.....	38
Şekil 3.6 Epitop olarak adlandırılan ve antikor tarafından tanıma antijen parçası.....	41
Şekil 4.1 ARB oluşumu için Hiper-Mutasyon.....	48
Şekil 4.2 Mutasyon Rutini.....	49
Şekil 4.3 Stimulation, Kaynak tahsisi ve ARB çıkarılması.....	51
Şekil 4.4 Hayatta kalan ARB' nin Mutasyonu.....	52
Şekil 4.5 Bellek Hücresi Takdimi.....	53
Şekil 4.6 Özellik seçme işleminin blok diyagramı.....	57
Şekil 4.7 Wrapper yaklaşımı blok diyagramı.....	58
Şekil 4.7 Filter yaklaşımı blok diyagramı.....	59
Şekil 6.1 Göğüs kanseri veritabanında özelliklerin sınıf değişkenine göre veri dağılımları.....	72
Şekil 6.2 Göğüs kanseri veritabanı sınıf dağılımları.....	73
Şekil 6.3 Kalp hastalığı veritabanındaki birkaç özelliğin sınıf değişkenine göre veri dağılımları.....	76
Şekil 6.4 Kalp hastalığı veritabanı sınıf dağılımları.....	76
Şekil 6.5 BUPA karaciğer rahatsızlığı veri setinde bulunan özelliklere ait veri dağılımları.....	79
Şekil 6.6 Karaciğer rahatsızlığı veritabanı sınıf dağılımları.....	79

Şekil 6.7 Lenf hastalığı veritabanındaki birkaç özelliğin sınıf değişkenine göre veri dağılımları.....	82
Şekil 6.8 Lenf hastalığı veritabanı sınıf dağılımları.....	82



TABLolar LİSTESİ

Sayfa no:

Çizelge 5.1 Göğüs kanseri hastalığı teşhisi için sınıflama performansı karşılaştırılması.....	62
Çizelge 5.2 Kalp hastalığı teşhisi için sınıflama performansı karşılaştırılması.....	64
Çizelge 5.3 Karaciğer rahatsızlığı teşhisi için sınıflama performansı karşılaştırılması.....	66
Çizelge 5.4 Lenf hastalığı teşhisi için sınıflama performansı karşılaştırılması.....	69
Çizelge 6.1 Göğüs kanseri veritabanında her bir durum için çıkış değerlerine bağlı özellikler.....	72
Çizelge 6.2 Göğüs kanseri hastalığı teşhisi için sınıflama performansı karşılaştırılması.....	74
Çizelge 6.3 Göğüs kanseri veritabanını sınıflamada kullanılan AIRS parametreleri.....	74
Çizelge 6.4 Kalp hastalığı veritabanında her bir durum için çıkış değerlerine bağlı özellikler.....	75
Çizelge 6.5 Kalp hastalığı teşhisi için sınıflama performansı karşılaştırılması.....	77
Çizelge 6.6 Kalp hastalığı veritabanını sınıflamada kullanılan AIRS parametreleri.....	78
Çizelge 6.7 Karaciğer rahatsızlığı veritabanında her bir durum için çıkış değerlerine bağlı özellikler.....	79
Çizelge 6.8 Karaciğer rahatsızlığı teşhisi için sınıflama performansı karşılaştırılması.....	80
Çizelge 6.9 Karaciğer rahatsızlığı veritabanını sınıflamada kullanılan AIRS parametreleri.....	80
Çizelge 6.10 Lenf hastalığı rahatsızlığı veritabanında her bir durum için çıkış değerlerine bağlı özellikler.....	81
Çizelge 6.11 Lenf hastalığı teşhisi için sınıflama performansı karşılaştırılması.....	83
Çizelge 6.12 Lenf hastalığı veritabanını sınıflamada kullanılan AIRS parametreleri.....	83

SİMGELER VE KISALTMALAR

Ab	: Antibody (Antikor)
Ag	: Antigen (Antijen)
APC	: Antigen Presenting Cell (Antijen Sunan Hücreler)
AIRS	: Artificial Immune Recognition System (Yapay Bağışıklık Tanıma Sistemi)
ARB	: Artificial Recognition Ball (Yapay Tanıma Topu)
AT	: Affinity Threshold (Uygunluk Eşiği)
ATS	: Affinity Threshold Scalar (Uygunluk Eşiği Ölçeği)
GA	: Genetic Algorithm (Genetik Algoritma)
FS	: Feature Selection (Özellik Seçme)
FS-AIRS	: Feature Selection- Artificial Immune Recognition System (Özellik Seçme ile Yapay Bağışıklık Tanıma Sistemi)
Ig	: Immunoglobulin
MHC	: Major Histocompatibility Complex
TCR	: T-Hücre Reseptörü (T-cell reseptor)
sl	: Stimulation Threshold (Uyarılma Eşiği)
YSA	: Yapay Sinir Ağları (Artificial Neural Network)

1. GİRİŞ

Modern hastaneler ve medikal enstitüler’deki medikal bilgi sistemleri büyümesine bağlı olarak, karar desteği için kullanışlı bilgiyi belirlemede büyük zorluklar meydana gelmekte ve bu yüzden verimliliği sağlamak için bilgisayar tabanlı analizlerin hastalık teşhisinde kullanılması kaçınılmaz olmuştur. Bu yüzden, karar desteği için medikal analizde modern, etkili ve verimli bir bilgisayar tabanlı bir yöntemin oluşturulması gereklidir. Medikal veri analizinin birçok farklı biçimi vardır; özellikle statiksel, makine öğrenmesi ve veri soyutlama sık kullanılan yöntemlerdir.

Medikal analiz veya teşhis, görüntü teşhisi ve veri analizidir. Görüntü teşhisi, görüntüyü analiz etmek ve birkaç görüntüyü tutmak için medikal testleri görüntülemek için kullanılır. X-ışınları, Bilgisayar Yardımlı Tomografi (*CAT-Computer Aided Tomography*), Tek Işınlı Emisyon Tomografisi (*SPECT- Single Photon Emission Tomography*) gibi yöntemler örnek olarak verilebilir. Diğer taraftan, veri analizi, elde edilen belirtiler veya ölçülmüş test sonuçlarından hastanın durumunun analizine odaklanır. Üstelik, elde edilen belirtiler veya ölçülmüş test sonuçları iki gruba bölünebilir: demografik veri ve sensor verisi. Demografik veri, hastanın yaşı, cinsiyeti, ağırlığı ve uzunluğu gibi kısa bir periyot üzerinden alınan hızlı olarak değişmeyen özelliklerin hastadan direk olarak elde edilmesiyle oluşabilir. Diğer taraftan, sensor verisi, kan akış hızı ve kalp atış sinyali gibi verileri içeren ve kısa bir periyotta değişmeyen özelliklerden oluşur.

Bu tez çalışmasında, medikal analiz için Yapay Bağışıklık Sistemleri ve Özellik Seçme kullanıldı. İnsanoğlu çok eskilerden beri gerek doğada gerçekleşen olayları anlamak ve bunları taklit ederek yaşamını kolaylaştırmak üzere çoğu problemin çözümünü çoğunlukla doğada aramıştır. Teknolojinin ilerlemesi, beraberinde çözümü karmaşık problemleri getirmiş ve bilinen yöntemler, karmaşıklaşan bu problemlerin çözümünde yetersiz kalmaya başlayınca, farklı çözüm yolları araştırılmaya başlanmıştır. Yapay Sistemler de bu tür problemler için geliştirilmiş ve özellikle 1900’lü yılların ortalarından itibaren adını sıkça duyurmaya başlamıştır.

Arı kolonilerindeki iletişimden, karıncalar arasındaki işbirliğinden, sinir sistemindeki işleyişten ve genetik olaylardan esinlenilerek oluşturulmuş sistemler, Yapay Sistemler adı altında örneklenebilecek sistemlerdir. Yapay sistemler içerisinde en büyük yeri, amacı insandaki kavrama, yorumlama, öğrenme gibi bir çok özelliği taklit etmek olan Yapay Zeka teknikleri kaplamaktadır. Günümüzde mühendislikte, uzay bilimlerinde, fen ve doğa bilimlerinde, tıpta ve daha bir çok alanda geniş bir uygulama alanına sahip olan Yapay Zeka tekniklerinden başlıcaları Yapay Sinir Ağları (*YSA*) ve Genetik Algoritmalarıdır (*GA*). Özellikle *YSA* ile ilgili yapılan çalışmalar büyük bir hızla ilerlemiş ve bilimin bir çok farklı alanında elde ettiği başarılarla ses getirmiştir.

Yapay Zeka teknikleri arasında yeni bir sistem, Yapay Bağışıklık Sistemleri (*YBS*), bir çok özelliği ile etkin sonuçlar elde etme anlamında umut vadeden bir sistem olarak belirginleşmektedir. Diğer Yapay Zeka tekniklerinin ortaya çıkışında olduğu gibi *YBS*, gelişimini Bağışıklık Sisteminde gerçekleşen olayları kavramaya yönelik oluşturulan Bağışıklık Sistemi modellerine borçludur. Bu modellerden yola çıkılarak oluşturulan Yapay Bağışıklık Sistemleri kısa zamanda araştırmacıların katkıları sayesinde pek çok alanda uygulanabilirliğe yönelik çalışmalar ile gelişimine hız katmıştır.

Tez çalışmasında kullanılan özellik seçme işlemi; ham veri setindeki örneklerin sayısı ve özelliklerin sayısı, oldukça büyük olabilir. Bu büyüklük, birçok veri madenciliği sisteminde ciddi problemlere sebep olabilir. Özellik seçme, bu problemleri çözmeye ilgili var olan yöntemlerden birisidir. Özellik seçme işleminin amacı, birkaç uygun kritere göre özelliklerin minimum sayısını seçer. Özelliklerin minimum sayısının seçilmesiyle, ilgisiz ve fazla özellikler kritere uygun olarak uzaklaştırılır. Özellik seçmede sonra oluşan basitleşen veri, daha açık ve daha anlaşılabilir sonuçlar verebilecektir.

Bu tez çalışmasında, medikal alanda önemli bir yere sahip göğüs kanseri, karaciğer rahatsızlığı, kalp hastalığı ve lenf hastalığını teşhis etmek ve sınıflamak için Yapay Zekâ alanında yeni bir alan olan *YBS (Yapay Bağışıklık Sistemi)* ve *FS (Özellik Seçme)* uygulandı. Bu hastalık veritabanları *UCI machine-learning database*' den alınmıştır. Bu veritabanları; göğüs kanseri veritabanı 683 hastadan,

karaciğer hastalığı veritabanı 345 hastadan, kalp hastalığı 303 hastadan ve lenf hastalığı 148 hastadan oluşmaktadır.

Bu çalışmanın giriş bölümünde tez konusu tanıtılarak konu ile ilgili literatürde yapılan çalışmalar yer almaktadır. İkinci bölümde, Yapay Bağışıklık Sistemlerinin model aldığı doğal Bağışıklık Sistemi hakkında bilgiler verilmiştir. Üçüncü bölümde anlatılan Yapay Bağışıklık Sistemleri, tarihsel gelişimi, temel nitelikleri ve tasarım modelleri gibi alt gruplarda tanıtılmıştır. Dördüncü bölümde, AIRS algoritması ve Özellik Seçme (FS) işlemi incelenmiştir. Beşinci bölümde, göğüs kanseri, kalp hastalığı, karaciğer rahatsızlığı ve lenf hastalıklarının bulunduğu bir medikal veritabanları tanıtılmıştır. Altıncı bölümde, AIRS ve FS-AIRS' nin medikal veri seti üzerindeki sınıflama performansları incelenmiştir ve son olarak yedinci bölümde sonuçlar ve öneriler bulunmaktadır.

1.1. Tez Konusunun Tanıtılması

Hafıza oluşumu, tanıma, dağıtılmış mekanizma, hata toleransı gibi daha bir çok özelliği bünyesinde barındıran Bağışıklık Sistemleri, Yapay Zeka uygulamalarına elverişli bir model zemini oluşturmakla Yapay Bağışıklık Sistemlerine (*YBS*) temel hazırlamıştır. Bağışıklık sistemindeki pek çok özelliğin modellenmesi, bir çok farklı alanda yeni *YBS* algoritmalarının oluşturulmasına vesile olmuştur. Ortaya atılan *YBS* algoritmaları ile optimizasyon, robotik, hata tespiti, medikal teşhis gibi bazı alanlarda yüksek performansların elde edilmesine karşın sınıflama ve tanıma sistemlerinde henüz çok yüksek başarılarla ulaşılamamıştır. Araştırmacıların bu kategorideki *YBS* çalışmalarına yapabilecekleri katkılar iki şekilde gerçekleştirilebilir. Bunlardan birisi oluşturulmuş mevcut algoritmaların eksikliklerini tespit edip bu eksiklikleri giderici öneriler sunarak sistemlerin performanslarını artırmaya çalışmaktır. Yapılabilecek diğer katkı şekli de yeni modellemelerin oluşturulması ile bağışıklık tabanlı yeni sınıflama algoritmalarının öne sürülmesidir.

Bu tez çalışmasında, medikal teşhiste karar verme işlemini gerçekleştirmede önemli olan ve insan yardımlarını ve maliyetleri azaltmak için ve ayrıca sınıflama doğruluğunu arttırmak için medikal veritabanlarına *AIRS* ve *FS-AIRS* uygulanmıştır.

FS-AIRS' deki *FS* işlemi, *AIRS* sınıflama algoritmasında kullanılacak verilerin özelliklerini değerlendirmek için, bir veriden ilgisiz veya fazla özellikleri çıkartarak bir veri ön-işleme işlemi için kullanılmıştır. *FS-AIRS* algoritması, bu dört medikal veritabanından göğüs kanseri ve kalp hastalığı veritabanlarına uygulandı, diğer veritabanlarındaki özellikler aynen kullanıldı. Bu çalışmada kullanılan göğüs kanseri veritabanındaki verilerin dokuz özelliği bulunmaktadır, bu verilere *FS* işlemi uygulandıktan sonra özellik sayısı altıya indirildi ve bu işlem sonucunda göğüs kanserinin bu altı özelliği kullanılarak *FS-AIRS* sınıflama algoritması sonunda teşhis doğruluğu arttığı gözlenmiştir.

1.2. Tez Çalışmasının Amacı ve Önemi

Makine öğrenmesi literatüründe kullanılan medikal veri kümeleri üzerinde danışmanlı bir yapay bağışıklık sınıflandırma algoritmasının performansını değerlendirmek, algoritmanın detaylı sunumunu göstermek, sınıflandırma yeteneklerini araştırmak ve elde edilen sonuçlarla ispat etmektir. Bu amacı başarmak için sadece bağışıklık sistemi prensiplerini temel alan bir sınıflandırma algoritmasını göstermek yeterli değildir aynı zamanda bu geliştirilmiş sistemin doğru sınıflandırmayı sağladığını göstermek gereklidir. Yapay bağışıklık tanıma sistemi (*AIRS*) olarak geliştirilen algoritma makine öğrenmesi için oldukça arzu edilebilir karakteristikler sergilemektedir (Watkins (2001)). Bilgisayar amaçlı medikal teşhis sistemlerinde, genel olarak, teşhis, makine öğrenmesi algoritmaları kullanılarak gerçekleştirilir. Danışmalı öğrenme algoritması olan *AIRS* de, bir makine öğrenmesi algoritması olarak kullanılır.

Makine öğrenmesi, bir elektronik veritabanından en iyi problemleri açıklayabilen bilgisayar modellerini veya kuralları bulmak için geliştirilen bilgisayar programları ile kullanılarak oluşturulur. Makine öğrenmesi teknikleri iki kategoriye

ayrılır: danışmalı öğrenme ve danışmasız öğrenme. Danışmalı öğrenme, doğru sınıf değişkeniyle bir veritabanından en iyi tanımlanan bilgisayar modelini bulan bir makine öğrenmesi tekniğidir. Sınıf değişkeni, sınıflamaya gerek duyulan bilgisayar modelinde bir veritabanındaki değişkendir. Diğer taraftan, danışmasız öğrenme, bir sınıf değişkenini kullanmaksızın bir veritabanını kullanan bir makine öğrenmesi tekniğidir. Makine öğrenmesi tekniğinin birçok çeşidi vardır: Yapay Bağışıklık Sistemleri, sinir ağları, karar ağaçları, Bayessian networkları ve genetik algoritmalar v.b.

Her bir makine öğrenmesi tekniğinde iki farklı aşama vardır: eğitime aşaması ve test aşamasıdır. Bir bilgisayar modeli veritabanında örnek bir veri setinden oluştuğu zamanki safhaya eğitime aşaması denir. Oluşturulan bilgisayar modeli, veritabanında öncelikle bilinmeyen örneklerden oluşan bir veritabanından test edildiği zamanki safhaya test aşaması denir.

Bu tez çalışmasında, *YBS*’ de danışmalı öğrenme algoritması olan *AIRS* ve veri ön-işleme işlemi için kullanılan özellik seçme (*FS*) teknikleri medikal teşhis için kullanılmıştır.

1.3. Bu Konuda Yapılan Çalışmaların Tarihsel Gelişimi

Yapay bağışıklık alanında ilk olarak bağışıklık sistemlerini temel alan hesaplama modelleri inşa etme amacıyla teorik tartışmalar ve geniş bir uygulamalar gerçekleştirilmiştir. Araştırmacıların bazıları tam olarak bağışıklık sistemini simule etmeye çalışmışlar bazıları da doğadan esinlenen yapay sinir ağları ve genetik algoritmalar alanları gibi bilgi işleme mantığıyla bağışıklık sistemini kullanmaya çalışmışlardır. Uygulama alanlarına, örüntü tanıma ve veri madenciliği (*Hunt and Cooke (1996); Timmis, Neal and Hunt (1999); Carter(2000); Timmis (2001)*), fonksiyon optimizasyonu (*De Castro and Von Zuben (2001b)*), anormalliği algılama ve bilgi güvenliği (*Forrest et al. (1994); Hofmeyr and Forrest (1999)*), robotik (*Jun, Lee and Sim (1999); Watanabe, Ishiguro and Uchikawa(1998)*) örnek verilebilir.

Yapay sinir ağlarının insan beynini ve genetik algoritmaların doğadaki evrimi model aldığı gibi Yapay Bağışıklık Sistemleri bağışıklık sistemlerinden esinlenilerek hesaplama için bir model olarak ortaya çıkarılmıştır. Doğal bağışıklık sistemleri çok fazla karmaşıktır ve çoğu parçası tam olarak anlaşılammıştır. Bununla beraber anlaşıldığı kadarıyla doğal bağışıklık sistemleri lenfositlerin iki farklı türünü içermektedir. Bunlar *B* ve *T* hücreleridir. Bu hücreler vücudumuz içerisinde dolaşırlar ve bağışıklık tepkisini sağlamak için birbirleri ile etkileşime girerler. Bağışıklık sisteminde saldırıya tepki veren diğer hücre tiplerinden çok sayıda mevcuttur fakat yapay bağışıklık sistemleri için model olarak çok sık kullanılabilen *B* ve *T* hücreleri genel olan hücrelerdir. *T* hücreleri olgunlaşma sürecinde timüste (*Thymus*) te dolaştığı için, *B* hücreleri kemik iliğinde (*Bone marrow*) olgunlaştığı için bu adı almıştır. *YBS* çalışmalarının çoğu *B* hücrelerine odaklanarak makine öğrenmesi için *YBS* kullanmayı araştırmaktadırlar. *B* hücreleri doğal bağışıklık sistemlerinin belleği olarak kullanılmaktadır. *YBS*'de makine öğrenme toplulukları tarafından kullanılan *B* hücreleri *Jerne(1974)* tarafından önerilen ve *Perelson (1989)* tarafından detaylı açıklanan bir bağışıklık ağını oluşturmaktadır. *Hunt and Cooke (1996)*, *Timmis, Neal and Hunt (1999)*, *Timmis, Neal and Hunt (2000)*, *Timmis and Neal (2000)* ve *Timmis(2001)*'deki çalışmalar bağışıklık ağ teorisinin uygulamalarıdır.

Çok ilginç fikirlere biri olan kaynak sınırlı yaklaşım kavramı, *Timmis* tarafından çalışılan *Yapay Bağışıklık Ağından (AIN-Artificial Immune Network)* esinlenilerek geliştirilmiştir. Bu kavramın temelinde aynı özelliklere sahip *B* hücrelerinin bir topluluğunu temsil eden *Yapay Tanıma Topları (ARB-Artificial Recognition Ball)* kullanılmaktadır. Temel kavram oldukça basittir. Verilen bir antijen ile bağışıklık hücresinin her bir karşılaşmasında her bir hücre uyarma seviyesini temel alarak kaynaklar elde etmektedir. Bununla beraber burada sistemin geniş kaynakları sınırlıdır. Eğer sistemde izin verilenden daha fazla kaynak tüketilirse kaynaklar hücrelerden silinir. Bu işleme en az uyarılanlara öncelik verilerek başlanır ve bu silme işlemine kaynakların sayısı sistemde izin verilen kaynakların sayısına eşit oluncaya kadar devam edilir. Kaynaklar için olan bu rekabet sadece en uygun hücrelerin hayatta kalmasını teşvik eden sistemi garantiler.

Timmis and Neal (2000) ve *Timmis (2001)* bağışıklık sistemlerinin bağışıklık ağı teorisini temel alan danışmansız makine öğrenmesi algoritmasını sunmuşlar ve kaynak sınırlama kavramını düzenlemişlerdir. Bu yaklaşımların veri kümelerinde bilinen örüntüleri tanımada başarısı gösterilirken sonuçların kalitesini değerlendiren mekanizmalar geliştirilmiştir. Yalnızca insan değerlendirmesine bağı olan bu mekanizmalar bağışıklık ağını grafiksel olarak göstermektedirler. Bu çalışmalarda insan etkileşiminden ayrı olarak çıkışın kalitesini tahmin etme veya veri kümesi için sınıf etiketleri atamada otomatik bir yol sunmamışlardır.

Bu tez çalışmasında, yapay bağışıklık ağı ve ona bağı çalışmalar üzerinde daha önceden yapılan bu çalışmalardan bazı özellikler toplanarak tartışılmıştır. Özellikle bağışıklık hücresi olarak yapay tanıma topları kavramı ve kaynak sınırlama fikri kullanılmıştır. *AIRS* aynı zamanda sistemdeki yapay tanıma topları (*ARB*) veya hücreler için sınıf etiketleri atama metodu sağlayarak yapay bağışıklık ağı üzerindeki ilk çalışmaları genişletmiştir. Şu ana kadar *Carter (2000)* tarafından sunulan çalışmadan başka yapay bağışıklık sistemleri için danışmanlı bir öğrenme alanında araştırma pek yaygın değildir. Bununla beraber *YBS* için *Carter (2000)* çalışmasındaki yaklaşım bu çalışmadaki yaklaşımdan önemli derecede farklıdır ve daha yüksek seviyede bir karmaşıklığa sahiptir.

Özellik Seçme, özellikler arasındaki bağılılık olarak da adlandırılır. Özellikler arasındaki bağılılık, sınıflama doğruluğunu arttırmak için özellikler arasındaki gerçek olmayan ve farz edilen bağılılığın serbest olması anlamına gelir. 1991’ de *Koronenka*, 1995’ de *Pazzani* ve 1997’ de *Friedman, Geiger & Goldszmidt* *Özellik Seçme* alanında yapılan çalışmalar ile literatürde yerlerini almışlardır.

2. BAĞIŞIKLIK SİSTEMİ

2.1. Genel Bilgiler

Vücudumuz sürekli olarak dış çevreden gelen mikropların, virüslerin, bakterilerin,...vb. saldırısına uğramaktadır. Bağışıklık sistemi, bu mikroorganizmaların vücuda zarar vermesini engellemek ve bir an önce vücudu terk etmelerini sağlamak için gerekli işlemleri başlatır. Bu işlemlerin hepsine birden ‘*bağışıklık cevabı*’ adı verilir. Bağışıklık cevabının gerçekleştiği süre içerisinde “*hastalık*” devam eder. Bu süre, mikroorganizmanın yapısına ve vücudun direncine göre farklılık gösterdiği için hastalık süreçleri de farklılık arz etmektedir. Bağışıklık cevabının sonunda vücut mikroorganizmayı yenmeyi başardıysa bu mikroorganizmaya duyarlı hücreler hafıza hücreleri olarak saklanır. Buradaki amaç, vücudun söz konusu mikroorganizmaya karşı bağışıklık kazanmasını sağlamaktır. Bağışıklık cevabında olumlu bir yanıt oluşmayabilir. Bu durumda hastalığın ilerlemesi söz konusudur ve vücut hastalıkla savaşmaya devam eder.

Bağışıklık sisteminde görev alan organlar hemen hemen sınırsız çeşitlilikteki bulaşıcı yabancı hücreleri ve maddeleri (*nonsel element*) tanıyabilir ve bulaşıcı olmayan doğal hücreleri (*self element*) ayırt edebilir. Bir patojen (bulaşıcı yabancı eleman) vücuda girdiği zaman yok edilmek için taranır ve yok edilmeye çalışılır. Sistemin her enfeksiyonu hatırlama yeteneği vardır. Aynı patojene ikinci kez bulaşınca daha etkili bir şekilde yok edilir.

