

T.C.

SELÇUK ÜNİVERSİTESİ

FEN BİLİMLERİ ENSTİTÜSÜ

**BİLGİ KAZANCI TABANLI
YAPAY BAĞIŞIKLIK TANIMA SİSTEMİ**

Halife KODAZ

DOKTORA TEZİ

**ELEKTRİK-ELEKTRONİK MÜHENDİSLİĞİ
ANABİLİM DALI**

Konya, 2007

**T.C.
SELÇUK ÜNİVERSİTESİ
FEN BİLİMLERİ ENSTİTÜSÜ**

**BİLGİ KAZANCI TABANLI
YAPAY BAĞIŞIKLIK TANIMA SİSTEMİ**

Halife KODAZ

DOKTORA TEZİ

ELEKTRİK-ELEKTRONİK MÜHENDİSLİĞİ ANABİLİM DALI

Bu tez 28.12.2007 tarihinde aşağıdaki jüri tarafından oybirliği/oyçokluğu ile kabul edilmiştir.

Prof. Dr. Mehmet BAYRAK

Üye

Prof. Dr. Saadetdin HERDEM

Üye

Prof. Dr. Şirzat KAHRAMANLI

Üye

Doç. Dr. Şeref SAĞIROĞLU

Üye

Yrd. Doç. Dr. Salih GÜNEŞ

Danışman

ÖZET

Doktora Tezi

BİLGİ KAZANCI TABANLI YAPAY BAĞIŞIKLIK TANIMA SİSTEMİ

Halife KODAZ

Selçuk Üniversitesi

Fen Bilimleri Enstitüsü

Elektrik-Elektronik Mühendisliği Anabilim Dalı

Danışman: Yrd. Doç. Dr. Salih GÜNEŞ

2007, 197 sayfa

Jüri: Prof. Dr. Mehmet BAYRAK

Prof. Dr. Saadetdin HERDEM

Prof. Dr. Şirzat KAHRAMANLI

Doç. Dr. Şeref SAĞIROĞLU

Yrd. Doç. Dr. Salih GÜNEŞ

Tez çalışmasında bir yapay bağışıklık sınıflandırıcı sistemi olan *Yapay Bağışıklık Tanıma Sistemi (YBTS)* incelenmiş ve tespit edilen eksiklikleri gidermek amacıyla *Bilgi Kazancı Tabanlı Yapay Bağışıklık Tanıma Sistemi (BK-YBTS)* geliştirilmiştir. Bu eksiklikler arasında, *YBTS'nin* çalışma süresinin çok zaman alması ve sınıflandırma performansının düşük olması sayılabilir. Tüm bu eksikliklere, hücreler arası duyarlılık hesabında bütün özelliklerin eşit ağırlıkta kabul edilmesi, mutasyon mekanizmasının rasgele değer atama şeklinde gerçekleşmesi ve duyarlılık eşik değeri hesabı ile kaynak tahsisi mekanizmasında eğitim verilerinin sınıf değerlerinin kullanılmaması sebep olmaktadır.

YBTS'den kaynaklanan bu eksiklikleri gidermek amacıyla geliştirilen sistemde; hücreler arası duyarlılık hesabında *bilgi kazancı* kavramı kullanılarak özelliklere ağırlık verilmiş, antikorlara mutasyon işlemi sırasında antikorun antijenle olan duyarlılığını temel alan bir mutasyon mekanizması kullanılmış ve duyarlılık eşiği hesabında ve kaynak tahsisi mekanizmasında eğitim verilerinin sınıf değerleri

temel alınmıştır. *BK-YBTS* ile gerçekleştirilen tüm bu değişikliklerin etkinliğini göstermek amacıyla çeşitli veri kümeleri üzerinde *BK-YBTS* ve *YBTS* parametrelerinin değişik değerleri ile iki sistem karşılaştırılmıştır.