Vücuttaki yabancı materyalleri tanıyan birbirleriyle ilişkili 2 sistem söz konusudur. Bunlar, doğal (*innate*) bağışıklık sistemi ve adaptif bağışıklık sistemidir.

Doğal bağışıklık sistemi adı verilmesinin sebebi vücudun doğumla birlikte mikropları tanıma ve hemen yok etme yeteneğinin olmasıdır. Bizim doğal bağışıklık sistemimiz ilk karşılaşmada çoğu patojenleri yok edebilir. Doğal bağışıklık sisteminin önemli bir bileşeni “*complement*” olarak bilinen bir kan proteini sınıfıdır. Complement antikorların aktif olmasına yardım etmektedir. Doğal bağışıklık, pattern associated molecular patterns (*PAMP*) olarak adlandırılan mikrobik

patojenlerle birleştirilmiş moleküler patternleri (örüntüleri) tanımak için pattern recognition receptors (*PRR*) olarak bilinen germinal merkezde kodlanmış alıcıların bir kümesi esasına dayanmaktadır. *PAMP'* lar sadece mikroplar tarafından üretilir ve asla organizma tarafından üretilmez; bu yüzden *PRR* ile onların tanınması, patojenik maddelerin bulunduğu işaret eden bir sinyalle sonuçlanabilir. Bu şekilde bağışıklık tanınması ile ilgili yapılar bulunduğu vücuda zarar vermekten kaçınmak için özel hücre ve moleküllerden tamamen bağımsız olmak zorundadır. Bu mekanizmanın sonucu olan doğal bağışıklık aynı zamanda *self/nonself* ayırım yeteneğine sahiptir ve adaptif bağışıklığın desteklenmesinde rol oynar.

Doğal bağışıklık tanınmasının en önemli aşaması, adaptif bağışıklık tepkisinin başlamasını sağlayan T hücrelerinin aktivasyonunda rol oynayan antijen sunan hücrelerde (*APC*) uyarıcı sinyal ifade etmesidir. Bu şekilde doğal bağışıklık tanınması olmaksızın adaptif bağışıklık tanınması, adaptif tanımda reseptörleri ifade eden lenfositlerin negatif seçimi ile sonuçlanabilir.

Adaptif bağışıklık sistemi, B ve T hücreleri şeklindeki iki tip lenfosit klonal olarak dağıtılmış antijen reseptörlerinin somatik olarak oluşturulmasın da kullanır. Bu antijen reseptörleri rastgele proseslerle oluşturulur ve sonuç olarak adaptif bağışıklık tepkisinin genel dizaynı, belirli özellikleri olan reseptörleri ifade eden lenfositlerin klonal seleksiyonuna dayanmaktadır (Burnet 1959-1978). Antikor molekülleri (*Ab*) adaptif bağışıklık sisteminde önemli bir rol oynar. Adaptif bağışıklık tepkisinde kullanılan reseptörler gen segmentleri ile birlikte parçalanarak düzenlenmiştir. Her hücre farklı olarak tek bir reseptör yapmak için uygun parçaları kullanır. Müşterek olarak bulaşıcı organizmayı tanımak için aktif edilen hücrelerle bir yaşam boyunca karşılaşılır (Tonegawa, 1983). Adaptif bağışıklık daha önce saldırganla hiç karşılaşılmasa bile vücudun herhangi bir mikrobu tanınmasını ve tepki vermesini sağlar.

2.1.1. Baęışıklıkta rol alan temel birimler

Baęışıklık sistemindeki hücreler temel olarak üç gruba ayrılır (Özbal 1994):

1. Lenfositler
2. Monositler
3. Polimorfo nükleer granüositler

Bu hücre grupları içinde lenfositlerin görevi, vücuda giren yabancı maddelerin (*Antijen*) uyarımıyla farklılaşarak baęışıklık cevabını oluşturmak ve Antijenleri yok etmektir. Diğer iki gruptaki hücrelerin görevi ise fagositozdur (Fagositoz: Hücreleri ve mikroorganizmaları sindirme işlemi). Bu hücreler, vücuda giren Antijenik özellikteki molekülleri fagositte ettikleri gibi, baęışıklık cevabından kalan hücre artıklarını da fagositoz yoluyla temizlerler. Şekil 2.1' de, baęışıklık sistemi tarafından üretilen hücreler ve salgılar arasındaki yapısal bölünmeyi göstermektedir.



Şekil 2.1 Baęışıklık sisteminin hücreleri ve salgılarının yapısal bölümü

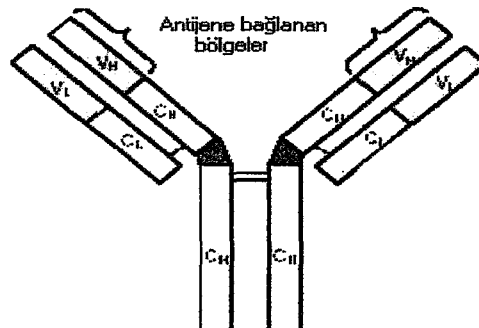
2.1.1.1. Lenfositler

Lenfositler, baęışıklık cevabını oluşturan asıl birimlerdir. Lenfositlerin vücuttaki diğer hücrelerle (self hücreler) ve Antijenlerle (nonself hücreler)

etkileşimi, yüzeylerinde bulunan reseptörler yardımıyla olur. Bu reseptörlerin değişiklik arz etmesi üç tip lenfositimize çıkarır: *T* tipi lenfositler (*T* hücreleri), *B* tipi lenfositler (*B* hücreleri) ve *NK* hücreler (Natural Killer).

T tipi hücrelerin bir kısmı aktivatör, bir kısmı da baskılayıcı olarak çalışır ve bu şekilde bağışıklık cevabının istenen seviyede olmasını sağlarlar. Aktivatör *T* hücreleri yaydıkları kimyasal sinyaller aracılığı ile bağışıklık cevabını direk oluşturur ($CD8^+$ CTL -*T*- lenfositleri ile) ya da oluşmasına katkıda bulunurlar ($CD4^+$ yardımcı *T* hücreleri ile). Baskılayıcı *T* hücreleri ise yine yaydıkları kimyasal sinyaller ile bağışıklık cevabını oluşturmakta olan *T* ve *B* hücrelerinin çalışmalarını frenleyen bir etki gösterirler. *T* lenfositlerinin yüzeylerindeki reseptörlere *TCR* (*T Cell Receptor*), *CD4* ve *CD8* adı verilir. *TCR*, *T* lenfositlerinin Antijenik yapılara bağlanmasını sağlar. *CD4* ve *CD8* reseptörleri ise bağışıklık yanıtı esnasında gerçekleşen etkileşimlerde görev alan ara-reseptörlerdir (co-reseptörler).

B hücreleri, yüzeylerinde *Immunoglobulin* molekülleri taşırlar (De Castro ve Von Zuben 1999a). *Immunoglobulin* molekülleri birbirine benzer yapıda Antikorlardan oluşur (Şekil 2.2). Şekil 2.2’de görüldüğü gibi Antikorlar, protein dizilimlerinin yer aldığı bir ağır zincir (*Heavy Chain*) bir de hafif zincir (*Light Chain*)’den oluşurlar. Antikorların *V* (*Variable*) bölgeleri, Antijenlerle ya da diğer Antijenik yapılarla birleşme yerleridir (reseptörleri). Bir birey yaklaşık 10^9 farklı çeşitte Antikor üretebilir. Antikorlar *C* (*Constant*) bölgelerine göre beş farklı sınıfa ayrılırlar; *IgM*, *IgD*, *IgG*, *IgE* ve *IgA*. Antikorların efektör fonksiyonlarının çoğu ağır zincirlerinin *C* bölgeleri ile başlar. Efektör fonksiyonlarının başlayabilmesi için de *V* bölgelerinin Antijen reseptörlerine bağlanmaları gerekmektedir.



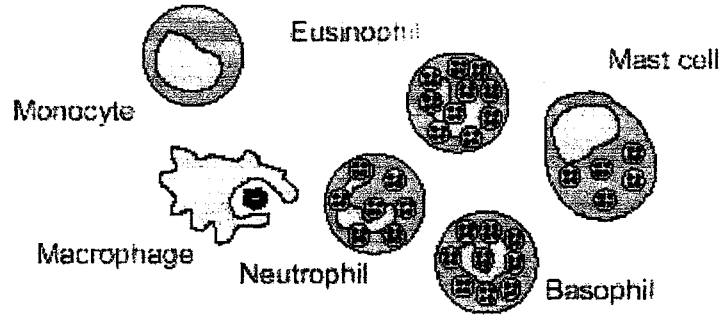
B hücreleri aktive olduklarında plazma hücrelerine dönüşürler ve yüzeylerinde bulunan *Immunoglobulin* moleküllerinden Antikor salgırlar. Salgıladıkları bu Antikorlar Antijenler ile birleşerek bir Antijen-Antikor kompleksi oluştururlar. Bu kompleksin oluşması, çevreye bazı kimyasal sinyaller yayılmasını indükler ve bu sinyalleri algılayan öldürücü hücreler bu kompleksi parçalayarak yok eder (De Castro ve Von Zuben 1999a).

2.1.1.2. Fagositer hücreler

Fagositerler (*hücre yiyenler*) mikro organizmaları ve antijenik tanecikleri yeme ve sindirme yeteneği olan beyaz kan hücreleridir. Bazı fagositerlerin aynı zamanda lenfositler için antijen sunma yeteneği vardır. Bu fagositler antijen sunan hücreler (*APC'ler*) olarak adlandırılmaktadır.

Önemli fagositerler, monositler ve makrofajlardır. Monositler kan içinde dolaşır ve makrofaj (*büyük yiyiciler*) haline geldikleri yerde dokular içine göç ederler. Makrofajlar çeşitli fonksiyonları uygulayan çok yönlü hücrelerdir. Onlar T lenfositleri için antijen sağlarlar sonra onları yerler ve sindirirler. Onlar bağışıklık tepkisinin başlamasında önemli bir rol oynarlar.

Nötrofiller ve eozinofiller aynı zamanda makrofajlara benzer fonksiyonları olan fagositlerdir. Basopiller kan akımında bulunur ve ayrı nesillerden türemelerine rağmen mast hücrelerine benzerler. Onlar alerjiye yanıt vermek için önemlidir ve güçlü kimyasal maddelerle doldurulmuş tanecikler içerirler. Bu kimyasal maddeler mikro organizmaları yok eder, tahrik edici reaksiyonla katkıda bulunur. Şekil 2.3 'de en önemli fagositler gösterilmiştir.



Şekil 2.3 Fagositik hücreler

2.2. Lenfositlerin Olgunlaşması, Aktivasyonu ve Regülasyonu

Lenfositlerin Bağışıklık Sisteminde görev yapabilecek olgun hücreler haline gelmesine olgunlaşma (*maturation*) adı verilir. *B* ve *T* lenfositlerin üretimi kemik iliğinde başlar. *B* hücresi, olgunlaşmasını kemik iliğinde tamamlar fakat *T* hücresi olgunlaşma için *Timus*'a geçer. Her iki hücre tipinin olgunlaşmasında göze çarpan şey, olgunlaşma esnasında yüksek bir mitotik aktivitenin yani bölünerek çoğalma işleminin olmasıdır. Bu da çok çeşitli bir repertuarın oluşmasını öncüllemiş olur. Burada repertuar ile kastedilen, *B* ve *T* lenfositlerinin spesifiklikleridir.

Fonksiyonel reseptör genleri, olgunlaşmamış *B* hücreleri için kemik iliğinde, *T* hücreleri için ise *Timus*'ta üretilir. Olgunlaşmadaki bir diğer önemli nokta, olgunlaşma sırasında reseptörlerde meydana gelen gen rekombinasyonudur (Abbas ve ark. 1994). Her bir *B* ve *T* lenfosit klonu, tek bir yapıda reseptör üretir ve bir bireyde 10^9 ya da daha fazla farklı *B* ve *T* lenfosit klonu bulunabilir. Fakat gerçekte bu mümkün değildir çünkü böyle olması halinde vücuttaki hücrelerin çoğunluğu *B* ve *T* hücrelerinden (lenfositlerinden) oluşması gerekir. Bu durumda repertuar çeşitliliği, olgunlaşma esnasında gerçekleşen somatik rekombinasyonlar ile gerçekleştirilir. Somatik rekombinasyonlarda farklı gen parçaları birleşerek repertuar çeşitliliğini sağlamış olurlar.

Lenfositlerdeki Antijen reseptörleri, rasgele *DNA* rekombinasyonu ile üretilirler ve Antijen varlığına bağlı değildirler. Bir başka deyişle, lenfosit

reseptörleri, Antijen ile karşılaşılmeden üretilirler. Somatik rekombinasyondan sonra gerçekleşen en önemli olay, seçme işlemidir. Rekombinasyon çoğunlukla başarısızlıkla sonuçlanır ve işe yaramayan bir çok *B* ve *T* hücresi *apoptosis* adı verilen bir mekanizma ile ölür.

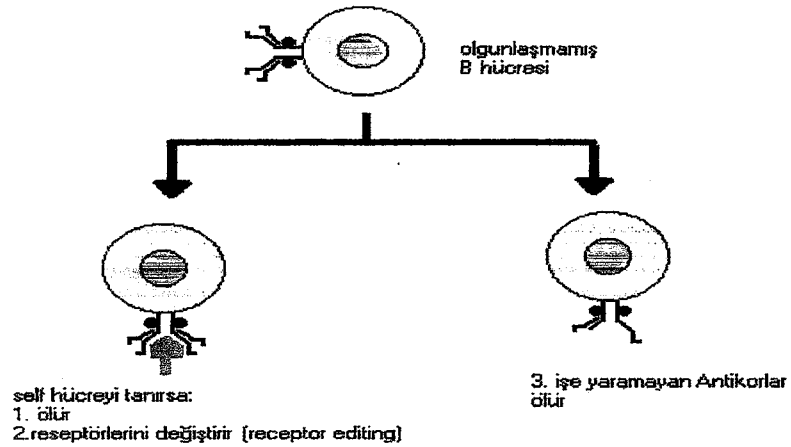
Olgunlaşma işleminden sonra, mitotik aktivite ile üretilen hücrelerin ancak %10'dan azı ikincil lenfoid organlara ve dokulara (bağışıklık cevabının gerçekleştiği organlar) geçebilirler.

2.2.1. B Lenfositlerinin olgunlaşması

Olgun *B* hücrelerinin yabancı Antijenlere karşı çoğalarak yanıt üretmesi ikincil lenfoid dokularda gerçekleşir. *B* hücreleri bir kez uyarıldıktan sonra aktive olurlar ve salgılanmış haldeki *Ig*'lerin hücre duvarına bağlı *Ig*'lere oranı artar. Aktive olan *B* hücrelerinin bazıları da hafıza hücresine dönüşürler. Hafıza hücreleri uzun süreler hayatta kalır ve kan, lenf ve lenfoid organlarda aktif olarak sirkülasyon halinde bulunurlar. Hafıza hücrelerinin Antijenler tarafından uyarılması, ikincil Antijen yanıtlarının oluşmasına neden olur. İkincil yanıtlar hem süre hem de miktar bakımından ilk bağışıklık yanıtından daha etkindirler.

Yaşamları süresince her bir *B* hücresi klonunun üyeleri aynı *V* bölgesine sahiptirler ve aynı Antijen spesifikliğinde kalırlar. Fakat Antijene karşı yanıtlarda, Antikorların *V* bölgelerinde açık değişimler meydana gelir. Antijenik uyarımdan sonra Antijen spesifikliği (duyarlılık) artar. Buna duyarlılık olgunlaşması (affinity maturation) adı verilir ve bu, humoral bağışıklığın bir karakteristiğidir. Duyarlılık olgunlaşmasında *V* bölgelerinin *DNA* kodlamasında mutasyonlar gerçekleşir ve bu mutasyonları bir seçme işlemi takip eder. Böylelikle sadece yüksek duyarlılıklı hücreler hayatta kalır.

Olgun *B* hücrelerinin seçimi, olgunlaşmamış hücrelerin arasından, bu hücreler kemik iliğini terk etmeden önce yapılır. Seçme aşamasında kemik iliğinde üç olay gerçekleşebilir (Şekil 2.4).



Şekil 2.4 B lenfositinin olgunlaşmasında kemik iliğindeki seçme işlemi

2.3. T Lenfositlerinin Aktivasyonu

Vücutta lenf sıvısı aracılığı ile sirkülasyon halinde bulunan ve herhangi bir Antijen ile uyarılmamış *T* hücrelerine 'naive *T* lenfositleri' adı verilir. Naive *T* lenfositlerinin Antijeni tanıması (peptid-*MHC* kompleksi şeklinde) bir dizi olayın gerçekleşmesine neden olur. Antijeni tanıyan *T* lenfositleri çoğalır ve farklılaşarak efektör *T* lenfositlerine dönüşürler. Efektör *T* lenfositlerinin, yine aynı ortam içinde bulunan Antijenleri tanımaları (peptid-*MHC* kompleksi şeklinde), bu hücrelerin efektör fonksiyonlarını gerçekleştirmelerine neden olur.

Protein Antijenleri vücuda girdiğinde, ikincil (*peripheral*) lenfoid organlara götürülürler. Burada bu Antijenler, Antijen sunan hücreler (*APC-Antigen Presenting Cells*) tarafından işlenir ve *MHC*'ye bağlı peptidler şeklinde bu hücrelerin yüzeylerine taşınırlar. Aynı organlarda hareket etmekte olan naive *T* lenfositleri, bu *MHC*-peptid kompleksleri ile karşılaşır ve uyarılırlar. Bu uyarım sonunda aktive olarak çoğalıp (*clonal expansion*) efektör ya da hafıza hücrelerine dönüşürler. Efektör hücrelere dönüşen *T* hücreleri sirkülasyona katılır ve *MHC*-peptid komplekslerini tanıyarak aktive olup efektör fonksiyonlarını yerine getirirler.

Buradaki efektör hücreleri iki sınıfta toplanırlar. $CD4^+$ alt sınıfındaki efektör hücreler aktive olduklarında sitokin salgırlarlar. $CD8^+$ CTL alt setindeki efektör T hücreleri ise *smf I MHC*'lere bağılı konumdaki yabancı Antijenleri öldürürler. T hafıza hücreleri, devam eden bir Antijen karşılaşmasına gerek duymadan uzun süreler hayatta kalabilirler. T hafıza hücreleri, Antijen ile yeniden karşılaşmaya kadar sessiz konumda kalırlar (Abbas ve ark. 1994).

Antijen Elendikten sonra T hücresi sayısı düşer çünkü Antijen ile aktive olmuş T hücrelerinin büyük çoğunluğu *apoptosis* ile ölür.

2.3.1. Tanıma

T Lenfositlerinin aktivasyonu için iki olay gerçekleşmelidir: T hücresi reseptörlerinin (TCR 'ler), MHC -peptid komplekslerini tanımaları ve T hücrelerinin yardımcı molekülleri ile Antijen sunan hücreler üzerindeki *ligand*'ların etkileşmesi (costimulator) gerekmektedir.

Co-stimulatörler, T hücresi aktivasyonu için "ikincil sinyaller" i sağrlarlar. Co-stimulatörler yoksa Antijen ile karşılaşan T hücreleri ya *apoptosis* ile ölür ya da *anergy* adı verilen yanıtızsızlık durumuna geçerler. Dolayısıyla co-stimulatörler sayesinde T Lenfositlerinin yanıtları doğru zaman ve doğru yerde başlatılmış olur.

T hücresi ve Antijen tanınmasının sonucu, TCR -Antijen etkileşiminin süresi (ya da yarı ömrü) ve duyarlılığı ile belirlenir. TCR 'lerin peptid- MHC kompleksi halindeki Antijenlere olan duyarlılıkları, çoğu Antikorun Antijenlere olan duyarlılıklarına göre oldukça düşüktür. TCR -Antijen etkileşiminin düşme oranı çok hızlıdır. Tek bir TCR 'nin MHC bağımlı peptidi 10 saniyeden az meşgul ettiği hesaplanmıştır. Bir T hücresinin aktivasyonu için etkileşim esnasında bir çok sıralı olay meydana gelmelidir. Bu nedenle T hücresi, TCR reseptörlerinin peptid- MHC komplekslerine bağılanış sayısını ve süresini toplayan bir sinyal toplayıcı (integratör) gibi davranır. Aktivasyonun olabilmesi için, TCR kaynaklı sinyallerin bir eşik seviyesine ulaşması gerekmektedir. Bu eşik seviyesi aşılamazsa; ya bir yanıt üretilmez, ya da sadece bazı yanıtların üretildiği bir yarı aktivasyon gerçekleşir.

Hatta *T* hücrelerinin fonksiyonel inaktivasyonu bile gerçekleşebilir. Hafıza ve efektör *T* hücrelerinin aktivasyonu için gerekli eşik seviyesi, *naive T* hücrelerinden daha azdır (Abbas ve ark. 1994).

2.3.2. Çoğalma

Naive T hücrelerinin çoğalmasına *clonal expansion* denir ve bunun sonucunda Antijeni elemek için gerek duyulan Antijen-spesifik hücrelerden oluşan bir populasyon meydana gelir.

2.3.3. Farklılaşma: efektör hücrelere dönüşme

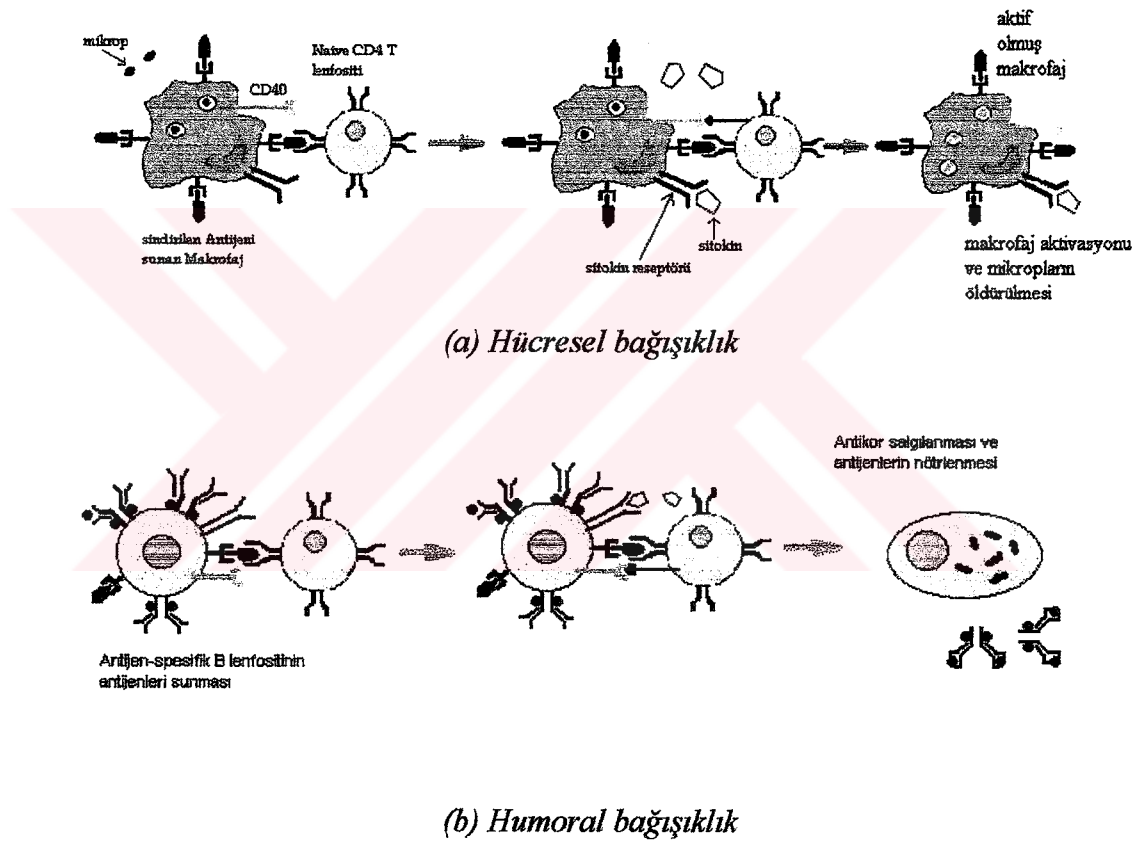
Olgun *T* hücrelerinin farklılaşması, *naive T* lenfositlerinin efektör hücrelere dönüşmesi işlemidir. Bu işlem, hem *T* hücrelerinin Antijen ile indüklenen aktivasyonunu hem de diğer uyarımları gerektirir. Burada bahsedilen diğer uyarımlar, genelde Antijen tanıma tarafında üretilen *interlökin 2 (IL-2)* adı verilen salgılardır (Baysal 2001).

2.3.4. Efektör fonksiyonları

Efektör hücreleri, *CD4* ve *CD8* alt sınıflarına ait *CD4⁺* yardımcı *T* hücreleri ve *CD8⁺ CTL*'lerden oluşmaktadır. *CD4⁺* yardımcı *T* hücrelerinin en büyük efektör fonksiyonu, fagositleri ve çeşitli Lenfositleri, çoğunlukla sitokinler salgılayarak aktif

hale geçirmektedir. $CD8^+$ CTL'lerin en büyük efektör fonksiyonu ise Antijen taşıyan hedef hücreleri öldürmektir (Abbas ve ark. 1994).