Genel olarak *BK-YBTS*, *YBTS'den* daha yüksek sınıflandırma doğruluklarına ulaşmıştır. Özellikle mutasyon ve kaynak tahsisi mekanizmasının değiştirilmesiyle birlikte *BK-YBTS*, *YBTS'den* daha kısa sürede çalışmasını tamamlamıştır. Literatürdeki sınıflandırıcılarla karşılaştırıldığında *BK-YBTS* ile beraber kabul edilebilir sınıflandırma doğruluklarına ulaşılmıştır.

Bu çalışmada sunulan yaklaşımların gelecekte farklı uygulamalara ve yeni *Yapay Bağışıklık Sistemleri* tasarlanmasına katkılar sağlayacağı değerlendirilmektedir.

Anahtar Kelimeler – Yapay Bağışıklık Sistemleri, Sınıflandırma, Bilgi kazancı, Antijen, Antikor, Duyarlılık, Kaynak rekabeti, Bellek hücresi.

ABSTRACT

PhD Thesis

INFORMATION GAIN BASED ARTIFICIAL IMMUNE RECOGNITION SYSTEM

Halife KODAZ

Selcuk University

Graduate School of Natural and Applied Sciences

Department of Electrical-Electronics Engineering

Supervisor: Assist.Prof.Dr. Salih GÜNEŞ

2007, 197 pages

Jury: Prof. Dr. Mehmet BAYRAK

Prof. Dr. Saadetdin HERDEM

Prof. Dr. Şirzat KAHRAMANLI

Assoc. Prof. Dr. Şeref SAĞIROĞLU

Assist. Prof. Dr. Salih GÜNEŞ

This thesis detects some deficiencies of *Artificial Immune Recognition System (AIRS)* which is an artificial immune classifier system. *Information Gain Based Artificial Immune Recognition System (IG-AIRS)* is developed in order to avoid these deficiencies. Some of the deficiencies of *AIRS* are the long execution time and the poor classification performance. These deficiencies are caused by the assumption of all the data attributes to have the same weight in the affinity calculation, random value assignment in the mutation mechanism and the exclusion of the class values of training data in the affinity threshold value calculation and resource allocation mechanism.

In order to eliminate these drawbacks, *IG-AIRS* deployed the *information gain* concept in affinity calculation by assigning weights to data attributes. A mutation mechanism based on affinity is also used during mutation process of antibodies. Furthermore, the affinity threshold criterion and the resource allocation

mechanism are based on class values of training data. In order to show the efficiency of these modifications with proposed *IG-AIRS*, two systems are compared with various parameter values on several datasets.

In general, classification accuracies of *IG-AIRS* are found to be higher than *AIRS*. Especially, with the modification of mutation and resource allocation mechanisms, the execution time of *IG-AIRS* was shorter than *AIRS*. When compared with other classifiers in the literature, *IG-AIRS* have reached acceptable classification accuracies.

The presented approaches in this thesis are envisaged to contribute to new Artificial Immune Systems and diverse applications in the future.

Keywords - Artificial Immune Systems, Classification, Information gain, Antigen, Antibody, Affinity, Resource competition, Memory cell.

ÖNSÖZ

Çalışmama verdiği destek ve gösterdiği anlayıştan dolayı danışmanım sayın Yrd.Doç.Dr. Salih GÜNEŞ'e, tez süresince belirttikleri görüşler ve önerilerle tezin yönlenmesinde yardımcı olan sayın Prof.Dr. Mehmet BAYRAK'a ve sayın Doç.Dr.Şeref SAĞIROĞLU'na, Yapay Bağışıklık Sistemleri alanında çalışmama vesile olan sayın Prof.Dr. Ahmet ARSLAN'a, biyolojik olarak bağışıklık sistemi hakkında temel bilgiler veren sayın Prof.Dr. Bülent BAYSAL'a, beraberce yapmış olduğumuz çalışmalardan dolayı sayın Arş.Gör. Seral ÖZŞEN'e ve sayın Arş.Gör. Kemal POLAT'a, Selçuk Üniversitesi Mühendislik-Mimarlık Fakültesi Bilgisayar Mühendisliği Bölümündeki çalışma arkadaşlarıma ve özellikle manevi desteğini hiç bir zaman esirgemeyen her zaman ve her konuda hep yanımda olan aileme desteklerinden dolayı teşekkür ederim.