Yardımcı T hücrelerinin efektör fonksiyonlarının önemli bir özelliği; T hücrelerine Antijeni sunan hücrelerin, aynı zamanda T hücrelerinin aktivasyonundan sonra onlara yanıt veren hücreler olmalarıdır. Hücresel bağışıklıkta makrofajlar, humoral bağışıklıkta ise B lenfositleri, yardımcı T hücrelerine Antijen sunan ve aynı yardımcı T hücrelerine yanıt vererek onlar tarafından aktive edilen hücrelerdir (Şekil 2.5).



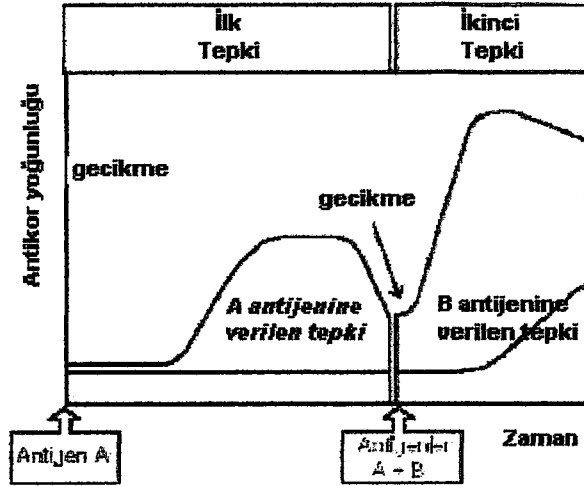
Şekil 2.5 Antijen sunan hücreler ile T hücresi etkileşimleri. (a) Hücresel bağışıklık, (b) Humoral bağışıklık.

2.4. B Hücreleri Aktivasyonu ve Antikor (Ab) Üretimi

B hücrelerinin aktif olarak rol oynadığı humoral bağışıklık, Antikorlar tarafından kontrol edilir. Antikor yanıtları, *T* hücrelerinin varlığı ile yokluğuna göre ve anatomik bölgeye bağlı olarak değişir. Protein Antijenlerine karşı gösterilen Antikor yanıtları hem *B* hücrelerini hem de *T* hücrelerini gerektirir. Bu nedenle protein Antijenleri *T*-bağımlı Antijenler olarak sınıflandırılır. *B* hücrelerinin yanıtında rol alan *T* hücreleri, *B* hücrelerinin çoğalmasında ve farklılaşmasında sağlayan yardımcı *T* hücreleridir. Polisakkarit, lipid gibi nonprotein Antijenlerde *T* hücrelerine gerek duyulmaz. Bu tür Antijenlere de *T*-bağımsız (*TI:Thymus Independent*) Antijenler denir. Yardımcı *T* hücrelerinin yer aldığı bağışıklık yanıtında spesiflik çok fazladır. Bu tür yanıtlarda, protein Antijenlerini tanıyan yardımcı *T* hücreleri, *B* hücrelerinin yüksek duyarlılıklı Antikorlar üretebilmesi için gerekli sinyalleri sağlarlar. *TI* Antijenlerine karşı gösterilen bağışıklık yanıtı ise oldukça basittir ve çoğunlukla düşük duyarlılıklı Antikorlardan oluşur (Abbas ve ark. 1994).

Bağışıklık belleğini göstermek için düşünelim ki bir *A* antijeni ortaya çıksın ve organizma içerisinde birkaç adet antikor bulsun. Bir gecikme (*lag*) fazından sonra antikor düşmanı *A* antijenini görür ve onun yoğunluğu belirli bir seviyeye kadar yükselir ve sonra azalmaya başlar (ilk tepki). Bir başka *B* antijeni ortaya çıktığı zaman hiçbir antikor bulunmaz bu da antikor tepkisinin özgüllüğünü (*specificity*) gösterir (*Janeway Jr. & Travers (1997)*). Bir başka deyişle bağışıklık belleğinin önemli bir karakteristiği *ilişkili (associative)* olmasıdır. *B* hücreleri sadece *A* antijeni değil aynı zamanda yapısal olarak *B* antijeniyle ilgili daha hızlı ve etkili ikinci tepkiyi vermek için uyarılır. Bu durum *immunological çapraz reaksiyon* veya *çapraz tepkimeli tepki* olarak adlandırılır (*Hoffman (1986), Ada & Nossal (1987), Sprent (1994), Smith et al. (1997), Hodgkin (1998), Mason (1998)*). Bu *ilişkili bellek (associative memory)* genelleme yeteneği ile diğer yapay zeka alanlarından sinir ağları olarak adlandırılır (*Haykin (1999)*). Şekil 2.6 birincil, ikincil ve çapraz tepkimeli bağışıklık tepkilerini göstermektedir. Şekil 2.6' tan da anlaşılacağı üzere ilk tepki ile karşılaştırılınca ikinci tepki yüksek bir oranda daha kısa bir gecikme fazı

ile karakterize edilmektedir. Üstelik antijenin dozu aslında ikinci tepkiye sebep olan başlangıç tepkisine gerekenden daha düşüktür.



Şekil 2.6 Birincil ve ikincil humoral yanıtın kinetikleri

2.4.1. Antijen tanıma ve B hücresi aktivasyonu

Tanıma fazı, Antijenler ile her Antijene özel küçük bir sayıda olgun *IgM* ve *IgD*'lere sahip *B* hücrelerinin etkileşmesi ile başlar. Antijenin belirli *B* hücrelerindeki *IgM* ve *IgD* moleküllerine bağlanması, aktivasyon fazını başlatır. *B* hücrelerinin Antijen reseptörleri *B* hücresi aktivasyonunda iki anahtar rolü oynar: Birincisi, *B* hücrelerine kimyasal sinyaller dağıtarak aktivasyon işlemi başlatırlar. İkincisi, bu reseptörler Antijenlere bağlanarak Antijenleri endosomal vesicle'lara dönüştürürler. Eğer Antijen bir protein ise, *B* hücresi reseptörleri onu peptidlere parçalarlar ki bu peptidler yardımcı *T* hücrelerine sunulmak üzere *B* hücrelerinin yüzeyine taşınırlar. Buradan anlaşılmaktadır ki, protein Antijenlerine karşı *T*-bağımlı bağışıklık cevabında Antijen sunan hücreler (*APC*) *B* hücreleridir. *B* hücrelerinin aktivasyonu ve Antikor üreten hücrelerin oluşumu birbirinden ayrı ardışıl fazları izler (Abbas ve ark. 1994).

2.4.2. Protein antijenlerine karşı gösterilen T-bağımlı B hücresi yanıtı

Protein Antijenlerine karşı gösterilen *B* hücresi yanıtları, Antijenin yardımcı *T* hücreleri tarafından tanınması ve Antijen spesifik *B* lenfositleri ile *T* lenfositlerinin ortak çalışmasını gerektirir.

Vücuda giren protein Antijeni *B* hücrelerinin reseptörleri tarafından tanınır. Antijeni tanıyan *B* hücrelerinin reseptörleri bu Antijeni peptidlere parçalar. Parçalanan peptidler de *B* hücrelerinin yüzeylerine taşınarak ilgili *T* hücrelerine sunulurlar. Antijenin yüzey reseptörlerine bağlanması aynı zamanda bir takım co-stimülatörlerin salgılanmasına da neden olurlar. Bu co-stimülatörler, *B* hücresinin yardımcı *T* hücrelerini aktive etme olasılığını artırır. Burada görev alan *B* hücreleri, *T* hücrelerine Antijen sunan APC'ler olarak çalışır (hücre sel bağışıklıkta bu görevi makrofajlar üstlenir). Yardımcı *T* hücreleri *B* hücrelerinin yüzeyinde sunulan peptidleri tanıyarak aktive olurlar. Aktivasyon ile *T* hücreleri tarafından salgılanan moleküller (sitokinler) yine ilgili *B* hücrelerini aktive etmek için kullanılır. *B* ve *T* hücreleri arasında gerçekleşen bu etkileşimin sonucunda ise *B* hücreleri uyarılmış olur. Buradaki sitokinler çok önemlidir çünkü bu uyarım sayesinde *B* hücreleri çoğalır ve Antikor salgılayan plazma hücrelerine dönüşürler. Antijenin vücuda girişinden 3 ila 7 gün sonra, aktive olmuş *T* hücrelerinin çok yakınında Antikor üreten *B* hücreleri görülür. *T* hücresi-*B* hücresi etkileşiminin temel bir özelliği, etkileşen lenfositlerin Antijen ile uyarılmış lenfositler olmalarıdır. Bu da, Antijen-spesifik yanıtın oluşmasını garanti eder (Abbas ve ark. 1994).

2.4.3. Yardımcı T hücrelerine bağımlı B hücreleri yanıtındaki son olaylar: germinal center reaksiyonları

T hücrelerine bağımlı *B* hücreleri yanıtlarında gerçekleşen son olaylar duyarlılık olgunlaşması ve hafıza *B* hücrelerinin üretilmesidir. Bu olaylar lenfoid foliküllerin Germinal Center adı verilen bölgelerinde gerçekleşir.

Antijen ile karşılaşmadan sonraki 4 ila 7 gün içinde, aktive olmuş *B* hücrelerinin bir kısmı foliküllerin derinliklerine gider ve orada hızlı bir şekilde çoğalırlar. Çoğalma mitotik bir şekilde gerçekleşir ve buradaki *B* hücrelerinin ikiye katlanma süresi yaklaşık 6-12 saat olarak hesaplanmıştır. Dolayısıyla bir Lenfosit, 5 gün sonra 5000 adet döl (progeny) oluşturabilir. Çoğalma sonucunda Germinal Center'larda oluşmuş olan hücrelerin hepsi, tek bir ya da birkaç Antijen-spesifik *B* klonlarından oluşmuşlardır. Çoğalmadan sonra Germinal Center'larda oluşmuş olan dölleri çok küçük hücrelerdir. Bu dölleri daha sonra farklılaşma ve seçim işlemlerine tabi tutulurlar (Abbas ve ark. 1994).

2.4.3.1. Duyarlılık olgunlaşması

Duyarlılık olgunlaşması, Germinal Center'larda gerçekleşen ve *Ig* genlerinin somatik mutasyonu ile takibinde yüksek duyarlılıklı *B* hücrelerinin seçimini kapsayan bir işlemdir. Duyarlılık olgunlaşması işlemi sadece yardımcı *T* hücrelerine bağımlı *B* hücreleri yanıtlarında gerçekleşir. Duyarlılık olgunlaşması işleminde gerçekleşen ilk olay somatik mutasyon olayıdır. Yine Germinal Center'larda gerçekleşen mutasyon sırasında *B* hücrelerinin *Ig* genlerinde mutasyonlar gerçekleşir. Bu esnada *T*-bağımlı protein Antijenlerinden gelen kalıcı ve tekrar eden uyarımlar, *Ig* genlerindeki mutasyon sayısını artırır. Bu mutasyonların bazıları işe yarar fakat çoğu kullanışsız *B* hücrelerinin oluşumu ile sonuçlanır. İşe yarayan *B*

hücreleri, bir sonraki aşamada gerçekleşecek olan seçim işlemi ile varlıklarını sürdürebilirler.

Seçme işlemi: vücuda giren Antijenlerin bir kısmı Germinal Center'lara giderler. Germinal Center'larda bulunan foliküler dentrik hücreler bu Antijenleri *B* hücrelerine sunarlar. Sunulan Antijenlere yüksek duyarlılıkla bağlanan *B* hücreleri hayatta kalmak için seçilirler, diğerleri de *apoptosis* ile ölürlür. *B* hücresi yanıtı geliştikçe daha çok Antijen yok edilir ve Germinal Center'lara daha az Antijen gider. Bu nedenle, buradaki Antijenlere bağlanabilecek olan *B* hücreleri birbirleri ile daha çok yarışır ve hayatta kalabilmek için daha yüksek bir duyarlılığa sahip olmaları gerekir. Bunun sonucunda da seçme işlemi bittikten ve Antijen vücuttan atıldıktan sonra, Antijen için ilk başta olduğundan daha çok duyarlılıklı Antikorlara sahip bir *B* hücresi popülasyonu hayatta kalır. Bu popülasyonda bulunan *B* hücrelerinin bir kısmı hafıza hücresi olarak saklanır ve uzun süreler hayatta kalır. Antijenin her vücuda girişi bir duyarlılık olgunlaşması işlemine vesile olacağı için, hafıza hücrelerinin duyarlılıkları her seferinde artar. Antijenin vücuda sonraki girişlerinde üretilen bağışıklık yanıtına ikincil bağışıklık yanıtı adı verilir.

2.4.4. T-Hücresi bağımsız antijenlere B hücresi yanıtları

Polisakkaritler ve lipidler gibi non-protein Antijenlerinin çoğu, yardımcı *T* hücrelerine gerek duymadan Antikor üretimi için *B* hücrelerini uyarırlar. Bu tür yanıtlarda, çoğunlukla *IgM* ve bazı *IgG* alt tiplerinin oluşturduğu düşük duyarlılıklı Antikorlar üretilir. Bu non-protein Antijenlerine karşı oluşturulan yanıtlarda hafıza *B* hücrelerinin üretilip üretilmediği bilinmemekle beraber, duyarlılık olgunlaşması olayı gerçekleşmemektedir (Abbas ve ark. 1994).

3. YAPAY BAĞIŞIKLIK SİSTEMLERİ

3.1. Yapay Bağışıklık Sistemlerinin Tarihçesi

Yapay Bağışıklık Sistemleri gibi Bağışıklık ta aslında tıp bilimi için oldukça yeni bir alandır. Bağışıklık (İmmünoloji) biliminin başlangıcı Edward Jenner'in aşığı bulduğu 1876 yılına dayanır. O tarihten günümüze Bağışıklık alanında yapılan keşifler, Bağışıklık Sisteminin işleyişindeki çoğu karanlık noktayı aydınlatmıştır. Bağışıklık Ağı Teorisi ile Niels K. Jerne 1974 yılında, reseptör çeşitliliği ve yapısal Bağışıklık üzerine yaptığı çalışmalar ile de Susumu Tonegawa 1983 yılında Nobel ödülüne layık görülmüştür. Bağışıklık biliminde yapılan çalışmalara paralel olarak oluşturulmaya başlanan Bağışıklık Sistemi modelleri, sistemin işleyişini kavrama açısından oldukça büyük katkılar sağlamışlardır.

1900'lü yılların ikinci yarısından itibaren gelişmeye başlayan Yapay Zeka, insan gibi düşünen ve davranan makinelerin ortaya çıkarılması gibi amaçlara yönelik çalışmalarda büyük ölçüde insanın biyolojik sisteminin işleyişini referans almıştır. Bu alanda en çok kabul gören ve Yapay Sistemlerde bir çığır açan Yapay Sinir Ağları, uygulama alanını sadece insanı taklit eden makinelerin oluşturulması ile sınırlı tutmamış, bilimin birçok alanında karşılaşılan karmaşık problemlerde başvurulabilir etkin bir çözüm sistemi olarak kendini ispatlamıştır. Sinir sistemindeki bilinen işleyişleri model alarak oluşturulmaya başlanan Yapay Sinir Ağları'nın böylesi ses getirmesi şaşırtıcı değildir çünkü insandaki biyolojik sistem, bilinen işleyişi ile bile çok karmaşık bir mükemmellik arz etmektedir. Bağışıklık sistemindeki olayların aydınlatılmaya başlanması ve Yapay Zeka' ya duyulan ilginin artması, Yapay Bağışıklık Sistemlerinin ortaya çıkışının ardında yatan temel nedenler arasında yer almaktadır. 1974 yılında Jerne' nin ortaya attığı Bağışıklık Ağı Teorisi ve Bağışıklık Sistemindeki işleyişleri daha anlaşılabilir kılan Bağışıklık

Sistemi Modelleri, Yapay Zeka arařtırmacılarının ilgisini çekmiş ve oluşturulan yeni algoritma ve uygulamalarla Yapay Bağıřıklık Sistemleri yavaş yavaş adından söz ettirmeye başlamıştır. Gelişiminin Yapay Sinir Ağları kadar hızlı olmaması ve onun kadar çok ilgi toplamamış olmamasına karşın son yıllarda bu durum deęişmeye başlamıştır. Özellikle 1990'lı yılların sonlarından itibaren Yapay Zeka alanında çalışmakta olan çoęu arařtırmaçı, dikkatini Yapay Bağıřıklık Sistemlerine yöneltmiş ve bu alanda en az Yapay Sinir Ağları kadar etkin yeni algoritmalar oluşturulmaya başlanmıştır.

Yapay Bağıřıklık Sistemlerinin uygulama alanları řu şekilde özetlenebilir: Robotik, kontrol, optimizasyon, Yapay Sinir Aęı yaklaşımları, anormallik tespiti, etmen tabanlı yaklaşımlar, öğrenme, indüktif problem çözme, örüntü tanıma, bilgisayar modelleri, dięer yaklaşımlar (Açık Web sunucusu koordinasyonu, listeleme, veri madencilięi, sınıflama sistemleri, sensör tabanlı teşhis, gen kütüphanelerinin geliřimi, self-tanıma işlemleri).

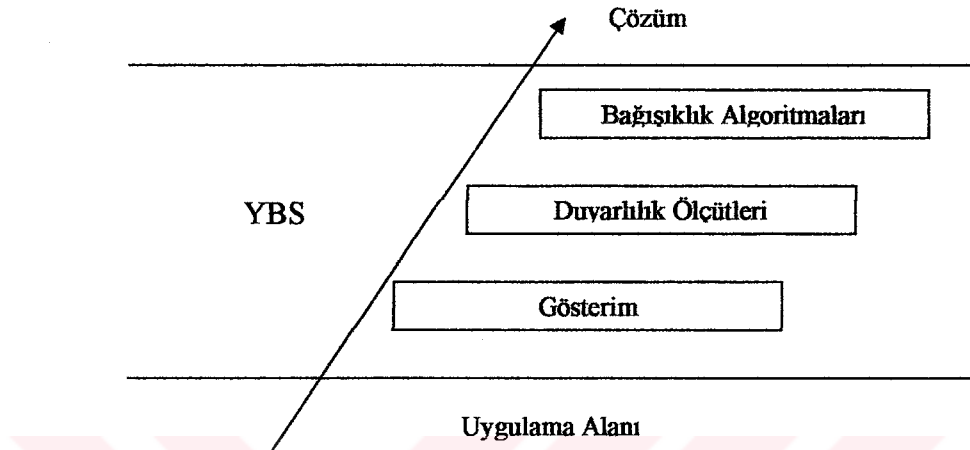
3.2. Bir Yapay Bağıřıklık Sisteminde Kullanılan Gereklil Kavramlar

Bağıřıklık sistemindeki etkileşimlerin sistem bazında ifade edilerek bir Yapay Bağıřıklık Sistemi oluşturulması için ařaęıdakilere ihtiyaç duyulur (Şekil 3.1):

- Sistemi oluşturan birimlerin gösterimi
- Sistemdeki birimlerin birbirleri ve çevre ile olan etkileşimlerini hesaplamak için bir mekanizma
- Bazı adaptasyon prosedürleri

Bir YBS tasarımında öncelikle sistemi oluşturan birimlerin Bağıřıklık Sisteminin hangi elemanlarının modellenmesinden ortaya çıkacaęının belirlenmesi gereklidir. Sonrasında bu birimlerin sistemdeki dięer birimlerle ve çevre ile olan

etkileşimlerinin türü ve boyutunun hesaplanmasında kullanılacak bazı duyarlılık ölçütleri oluşturulmalı ya da kullanılmalıdır. Son olarak ta sistem birimlerinin fonksiyonlarını belirleyen ve sistem birimlerinin hangi durumlarda nasıl davranması gerektiğini karakterize eden bir takım adaptasyon prosedürleri kullanılmalıdır.



Şekil 3.1 Yapay Bağışıklık Sisteminin katmanlı yapısı

Yapay Bağışıklık Sistemleri alanında oluşturulacak her sistemin temelini uygulama alanı oluşturur. Uygulama alanının karakteristiğine uygun bir gösterim seçilmeli, duyarlılık ölçütleri de bu gösterimler baz alınarak oluşturulmalıdır. Gösterim şekli ve duyarlılık ölçütleri belirlendikten sonra oluşturulacak bağışıklık algoritmaları yukarıda da bahsedildiği gibi sistem birimlerinin fonksiyonlarını belirler. Örneğin bir tasarımcı optimizasyon problemi için bir YBS sistemi oluşturmak istemektedir. Tasarımcı öncelikle bir gösterim şekli belirlemelidir. Bunun için bağışıklık sistemindeki Antikorları sistem birimleri olarak ifade etmeyi seçebilir ve bu Antikorların her birini binary diziler yardımıyla ifade edebilir. Bu durumda tasarımcının seçmiş olduğu gösterim şekli binary dizilerdir ve duyarlılık ölçütlerini bu gösterim şeklini dikkate alarak belirlemelidir. Tasarımcının belirleyeceği duyarlılık ölçütü binary diziler söz konusu olduğu için hamming uzaklık ölçütü olarak seçilebilir. Son ve en önemli aşama olan bağışıklık algoritmasının oluşturulmasında ise tasarımcı bağışıklık sistemindeki bazı mekanizmaları seçmiş olduğu gösterim şekli ve duyarlılık ölçütünü dikkate alarak

modelleyebilir. Bu modellemeler bire bir bağışığıdaki bazı mekanizmaları kullanabileceđi gibi bu mekanizmalardan esinlenilerek oluşturulmuş ve sırf bağışıklık metaforlarından oluşmayan modellemeler de olabilir. Örneđin bağışıklık sistemindeki duyarlılık olgunlaşması mekanizmasından yola çıkılarak bir bağışıklık algoritması oluşturulabilir.

Yapay Bağışıklık Sistemleri alanında, yukarıda bahsedilen işlemlerin her biri için çeşitli yöntemler geliştirilmiştir. Sistemin oluşmasında ilk yapılması gereken iş olan sistemi oluşturan birimlerin gösterimi için Perelson ve Oster'in 1979 yılında ortaya attıkları şekil-uzay yaklaşımı, en çok kabul gören ve kullanılan yaklaşımdır (Perelson ve Oster 1979). Antikor ve Antijen arasındaki etkileşimler de duyarlılık ölçütünün kullanılması aşamasında en çok tercih edilen modellemelerden birisidir. En son aşamada da klonsal seçme algoritması (Ada ve Nossal 1987), negatif seçim algoritması (Forrest ve ark. 1994), bağışıklık ađları, deđişik uygulama alanları için oluşturulmuş ve oldukça çok kullanılan yöntemlerdir.

3.3. Şekil Uzay Modeli

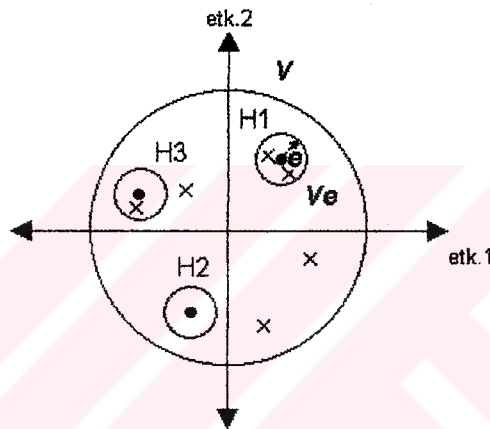
Antijen tanıma için gerekli olan tüm hücrelerin oluşturduđu kümeye, *bađışıklık repertuarı* adı verilmektedir. Vücudumuzun tüm antijenlerden korunabilmesi için bu repertuarın tam olması gereklidir. Vücut, herhangi bir antijeni tanıyabilecek reseptör molekülerini oluşturacak protein genlerine sahiptir ve bu genlerin farklı kombinasyonlarda dizilimi ile doğada bulunan hemen hemen tüm antijenleri tanıyabilir. Bir başka ifade ile *vücudumuzun bağışıklık repertuarı tamdır*.

Bađışıklık hücreleri kümesi zararlı saldırganlardan vücudu uygun şekilde korumak için bağışıklık tepkisi olarak adlandırılan antijenleri tanımaya çalışırlar. Nicel olarak bağışıklık hücresi molekülleri ve antijenler arasındaki etkileşimleri açıklamak için *Perelson & Oster (1979) şekil uzay (S- shape space)* kavramını ortaya çıkarmıştır.

Şekil uzay modelinde bir reseptör ve bir molekül arasında bağlayıcılıđın derecesi temel alınır. L adet parametreden oluşan antikor bağlanma yerinin

(*paratop*) olduğu kabul edilir. Parametrelerin kesin sayısı veya onların değerleri burada önemli değildir. Şekil uzay olarak adlandırılan L boyutlu uzayda bir nokta antijene bağlanan bölgenin oluşturulmuş şeklini belirtir.

Bağışıklık sisteminde B ve T lenfositlerinin antijenik yapıları tanıyabilmeleri için bu antijenik yapılardan gelen uyarımların belirli bir eşik seviyesini aşması gerekmektedir. Bu özellik ise şekil uzayında *tanıma topu* adı verilen modelleme ile ifade edilmiştir. Şekil uzayında gösterilen her bir self hücrenin (B veya T hücresi) yarıçapı bir eşik seviyesi ile belirlenen bir tanıma çemberi vardır. Her bir self hücre sadece bu çemberin içinde kalan antijenleri tanıyabilir. Şekil 3.2’te tanıma çemberi kavramı ve şekil uzayında repertuar tamlığı gösterilmektedir.