Halife KODAZ
Aralık 2007, Konya

İÇİNDEKİLER

ÖZET	i
ABSTRACT	iii
ÖNSÖZ	v
İÇİNDEKİLER	vi
SİMGELER VE KISALTMALAR	viii
ŞEKİLLER LİSTESİ	x
TABLolar LİSTESİ	xv
1. GİRİŞ	1
1.1. Literatür Araştırması	2
1.2. Çalışmanın Temel Amacı ve Literatüre Katkıları	11
1.3. Tezin Organizasyonu	13
2. BAĞIŞIKLIK SİSTEMİ	15
2.1. Temel Bileşenler	15
2.1.1. Bağışıklık sisteminin fizyolojisi	18
2.1.2. Doğal bağışıklık sistemi	21
2.1.3. Adaptif bağışıklık sistemi	22
2.2. Bağışıklık Mekanizmaları	23
2.2.1. Örüntü tanıma	23
2.2.2. Klonsal seçim prensibi	25
2.2.3. Bağışıklık öğrenimi ve bellek	26
2.2.4. Duyarlılık olgunlaşması	28
3. YAPAY BAĞIŞIKLIK SİSTEMLERİ	30
3.1. B Hücrelerinin Modellenmesi	30
3.2. Yapay Tanıma Toplari ve Kaynak Rekabeti	32
3.3. Bellek Hücreleri, Mutasyon ve Klonsal Seçim	32
3.4. Bağışıklık Ağları ve Hücreler Arası Etkileşimler	33
3.5. Klonsal Seçim Algoritması	34
3.6. Takviyeli Öğrenme ve Bağışıklık Belleği	36
3.7. Örüntü Tanıma	38
3.8. Şekil Uzay Modeli	38
3.9. Antijen-Antikor Gösterimleri ve Duyarlılık	40
4. KARAR AĞAÇLARI VE BİLGİ KAZANCI KAVRAMI	42
4.1. Standart Bir Veri kümesi Üzerinde Karar Ağacı Uygulaması	44
4.2. Özelliklerin Değerlerinin Ayrık Hale Getirilmesi	47
4.2.1. Danışmansız ayırım	47

4.2.1.1. Eşit aralık genişliği ayırım yaklaşımı.....	48
4.2.1.2. Eşit sıklık ayırım yaklaşımı.....	48
4.2.2. Danışmanlı ayırım.....	49
4.2.2.1. Holte'un 1R ayırıcısı.....	49
4.2.2.2. C4.5 ayırıcısı.....	50
4.2.2.3. Fayyad ve Irani'nin entropi tabanlı MDL metodu.....	50
5. GELİŞTİRİLEN SİSTEM.....	52
5.1. Tanımlamalar.....	52
5.2. Yapay Bağışıklık Tanıma Sistemi (YBTS).....	55
5.2.1. YBTS için normalizasyon.....	56
5.2.2. YBTS için antikor oluşumu ve bellek hücresi belirleme.....	56
5.2.3. YBTS için bir aday bellek hücresinin gelişimi ve kaynaklar için rekabet.....	58
5.2.4. YBTS için bellek hücresinin değerlendirilmesi.....	62
5.2.5. YBTS için sınıflandırma.....	63
5.3. YBTS'deki Bazı Eksiklikler.....	63
5.4. Bilgi Kazancı Tabanlı Yapay Bağışıklık Tanıma Sistemi (BK-YBTS).....	66
5.4.1. BK-YBTS için veri özelliklerinin bilgi kazancı değerlerinin hesaplanması.....	67
5.4.2. BK-YBTS için normalizasyon.....	70
5.4.3. BK-YBTS için antikor oluşumu ve bellek hücresi belirleme.....	71
5.4.4. BK-YBTS için aday bellek hücresinin gelişimi ve kaynaklar için rekabet.....	74
5.4.5. BK-YBTS için bellek hücresinin değerlendirilmesi.....	79
5.4.6. BK-YBTS için sınıflandırma.....	79
5.5. BK-YBTS Ayrıntıları.....	80
5.6. BK-YBTS'nin Karmaşıklık Analizi.....	84
6. UYGULAMALAR.....	87
6.1. BK-YBTS'de Kullanılan Parametreler.....	87
6.1.1. Çekirdek hücre sayısı.....	88
6.1.2. Kaynak sayısı.....	88
6.1.3. Uyarma eşiği.....	88
6.1.4. Mutasyon oranı.....	89
6.1.5. Duyarlılık eşik ölçüsü.....	89
6.1.6. k değeri.....	90
6.1.7. Parça değeri.....	91
6.2. YBTS ve BK-YBTS'nin Sınıflandırma Performanslarının Analizi.....	91
6.2.1. Iris veri kümesi uygulamaları.....	92
6.2.2. Thyroid veri kümesi uygulamaları.....	113
6.2.3. Damar tıkanıklığı veri kümesi uygulamaları.....	137
6.2.4. Bakteri veri kümesi uygulamaları.....	159
6.3. Bütün Veri Kümeleri İçin Elde Edilen Sınıflandırma Sonuçları.....	181
7. SONUÇLAR VE ÖNERİLER.....	183
8. KAYNAKLAR.....	189