Şekil 3.2 Tanıma topu kavramı ve repertuar gösterimi.

Şekil 3.2’te üç tane self hücre ve antijenler iki boyutlu bir şekil uzayında gösterilmiştir. Her bir hücrenin tanıma çemberi e eşik seviyesi ile belirlenen V_e çemberidir. Şekilde $H1$ tanıma çemberi içinde kalan iki antijeni, $H3$ ise sadece bir antijeni tanır. $H2$ 'nin tanıma çemberinde herhangi bir antijen olmadığı için $H2$ antijen tanımamaktadır. Doğada mevcut olabilecek antijen türleri, protein dizilimleri ile belirlendiği için belirli bir çeşitlilikte bulunabilirler. Şekil 3.2’teki iki boyutlu şekil uzayında bulunabilecek tüm antijen türleri bir V hacmi ile ifade edilmiştir. Başka bir deyişle bulunabilecek tüm antijenler mutlaka bu hacim içerisinde gösterilirler. Yine aynı şekil uzayında bulunan her bir antikorun bir V_e tanıma hacmi olduğuna göre, şekil uzayında V hacmi içerisinde tüm noktaları kapsayacak biçimde bir antikor popülasyonu olduğunda bu antikor popülasyonu tüm antijenleri

tanıyabilir. Bu durumda bu antikor popülasyonunda repertuar tamlığından söz edilebilir.

3.4. Antijen-Antikor Gösterimleri ve Duyarlılıklar

Antijen-Antikor gösterimi kısmen etkileşim derecesini (tamamlayıcılık) hesaplamada kullanılan uzaklık ölçüsüne karar verecektir. Matematiksel olarak bir m molekülünün genelleşmiş şekli ya bir antikordur (Ab) ya da bir antijendir (Ag) ve L boyutlu gerçekte değer uzaylı ($m \in S^L \subseteq R^L$, S şekil uzayı ve L boyutunda sunulur) bir nokta olarak kabul edilen $m = \langle m_1, m_2, \dots, m_L \rangle$ gerçekte değerli koordinatlarının bir kümesi ile gösterilebilir.

Bir antijen ve bir antikor arasındaki *duyarlılık (affinity)* iki string (veya vektör) arasındaki her hangi uzaklık ölçüsü yoluyla tahmin edilebilen uzaklık ile ilişkilidir.

Antikor : $Ab = \langle Ab_1, Ab_2, \dots, Ab_L \rangle$;

Antijen : $Ag = \langle Ag_1, Ag_2, \dots, Ag_L \rangle$;

Ab ve Ag arasındaki Öklid uzaklığı:

$$D = \sqrt{\sum_{i=1}^L (Ab_i - Ag_i)^2} \quad (3.1)$$

Ab ve Ag arasındaki Manhattan uzaklığı:

$$D = \sqrt{\sum_{i=1}^L |Ab_i - Ag_i|} \quad (3.2)$$

Gerçel (real) değerli koordinatlar kullanan ve (3.1) eşitliğindeki gibi uzaklığı ölçen şekil-uzayları *Euclidean şekil uzayı* olarak adlandırılır (Segel & Perelson (1988), DeBoer et al. (1992), Smith et al. (1997)). *Gerçel* değerli koordinatlı fakat *Euclidean* uzaklığı yerine *Manhattan* uzaklığı kullanan şekil-uzayı *Manhattan şekil uzay* olarak adlandırılır.

Antijenler ve antikorlar *gerçel* değerli vektörler yerine *ikili(binary)* sembollerle ifade edilirse farklı bir uzaklık ölçütü olan *Hamming* uzaklık ölçütü kullanılır. Bu durumda antijen ve antikorların gösterildiği vektör elemanları *gerçel* değerli sayılardan değil, ikili sayılardan oluşur. Kullanılan *Hamming uzaklık ölçütü* (3.3) eşitliğinde gösterilmiştir.

$$D = \sum_{i=1}^L \delta, \quad \delta = \begin{cases} 1 & Ab_i \neq Ag_i, \\ 0 & \text{diğer} \end{cases} \quad (3.3)$$

Şekil uzayında oluşabilecek toplam Antikor sayısı k^L ile ifade edilir. Burada k , kullanılan alfabenin büyüklüğü, L ise vektör uzunluğudur. Şekil uzayındaki her bir Antikoron tanıma çemberi içerisinde kalabilecek Antijen sayısına o Antikoron *kapsamı* denir ve şu eşitlik ile hesaplanır:

$$C = \sum_{i=0}^{\varepsilon} \binom{L}{i} = \sum_{i=0}^{\varepsilon} \frac{L!}{i!(L-i)!} \quad (3.4)$$

L : vektör uzunluğu,

ε : duyarlılık eşiği (tanıma çemberi yarıçapı).

L uzunluğundaki bir vektör ve k büyüklüğündeki bir alfabe ile toplam oluşabilecek Antijen sayısının k^L olduğu göz önüne alındığında, yukarıdaki eşitlikten de yararlanılarak, şekil uzayındaki tüm Antijenlerin tanınması için gerekli minimum Antikor sayısı şu şekilde hesaplanabilir (De Castro ve Arkadaşları, 2000):

$$N = \text{ceil}\left(\frac{k^L}{C}\right) \quad (3.5)$$

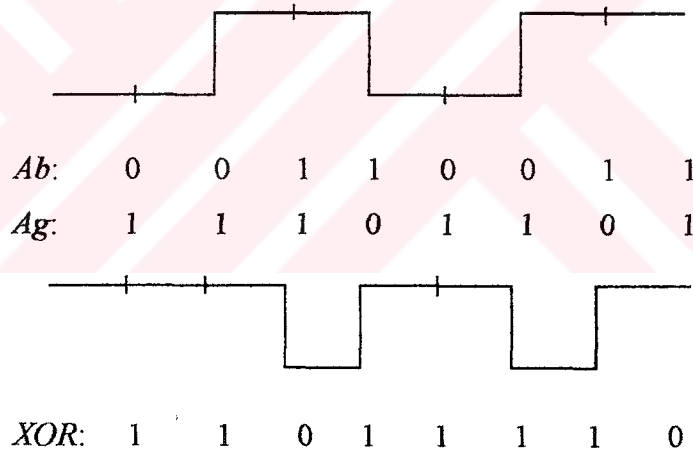
Bu eşitlikteki *ceil* operatörü, parantez içindeki ondalıklı sayıyı en yakın büyük tam sayıya yuvarlar.

Antikorların Antijenlere yanıt üretebilmeleri için onları tanımaları gerekmektedir. Tanıma işlemi için ise Antikorların Antijene olan duyarlılıklarının belirli bir eşik seviyesini (ϵ) aşması gereklidir ki bu duyarlılıklar da uzaklık ölçütleri kullanılarak hesaplanır. Binary gösterimin olduğu bir örnekle tanıma işlemi şu şekilde açıklanabilir:

Antikor: $Ab=[0\ 0\ 1\ 1\ 0\ 0\ 1\ 1]$;

Antijen: $Ag=[1\ 1\ 1\ 0\ 1\ 1\ 0\ 1]$;

Etkileşim:



Duyarlılık (eşleşme skoru)=6

Şekil 3.3 Hamming uzaklığı ve duyarlılık hesabı

Söz konusu binary vektörler için uzaklık, Hamming uzaklık ölçütü kullanılarak hesaplandığında (*XOR* işlemi ile) 6 olarak bulunur. Farklı bir ifade ile, bu iki vektörün 6 tane elemanı birbirini tamamlar nitelikte olduğu için aralarındaki uzaklık 6'dır.

3.5. Yapay Tanıma Toplari ve Kaynak Rekabeti

AIRS algoritmasındaki bir B hücresi tek bir antikoru temsil edecek şekilde basitleştirilmiştir. Bu antikor stringdeki her bir elemanın farklı bir niteliği temsil ettiği problemin nitelik arama uzayında bir string olarak düşünülebilir. Bundan dolayı bir B hücresi tüm özellik uzayında belirli bir noktayı ifade etmektedir. Bu anlamda *Timmis and Neal (2000)* tarafından, *Yapay Tanıma Toplari (ARB)* kavramı ilk kez önerilmiş, *Timmis (2001)* ve *Knight and Timmis (2001)* çalışmalarında daha da geliştirmişlerdir.

ARB aynı antikora sahip çok sayıda B hücresi olarak tanımlanabilir. *Timmis and Neal (2000)* kaynaklar olarak *ARB*'nin B hücrelerinin sayısını ifade etmişlerdir. Bu yaklaşımın devamında bir yapay bağışıklık sistemi sistemin temel bellek mekanizmasını oluşturmak için çok sayıda *ARB*'yi içermektedir. Uygun kaynakların sayısı belirli bir sayı ile sınırlıdır. Öğrenme kaynakların rekabeti ile meydana gelmektedir. *ARB*'ler saldıran antijene tepki verecek şekilde uyarılır. Belirli bir antijen (veya antijen popülasyonu) sisteme geldikten sonra her bir *ARB* uyarma seviyesini temel alarak kaynakları tüketmeye teşebbüs edecektir. Bununla beraber kaynakların sayısı sınırlı olduğu için sadece en çok uyarılan *ARB*'ler kaynakları tüketecektir. Kalan *ARB*'ler (örneğin her hangi bir kaynağı olmayanlar) sistemden çıkartılır. Kaynaklar için bu rekabet sistemde sadece en güçlü *ARB*'lerin kalmasını garantilemektedir.

3.6. Bellek Hücreleri, Mutasyon ve Klonlama Seçimi

B hücrelerinin bir popülasyonu bilinmeyen bir antijenle karşılaştığı zaman, en fazla uyarılan hücreler saldırganla savaşıyan yeni ürünler oluşturmak için klonlanacak ve mutasyona uğratılacaktır. Bu bağışıklık sisteminin ilk tepkisi olarak düşünülebilir. Bu ilk tepki boyunca belirli hücrelerin özellikle önlenmesine belirli

antijenlerin veya yapısal olarak benzer antijenlerin gelecekte karşılaşma olasılığına karşı sistemde kalmasına izin verilir. Bu kalıcı hücreler bundan sonra bellek hücresi (*De Castro and Von Zuben 2001b*) olarak adlandırılır ve bağışıklık sisteminin ikincil tepkisi için temel teşkil ederler. Bağışıklık sistemi daha önceden karşılaşmış antijenlerle ya da benzer antijenlerle karşılaştığı zaman bellek hücreleri hızlıca uyarılır ve daha büyük oranda klonlanır ve mutasyonlarla yeni ürünler üretilir. Bu bellek hücreleri bağışıklık sisteminin bilinen tehlikeleri tanıma ve uygun olarak tepki verme yeteneğini sağlamaktadır. Burada önemli olan bellek hücrelerinin sadece orijinal antijene tepki vermekle kalmaması aynı zamanda yapısal olarak benzer antijenlere tepki verme yeteneklerinin bulunmasıdır.

AIRS algoritması bu bellek hücrelerinin davranışından sonra nesnelere modellemektedir. *De Castro and Von Zuben (2001a) ve (2001b)* tarafından *YBS* için bellek hücrelerinin gelişiminde, klonlama seçimi prensibinin ve duyarlılık olgunlaşmasının rolü tartışılmıştır. *YBS*'de klonlama seçimi temel olarak bir antijen için en büyük duyarlılığa sahip veya en çok uyarılan hücrelerin yeni ürün oluşturmak için seçilmesidir. *ARB* kavramı kullanıldığından dolayı üretilen ürün gerçekte sadece mutasyona uğramış ürün üretimidir. *AIRS* algoritmasında mutasyon, rastgele nitelik mutasyonu şeklindedir. Eğer bir *ARB* antikorunun belirli bir niteliği mutasyon için seçilirse, o nitelik belirli bir aralık içerisinde herhangi bir rastgele değer ile mutasyona uğratılır. *AIRS* algoritmasında duyarlılık olgunlaşması antijen tarafından en fazla uyarılan ürünleri üreten hücrelerin hayatta kalması ile sağlanır. Bu işlem bölüm 3.5' de tartışılan kaynaklar için rekabeti temel almaktadır. Sadece başarılı olarak elde edilmiş kaynaklara klonlama için izin verilir. Sonuçta klonlama seçimi, nitelik mutasyonu ve duyarlılık olgunlaşması sürecinde belirli bir antijen tarafından en çok uyarılan tek bir *ARB* bellek hücresi olarak sistemde kalması için seçilir. Antijen popülasyonu bir kez tamamlanınca, geliştirilen bellek hücresi popülasyonu yeni antijenler için sınıflandırmada temel oluşturacak duruma gelmiş olur.

3.6. Baęışıklık Ağları ve Hücreler Arası Etkileşimler

AIRS algoritmasının gelişmesinde rol oynamış olan son baęışıklık sistem prensibi baęışıklık ağ teorisidir. *Jerne (1974)* tarafından sunulan ve *Perelson (1989)* tarafından geliştirilen baęışıklık ağ teorisinde baęışıklık tepkileri sadece *B* hücreleri ile patojenlerin etkileşimini temel almaz aynı zamanda *B* hücrelerinin dięer *B* hücreleri ile etkileşimini de temel almaktadır. *B* hücreleri birbirlerini hem uyarırlar hem de bastırırlar ve bu etkileşim sonucunda bellek baęışıklık sisteminde alıkonulur. *Timmis'* in *baęışıklık ağ teorisi* danışmansız öğrenme, veri analizi ve veri görselleştirmede başarılı bir metottur. Bu çalışmanın önemli noktalarından biri hücreler arası duyarlılık ve bağlantı kuvvetlerinin düzenlenmesi fikridir. *Timmis'*in çalışmasında *ARB'* lerin bir ağı *ARB'* ler arasındaki benzerlięi veya duyarlılıęı temel alan formdadır. Eęer iki *ARB* birbirleri için yeterli duyarlılıęı gösterirse bu iki *ARB* birbirleri ile bağlantılır. Bu bağlantılar bağlantılı *ARB'* lerin desteklendięi bir mekanizma saęlarlar ve gelen antijenlere tepki verirler.

Bu ağ fikri *AIRS* algoritmasında kullanılmazken hücreler arası duyarlılık kavramı önemli bir rol oynamaktadır. *AIRS* algoritmasının özelliklerinden biri veri azaltılmasını uygulama yeteneęidir. *Eęitilmiş AIRS sınıflandırıcısı* tipik olarak orijinal eğitim örneklerinden çok daha az bellek hücrelerine sahiptir. Bu durum bellek hücreleri yer deęişimi ile başarılı ve bellek hücreleri yer deęişimi için anahtar kriter yeni geliştirilmiş bellek hücreleri ve mevcut bellek hücreleri arasındaki hücre etkileşiminin temel alınmasıdır.

3.7. Sınıflandırma

AIRS için temel teşkil eden prensiplerin anlatıldıęı bu bölümü bitirmeden önce sınıflandırma veya danışmanlı öğrenme için *YBS'* nin özet bir ifadesinden bahsedilmesi gerekir. Bugüne kadar bir danışmanlı öğrenme sistemi geliştirmek için baęışıklık sistemi prensiplerini kullanılmaya sadece bir kez teşebbüs edildięi

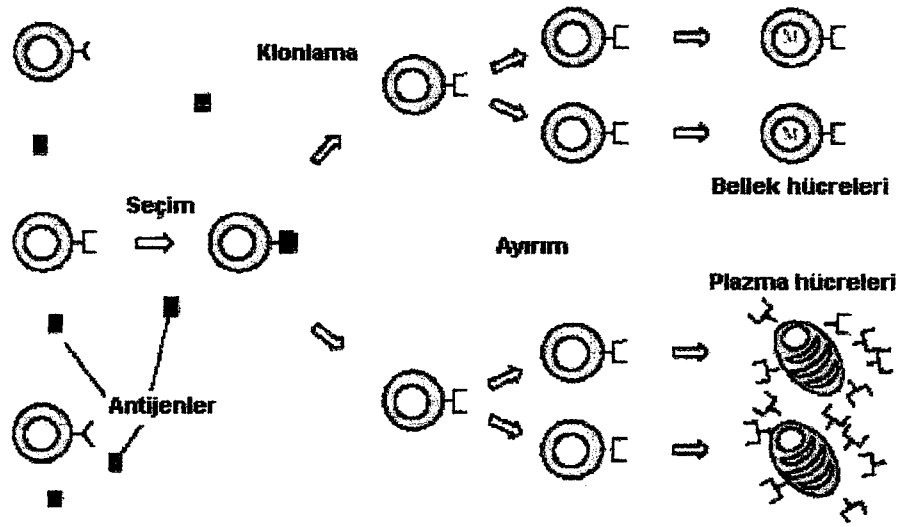
bilinmektedir. *Carter (2000)* bağışıklık sistemi prensiplerini modelleyen danışmanlı bir sınıflandırma metodu sunmuştur. Bu bölümde tartışılan diğer *YBS*'ten farklı olarak *Carter* hem *T* hücreleri hem de *B* hücrelerini modellemiştir ve bu iki hücre arasındaki etkileşimler sınıflandırmanın geliştirilmesinde önemli bir rol oynamıştır.

3.8. Klonlama Seçimi Algoritması

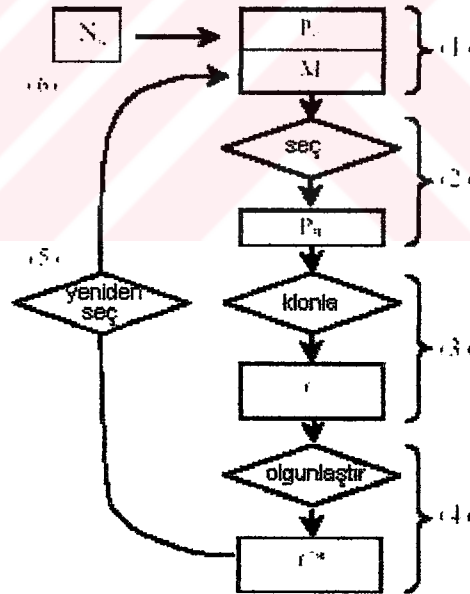
Klonlama seçimi bir antijenik uyarıcı için bağışıklık tepkisinin temel özelliklerini açıklamak için bağışıklık sistemini kullanan bir algoritmadır. Bu prensip sadece antijenlerin hızlıca çoğalmasını sağlayacak hücrelere uygulanır. *Klonlama seçimi* hem *T* hücreleri hem de *B* hücrelerine uygulanır.

Bağışıklık sisteminde *T* ve *B Lenfositlerinin* olgunlaşması sırasında (*duyarlılık olgunlaşması*) birçok seçme mekanizması devreye girmektedir. *De Castro ve Von Zuben* duyarlılık olgunlaşmasındaki işlemleri temel alarak *Klonlama Seçimi Algoritmasını* ortaya atmış ve oluşturdukları algoritmayı karakter tanıma, optimizasyon gibi problemlere uygulayarak performansını analiz etmişlerdir (*De Castro and Von Zuben (1999) ve (2000a)*).

Klonlama Seçimi Algoritması iki temel esas üzerine şekillendirilmiştir: birincisi sadece antijeni tanıyan hücreler çoğalma için seçilirler, ikincisi seçilen ve çoğalan hücreler duyarlılık olgunlaşması işlemine tabi tutularak, antijene olan duyarlılıkları artırılır. *Klonlama Seçimi Algoritmasının* biyolojik olarak gösterimi Şekil 3.4' de blok diyagramı ise Şekil 3.5'de verilmiştir.



Şekil 3.4 Klonlama seçimi algoritmasının biyolojik olarak gösterimi.



Şekil 3.5 Klonlama Seçimi Algoritmasının blok diyagramı

Algoritmanın çalışması şu şekildedir:

- Aday çözümlerden oluşan bir populasyon oluşturulur (P). Bu populasyon, hafıza hücreleri (M) ve geri kalan populasyondan (P_r) oluşur ($P=P_r+M$).
- Duyarlılık ölçütüne bağlı olarak populasyondaki en iyi n tane eleman seçilir ve P_n populasyonu oluşturulur.
- Seçilen bu en iyi bireyler antijene olan duyarlılıklarına bağlı olarak klonlanırlar (çoğalırlar). Duyarlılığı yüksek olan bireylerin oluşturacağı klon sayısı yüksek, duyarlılığı düşük olan bireylerin klon sayısı ise düşük olacaktır. Böylelikle bir C klon kümesi oluşturulmuş olur.
- Klondan oluşan C populasyonu bir hipermutasyon işlemine tabi tutulur. Hipermutasyon işlemi de duyarlılık ile doğru orantı söz konusudur. Hipermutasyondan sonra oluşan sete ise C^* populasyonu denir.
- Klonlama ve hipermutasyondan sonra gelişmiş bireyleri popülasyona eklemek için tekrar bir seçme işlemi yapılır ve M hafıza seti oluşturulur. Bu seçme işleminden sonra P popülasyonundaki bazı hücrelerin yerlerini C^* setindeki bazı hücelere bırakmaları olasıdır.
- Popülasyonda çeşitliliği sağlamak amacıyla P popülasyonundaki d tane birey yeni üretilen bireyler ile yer değiştirir. Düşük duyarlılıklı bireylerin yer değiştirme ihtimali daha fazladır.

Yapılan çalışmalarda görülmüştür ki, *Klonlama Seçimi Algoritması* standart bir *Genetik Algoritma* ile karşılaştırıldığında lokal optimumlardan oluşan çok çeşitli bir çözüm seti üretebilir. Genetik Algoritmalar ise tüm popülasyonu en iyi bireye benzetmeye çalışırlar. Gerçekte iki algoritmanın da kodlama ve hesaplama yöntemleri çok farklı değildir fakat araştırma işlemlerini gerçekleştirirken esinlendikleri kaynak, kullandıkları notasyon ve gerçekleştirdikleri işlemlerin sırası bakımından farklılık arz etmektedirler.

3.9. Takviye ile Öğrenme (Reinforcement Learning) ve Bağışıklık Belleği

Korunmak için antijeni tanıma yeterli değildir, bağışıklık sistemi aynı zamanda patojenlere karşı etkin bir tepki verecek yeterli kaynaklara sahip olmalıdır. *Patojen* popülasyonunun boyutu ile ilişkili olarak patojen için *lenfositin* popülasyonunun boyutu enfeksiyonun sonucuna karar vermek çok önemlidir. Bağışıklık sistemindeki *lenfositlerin* toplam sayısı düzenlendiği için bazı klonların boyutu artarken diğer klonların boyutu azalabilir. Bununla beraber *lenfositlerin* toplam sayısı tamamen sabit kalmaz. Eğer bağışıklık sistemi sadece belirli *lenfositlerin* popülasyon boyutu artınca öğreniyorsa, ya boyutu arttığı için önceden öğrendiği antijenleri unuttur ya da yeni antijenlere tepki vermeden sorumlu ve rastgele oluşturulmuş *repertuar(repertoire)* sürekli olarak azalır (*Perelson & Weisbuch (1997)*).

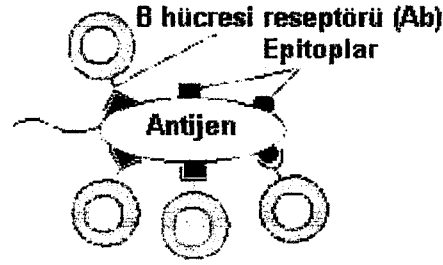
Bağışıklık sisteminin gelişiminde bir organizmanın hayatı boyunca defalarca belirli bir antijenle karşılaşması beklenebilir. Adaptif bağışıklık tepkisini uyaran bir antijen ilk açığa çıktığında, her bir farklı eğilim için antikör üreten *B* hücrelerinin küçük klonları tarafından ele alınır. İkinci kez karşılaşınca bağışıklık tepkisi ilk tepkideki yüksek eğilimli antikör üreten hücreler (bellek hücreleri) depolandığı için gelişmiş olur (*Ada & Nossal (1987)*). Her zaman sıfırdan başlamaktan ziyade bu strateji uygulanarak her bulaşmadan sonra bağışıklık tepkisinin hem hızı hem de doğruluğunun artması sağlanır (*Perelson et al. (1978), Farmer et al. (1986)*). Bu yapı takviye ile öğrenme stratejisine özgüdür (*Sutton&Barto (1998)*). Sistem sürekli olarak görevin uygulanma yeteneğini geliştirir.

Bazı yazarlar (*Allen (1987), Coutinho (1989)*) önceki antijene tepki vermiş uzun ömürlü *B* hücrelerinin durumunu koruduğunu ve ikinci antikör tepkisinde önemli bir rol oynadıklarını ileri sürmüştür. Bu bellek hücreleri diğer hücrelerden en azından fonksiyonel olarak bağlantısızdır ve bellek en azından ikinci tepkide klonal bir özelliktir.

Burada önemli bir nokta en yüksek eğilimli hücrelerin en yüksek kalitedeki aday çözümleri olarak korunmalarının gerekliliğidir.