SİMGELER VE KISALTMALAR

ab	: Tek bir antikoru ifade etmektedir.
AB	: Antikor kümesini ifade etmektedir.
ab.c	: Tek bir <i>ab</i> antikorunun sınıf değerini temsil etmektedir.
ab.uyarım	: Tek bir <i>ab</i> antikorunun uyarım seviyesini ifade etmektedir.
ab.kaynak	: Tek bir <i>ab</i> antikoruна tahsis edilen kaynak sayısını ifade etmektedir.
ag	: Tek bir <i>ag</i> antijenini ifade etmektedir.
ag.c	: Tek bir <i>ag</i> antijenin sınıfını ifade etmektedir.
ag.f	: Tek bir antijenin özellik vektörünü ifade etmektedir.
ATS	: Duyarlılık eşik ölçüsü (Affinity Threshold Scalar).
BK-YBTS	: Bilgi Kazancı Tabanlı Yapay Bağışıklık Tanıma Sistemi.
Çekirdek hücre	: Eğitim başlangıcında bellek hücresi ve antikor popülasyonlarını başlatmak için kullanılan ve eğitim kümesinden alınan bir antikor.
duyarlılık(x,y)	: <i>x</i> ve <i>y</i> özellik vektörü arasındaki bilgi kazancı tabanlı uzaklık.
Entropi	: Verideki rasgeleliği, belirsizliği, kirliliği, beklenmeyen durumun ortaya çıkma olasılığını karakterize eden bir ölçüdür.
kaynak_sayısı	: İzin verilen sistem kaynaklarının toplam sayısını ifade etmektedir.
Kazanç_i	: Veri kümesindeki <i>i</i> . özelliğin bilgi kazancı değerini ifade etmektedir.
L	: Özellik vektörlerinin boyutunu ifade etmektedir.
MC	: Bellek hücrelerinin kümesini ifade etmektedir.
mc	: <i>MC</i> kümesinin tek bir üyesini ifade etmektedir.
mc.c	: Belirli bir <i>mc</i> bellek hücresinin sınıfını temsil etmektedir.
nc	: Sınıf sayısını ifade etmektedir.
mc.f	: Bellek hücresinin özellik vektörünü ifade etmektedir.
mc_{ilk}	: Antijene ilk tepkide meydana gelen başlangıç bellek hücresini ifade etmektedir.
mc_{aday}	: Eğitim işleminin sonunda bellek hücreleri arasındaki antijene en çok tepki veren bellek hücresini ifade etmektedir.
MU	: Antikorların mutasyona uğramış klonlarının bir kümesini ifade etmektedir.