3.10. Örüntü Tanıma (Pattern Recognition)

Bağıışıklık sisteminde örüntü tanımanın bakış açısından hem *B* hem de *T* hücrelerinin en önemli özelliđi antijenleri tanıyabilen ara yüzlerindeki reseptör moleküllerine sahip olmasıdır (Şekil 3.6).



Şekil 3.6 Epitop olarak adlandırılan ve antikor tarafından tanınan antijen parçası.

Bağıışıklık sisteminde tanıma, moleküler seviyede meydana gelir ve reseptörün bağlandığı yer ile *epitop* olarak adlandırılan antijen parçasının birbirini tamamladığına dayanmaktadır. Antikorlar tek bir reseptör türüne sahip iken antijenler birden fazla *epitop*'a sahip olabilir. Bunun anlamı tek bir antijenin farklı antikor molekülleri tarafından tanınabildiğidir.

B ve *T* hücre reseptörleri bir antijenin farklı özelliklerini tanırlar. *B* hücre reseptörleri bozulmamış antijen moleküllerinde bulunan *epitop*larla birlikte birbirlerini etkiler. Antijen moleküllerini eritilebilir. *T* hücre reseptörleri sadece hücre arayüz molekülleri ile etkileşimdedir. *T* hücreleri bağıışıklık tepkisinin düzenlenmesinde çok önemli bir rol oynayan diđer hücreleri öldürebilen veya onların gelişmesini ilerleten kimyasal maddeler salgılar.

4. AIRS ALGORİTMASI VE ÖZELLİK SEÇME (FS) İŞLEMİ

4.1. AIRS Algoritmasının Tanımı ve Algoritmada Kullanılan Tanımlamalar

AIRS, doğal ve yapay bağışıklık sistemlerinde gösterilen mekanizmalar tarafından önerilen kaynak sınırlamalı bir yapay bağışıklık sınıflayıcısı olan yeni bir danışmalı öğrenme sistemidir. Bu bölümde, *AIRS* algoritmasının uygulanmasında kullanılan önemli terimler ve kavramların tanımlaması sunulmuştur.

- *Duyarlılık(affinity)*: İki *antikor* veya *antijen* arasındaki *yakınlığın* veya *benzerliğin* bir ölçüsüdür. Uygulamada bu değerin 0 ile 1 aralığını kapsayacak şekilde kalması garantilenmiştir ve bu değer iki nesnenin özellik vektörlerinin öklit (Euclidean) uzaklığı olarak basitçe hesaplanmıştır. Böylece küçük uzaklık değerleri güçlü duyarlılığa işaret etmektedir.
- *Duyarlılık eşiği (AT-Affinity Threshold)*: Eğitim kümesindeki antijenlerin veya bu eğitim antijenlerinin seçilmiş olan alt kümesindeki antijenler arasındaki ortalama duyarlılık değeridir.
- *Duyarlılık eşik ölçüsü (ATS-Affinity Threshold Scalar)*: 0 ile 1 arasında bir değer olan duyarlılık eşiği ile çarpıldığı zaman *AIRS* eğitim rutininde bellek hücresi yer değişimi için bir kesme değeri sağlamaktadır.
- *Antikor(Antibody)*: Çıkış veya sınıf ile bir özellik vektörünün birleştirilmiş halidir. *Özellik vektörü-çıkış* kombinasyonu bir *ARB'* nin veya bellek hücresinin bir parçası olduğu zaman bir antikor olarak kullanılır.
- *Antijen*: Antikor ile aynı şekilde temsil edilir. Bununla beraber *özellik vektörü-sınıf* kombinasyonu uyarma veya tepki için sisteme sunulduğu zaman bir antijen olarak kullanılır.

- *Yapay Bağışıklık Tanıma Sistemi (AIRS-Artificial Immune Recognition System)*: Doğal bağışıklık sistemlerinden esinlenen bir sınıflandırma algoritmasıdır.
- *Yapay Tanıma Topu (ARB-Artificial Recognition Ball)*: Aynı zamanda B hücresi olarak bilinmektedir. ARB bir antikoru, hücre tarafından tutulan kaynakların sayısını ve hücrenin o anki uyarma değerini içermektedir.
- *B hücresi*: *Yapay Tanıma Topu (ARB)* olarak kullanılmaktadır.
- *Aday Bellek Hücresi (Candidate Memory Cell)*: Bir antijen sisteme sunulduktan sonra en çok uyarılan eğitim antijeni ile aynı sınıftaki bir ARB' nin antikoru.
- *Sınıf*: Verilmiş olan bir özellik vektörünün kategorisidir. Bu aynı zamanda bir hücrenin çıkışı olarak da kullanılmaktadır.
- *Klonlama oranı*: Belirli bir ARB' nin üretmeye izin verilen kadar mutasyona uğramış klonlarının sayısına karar vermek için kullanılan bir tamsayı değeridir. Uygulamada seçilmiş bir ARB' nin belirli bir antijen verdikten sonra (*klonlama_oranı*uyarma değeri*) kadar mutasyona uğramış klonların üretilmesine izin verilir. Bu üretim aynı zamanda bir ARB için kaynak tahsisinde kullanılır. Bundan dolayı *klonlama oranı* hücre popülasyonu için *kaynak tahsisi faktörü* ve *klonal mutasyon faktörü* olarak iki rolde hizmet sunmaktadır.
- *Yerleşmiş Bellek Hücreleri*: Kaynakların rekabeti ile hayatta kalan ve belirli bir eğitim antijeni için en çok uyarılan bir ARB' nin antikoru ve bellek hücrelerinin geliştirilmiş kümesine eklenir.
- *Özellik vektörü*: Verinin bir örneği, değerlerin bir sırası olarak ifade edilir. Sıradaki her pozisyon veri içerisindeki farklı bir özelliği ifade etmektedir. Her bir özellik uygun değer aralıklarına sahiptir.
- *Hipermutasyon oranı*: Belirli bir hücre popülasyonu içerisine eklenmesine izin verilen bellek hücrelerinin mutasyona uğramış klonlarının sayısına karar vermek için kullanılan bir tam sayı değeridir. Uygulamada seçilmiş bellek hücrelerine en az *hipermutasyon oranı*klonlama oranı*uyarma değeri* kadar hücre popülasyonu içerisine mutasyona uğramış klon eklenir.

- *K en yakın komşu (KNN)*: Daha önceden gözlemlenmemiş veriyi k adet en yakın veri noktası arasında bir çoğunluk oylaması ile karar vererek sınıflandıran bir sınıflandırma yöntemidir. *AIRS* algoritması için k adet en yakın veri noktaları gerçekte belirli bir test antijeni ile k adet en çok uyarılan bellek hücreleridir.
- *K değeri*: Belirli bir test verisinin sınıflandırmasında kaç adet bellek hücresi kullanılacağına karar vermek için kullanılan parametredir.
- *Bellek hücresi (mc)*: Antijenin sisteme sunulması sonunda eğitim antijeni tarafından en fazla uyarılan bir *ARB*' nin antikordur. Bellek hücresi gelen eğitim antijenlerine tepki de *hiper mutasyon* için kullanılır. Bir *mc* yenisi ile değiştirilebilir. Bu sadece bir aday *mc*, eğitim antijenini en çok uyararak yerleşmiş *mc*' den daha fazla uyardığı ve yerleşmiş *mc* ve aday *mc* arasındaki duyarlılık, duyarlılık eşiği ve duyarlılık eşik ölçüsünden daha az olduğu zaman meydana gelir.
- *Mutasyon oranı*: Bir *ARB*' nin herhangi bir özelliğinin (veya çıkışın) mutasyona uğrama olasılığını işaret eden 0 ile 1 arasında bir parametredir.
- *Çıkış*: Bir hücre ile birleştirilmiş sınıflandırma kategorisidir. Hücreye uygun özellik vektörünün sınıfı ile aynıdır.
- *Kaynaklar*: Sistemde izin verilen *ARB*' lerin sayısını sınırlayan bir parametredir. Her bir *ARB*' ye uyarma değerini ve klonlama oranını temel alarak kaynak ayrılır. Sistemdeki kaynakların toplam sayısı belirli bir sınırla belirlenmiştir. Eğer sistemde bulunması için izin verilenden daha fazla kaynak tüketilirse kaynaklar sistemde izin verilen kaynak sayısına ulaşana kadar en az uyarılan *ARB*' lerden başlayarak çıkarılır. Eğer belirli bir *ARB*' nin kaynaklarının tamamı çıkarılırsa *ARB* hücre popülasyonundan çıkarılır.
- *Çekirdek hücre(seed cell)*: Eğitim başlangıcında bellek hücresi ve *ARB* popülasyonlarını başlatmak için kullanılan ve eğitim kümesinden alınan bir antikordur.
- *Uyarma fonksiyonu(stimulation funstion)*: Bir *ARB*' nin bir antijene veya bir başka *ARB*' ye tepkisinin ölçüsü için kullanılır. *AIRS*

sınıflandırıcısının formülasyonunda bu fonksiyon 0 ile 1 aralığını kapsayacak şekilde bir değer döndüren bir fonksiyondur. Bu çalışmada sunulan *AIRS* uygulaması için uyarma fonksiyonu *ARB*'nin özellik vektörleri ve antijen arasındaki öklit uzaklığı ile ters orantılıdır.

- *Uyarma değeri(stimulation value)*: Uyarma fonksiyonu tarafından döndürülen değerdir.
- *Uyarma eşiği(stimulation threshold)*: Spesifik bir antijenin eğitimi için bir durma kriteri olarak kullanılan 0 ile 1 arasında bir parametredir. Uygulama için sadece her bir sınıfın *ARB*'lerinin ortalama uyarma değeri uyarma eşiğinden daha yüksek olduğu zaman o antijenin eğitimi durdurulur.
- *Test kümesi*: Eğitilmiş *AIRS* sınıflandırıcısının sınıflandırma performansını değerlendirmek için kullanılan antijenlerin topluluğudur.
- *Eğitim kümesi*: *AIRS* sınıflandırıcısını eğitmek için kullanılan antijenleri

4.1.1. Algoritmanın rutinleri

Bu bölümde *AIRS* algoritmasının temel rutinleri sunulmuştur. Özellikle bu bölümde sınıflandırıcıyı temel alan bir bağışıklık sisteminin eğitimi ve meydana getirilmesinde kullanılan başlıca rutinler, metotlar ve eşitlikler gözden geçirilmiştir. Burada *AIRS* algoritmasının başlıca dört aşaması söz konusudur. İlk aşama veri normalizasyonu ve başlangıç aşamasıdır. İkinci aşama bellek hücresi belirleme ve *ARB* oluşumudur. Üçüncü aşama aday bellek hücresinin gelişmesinde kaynaklar için rekabettir. Eğitim algoritmasının son aşaması, yerleşmiş bellek hücrelerinin kümesi içerisine aday bellek hücrelerinin potansiyel sunumudur.

Çalışmada kullanılan işaretlemeler şunlardır:

- *MC* bellek hücrelerinin kümesini ifade etmektedir ve *mc* bu kümenin tek bir üyesini ifade etmektedir.

- $ag.c$, belirli bir ag antijenin sınıfını ifade etmektedir. Burada $ag.c \in C = \{1, 2, \dots, nc\}$ ve nc veri kümesindeki sınıfların sayısıdır.
- $mc.c$, belirli bir mc bellek hücresinin sınıfını temsil etmektedir. Burada $mc.c \in C = \{1, 2, \dots, nc\}$ dir.
- $MC_c \subseteq MC = \{MC_1 \cup MC_2 \cup \dots \cup MC_{nc}\}$ ve eğer $mc.c \equiv c$ ise $mc \in MC_c$ dir.
- $ag.f$ ve $mc.f$, belirli bir antijen ve bellek hücresinin özellik vektörünü temsil etmektedir. $ag.f_i$ ifadesi $ag.f$ deki i . özelliğin değerini ve $mc.f_i$ ifadesi $mc.f$ deki i . özelliğin değerini ifade etmektedir.
- AB , ARB ' lerin kümesini veya mevcut hücrelerin popülasyonunu temsil etmektedir. MU , ARB ' lerin mutasyona uğramış klonlarının bir kümesini temsil etmektedir. Bundan başka ab , tek bir ARB ' yi temsil etmektedir. Burada $ab \in CAB$ ' dir.
- $ab.c$ verilen bir ARB ' nin sınıfını temsil etmektedir. Burada $ab.c \in C = \{1, 2, \dots, nc\}$ dir.
- $AB_c \subseteq AB = \{AB_1 \cup AB_2 \cup \dots \cup AB_{nc}\}$ ve $ab \in AB_c$ if $ab.c=c$
- $ab.stim$ ifadesi ab ARB ' sinin uyarma seviyesini temsil etmektedir.
- $ab.resources$ ifadesi ab ARB ' si tarafından tutulan kaynakların sayısını temsil etmektedir.
- $TotalNumResources$, izin verilen sistem kaynaklarının toplam sayısını temsil etmektedir.

4.1.1.1. Başlangıç aşaması

Algoritmanın ilk aşaması aslında bir veri ön işleme aşaması olarak düşünülebilir. Başlangıç aşaması boyunca öncelikle veri kümesindeki herhangi iki veri arasındaki bütün nitelikler öklit uzaklığı $[0,1]$ aralığında olacak şekilde normalize edilir. öklit uzaklığı hem duyarlılık hem de uyarmayı (*stimulation*) hesaplarken başlıca ölçektir. Normalizasyondan sonra duyarlılık eşiği (AT) hesaplanır. Eğitim kümeleri için duyarlılık eşiği bütün eğitim verilerinin özellik

vektörleri üzerindeki ortalama duyarlılıktır. Duyarlılık eşiği (4.1) eşitliğinde açıklandığı gibi hesaplanır.

$$\text{Duyarlılık eşiği} = \frac{\sum_{i=1}^{n-1} \sum_{j=i+1}^n \text{duyarlilik}(ag_i, ag_j)}{\frac{n(n-1)}{2}} \quad (4.1)$$

Burada n eğitim verilerinin (antijenler) sayısıdır. ag_i ve ag_j ifadeleri i . ve j . öklit uzaklığıdır.

Başlangıç aşamasında son adım bellek hücrelerinin ve başlangıç ARB popülasyonunun oluşturulmasıdır. Bu rastgele olarak 0 veya daha fazla antijenin bellek hücrelerine ve ARB ' lerin kümesine eklenmesiyle yapılır.

4.1.1.2. Bellek hücresi belirleme ve ARB oluşumu

Başlangıç aşaması tamamlanınca eğitim döngüsel bir algoritma olarak devam eder. Algoritmanın bu aşamasının ilk adımı bellek hücresinin belirlenmesi ve ARB oluşumudur. Belirli bir özellikli bir ag eğitim antijeni, $mc_{match} = \arg \max_{mc \in MC_{ag,c}} \text{uyarma}(ag, mc)$ olan bellek hücresi mc_{match} 'i bulur. Burada $\text{uyarma}(x,y)$ (4.2) eşitliğinde tanımlandığı gibi hesaplanmaktadır:

$$\text{uyarma}(x,y) = 1 - \text{öklit_uzaklığı}(x.f, y.f) \quad (4.2)$$

Eğer antijen ile aynı sınıftaki bellek hücrelerinin kümesi boş ise bellek hücresinin kümesine antijen eklenir ve *match memory cell* olarak mc_{match} yeni eklenen bellek hücresini gösterilir. Bu çalışmada uyarma fonksiyonu olarak yalnızca *Euclidean* uzaklığı kullanılmıştır.

Bir kez mc_{match} tanımlandınca, bu bellek hücresi, önceden mevcut ARB 'lerin (örnek: Önceki antijenlere maruz kalan, sistemdeki kalan ARB ' ler) popülasyonuna

(muhtemelen) yeni *ARB*' ler eklemek için kullanılır. Bu Şekil 4.1' de gösterilen metotla gerçekleştirilir. Burada *makeARB(x)* fonksiyonu bu *ARB*' nin antikorunu olarak *x* ile birlikte bir *ARB* döndürür. *Mutate(x,b)* Şekil 4.2' de tanımlanmıştır. Şekil 4.2'deki *drandom()* fonksiyonu $[0,1]$ aralığında rastgele bir değer döndürmektedir ve $(lrandom() \bmod nc)$ $\{1,nc\}$ aralığında rastgele bir değer döndürmektedir.

```

MU ← ∅
MU ← MU ∪ makeARB(mcmatch)
stim ← uyarma(ag, mcmatch)
NumClones ← hiper_mutasyon_oranı*klonlama_oranı*stim
While(|MU| < NumClones)
do
    mut ← false
    mcclone ← mcmatch
    mcclone ← mutate(mcclone, mut)
    if (mut ≡ true)
        MU ← MU ∪ makeARB(mcclone)
done
AB ← AB ∪ MU

```

Şekil 4.1 *ARB* oluşumu için Hiper-Mutasyon

```

Mutate(x,b)
{
  foreach (x.fi in x.f)
  do
    change ← drandom()
    change_to ← drandom()
    if(change < mutasyon_oranı)
    {
      x.fi ← change_to * normalizasyon_değeri
      b ← true
    }
  done
  if(b = true)
  {
    change ← drandom()
    change_to ← (lrandom() mod nc)
    if (change < mutasyon_oranı)
      x.c ← change_to
  }
  return x
}

```

Şekil 4.2 Mutasyon rutini

4.1.1.3. Bir aday bellek hücresinin gelişimi ve kaynaklar için rekabet

Bu noktada *ARB*'lerin (*AB*) bir kümesi mC_{match} , mC_{match} 'ten mutasyonlar ve (muhtemelen) önceki karşılaşılan antijenlere olan tepkiden oluşan kalan *ARB*'leri içermektedir. *AIRS* olarak adlandırılan algoritma dögüsel bir algoritmadır.

Algoritma ayrı aşamalara bölünmüştür ve her bir antijen kesin olarak en az bir kez bu sürece katılmaktadır. Algoritmanın bir sonraki parçasının amacı bir aday bellek hücresi geliştirmektir. Bu hücre belirli bir *ag* antijenini doğru olarak sınıflandıran en başarılı hücredir. Bu başlıca üç mekanizma ile gerçekleştirilir. Birinci mekanizma sistemin geniş kaynakları için rekabettir. Diğer metotlar *Timmis and Neal (2000)*' de taslağı çizilmiş ve *Knight and Timmis (2001)*' de denemiştir. Kaynaklar *ag* antijeninin tanınması için *ARB*' nin uygunluğuna işaret eden *normalize* edilmiş uyarılma değerini temel alan bir *ARB* için yerleştirilir. İkinci mekanizma *çeşitlilik* ve *şekil-uzayı araştırması* için mutasyonun kullanılmasıdır. Üçüncü mekanizma *ag* üzerinde eğitimi durdurma zamanına karar vermek için bir kriter olarak *ortalama uyarılma eşiğinin* kullanılmasıdır.

Genetik algoritmaların prensiplerine benzer olarak *AIRS* algoritması *ARB* popülasyonu içerisinde bireylerin hayatta kalması için uygunluk kavramını kullanmaktadır. Bir *ARB*' nin hayatta kalmasına birbiri ile bağlantılı olarak iki aşamada karar verilir. Önce *AB* popülasyonunda her bir *ARB*, *ARB*' nin *uyarma* seviyesine karar vermek için *ag* antijeni ile birlikte sunulur. Bu *uyarma* hem ham olarak *uyarma* seviyesini hem de *ARB*' nin sınıfını (*ab.c*) temel alan *ARB* popülasyonunda *normalize* edilir. Bu *normalize* edilmiş *uyarma* değerini temel alan her bir $ab \in AB$ kaynakların sınırlı sayısında yerleştirilir. Eğer kaynakların bu yerleşimi popülasyon için izin verilen kaynaklardan fazla olarak sonuçlanırsa, sistemdeki kaynakların sayısı izin verilen kaynak sayısına ulaşıncaya kadar kaynaklardan en zayıf *ARB*' ler (en az uyarılan) silinir. Sıfır kaynağa sahip *ARB*' ler *ARB* popülasyonundan silinir. Bu süreç Şekil 4.3' te gösterilmiştir.

```

minStim ← 2.0
maxStim ← 0.0
foreach(ab ∈ AB)
do
    stim ← uyarma(ag, ab)
    if(stim < minStim)
        minStim ← stim
    if(stim > maxStim)
        maxStim ← stim
    ab.stim ← stim
done
foreach(ab ∈ AB)
do
    if(ab.c ≡ ag.c)
        ab.stim ←  $\frac{ab.stim - minStim}{maxStim - minStim}$ 
    else
        ab.stim ←  $1 - \frac{ab.stim - minStim}{maxStim - minStim}$ 
    ab.kaynak ← ab.stim * klonlama_orani
done
i ← 1
while(i ≤ nc)
do
     $|AB|$ 
    resAlloc ←  $\sum_{j=1}^{|AB|} ab_j.kaynak, ab_j \in AB_i$ 
    if(i ≡ ag.c)
        izin_verilen_kaynak ← kaynakların_sayısı / 2
    else
        izin_verilen_kaynak ← kaynakların_sayısı / 2 * (nc - 1)
    while(resAlloc > izin_verilen_kaynak)

```

```

do
  silinecek_kaynak_sayısı ← resAlloc – izin_verilen_kaynak
  ab_sil ← argminab ∈ ABi (ab.stim)
  if(ab_sil.kaynak ≤ silinecek_kaynak)
    ABi ← ABi - ab_sil
  else
    ab_sil.kaynak ← ab_sil.kaynak – silinecek_kaynak
done
i=i+1
done

```

Şekil 4.3 Stimulation, kaynak tahsisi ve ARB çıkarılması.

Bu kaynak tahsisi rutininin iki önemli bakış açısı söz konusudur. Birincisi bir ARB' nin uyarma değerine sadece (4.2) eşitliğindeki uyarma fonksiyonu tarafından karar verilmez aynı zamanda ARB' nin sınıfı temel alınarak karar verilir. Şekil 4.3' te gösterilen uyarma hesaplama metodu hem *ag* tarafından yüksek derecede uyarılan *ag* ile aynı sınıfın ARB' leri, hem de *ag* ile güçlü pozitif reaksiyon sergilemeyen *ag* ile farklı sınıfın ARB' leri ile takviye edilir. İkincisi kaynakların dağılımının ARB' nin sınıfını temel almasıdır. Bu gelecek antijenlere reaksiyon için kalan ARB' lerin potansiyel olarak pozitif niteliklerini kaybetmeksizin *ag* ile aynı sınıftaki ARB' ler ayrıca takviye edilerek yapılır.

Algoritmada bu noktada AB sadece *ag* antijeni tarafından en çok uyarılan ARB' leri içermektedir veya daha özellikli olarak AB, kaynaklar için başarılı olarak rekabet edebilen ARB' ler içermektedir. Algoritma önce AB deki ARB' ler bu verinin eğitimini durdurmak için *ag* tarafından yeterince uyarılmış mı karar vererek devam eder. Bu AB' nin her bir alt sınıfı için ortalama uyarma değerini içeren *nc* boyutlu bir *s* vektörü tanımlanarak yapılır (4.3).

$$s_i \leftarrow \frac{\sum_{j=1}^{|AB_i|} ab_j.stim}{|AB_i|}, ab_j \in AB_i \quad (4.3)$$

Durma kriteri $s=\{s_1, s_2, \dots, s_{nc}\}$ deki bütün elemanlar için $s_i \geq \text{uyarma eşiği}$ (*stimulation_threshold*) olmasıdır.

Durma kriterine ulaşıp ulaşılmadığına bakmaksızın algoritma mutasyona uğratarak oluşturulmuş ürünlerle oluşturulan Ab ' deki her bir ARB hesaba katılarak ilerler. Burada Şekil 4.2' de gösterilen metot ile benzer mutasyona uğramış ürünler oluşturulurken Şekil 4.2'deki metottan birkaç farklılık söz konusudur. Bu değişime uğramış mutasyon oluşturma rutini Şekil 4.4' de sunulmuştur.

```

MU ←  $\phi$ 
foreach(ab ∈ AB)
  Do
    rd ← drandom()
    if(ab.stim < rd)
      NumClones ← ab.stim * klonlama_oranı
      i ← 1
      while (i ≤ NumClones)
        Do
          mut ← false
          ab_clone ← ab
          ab_clone ← mutate (ab_clone, mut)
          if(mut = true)
            MU ← MU ∪ ab_clone
          i = i + 1
        Done
      Done
    AB ← AB ∪ MU

```

Şekil 4.4 Hayatta kalan ARB ' nin mutasyonu.

Hayatta kalan her bir ARB ' den mutasyona uğramış ürünler üretildikten sonra *durma kriteri* kontrol edilir. Eğer *durma kriteri* ile karşılaşılmışsa bu antijen

için eğitim durur. Eğer durma kriteri ile karşılaşılmamışsa bu sürecin tamamı Şekil 4.3' de gösterilen metottan başlayarak durma kriteri ile karşılaşılan kadar tekrar edilir. Bu tekrar işleminde her tekrarda tek istisna *uyarma (stimulation)* ve *kaynak tahsisi* aşamasından sonra eğer durma kriteri ile karşılaşırsa mutasyona uğramış ürün üretimi uygulanmaz. *Durma kriteri* ile bir kez karşılaşıncaya *aday bellek hücresi* seçilir. *Aday bellek hücresi* $mc_{candidate}$, *ag* eğitim antijeninin sınıfı ile aynı sınıftaki en fazla uyarılan *ARB* olan mutasyondan önce sistemde bulunan *ARB*' nin sınıfına sahiptir ve özellik vektörüdür.