- Mutasyon oranı** : Bir antikorun herhangi bir özelliğinin mutasyona uğrama olasılığını işaret eden 0 ile 1 arasında bir parametredir.
- Parça** : Verinin sürekli özellik değerlerinin alabileceği ayrık değer sayısını ifade etmektedir.
- sınıf_i** : i . sınıfa ait verilerin oluşturduğu kümeyi ifade etmektedir.
- S** : Her bir sınıfa ait uyarma değerlerinin ortalamasını ve eğitim işleminin durma kriterini ifade etmektedir.
- Uyarma eşiği** : Antijenin eğitimi için bir durma kriteri olarak kullanılan 0 ile 1 arasında bir parametredir.
- YBS** : Yapay Bağışıklık Sistemi
- YBTS** : Yapay Bağışıklık Tanıma Sistemi

ŞEKİLLER LİSTESİ

Şekil 2.1 Bağışıklık sistemi bileşenleri (De Castro ve Timmis 2002a)	16
Şekil 2.2 Bağışıklık tanıma ve aktivasyon mekanizmaların basitleştirilmiş gösterimi (De Castro ve Timmis 2002a)	18
Şekil 2.3 Bağışıklık sisteminin anatomisi (De Castro ve Timmis 2002a)	19
Şekil 2.4 Bağışıklık sistemi koruma katmanları	21
Şekil 2.5 (a) <i>B</i> hücresi alıcısı (b) Antikor-Antijen etkileşimi	24
Şekil 2.6 (a) <i>T</i> hücresi alıcısı (b) Tanıma olayı	24
Şekil 2.7 Klonsal seçim prensibinin basitleştirilmiş açıklaması	26
Şekil 2.8 Birincil, ikincil ve çapraz bağışıklık tepkisi	28
Şekil 3.1 Klonsal Seçim Algoritmasının blok diyagramı (De Castro ve Von Zuben 2000a)	36
Şekil 3.2 Epitop olarak adlandırılan ve antikor tarafından tanınan antijen parçası	38
Şekil 3.3 Tanıma topu kavramı ve repertuar gösterimi	40
Şekil 4.1 Sınıf dağılımına göre belirsizlik değişimi	43
Şekil 4.2 Tenis örneği için oluşan karar ağacı	47
Şekil 5.1 YBTS'de kullanılan temel rutinler	55
Şekil 5.2 YBTS'de antikor oluşumu için mutasyon	57
Şekil 5.3 YBTS'de mutasyon rutini	58
Şekil 5.4 YBTS'de uyarma, kaynak tahsisi ve antikor çıkarılması	60
Şekil 5.5 YBTS'de hayatta kalan antikorların mutasyonu	61
Şekil 5.6 YBTS'de bellek hücresi seçimi	63
Şekil 5.7 BK-YBTS'nin temel işleyiş adımları	67
Şekil 5.8 BK-YBTS'de antikor oluşumu için mutasyon	72
Şekil 5.9 BK-YBTS'de antikor oluşumu için mutasyon mekanizmasının akış diyagramı	73
Şekil 5.10 BK-YBTS'de mutasyon rutini	73
Şekil 5.11 BK-YBTS'de mutasyon rutininin akış diyagramı	74
Şekil 5.12 BK-YBTS'de uyarma, kaynak tahsisi ve antikor çıkarılması	75
Şekil 5.13 BK-YBTS'de uyarma, kaynak tahsisi ve antikor çıkarılması sürecinin akış diyagramı	76
Şekil 5.14 BK-YBTS'de hayatta kalan antikorların mutasyonu	77
Şekil 5.15 BK-YBTS'de hayatta kalan antikorların mutasyonunun akış diyagramı	78
Şekil 5.16 BK-YBTS'de bellek hücresi seçimi	79
Şekil 5.17 BK-YBTS'de bellek hücresi seçiminin akış diyagramı	80
Şekil 6.1 Iris veri kümesinde çekirdek hücre sayısı parametresinin değişimine göre YBTS ve BK-YBTS'nin çalışma sürelerinin grafiksel gösterimi	94
Şekil 6.2 Iris veri kümesinde çekirdek hücre sayısı parametresinin değişimine göre YBTS ve BK-YBTS'nin iterasyon sayılarının grafiksel gösterimi	94
Şekil 6.3 Iris veri kümesinde çekirdek hücre sayısı parametresinin değişimine göre YBTS ve BK-YBTS'nin bellek hücreleri sayılarının grafiksel gösterimi	95
Şekil 6.4 Iris veri kümesinde çekirdek hücre sayısı parametresinin değişimine göre YBTS ve BK-YBTS'nin sınıflandırma doğruluklarının grafiksel gösterimi	95
Şekil 6.5 Iris veri kümesinde kaynak sayısı parametresinin değişimine göre YBTS ve BK-YBTS'nin çalışma sürelerinin grafiksel gösterimi	97
Şekil 6.6 Iris veri kümesinde kaynak sayısı parametresinin değişimine göre YBTS ve BK-YBTS'nin iterasyon sayılarının grafiksel gösterimi	97
Şekil 6.7 Iris veri kümesinde kaynak sayısı parametresinin değişimine göre YBTS ve BK-YBTS'nin bellek hücreleri sayılarının grafiksel gösterimi	98

Şekil 6.8 Iris veri kümesinde kaynak sayısı parametresinin değişimine göre YBTS ve BK-YBTS'nin sınıflandırma doğruluklarının grafiksel gösterimi.....	98
Şekil 6.9 Iris veri kümesinde uyarma eşiği parametresinin değişimine göre YBTS ve BK-YBTS'nin çalışma sürelerinin grafiksel gösterimi.....	100
Şekil 6.10 Iris veri kümesinde uyarma eşiği parametresinin değişimine göre YBTS ve BK-YBTS'nin iterasyon sayılarının grafiksel gösterimi.....	100
Şekil 6.11 Iris veri kümesinde uyarma eşiği parametresinin değişimine göre YBTS ve BK-YBTS'nin bellek hücreleri sayılarının grafiksel gösterimi	101
Şekil 6.12 Iris veri kümesinde uyarma eşiği parametresinin değişimine göre YBTS ve BK-YBTS'nin sınıflandırma doğruluklarının grafiksel gösterimi	101
Şekil 6.13 Iris veri kümesinde mutasyon oranı parametresinin değişimine göre YBTS ve BK-YBTS'nin çalışma sürelerinin grafiksel gösterimi	103
Şekil 6.14 Iris veri kümesinde mutasyon oranı parametresinin değişimine göre YBTS ve BK-YBTS'nin iterasyon sayılarının grafiksel gösterimi	103
Şekil 6.15 Iris veri kümesinde mutasyon oranı parametresinin değişimine göre YBTS ve BK-YBTS'nin bellek hücreleri sayılarının grafiksel gösterimi.....	104
Şekil 6.16 Iris veri kümesinde mutasyon oranı parametresinin değişimine göre YBTS ve BK-YBTS'nin sınıflandırma doğruluklarının grafiksel gösterimi	104
Şekil 6.17 Iris veri kümesinde duyarlılık eşik ölçüsü parametresinin değişimine göre YBTS ve BK-YBTS'nin çalışma sürelerinin grafiksel gösterimi.....	106
Şekil 6.18 Iris veri kümesinde duyarlılık eşik ölçüsü parametresinin değişimine göre YBTS ve BK-YBTS'nin iterasyon sayılarının grafiksel gösterimi.....	106
Şekil 6.19 Iris veri kümesinde duyarlılık eşik ölçüsü parametresinin değişimine göre YBTS ve BK-YBTS'nin bellek hücreleri sayılarının grafiksel gösterimi	107
Şekil 6.20 Iris veri kümesinde duyarlılık eşik ölçüsü parametresinin değişimine göre YBTS ve BK-YBTS'nin sınıflandırma doğruluklarının grafiksel gösterimi	107
Şekil 6.21 Iris veri kümesinde k değeri parametresinin değişimine göre YBTS ve BK-YBTS'nin sınıflandırma doğruluklarının grafiksel gösterimi	109
Şekil 6.22 Iris veri kümesinde parça parametresinin değişimine göre BK-YBTS'nin çalışma sürelerinin grafiksel gösterimi	110
Şekil 6.23 Iris veri kümesinde parça parametresinin değişimine göre BK-YBTS'nin iterasyon sayılarının grafiksel gösterimi	111
Şekil 6.24 Iris veri kümesinde parça parametresinin değişimine göre BK-YBTS'nin bellek hücreleri sayılarının grafiksel gösterimi.....	111
Şekil 6.25 Iris veri kümesinde parça parametresinin değişimine göre BK-YBTS'nin sınıflandırma doğruluklarının grafiksel gösterimi	112
Şekil 6.26 Thyroid veri kümesinde çekirdek hücre sayısı parametresinin değişimine göre YBTS ve BK-YBTS'nin çalışma sürelerinin grafiksel gösterimi.....	116
Şekil 6.27 Thyroid veri kümesinde çekirdek hücre sayısı parametresinin değişimine göre YBTS ve BK-YBTS'nin iterasyon sayılarının grafiksel gösterimi.....	117
Şekil 6.28 Thyroid veri kümesinde çekirdek hücre sayısı parametresinin değişimine göre YBTS ve BK-YBTS'nin bellek hücreleri sayılarının grafiksel gösterimi .	117
Şekil 6.29 Thyroid veri kümesinde çekirdek hücre sayısı parametresinin değişimine göre YBTS ve BK-YBTS'nin sınıflandırma doğruluklarının grafiksel gösterimi	118
Şekil 6.30 Thyroid veri kümesinde kaynak sayısı parametresinin değişimine göre YBTS ve BK-YBTS'nin çalışma sürelerinin grafiksel gösterimi.....	119
Şekil 6.31 Thyroid veri kümesinde kaynak sayısı parametresinin değişimine göre YBTS ve BK-YBTS'nin iterasyon sayılarının grafiksel gösterimi.....	120
Şekil 6.32 Thyroid veri kümesinde kaynak sayısı parametresinin değişimine göre YBTS ve BK-YBTS'nin bellek hücreleri sayılarının grafiksel gösterimi	120
Şekil 6.33 Thyroid veri kümesinde kaynak sayısı parametresinin değişimine göre YBTS ve BK-YBTS'nin sınıflandırma doğruluklarının grafiksel gösterimi	121