4.1.1.4. Bellek hücresi takdimi

Eğitim rutininde son aşamada mevcut *MC* bellek hücreleri kümesi içerisindeki en çok geliştirilmiş aday bellek hücresi $mc_{candidate}$ belirlenir ve bu aşama boyunca önceden belirlenmiş olan mc_{match} ile $mc_{candidate}$ ' in yer değiştirilip değiştirilmeyeceğini belirten başlangıçta hesaplanan *duyarlılık eşiği* ile oluşturulur. *Aday bellek hücresi*, sadece *ag* eğitim antijeni tarafından mc_{match} 'ten daha fazla uyarılırsa bellek hücresinin kümesine eklenir. Buradaki *uyarma (stimulation)* (4.2) eşitliğinde tanımlanmıştır. Bu şart sağlanmışsa ve eğer $mc_{candidate}$ ve mc_{match} arasındaki duyarlılık, duyarlılık eşiği ve kullanıcı tarafından belirtilen duyarlılık eşik ölçüsünün (*ATS*) çarpımından daha az ise $mc_{candidate}$ bellek hücresi kümesindeki mc_{match} ile yer değiştirilir. Bu süreç Şekil 4.5' de gösterilmiştir.

$$CandStim \leftarrow uyarma(ag, mc_{candidate})$$

$$MatchStim \leftarrow uyarma(ag, mc_{match})$$

$$CellAff \leftarrow duyarlılık(mc_{candidate}, mc_{match})$$

$$if (CandStim > MatchStim)$$

$$if (CellAff < duyarlılık_eşiği * ATS)$$

$$MC \leftarrow MC - mc_{match}$$

$$MC \leftarrow MC \cup mc_{candidate}$$

Şekil 4.5 Bellek hücresi takdimi.

Aday bellek hücresi mevcut bellek hücreleri içerisinde değerlendirilince bu antijen için eğitim tamamlanmıştır. Eğitim kümesindeki sonraki antijen seçilir (veya sonraki antijen bir veri oluşturma fonksiyonu kullanılarak oluşturulur) ve eğitim süreci bellek hücresi belirlenmesi ve *ARB* oluşturulması ile ilerler. Bu süreç bütün antijenler sisteme sunulana kadar devam eder.

4.1.1.5. Sınıflandırma

Eğitim tamamlandıktan sonra geliştirilen bellek hücreleri sınıflandırmada kullanılmak için uygun durumdadır. Sınıflandırmada k en yakın komşu yaklaşımı uygulanmaktadır. Her bir bellek hücresi *uyarma* için bir veri olarak sunulur. Bir verinin sistem tarafından sınıflandırılmasına k adet en fazla uyarılan bellek hücrelerinin çıkışlarının bir *çoğunluk oylaması (majority vote)* kullanılarak karar verilir.

4.1.2. AIRS algoritmasının ayrıntıları

Bölüm 4.1' de *AIRS* algoritmasının bir özeti sunulurken bu bölümde bu sınıflandırma sisteminin önemli elemanlarından bazıları kavramsal olarak tartışılmıştır. Algoritmanın amacı sınıflandırma verisi olarak kullanılabilen bellek hücrelerinin bir kümesini geliştirmektir. Bu yapay bellek hücreleri, doğal bağışıklık sistemlerinde gözlemlenen çeşitli karakteristikleri kapsamaktadır. Esasen bellek hücreleri vücuttaki olgunlaşma sürecinden geçmiş B hücrelerini temel almaktadır. Memeli bağışıklık sistemlerinde bu bellek B hücreleri saldırgan antijenler tarafından kolaylıkla uyarılır ve saldırgan patojenleri tanımaya tepki olarak *hiper-somatik mutasyonu* prosesinden geçer. Bu kavramın düzenlenmiş hali *AIRS* algoritmasında bellek hücrelerinin fonksiyonu olarak gözlemlenebilir. Yapay bellek hücreleri aynı

zamanda T hücrelerinin ve bir derece *antijen sunan hücrelerin (APC)* rolünü aldığı söylenebilir. Doğal bağışıklık sistemlerinde T hücreleri B hücrelerinin spesifik bir popülasyonu ile birleşme eğilimindedir. Bir T hücresi bir antijeni tanıdığı zaman B hücresi ile bu antijeni sunar.

AIRS algoritmasının kalbi *ARB*'lerin bir popülasyonundan bellek hücrelerinin geliştirilmesi sürecidir. Bu evrimsel süreç burada bahsedilen çeşitli anahtar kavramlara sahiptir. Bellek hücrelerinin geliştirilmesinde *ARB*'lerin popülasyonu için evrimsel baskı sağlamak için başlıca mekanizma sistemin geniş kaynakları için rekabettir. *Timmis and Neal (2000)* ve *Knight and Timmis (2001)* tarafından esinlenen bu kavram, hücrenin hayatta kalmasına karar verilmesi ve sınıflandırma kalitesinin desteklenmesi ile sağlanır. Genetik algoritmalara benzer olarak kaynak rekabetinin amacı en uygun bireyleri geliştirmektir. *AIRS* algoritmasında uygunluğa başlangıçta bir antijen ile ayrı ayrı *ARB*'lerin uyarma tepkisi ile karar verilmektedir. *Jon Timmis* ve arkadaşlarının danışmansız öğrenme çalışmasında belirli bir hücrenin uyarma değeri kaynak tahsisi için yeterlidir. Bununla beraber o çalışmada, ne bir hücre ile bir antijenin uyarma tepkisini kontrol etmeye ne de, antijenin sınıfı ile hücrenin sınıfının hesaba katılmasına gerek vardır. Bu sebepten *AIRS* algoritmasında antijenle aynı sınıftaki hücreler yüksek uyarımla ve antijenle aynı sınıfa sahip olmayan hücreler düşük uyarımla ödüllendirilir. Ödül daha fazla sistem kaynakları tahsis edilmesine izin verir. Diğer taraftan düşük uyarma değerli fakat antijenle aynı sınıfta veya yüksek uyarma değerli fakat antijenle farklı sınıftaki hücreler tam olarak zayıf sınıflandırıcı hücreler olarak görülmez fakat potansiyel olarak zararlı sınıflandırıcı hücreler ve onlar böylece daha az ödüllendirilir. Elde edilmiş kaynaklar için en az yetenekli *ARB*'ler sistemden çıkarıldığı için burada en çok ödül sağlayacak arama uzayının olduğu yere doğru gelişme olacaktır. Uygulamada antijenle aynı sınıfa sahip *ARB*'lerin kaynakları antijenle farklı sınıfta olan *ARB*'lerin kaynaklarından daha fazla olması sağlanmıştır. Bu aslında eğitim antijeni civarındaki uzay araştırıldığı zaman daha büyük potansiyel çeşitliliğe izin verilerek yapılır. Bu kavramın gelecekteki bir uzantısı olarak veri kümesindeki sınıf oranlarını yansıtan bir şekilde *ARB*'ler arasındaki kaynaklar bölünebilir.

Kaynakların rekabetinden hayatta kalan *ARB*' ler mutasyona uğramış ürünler üretme fırsatı verilerek ödüllendirilir. Tekrar genetik algoritmalarda olduğu gibi hayatta kalmak için bu rekabet aynı şekilde devam eder. Belirli bir antijenin etrafındaki en uygun bireyler gerçekten bir bellek hücresi olarak hayatta kalmayabilirken onların mutasyona uğramış ürünleri hayatta kalabilir. Böylece algoritmanın desteklediği hücreler hayatta kalmaktadır. Bu, eğitim rutininde özelliklerin mutasyona uğratılması ile başarılmaktadır. *ARB* popülasyonu içerisindeki mutasyona uğramış ürünleri oluşturmak arama uzayının daha fazla araştırılmasını sağlamaktadır. Bu çalışma, belirli bir antijeni tanıdığı söylenebilen *ARB* popülasyonunun karşılamak zorunda olduğu bir uyarma eşiği kullanılarak geliştirilmiştir. Bu, bellek hücresi için arzu edilebilir nitelikler sergileyen popülasyon içerisindeki hücreler geliştirmek için evrimsel basıncı arttırmaktadır.

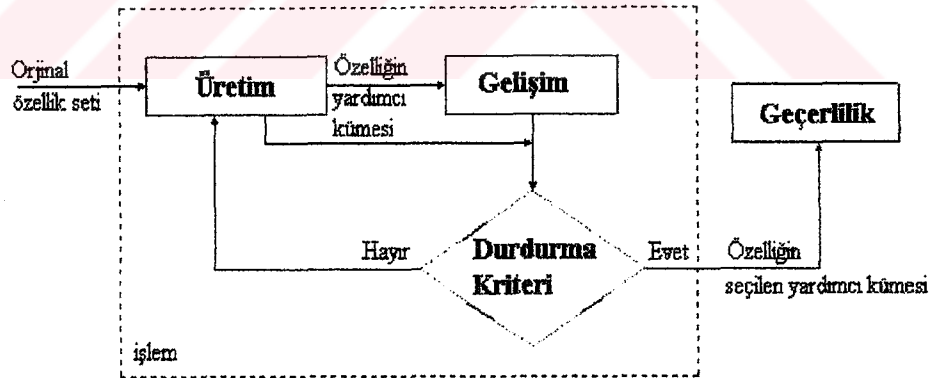
Bir *ARB* genel *ARB* popülasyonu arasındaki kaynaklar için başarılı olarak rekabet ettikten sonra eğer *ARB* eğitim antijeni için tepkide en çok uyarılan olursa ve eğitim antijeni ile aynı sınıfta ise *ARB*, bellek hücresi havuzuna eklenir. Bir antijen ilk kez sunulduğu zaman antijenle aynı sınıfta ve en çok uyarılan bellek hücresinin *ARB* genel popülasyonu içerisine mutasyona uğramış ürünlerini eklemesine izin verilir. Bununla beraber bu *ARB* popülasyonunun evrimsel süreci sonunda bu orijinal bellek hücresi potansiyel olarak geliştirilmiş bellek hücresi ile yer değiştirilir. Bu sadece geliştirilmiş bellek hücresi eğitim antijeni için arama uzayında başlangıçta yerleştirilmiş bellek hücresinden daha yakın olduğu zaman ve bu iki bellek hücresi birbirleri ile yeterince yakınsa (kullanıcı tarafından tanımlanmaktadır) meydana gelmektedir. Makine öğrenmesi bakış açısından bu süreç *AIRS* algoritmasının veri azaltma yeteneğini sağladığını göstermektedir. Mevcut bellek hücresi ile şekil uzayında yakın yerlerde bulunan daha iyi sınıflandıran bellek hücrelerinin yer değişimine izin verilerek *AIRS* algoritması, problemde sunulmak için gerekli olunan hücrelerin sayısını azaltmaktadır. Algoritma aynı zamanda genelleme yapmaktadır çünkü sistemde geliştirilmiş bellek hücrelerinin her hangi eğitim örneği ile aynı olması zorunlu değildir.

4.2. Özellik Seçme (*Feature Selection*) Prosedürü

4.2.1. Özellik seçme işleminin tanımı ve kullanılan yöntemler

Ham bir veri setinde örneklerin sayısı ve özelliklerin sayısı oldukça büyük olabilir. Bu büyüklük, birçok veri madenciliği sistemlerinde ciddi problemlere sebep olabilir. Özellik seçmenin amacı, birkaç kritere uygun özelliklerin minimum sayısını seçmektir. Özelliklerin minimum bir yardımcı kümesinin seçilmesiyle, ilgisiz ve fazla özellikler oluşturulan kritere göre uzaklaştırılır. Basitleştirilmiş veri, daha kararlı ve anlaşılabilir sonuçlar verecektir.

Özellik seçme işleminin yapılmasının sebebi, verinin bütün özelliklerini kullanmak oldukça zaman alır ve veri kümesinin bilgi kazancı (information gain) hesaplandıktan sonra bütün özellikler içermek gerekli olmadığı için bu işlem yapılır. Özellik seçme işlemi temel olarak dört adımdan oluşur. Özellik seçmenin blok diyagramı Şekil 4.6' da gösterilmiştir.



Şekil 4.6 Özellik seçme işleminin blok diyagramı

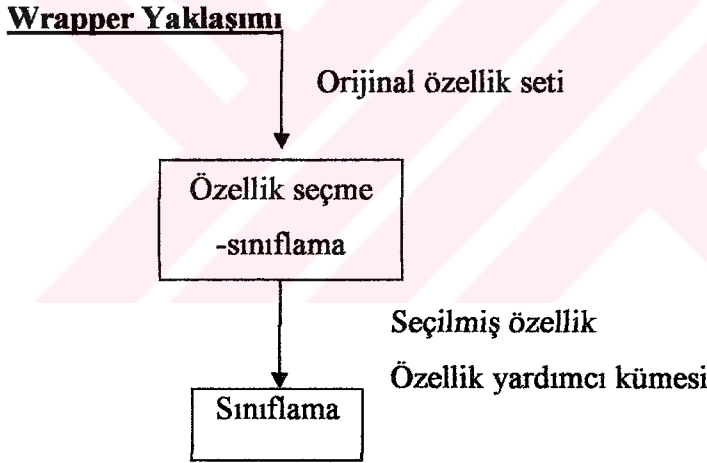
Özellik seçmede uygulanan temel işlemler:

- Üretim = aday yardımcı özelliğin seçilmesi
- Gelişim = yardımcı kümenin ilgili değerinin hesaplanması
- Durdurma kriteri = yardımcı kümenin ilgili kümeyle ilişkili olup olmadığını göstermek

- Geçerlilik = yardımcı kümenin geçerliliğini doğrulamak

Özellik seçme bazen özellik belirleme olarak bilinen işlem için birçok algoritma geliştirilmiştir. Özellik seçme, uygun olduğunda ilgisiz ve fazla bilgiyi uzaklaştırmak veya veritabanındaki ilgili bilgiyi tanımlamak için bir ön-işleme yöntemidir.

Özellik seçme algoritmasının Wrapper ve Filter olmak üzere iki kategorisi vardır. *Wrapper*, öğrenme sistemini oluşturan ve özelliğin iyi bir yardımcı kümesini seçmeyi sürekli olarak araştıran bir araştırma fonksiyonunu inceler ve gerçek bir hedefe bağlı olarak sınıflama performansını kullanır. Bu kategori yaklaşımları çok güçlüdür. Bununla birlikte, *Wrapper*, özelliklerin en iyi yardımcı kümesine ulaşamayabilir, fakat yerel bir minimuma ulaşabilir. *Wrapper* yaklaşımının blok diyagramı Şekil 4.7' de verilmiştir.

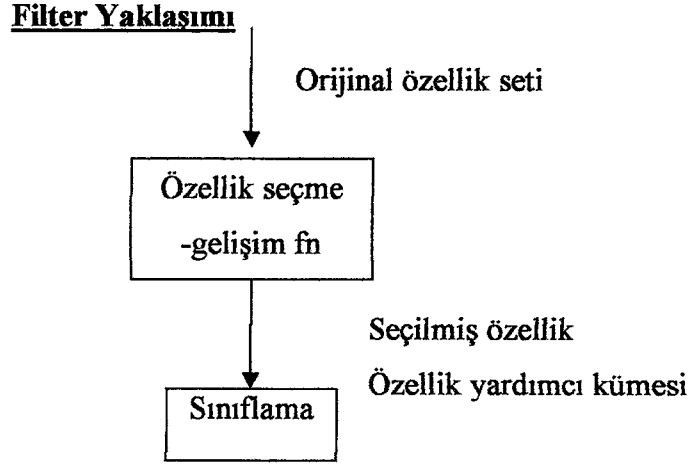


Şekil 4.7 Wrapper yaklaşımı blok diyagramı

Wrapper yaklaşımının özellikleri; yüksek doğruluk, düşük kayıp ve sınıflamayı dikkate alır.

İkinci kategori olan, *Filter*, öğrenme sistemini oluşturmadan önce, fazla özelliği uzaklaştıran bireysel bir algoritmadır. Bu yüzden, bu kategori maliyetlerin hesaplanmasında kullanılan birkaç terim içinde ilk kategoriden daha iyidir, çünkü

algoritmanın bu kategorisi öğrenme sistemi oluşturulmadan önce uygulanır. *Filter* yaklaşımının blok diyagramı Şekil 4.8’ de verilmiştir.



Şekil 4.8 *Filter* yaklaşımı blok diyagramı

5. MEDİKAL VERİTABANI

Bu bölüm dört medikal veritabanının detaylarını açıklar: bu tezde kapsamlı olarak kullanılan göğüs kanseri, kalp hastalığı, karaciğer hastalığı ve lenf hastalığı veritabanlarıdır. Veritabanının detayları, hastalığın temeli, veritabanı bilgisi ve makine öğrenmesi teknikleri, sinir ağları ve sinyal işleme tekniklerindeki geçmiş tahmin performansını içerir. Hastalığın temeli, hastalık hakkındaki biyolojik ve medikal bilgiye önem verir. Hastalığın özelliğini tespit etmede, en iyi öğrenen sistemi bulmak veya oluşturmak için hastalığın medikal doğasını bilmek gerekir. Veritabanı bilgisi, veritabanının büyüklüğü ve eksik özelliğe sahip örneklerin veritabanında olup olmadığını içerir. Sonra, veritabanı özellikleri, eğer gerekliyse kendi ölçtüğü birimi ve kendine uygun değerlerden oluşan listeden oluşur. Bütün medikal veritabanları, Irvine' deki California Üniversitesi, Bilgisayar Bilimi ve Bilgi Bölümünden, makine öğrenmesi repertuarından elde edildi.

5.1 Göğüs Kanseri Veritabanı

5.1.1. Göğüs kanseri hastalığının tanımı

Göğüs kanseri, genellikle bir kütle dokusuyla başlayan ve göğüste büyümeye başlayan kötü huylu bir dokudur. Kadınlarda en yaygın ikinci kanserdir. Üstelik erkeklerde de görülür, fakat çok küçük bir şansa sahiptir.

Göğüs kanserinin belirtileri, göğüs kütlesi, göğüsün büyüklüğünde ve şeklinde değişiklik, göğüsün derisinin hassaslığı veya rengindeki değişiklik, göğüs ağrısı, meme ucunun görünümündeki değişiklik ve meme ucu değişiklikleri v.b.

5.1.2. Göğüs kanseri veritabanı bilgisi

Göğüs kanseri veritabanı, Dr. William H. Wolberg tarafından Wisconsin Hastanesinden elde edildi. Veritabanı 683 örnekten oluşmaktadır. 9 tamsayı değerli özelliğe ve her bir veri 1' den 10' a aralık değerine sahiptir. Göğüs kanser tespitinde kullanılan özellikler;

- Kütle kalınlığı
- Hücre Büyüklüğünün Benzerliği
- Hücre Şeklinin Benzerliği
- Marjinal Yapışma
- Tek Epitel Hücreleri Büyüklüğü
- Çıplak Çekirdek Kütlesi
- Hafif Çekirdek Kütlesi
- Normal Çekirdek Kütlesi
- Mitoz

Bu özellikler, dokuz farklı ölçekte dış görünüm ve iç kromozom değişikliklerinden elde edilen ölçümlerdir. Göğüs kanserinin sınıf değişkeninde iki değer vardır: sırayla 0 ve 1 ile nümerik olarak gösterilen iyi huylu (benign) ve kötü huylu (malignant) sınıflarıdır.

5.1.3. Göğüs kanserinin geçmiş tahmin performansı

Bu tezde kullandığımız veritabanı, makine öğrenmesi, sinir ağları ve sinyal işlemede çok iyi bilinen bir veritabanıdır. Bu veritabanındaki sınıflama genellikle yaklaşık olarak %90 veya daha fazladır. Göğüs kanseri teşhisi için literatürde yapılan yöntem ve sınıflama doğrulukları Çizelge 5.1' de verilmiştir.

Çizelge 5.1 Göğüs kanseri hastalığı teşhisi için sınıflama performansı karşılaştırılması

Yazar (yıl)	Yöntem	Sınıflama Doğruluğu (%)
Quinlan(1996)	C4.5	94.74
Hamilton et al. (1996)	C4.5	96
Ster and Dobnikar (1996)	FLDA	96.8
Bennett and Blue (1997)	SVM (5xCV)	97.2
Nauck ve Kruse (1999)	Neuro-Fuzzy	95.06
Andres et al. (1999)	Fuzzy-GA	97.36
Setione (2000)	Neuro-Rule	98.1
Watkins (2002)	Optimized- LVQ	96.7
Watkins (2002)	Big- LVQ	96.8
Watkins (2002)	AIRS	97.2
Abonyi and Szeifert (2003)	SFC	95.57

5.2. Kalp Hastalığı Veritabanı

5.2.1. Kalp hastalığının tanımı

Kalp hastalığı, kalbin kan odacıklarını veya kalp kasını etkiler. Kalp rahatsızlığı, genellikle, kan odacık problemdir, örnek olarak kan odacığının darlaşması veya kan akışının blokajı gibi problemlerdir.

Kalp hastalığının belirtileri, göğüs ağrısı, aşırı yorgunluk, yüksek kan basıncı ve kan şekeri. İlave olarak, kalp hastalığının birkaç biçimi vardır: kalp saldırısı, kalp ağrısı ve kalp tümörü v.b.

5.2.2. Kalp hastalığı veritabanı bilgisi

Sadece Cleveland verisini içeren dört farklı yardımcı kümeden oluşan Kalp Hastalığı Veritabanı kullanıldı. Cleveland verisi, Dr. Robert Detrano' dan Long Beach ve Cleveland Klinik Kuruluşu Medikal Center' dan elde edildi. Veritabanı, 6 eksik örnek ve 297 tamamlanmış örnek olmak üzere 303 örnekten oluşmaktadır. Orijinal olarak, veritabanı 76 ham özellikten oluşmaktadır. Bununla birlikte, bütün yayınlanmış deneylerde bu özelliklerden sadece 13 özelliği kullanılır ve aşağıdakine bağlı olarak listelenir:

- Yaş – 20' dan 77' ye yaş arası
- Cinsiyet – erkek veya kız, sırayla 0 ve 1 olarak gösterilir.
- Göğüs ağrısı tipi- göğüs ağrısının dört tipi
 - Değer 1: angina tipi
 - Değer 2: atypical angina
 - Değer 3: non-anginal ağrısı.
 - Değer 4: asymptomatic
- Geriye kalan kan basıncı- mm Hg' de ölçülmüş.
- Serum kolesterolü- mg/dl' de ölçülmüş.
- Hızla artan kan şekeri > 120 mg/dl- doğru:1 veya yanlış:0;
- Geriye kalan elektrokardiyografik sonuçları
 - Değer 0: Normal
 - Değer 1: Anormal olan ST-T dalga oluşması
 - Değer 2: Sol ventrikular hypertrophy olasığının gösterilmesi
- Maksimum kalp oranının elde edilmesi
- Angina' ya neden olan uygulama
 - Değer 1: Evet;
 - Değer 0: Hayır;
- Geriye kalan uygulama tarafından oluşan ST depresyon.
- Zirve (doruk) ST uygulama parçasının eğimi
 - Değer 1: Eğimin artması

- Değer 2: Düz
- Değer 3: Eğimin azalması
- Fluoroscopy tarafından renklenen odacıkların sayısı – 0’ dan 3’ e değerlidir.
- Thal – Thalasemias
 - Değer 3: Normal
 - Değer 6: Değişmez hasar
 - Değer 7: Eski hasar

%50 çap daralmasından daha az veya %50 çap daralmasından daha fazla olmasından oluşan kalp hastalığının teşhisi için iki çıkış vardır ve sırayla 0 veya 1 ile gösterilir.

5.2.3. Kalp hastalığının geçmiş tahmin performansı

Sınıflama performansı geçmişte yapılan çalışmalarda yaklaşık olarak %75–85 olarak elde edildi. Veritabanının önemi, Dr. Robert Detrano tarafından bir lojistik regresyon ile üretilmiş ayrık bir fonksiyonla sınıflama performansı %77 elde edildi. Kalp hastalığı teşhisi için kullanılan sınıflamaların performans karşılaştırılması aşağıdaki Çizelge 5.2’ de verilmiştir.

Çizelge 5.2 Kalp Hastalığı teşhisi için sınıflama performansı karşılaştırılması

Yazar (yıl)	Yöntem	Sınıflama Doğruluğu (%)
Newton Cheung (2001)	C4.5	81.11
Newton Cheung (2001)	Naive Bayes	81.48
Newton Cheung (2001)	BNND	81.11
Newton Cheung (2001)	BNNF	80.96

5.3. Karaciğer Rahatsızlığı Veritabanı

5.3.1. Karaciğer rahatsızlığı tanımı

Karaciğer rahatsızlığı, vücudumuzda fazla alkolün bulunmasıyla meydana gelebilir. İnsanda alkolün nasıl bulunacağını dedekte eden yöntemlerden birisi, karaciğer rahatsızlığının hassas olduğu kandaki farklı enzimlerin ölçümüyle yapılır.

5.3.2. Karaciğer rahatsızlığı veritabanı bilgisi

Karaciğer hastalığı veritabanı, BUPA Medikal Araştırma Şirketinde Richard S. Forsyth tarafından sağlandı. Bu veritabanı toplam olarak 345 örnekten oluşmaktadır ve bütün veriler hiçbir eksik özelliğe sahip değildir. Veritabanı, 6 özellikten oluşmaktadır ve ilk beş özelliği kan testidir. Bütün kan testleri, fazla alkol tüketiminden meydana gelebilen karaciğer rahatsızlıklarına duyarlıdır. Bu özellikler aşağıda listelenmiştir:

- **Mev** (Mean corpuscular volume)- Ortalama korpüskulor hacim.
- **Alkphos** (Alkaline phosphatase)- Safra salgısının hücrelerinde bulunan protein
- **Sgpt** (Alanine aminotransferase) - Hepatocellular nekrozda kan seviyesi yükselen bir çeşit aminotransferase.
- **Sgot** (Aspartate aminotransferase) - Hepatocellular nekrozda kan seviyesi yükselen bir çeşit aminotransferase.
- **Gammagt** (Gamma-glutamyl tranpeptidase)- Kandaki GGT enzim miktarını ölçen bir test.
- **Drinks** (Günlük alınan alkol miktarı)

Çıkış sınıfları, karaciğer rahatsızlığı hastası ve normal hasta olmak üzere sırayla 0 ve 1 olarak gösterilmiştir.

5.3.3. Karaciğer rahatsızlığı veritabanının geçmiş tahmin performansı

Sınıflama performansı geçmişte yapılan çalışmalarda %55–65 olarak elde edildi. Karaciğer hastalığı (liver disorders) teşhisi için kullanılan sınıflamaların performans karşılaştırılması aşağıdaki Çizelge 5.3' de verilmiştir.

Çizelge 5.3 Karaciğer Rahatsızlığı (Liver Disorders) teşhisi için sınıflama performansı karşılaştırılması

Yazar (yıl)	Yöntem	Sınıflama Doğruluğu(%)
Newton Cheung (2001)	C4.5	65.59
Newton Cheung (2001)	Naive Bayes	63.39
Tony Van Gestel et. al. (2001)	VMC	69.7
Yuh-Jye Lee ve O.L. Mangasarian	SSVM	70.33
Yuh-Jye Lee ve O.L. Mangasarian	RSVM	74.86
D.T. Pham et. Al.	Inductive Learning	53.76
T. Yıldırım et. al. (2003)	MLP	74.56
T. Yıldırım et. al. (2003)	PNN	73.86
T. Yıldırım et. al. (2003)	GRNN	60.68

5.4. Lenf hastalığı Veritabanı

5.4.1. Lenf hastalığı bilgisi

Lenf sistemi, düğümlerle birlikte bağlantı kuran odacıklar ve düğümlerden oluşur. Lenf düğümleri, enfeksiyonla savaşmak veya korumak için beyaz kan hücreleri oluşturur.

5.4.2. Lenf hastalığı veritabanı bilgisi

Lenf hastalığı veritabanı, Yugoslavya Medikal Center Üniversitesi Onkoloji Enstitüsü Ljubljana bölümünden elde edildi. Bu veritabanı toplam 148 örnekten oluşmaktadır ve hiçbir kayıp özellik yoktur. Aşağıda gösterilen listeye göre 18 nümerik değerli özellikten oluşmaktadır:

- Lymphatic (lenfosit) - bütün lenf sistemi üzerinden bir test; normal için 1, arched (en baştaki) için 2, deforme olmuş için 3 ve yerleştirilen için 4' dür;
- Afferent' in bloğu – hayır için değer 1 ve evet için değer 2;
- C Lenfinin bloğu - hayır için değer 1 ve evet için değer 2;
- S Lenfinin bloğu - hayır için değer 1 ve evet için değer 2;
- By pass - hayır için değer 1 ve evet için değer 2;
- Extravasates – bir düğümden kovulması (çıkartılması) ve 1 ve 2 ile gösterilir;
- Regeneration - hayır için değer 1 ve evet için değer 2;
- Early uptake (Erken kalkmak) - hayır için değer 1 ve evet için değer 2;
- Lenf düğümlerinin boyutu – 0' dan 3' e aralık değeri;

- Lenf düğümlerinin uzunluğu – 1’ dan 4’ e aralık değeri;
- Lenf’ deki değişiklikler – çekirdek için değeri 1, oval için değeri 2 ve yuvarlak için değeri 3;
- Düğümdaki hatalar – hayır için değeri 1, lacunar’ lar (tekneli tavan) için değeri 2, marjinal lacunar’ lar için değeri 3 ve merkezi lacunar’ lar için değeri 4;
- Yapıdaki değişiklikler – Lenfosit sisteminin yapısı;
- Özel biçimler – hayır için değeri 1, chalices (kalis) için değeri 2 ve vesicles (vezikül) için değeri 3;
- Düğümün yerleştirilememesi - hayır için değeri 1 ve evet için değeri 2;
- Düğümün çıkarılması - hayır için değeri 1 ve evet için değeri 2;
- Düğümlerin sayısı – 0’ dan 80’ e aralık değeri;

Sınıf değişkenlerinde dört sınıf bulunmaktadır: normal, metastases, kötü lenf ve fibrosis olmak üzere, sırayla 1, 2, 3 ve 4 tamsayılarıyla gösterilmektedir.

5.4.3. Lenf hastalığı Veritabanının Geçmiş Tahmin Performansı

Bu veritabanı, makine öğrenmesinde çok iyi kullanılan bir veritabanıdır ve genellikle, bu verilerin daha önceki yapılan çalışmalarda sınıflama performansı yaklaşık olarak %80–82 olarak elde edilmiştir. 1997’ de, Friedman, Arttırılmış Ağaç Naive Bayes kullanılarak %85 sınıflama performansı elde etti. Lenf hastalığı teşhisi için kullanılan sınıflamaların performans karşılaştırılması aşağıdaki Çizelge 5.4’ de verilmiştir.

Çizelge 5.4 Lenf hastalığı teşhisi için sınıflama performansı karşılaştırılması

Yazar (yıl)	Yöntem	Sınıflama Doğruluğu(%)
Newton Cheung (2001)	C4.5	77.03
Newton Cheung (2001)	Naive Bayes	79.72
Newton Cheung (2001)	BNNF	80.88
Newton Cheung (2001)	BNND	81.08

6. AIRS VE FS-AIRS' NİN MEDİKAL VERİ SETİ ÜZERİNDEKİ SINIFLAMA PERFORMANSLARI

Bu bölümde, *AIRS* ve *FS-AIRS*' nin, bölüm 5' de anlattığımız bu dört medikal veri seti üzerindeki sınıflama performansları karşılaştırılacaktır. *FS-AIRS* algoritması bu dört medikal veri setinden sadece göğüs kanseri ve kalp hastalığı verilerine uygulanmıştır. Diğer veri setlerinde özellik seçme işlemi sonucu verilerin özellik sayısı değişmediğinden, bu veri setlerine *AIRS* algoritması uygulandı. Şimdi bu dört veri setinin özelliklerini ayrı ayrı grafik olarak gösterilecek ve aynı zamanda diğer sınıflama yöntemleriyle karşılaştırılacaktır.

6.1. Göğüs Kanseri Hastalığı Veritabanının AIRS ve FS-AIRS Sınıflama Performansı

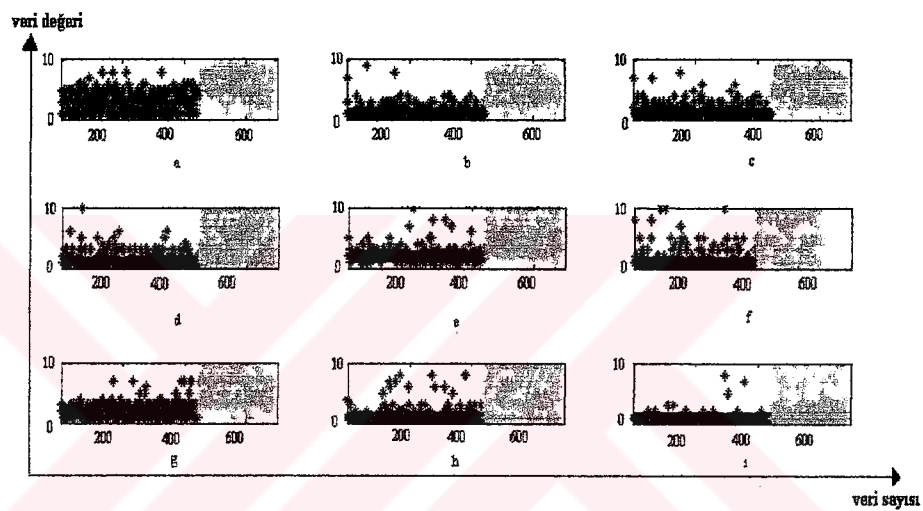
Göğüs kanseri veritabanı 683 örnekten oluşmaktadır. Tam sayı değerli dokuz özelliği bulunmaktadır ve her bir özellik 1' den 10' a aralıklı değerlere sahiptir. Göğüs kanseri veri setinin iki değişken sınıfı vardır: 0 ve 1 nümerik değerlerle sırayla, iyi huylu ve kötü huylu' dur.

Bu veritabanındaki 683 örnekten 444 tanesi iyi huylu sınıfa, 239 tanesi de kötü huylu sınıfa aittir. Göğüs kanseri veritabanında her bir durum için çıkış değerlerinin birkaçı Çizelge 6.1' de verilmiştir.

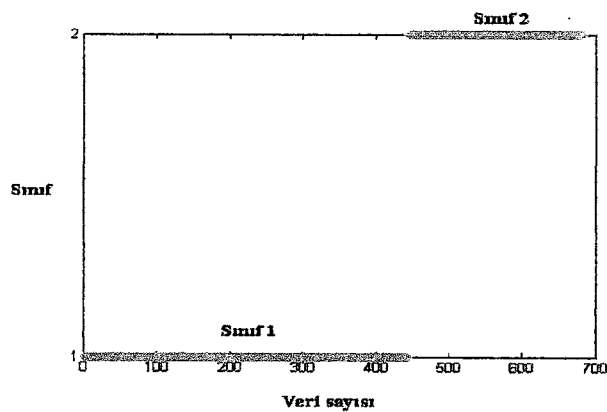
Bu veri setinin dokuz özelliği, sınıf değişkeni durumlarına göre Şekil 6.1' de verilmiştir. Şekil 6.2' de ise bu veri setinin sınıf dağılımı verilmiştir.

Çizelge 6.1 Göğüs kanseri veritabanında her bir durum için çıkış değerlerine bağlı özellikler

Durum	x_1	x_2	x_3	...	x_9	Teşhis
1	5	1	1	...	1	İyi huylu
2	5	4	4	...	1	İyi huylu
.
.
683	4	8	8	...	1	Kötü huylu



Şekil 6.1. Göğüs kanseri veritabanında özelliklerin sınıf değişkenine göre veri dağılımları. (a) Kütle kalınlığı, (b) Hücre büyüklüğünün benzerliği, (c) Hücre şeklinin benzerliği, (d) Marginal adhesion, (e) Tek epitel hücre büyüklüğü, (f) Yalın çekirdek, (g) Bland kromotin, (h) Normal çekirdek ve (i) Mitoz.



Şekil 6.2. Göğüs kanseri veritabanı sınıf dağılımları: Sınıf 1: 1–444, Sınıf 2: 445–683.

Göğüs kanseri veritabanına özellik seçme işlemi uygulandıktan sonra bu veritabanındaki özellik sayısı 9' dan 6' ya indirildi. Bu seçilen altı özellik, veritabanındaki verilerin dağılımına ve veriler arasındaki aralık farkına göre yapıldı. Bu seçilen altı özellik şunlardır:

- Hücre büyüklüğünün benzerliği
- Hücre şeklinin benzerliği
- Marginal Adhesion
- Tek epitel hücre büyüklüğü
- Yalın çekirdek
- Mitozlar

Göğüs kanseri veritabanındaki dokuz özellik kullanılarak yapılan *AIRS* sınıflama sonucu %97,2 başarı sağlanırken, bu veritabanına özellik seçme işlemi uygulandıktan sonra elde edilen bu altı özelliği göre sınıflama yapıldığında ise %98,51 başarı sağlanmıştır. Çizelge 6.2. ' de göğüs kanseri veritabanını değişik sınıflama algoritmasına göre performans karşılaştırılması verilmiştir. Kalp hastalığı veritabanını sınıflama işlemi sırasında, *AIRS* algoritmasında kullanılan parametreler Çizelge 6.3' de verilmiştir. Bu *AIRS* sınıflama algoritmasında kullanılan parametreler bölüm 4' de açıklanmıştır.

Çizelge 6.2 Göğüs kanseri hastalığı teşhisi için sınıflama performansı karşılaştırılması

Yazar (yıl)	Yöntem	Sınıflama Doğruluğu (%)
Quinlan(1996)	C4.5	94.74
Hamilton et al. (1996)	C4.5	96
Ster and Dobnikar (1996)	FLDA	96.8
Bennett and Blue (1997)	SVM (5xCV)	97.2
Nauck ve Kruse (1999)	Neuro-Fuzzy	95.06
Andres et al. (1999)	Fuzzy-GA	97.36
Setione (2000)	Neuro-Rule	98.1
Watkins (2002)	Optimized- LVQ	96.7
Watkins (2002)	Big- LVQ	96.8

Watkins (2002)	AIRS	97.2
Abonyi and Szeifert (2003)	SFC	95.57
Çalışmamız (2004)	FS-AIRS	98.51

Çizelge 6.3 Göğüs kanseri veritabanını sınıflamada kullanılan AIRS parametreleri

Kullanılan Parametre	Değeri
Mutasyon oranı	0.15
ATS (Affinity Threshold Scalar)	0.1
Uyarılma eşiği	0.91
Klonal oranı	10
Hiper klonal oranı	2.0
Kaynakların sayısı	250
İterasyon sayısı	10000
En yakın komşu için k değeri	1

6.2. Kalp Hastalığı Veritabanının AIRS ve FS-AIRS Sınıflama Performansı

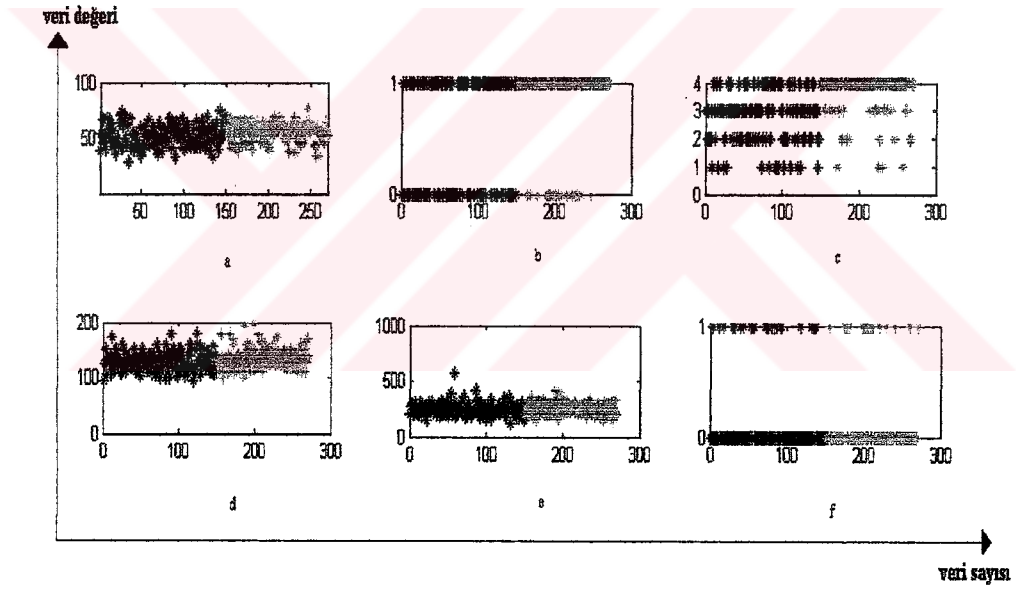
Kalp hastalığı veritabanı 270 örnekten oluşmaktadır. Orijinal olarak, veritabanı 76 ham özellikten oluşmaktadır. Bununla birlikte, bütün yayınlanmış deneylerde sadece bu özelliklerden 13 özellik kullanılmıştır. Bu özellikler ayrık değerli yani evet ve hayır şeklinde olan ve sürekli değerlerden oluşmaktadır. Kalp hastalığı veritabanını iki değişken sınıfı vardır: hastalık yok (% 55.56) ve hastalık var (%44.44).

Bu veritabanındaki 270 örnekten 150 tanesi hastalık yok sınıfa, 120 tanesi de hastalık var sınıfa aittir. Kalp hastalığı veritabanında her bir durum için çıkış değerlerinin birkaçı Çizelge 6.4' de verilmiştir.

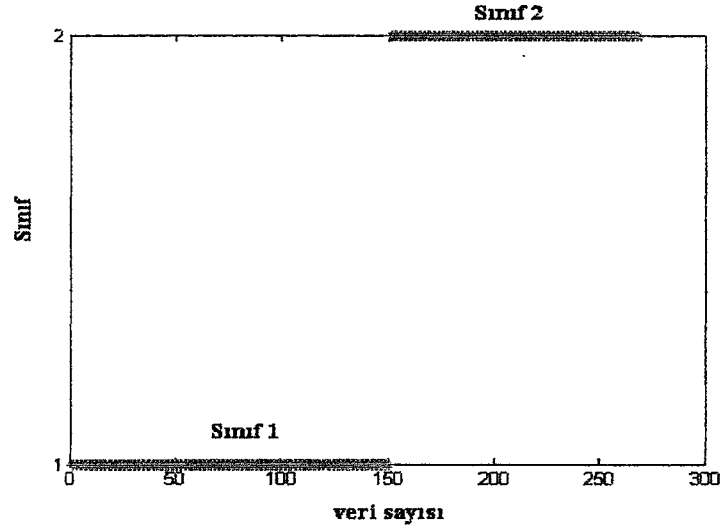
Bu veri setinin birkaç özelliği, sınıf değişkeni durumlarına göre Şekil 6.3' de verilmiştir. Şekil 6.4' de ise bu veri setinin sınıf dağılımı verilmiştir.

Çizelge 6.4 Kalp hastalığı veritabanında her bir durum için çıkış değerlerine bağlı özellikler

Durum	x ₁	x ₂	x ₃	x ₄	x ₅	x ₆	x ₇	x ₈	x ₉	x ₁₀	x ₁₁	x ₁₂	x ₁₃	Teşhis
1	44	1	3	120	226	0	0	169	0	0	1	0	3	Hastalık yok
2	44	1	3	140	235	0	2	180	0	0	1	0	3	Hastalık yok
...
...
270	63	0	4	108	269	0	0	169	1	1.8	2	2	3	Hastalık var



Şekil 6.3 Kalp hastalığı veritabanındaki birkaç özelliğin sınıf değişkenine göre veri dağılımları. (a) Yaş, (b) Cinsiyet, (c) Göğüs ağrısı tipi, (d) Kan basıncı, (e) Serum kolesterolü ve (f) Kan şekeri.



Şekil 6.4 Kalp hastalığı veritabanı sınıf dağılımları: Sınıf 1: 1–150, Sınıf 2: 151–270.

Kalp hastalığı veritabanına özellik seçme işlemi uygulandıktan sonra bu veritabanındaki özellik sayısı 13'den 9'a indirildi. Bu seçilen dokuz özellik, veritabanındaki verilerin dağılımına ve veriler arasındaki aralık farkına göre yapıldı. Bu seçilen dokuz özellik şunlardır:

- Yaş
- Cinsiyet
- Göğüs baş ağrısı tipi
- Maksimum elde edilen kalp atış oranı
- Angina' ya neden olan uygulama
- Angina' ya neden olan uygulama
- Flourosopy tarafından renklenen önemli odacıkların sayısı
- Thal – Talasemias

Kalp hastalığı veritabanındaki onüç özellik kullanılarak yapılan *AIRS* sınıflama sonucu %55,8 başarı sağlanırken, bu veritabanına özellik seçme işlemi uygulandıktan sonra elde edilen bu dokuz özelliği göre sınıflama yapıldığında ise %84,5 başarı sağlanmıştır. Çizelge 6.5. ' de göğüs kanseri veritabanını değişik sınıflama algoritmasına göre performans karşılaştırılması verilmiştir. Kalp hastalığı

veritabanını sınıflama işlemi sırasında, *AIRS* algoritmasında kullanılan parametreler Çizelge 6.6' de verilmiştir.

Çizelge 6.5 Kalp hastalığı teşhisi için sınıflama performansı karşılaştırılması

Yazar (yıl)	Yöntem	Sınıflama Doğruluğu (%)
Newton Cheung (2001)	C4.5	81.11
Newton Cheung (2001)	Naive Bayes	81.48
Newton Cheung (2001)	BNND	81.11
Newton Cheung (2001)	BNNF	80.96
Çalışmamız (2004)	<i>AIRS</i> (2004)	55.8
Çalışmamız (2004)	FS-<i>AIRS</i>	84.5

*Çizelge 6.6. Kalp hastalığı veritabanını sınıflamada kullanılan *AIRS* parametreleri*

Kullanılan Parametre	Değeri
Mutasyon oranı	0.1
ATS (Affinity Threshold Scalar)	0.2
Uyarılma eşiği	0.9
Klonal oranı	10
Hiper klonal oranı	2.0
Kaynakların sayısı	200
İterasyon sayısı	10000
En yakın komşu için k değeri	1

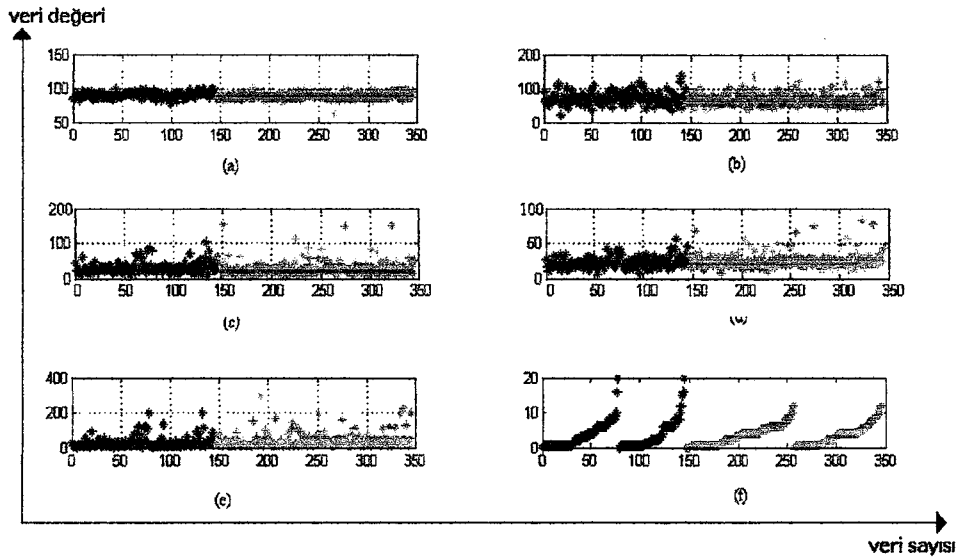
6.3. Karaciğer Rahatsızlığı Veritabanı AIRS ve FS-AIRS Sınıflama Performansı

BUPA TIP Araştırma Şirketi' nin hazırlamış olduğu BUPA Liver Disorders (Karaciğer bozuklukları) veri tabanı 6 özellik ve 2 sınıftan oluşan toplam 345 örnek içermektedir. Veri kümesindeki her örnek bekâr-erkek bir bireyden alınmıştır. Bu örneklerden 200' ü sınıflardan birine, kalan 145' i diğer sınıfta yer almaktadır. Karaciğer rahatsızlığı veritabanında her bir durum için çıkış değerlerinin birkaçı Çizelge 6.7' de verilmiştir.

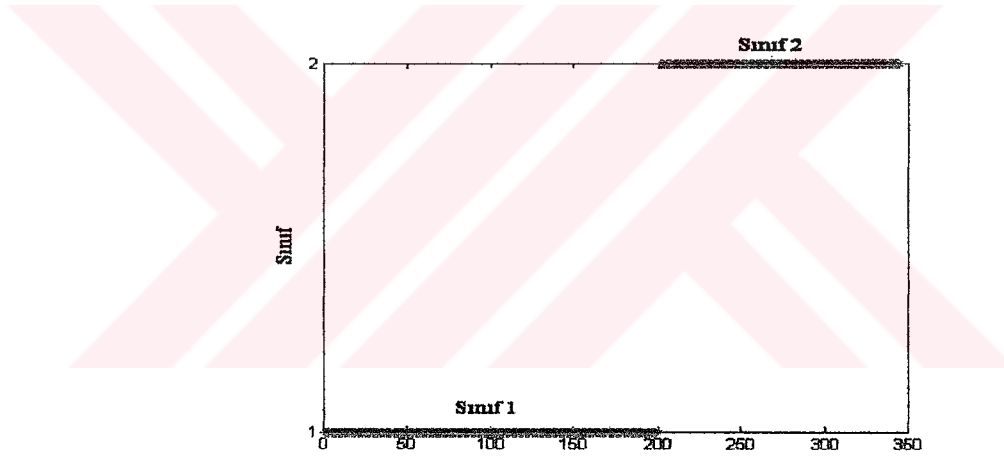
Bu veri setinin altı özelliği, sınıf değişkeni durumlarına göre Şekil 6.5' de verilmiştir. Şekil 6.6' de ise bu veri setinin sınıf dağılımı verilmiştir.

Çizelge 6.7 Karaciğer rahatsızlığı veritabanında her bir durum için çıkış değerlerine bağlı özellikler

Durum	x ₁	x ₂	x ₃	x ₄	x ₅	x ₆	Teşhis
1	85	92	45	27	31	0	Rahatsızlık var
2	88	62	20	17	9	0,5	Rahatsızlık var
...
...
345	85	64	59	32	23	0	Rahatsızlık yok



Şekil 6.5 BUPA karaciğer rahatsızlığı veri setinde bulunan özelliklere ait veri dağılımları



Şekil 6.6 Karaciğer rahatsızlığı veritabanı sınıf dağılımları: Sınıf 1: 1–200, Sınıf 2: 201–345.

Kalp hastalığı veritabanına özellik seçme işlemi uygulandıktan sonra bu veritabanındaki özellik sayısı değişmemiştir. Çünkü bu veritabanındaki veriler birbirine çok yakın olduğu için ve veriler arasındaki fark çok küçük olduğundan özellik sayısı indirilememiştir. Çizelge 6.8 ' de göğüs kanseri veritabanını değişik sınıflama algoritmasına göre performans karşılaştırılması verilmiştir. Karaciğer rahatsızlığı veritabanını sınıflama işlemi sırasında, *AIRS* algoritmasında kullanılan parametreler Çizelge 6.9' da verilmiştir.

Çizelge 6.8 Karaciğer rahatsızlığı teşhisi için sınıflama performansı karşılaştırılması

Yazar (yıl)	Yöntem	Sınıflama Doğruluğu (%)
Newton Cheung (2001)	C4.5	65.59
Newton Cheung (2001)	Naive Bayes	63.39
Tony Van Gestel et. al. (2001)	VMC	69.7
Yuh-Jye Lee ve O.L. Mangasarian	SSVM	70.33
Yuh-Jye Lee ve O.L. Mangasarian	RSVM	74.86
D.T. Pham et. Al.	Inductive Learning	53.76
T. Yıldırım et. al. (2003)	MLP	74.56
T. Yıldırım et. al. (2003)	PNN	73.86
T. Yıldırım et. al. (2003)	GRNN	60.68
Çalışmamız (2004)	AIRS	83.38

Çizelge 6.9. Karaciğer rahatsızlığı veritabanını sınıflamada kullanılan AIRS parametreleri

Kullanılan Parametre	Değeri
Mutasyon oranı	0.12
ATS (Affinity Threshold Scalar)	0.18
Uyarılma eşiği	0.84
Klonal oranı	10
Hiper klonal oranı	2.0
Kaynakların sayısı	200
İterasyon sayısı	10000
En yakın komşu için k değeri	1

6.4. Lenf hastalığı Veritabanının AIRS ve FS-AIRS Sınıflama Performansı

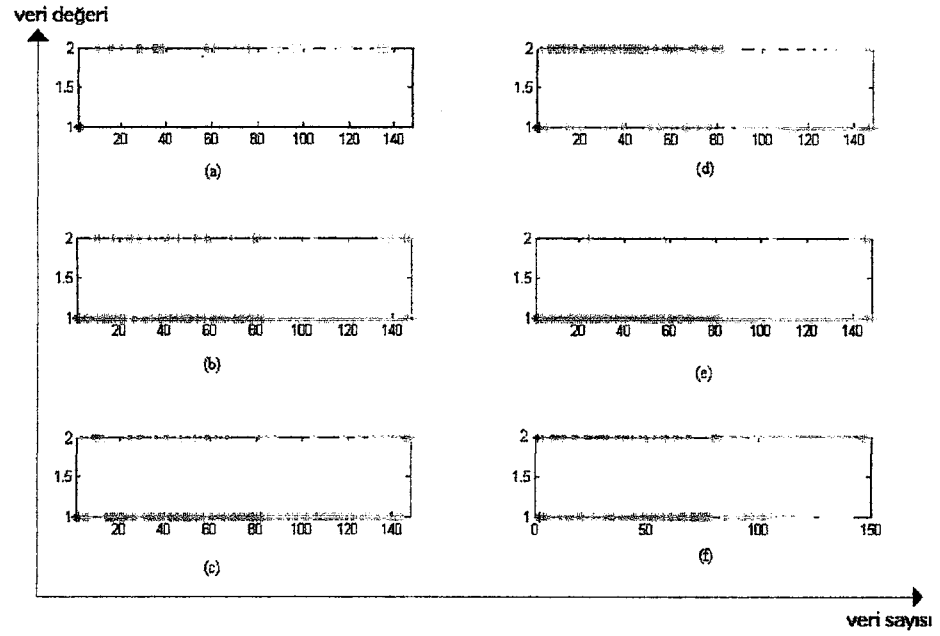
Lenf hastalığı veritabanı, Yugoslavya Medikal Center Üniversitesi Onkoloji Enstitüsü Ljubljana bölümünden elde edildi. Bu veritabanı toplam 148 örnekten oluşmaktadır ve hiçbir eksik veri yoktur. Bu veritabanı 18 nümerik özellikten

oluşmaktadır. Lenf hastalığı veritabanındaki sınıf değişkenlerinde dört sınıf bulunmaktadır: normal, metastases, kötü lenf ve fibrosis olmak üzere, sırayla 1, 2, 3 ve 4 tamsayılarıyla gösterilmektedir. Veritabanındaki 148 örneğin sınıf değişkenlerine göre dağılımları: normal 2 tane, metastases 81 tane, kötü lenf 61 tane ve fibrosis 4 tane olmak üzere 148 örnekten oluşmaktadır. Lenf hastalığı veritabanında her bir durum için çıkış değerlerinin birkaçı Çizelge 6.10' da verilmiştir.

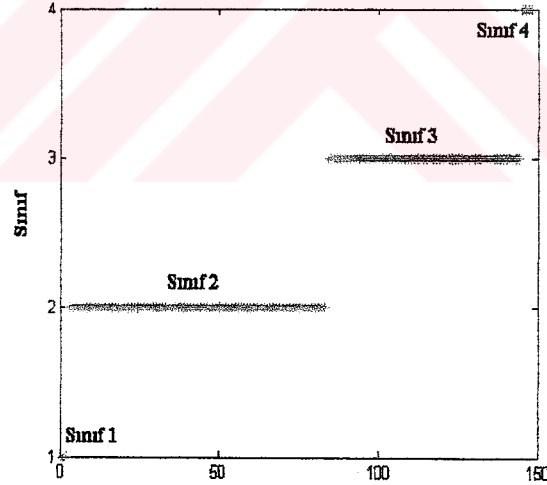
Bu veri setinin birkaç özelliği, sınıf değişkeni durumlarına göre Şekil 6.7' de verilmiştir. Şekil 6.8' de ise bu veri setinin sınıf dağılımı verilmiştir.

Çizelge 6.10 Lenf hastalığı veritabanında her bir durum için çıkış değerlerine bağlı özellikler

Durum	x ₁	x ₂	x ₃	...	x ₁₈	Teşhis
1	1	1	1	...	1	Normal
2	2	2	1	...	1	Metastates
3	3	1	1	...	3	Kötü lenf
4	3	1	1	...	4	Fibrosis
.
						...
148	1	1	1	...	2	Normal



Şekil 6.7 Lenf hastalığı veritabanındaki birkaç özelliğin sınıf değişkenine göre veri dağılımları. (a) Lymphatic (lenfosit), (b) Afferent' in bloğu, (c) C Lenfinin bloğu, (d) S Lenfinin bloğu, (e) Serum kolestrolü ve (f) By pass.



Şekil 6.8 Lenf hastalığı veritabanı sınıf dağılımları: Sınıf 1: 1–2, Sınıf 2: 3–83, Sınıf 3: 83–144 ve Sınıf 4: 144–148.

Lenf hastalığı veritabanına özellik seçme işlemi uygulandıktan sonra bu veritabanındaki özellik sayısı değişmemiştir. Çünkü bu veritabanındaki veriler birbirine çok yakın olduğu için ve veriler arasındaki fark çok küçük olduğundan

özelliik sayısı indirilememiştir. Çizelge 6.11' de göğüs kanseri veritabanını deęişik sınıflama algoritmasına göre performans karşılaştırılması verilmiştir. Karaciğer rahatsızlığı veritabanını sınıflama işlemi sırasında, AIRS algoritmasında kullanılan parametreler Çizelge 6.12' de verilmiştir.

Çizelge 6.11 Lenf hastalığı teşhisi için sınıflama performansı karşılaştırılması

Yazar (yıl)	Yöntem	Sınıflama Doğruluęu (%)
Newton Cheung (2001)	C4.5	77.03
Newton Cheung (2001)	Naive Bayes	79.72
Newton Cheung (2001)	BNNF	80.88
Newton Cheung (2001)	BNND	81.08
Çalışmamız (2004)	AIRS	83.139

Çizelge 6.12 Lenf hastalığı veritabanını sınıflamada kullanılan AIRS parametreleri

Kullanılan parametre	Deęeri
Mutasyon oranı	0.1
ATS (Affinity Threshold Scalar)	0.2
Uyarılma eşięi	0.9
Klonal oranı	10
Hiper klonal oranı	2.0
Kaynakların sayısı	200
İterasyon sayısı	10000
En yakın komşu için k deęeri	1

7. SONUÇLAR ve ÖNERİLER

Bu tez çalışmasında, Yapay Zeka alanında yeni alan olan Yapay Bağışıklık Sistemi algoritması olan *AIRS* ve Özellik Seçme işlemleri kullanarak medikal teşhis alanında uygulamalar yapılarak medikal teşhiste elde edilen sınıflama sonuçları, literatürde yapılan diğer çalışmalarla karşılaştırılmıştır. Ayrıca bu tez çalışmasında kullanılan medikal veritabanları detaylı bir biçimde açıklanmıştır.

Doğal bağışıklık sistemlerinde ve yapay bağışıklık sistemlerinde önceki çalışmaların gözlemlerini temel alan bir danışmanlı öğrenme metodu olan *AIRS* sistemi ayrıntılı bir biçimde açıklanmıştır. Bağışıklık sisteminin çeşitli anahtar bileşenleri *AIRS* yöntemini geliştirmek için kullanılmıştır. Bu bileşenlerden biri veri temsiliinde kullanılan *B* hücrelerinin soyutlanmasıdır. Saldırgan patojenin antijeni ve *B* hücresinin antikoru arasındaki reaksiyonlar *Euclidean uzaklığı* kullanılarak uyarılmaktadır. Kaynak sınırlı yapay bağışıklık sınıflandırıcısının gelişimi için kullanılan bileşenlerin çoğu daha önce araştırılmasına rağmen hiç birisi danışmanlı öğrenme için kullanılmamıştır. Danışmanlı öğrenme çeşitli şekillerde tanıtılmıştır. Birincisi, bir eğitim antijeni ile reaksiyonda olan sistem hücresinin uyarılması sadece o hücrenin antijenle olan duyarlılığını değil aynı zamanda antijenin sınıflandırılmasıyla kıyaslandığında hücrenin sınıflandırılmasını temel almaktadır. İkincisi, kaynaklar sınıflandırmayı temel alan hücreler için tahsis edilmektedir. Daha çok kaynağa sahip uyarma seviyesi antijenle aynı sınıftaki hücreler için tahsis edilmektedir. Üçüncüsü, hem uyarma seviyesini hem de sınıfı temel alan bir eğitimi durdurma kriteri yüksek kalitede sınıflandırma hücrelerinin sistemde kalmasını garantilemek için kullanılmaktadır. Dördüncüsü, bellek hücresi belirleme, hatırlama yeteneği ve yer değişim bellek hücresinin sınıflandırılmasında öncelikle temel alınmaktadır. Son olarak, daha önceden gözlemlenmemiş verinin sınıflandırılması bellek hücrelerinin geliştirilmiş kümesinin reaksiyonlarını ve sınıflandırılmasını temel almaktadır.

Bu tezde çalışmasında, *Özellik Seçme* işlemi *AIRS'* ye adapte edilerek geliştirilmiştir. Tek bir algoritmada hem veri ön-işleme hem de sınıflama yapma yeteneğine sahip danışmanlı bir öğrenme algoritmasıdır.

AIRS sınıflama algoritmasında ileride yapılacak yenilikler olarak, *AIRS* algoritmasında bellek hücresi yenileme için geçerli metot ağırlıkla iki alana eğilimlidir. Birincisi, bellek hücresi yenileme için eşige karar veren ölçüt eğitim hücrelerinin tamamı arasındaki ortalama duyarlılık değerini temel almaktadır. İkincisi, *AIRS* en çok son zamanlarda karşılaşılan antijene doğru bir eğilime sahiptir. Bu ikinci eğilim bir dereceye kadar doğa için doğru iken potansiyel olarak sınıflandırıcı hücreleri geliştirmek için zararlıdır. Bu eğilimlerin etkilerini araştırmak amacıyla *AIRS* için çeşitli değişiklikler yapılabilir. Potansiyel bir değişiklik olarak eğitim verilerinin tamamını kullanan geniş bir metottan ziyade aynı sınıfa sahip eğitim örneklerinin ortalama duyarlılığı temel alınarak duyarlılık eşigine karar verilebilir. Bu işlem veri kümesindeki her bir sınıfın muhtemelen daha artırılmış olarak ifade edilmesine sebep olacaktır. Bir başka değişiklik ikinci eğilimde yapılabilir. Bu değişiklik sisteme ikinci bir kez eğitim verisinin sunulması olabilir. Bununla beraber ikinci geçişte sadece bellek hücreleri geliştirilecektir. Düşünce olarak gelişmiş bellek hücrelerinin eğitim verisine tepkisine dikkat edilmiştir. Eğer bir bellek hücresi herhangi bir eğitim verisine hiç en yakın olmamışsa veya en çok uyarılmamışsa bu hücre bellek hücrelerinin kümesinden silinmelidir.

Bu tez çalışmasında kullanılan yöntemler geliştirilerek değişik uygulama alanlarında daha iyi sonuçlar elde edilebilir. Medikal teşhis, tıp alanında çok önemli bir yere sahip olduğu için bu konu üzerine çalışmalar devam edecektir.

8. KAYNAKLAR

- Ada, G. L. & Nossal, G. (1987), The Clonal Selection Theory, *Scientific American*, 257(2), pp. 50-57.
- Allen, D. et al. (1987), Timing, Genetic Requirements and Functional Consequences of Somatic Hypermutation During B-cell Development, *Imm. Rev.*, 96, pp. 5-22.
- Carter, J.H. (2000). The Immune System As a Model For Pattern Recognition And Classification. *Journal of the American Medical Informatics Association*, 7(1): 28-41.
- Coutinho, A. (1989), Beyond Clonal Selection and Network, *Immun. Rev.*, 110, pp.63-87.
- De Boer, R. J., Segel, L. A. & Perelson, A. S. (1992), Pattern Formation in One and Two dimensional Shape-Space Models of the Immune System, *J. theor. Biol.*, 155, pp. 295-333.
- Abbas, A. K., Lichtman, A. H., Pober, J. S. 1994. *Cellular and Molecular Immunology*, Fourth ed. W. B. Saunders Com., United States of America.
- Ada, G. L. & Nossal, G. J. V. 1987 The Clonal Selection Theory. *Scientific American* 257(2):50-57.
- Baysal, B. 2001. Temel İmmünoloji Ders Notları, Selçuk Üniversitesi Tıp Fakültesi, Mikrobiyoloji Anabilim Dalı, Konya.
- De Castro, L. N. & Von Zuben, F. J. (1999), *Artificial Immune Systems: Part I- Basic Theory and Applications*, Technical Report, December 1999.

De Castro, L. N. & Von Zuben, F. J. (2000a), The Clonal Selection Algorithm with Engineering Applications, submitted to GECCO'00.

De Castro, L. N. & Von Zuben, F. J. (2000b), Artificial Immune Systems: Part II- A Survey of Applications, Technical Report, December 2000.

De Castro, L. N., and F. J. Von Zuben. (2001a), Ainet: An Artificial Immune Network For Data Analysis.

<ftp://ftp.dca.fee.unicamp.br/pub/docs/vonzuben/lnunes/DMHA.pdf>

De Castro, L. N., and F. J. Von Zuben. (2001b). Learning and Optimization Using The Clonal Selection Principle.

ftp://ftp.dca.fee.unicamp.br/pub/docs/vonzuben/lnunes/ieee_tec01.pdf

Duch, W. (2000a). Datasets Used For Classification: Comparison of Results.

<http://www.phys.uni.torun.pl/kmk/projects/datasets.html> 5 Ocak 2004

Duch, W. (2000b). Logical Rules Extracted From Data.

<http://www.phys.uni.torun.pl/kmk/projects/rules.html> 20 Mayıs 2004

Farmer, J. D., Packard, N. H. & Perelson, A. S. (1986), The Immune System, Adaptation, and Machine Learning, Physica 22D, pp. 187-204.

Forrest, S., A. S. Perelson, L. Allen, and R. Cherukuri. (1994). Self-nonsel Discrimination in a Computer. In Proceedings of the IEEE computer society symposium on research in security and privacy held in Oakland, California, May 16-18, 1994, 202-212. IEEE Comp. Society Press.

Haykin S. (1999). Neural Networks – A Comprehensive Foundation, Prentice Hall, 2nd Ed., 1999.

- Hoffmann, G. W. (1986), A Neural Network Model Based on The Analogy with The Immune System, *J. theor. Biol.*, 122, pp. 33-67.
- Hofmeyr S. A. & Forrest, S. (1999), Immunity by Design: An Artificial Immune System, *Proc. Of GECCO'99*, pp. 1289-1296.
- Hodgkin, P. D. (1998), Role of Cross-Reactivity in the Development of Antibody Responses, *The Immunologist*, 6/6, pp. 223-226.
- Hunt, J. E. & Cooke, D. E. (1996), Learning Using an Artificial Immune System, *Journal of Network and Computer Applications*, 19, pp. 189-212.
- Hunt, J., J. Timmis, D. Cooke, M. Neal, and C. King. (1998). Jiys: The development of An Artificial Immune System For Real World Applications. In *Artificial Immune Systems and Their Applications*, Edited by D. Dasgupta, 157-186. Berlin: Springer.
- Janeway Jr., C. A. (1992), The Immune System Evolved to Discriminate Infectious Nonself From Noninfectious Self, *Imm. Today*, 13(1), pp. 11-16.
- Janeway Jr, C. A. & P. Travers (1997), *Immunobiology The Immune System in Health and Disease*, Artes Médicas (in Portuguese), 2nd Ed.
- Jerne, N. K. (1973), The Immune System, *Scientific American*, 229(1), pp. 52-60.
- Jerne, N. K. (1974). Towards a Network Theory of the Immune System, *Annals of Immunology. (Inst. Pasteur)* 125C, pp. 373-389.
- Jun, J. H., D. W. Lee, And K.B. Sim. (1999). Realization Of Cooperative Swarm Behavior In Distributed Autonomous Robotic Systems Using Artificial Immune System. In *Proceedings Of IEEE International Conference On Systems, Man, And Cybernetics, Held In Tokyo, Japan, November, 1999, Volume 6*, 614-619.

- Knight, T., And J. Timmis. (2001). Assessing The Performance Of The Resource Limited Artificial Immune System AINE. Computing Laboratory, University Of Kent At Canterbury. Technical Report 3-01.
- Mason, D. (1998), Antigen Cross-Reactivity: Essential in the Function of TCRs, *The Immunologist*, 6/6, pp. 220-222.
- Perelson, A. S., Mirmirani, M. & Oster, G. F. (1978), Optimal Strategies in Immunology II. B Memory Cell Production, *J. Math. Biol.*, 5, pp. 213-256.
- Perelson, A. S. & Oster, G. F. (1979), Theoretical Studies of Clonal Selection: Minimal Antibody Repertoire Size and Reliability of Self-Nonself Discrimination, *J. theor.Biol.*, 81, pp. 645–670.
- Perelson, A. S. (1989). Immune Network Theory. *Immunological Review*, 110, pp. 5–36.
- Perelson, A. S. & Weisbuch, G. (1997), “Immunology for Physicists”, *Rev. of Modern Physics*, 69(4), pp. 1219–1267.
- Segel, L. & Perelson, A. S. (1988), Computations in Shape Space: A New Approach to Immune Network Theory, In *Theoretical Immunology, Part Two*, (Ed.) A. S. Perelson, pp. 321-343.
- Smith, D. J., Forrest, S., Hightower, R. R. & Perelson, S. A. (1997), Deriving Shape Space Parameters from Immunological Data, *J. theor. Biol.*, 189, pp. 141-150.
- Sprent, J. (1994), T and B Memory Cells, *Cell*, 76, pp. 315-322.
- Sutton, R. S. & Barto, A. G. (1998), *Reinforcement Learning an Introduction*, A Bradford Book.

- Timmis, J., M. Nral, And J. Hunt. (1999). Data Analysis Using Artificial Immune Systems, Cluster Analysis And Kohonen Networks: Some Comparisons. In Proceedings Of IEEE International Conference On Systems, Man, And Cybernetics Held In Tokyo, Japan, November, 1999, Volume 3, 922-927.
- Timmis, J., And M. Neal. (2000). Invastigating The Evolution And Stability Of A Resource Limited Artificial Immune System. In Proceeding Of Genetic And Evolutionary Computation Conference (GECCO) 2000, Workshop Program, Held In Las Vegas, Nevada, July 2000, Edited By A.S. Wu, 40-41. AAAI Press.
- Timmis, J., M.Neal, And J. Hunt. (2000). An Artificial Immune System For Data Analysis. *Biosystems* 55(1/3): 143-150.
- Timmis, J. (2001). Artificial Immune Systems: A Novel Data Analysis Technique Inspired by the Immune Network Theory. Ph. D. Thesis, University Of Wales, Aberystwyth.
- Watanabe, Y., A. Ishiguro, And Y. Uchikawa. (1998). Decentralized Behavior Arbitration Mechanism For Autonomous Mobile Robot Using Immune Network. In *Artificial Immune Systems And Their Applications*, Edited By D. Dasgupta, 187-209. Berlin: Springer.
- Watkins, A., B. (2001). AIRS: A Resource Limited Artificial Immune Classifier, M.S. Thesis, Mississippi State University, December 2001.
- Xianglin L. Du, Charles R. Key, Cynthia Osborne, Jonathan D. Mahnken, and James S. G., Discrepancy Between Consensus Recommendations and Actual Community Use of Adjuvant Chemotherapy in Women with Breast Cancer Vol. (138) (2003) 90-97

- Qualin, J.R., Improved Use Of Continuous Features In C4.5. J., Artificial Intell. Res. Vol. (4) (1996) 77-90
- Hamilton, H.J., Shan, N. and Cercone, N., RIAC: A Rule Induction Algorithm Based on Approximate Classification, Tech. Rep. CS 96-06, Regina University (1996)
- Ster, B. and Dobnikar, A., Neural Networks in Medical Diagnosis: Comparison With Other Methods, Proceedings of the International Conference EANN '96, (1996) 427-430
- Bennett, K.P. and Blue, J., A Support A Support Vector Machine Approach to Decision Trees, Rensselaer Polytechnic Institute, Math Report, No. 97-100 (1997)
- Nauck, D. and Kruse, R., Obtaining Interpretable Fuzzy Classification Rules From Medical Data, Artificial Intelligence in Medicine. Vol. (16) (1999) 149-169
- Andress, C., Pena-Reyes and Sipper, M., A Fuzzy-Genetic Approach to Breast Cancer Diagnosis, Artificial Intelligence in Medicine. Vol. 17(1999) 131-155
- Setiona, R., Generating Concise and Accurate Classification Rules for Breast Cancer Diagnosis, Artificial Intelligence in Medicine. Vol. 18 (2000) 205-219
- Newton Cheung, Machine Learning Techniques for Medical Analysis, School of Information Technology and Electrical Engineering, BSc thesis, University of Queensland, 19 Oct. 2001.
- M. Yalçın and T. Yıldırım, Diagnosis of Liver Disorders by Artificial Neural Networks, Biyomedikal Mühendisliği Ulusal Toplantısı, BİYOMUT, İstanbul-Türkiye, pp.293-297, 2003.

Tony Van Gestel, Johan A.K. Suykens., Gert Lanckriet, Annemie Lambrechts, Bart De Moor and Joos Vandewalle, Bayesian Framework for Least Squares Supportvector Machine Classifiers, Gaussian Processes and Kernel Fisher Discriminant Analysis, August 2001.

Yuh-Jye Lee and O.L. Mangasarian, SSVN: A Smooth Support Vector Machine for Classification.

Yuh-Jye Lee and O.L. Mangasarian, RSVM: Reduced Support Vector Machines.

D.T. Pham, S.S. Dimov and Z. Salem, Technique for Selecting Examples in Inductive Learning.

BUPA Liver Disorders Dataset, UJI Repository of Machine Learning Databases, http://ftp.ics.uci.edu/pub/machine-learning-databases/liverdisorders/bupa_data, May, 2004