Şekil 6.34 Thyroid veri kümesinde uyarma eşiği parametresinin değişimine göre YBTS ve BK-YBTS'nin çalışma sürelerinin grafiksel gösterimi	122
Şekil 6.35 Thyroid veri kümesinde uyarma eşiği parametresinin değişimine göre YBTS ve BK-YBTS'nin iterasyon sayılarının grafiksel gösterimi	123
Şekil 6.36 Thyroid veri kümesinde uyarma eşiği parametresinin değişimine göre YBTS ve BK-YBTS'nin bellek hücreleri sayılarının grafiksel gösterimi	123
Şekil 6.37 Thyroid veri kümesinde uyarma eşiği parametresinin değişimine göre YBTS ve BK-YBTS'nin sınıflandırma doğruluklarının grafiksel gösterimi	124
Şekil 6.38 Thyroid veri kümesinde mutasyon oranı parametresinin değişimine göre YBTS ve BK-YBTS'nin çalışma sürelerinin grafiksel gösterimi	125
Şekil 6.39 Thyroid veri kümesinde mutasyon oranı parametresinin değişimine göre YBTS ve BK-YBTS'nin iterasyon sayılarının grafiksel gösterimi	126
Şekil 6.40 Thyroid veri kümesinde mutasyon oranı parametresinin değişimine göre YBTS ve BK-YBTS'nin bellek hücreleri sayılarının grafiksel gösterimi	126
Şekil 6.41 Thyroid veri kümesinde mutasyon oranı parametresinin değişimine göre YBTS ve BK-YBTS'nin sınıflandırma doğruluklarının grafiksel gösterimi	127
Şekil 6.42 Thyroid veri kümesinde duyarlılık eşik ölçüsü parametresinin değişimine göre YBTS ve BK-YBTS'nin çalışma sürelerinin grafiksel gösterimi	128
Şekil 6.43 Thyroid veri kümesinde duyarlılık eşik ölçüsü parametresinin değişimine göre YBTS ve BK-YBTS'nin iterasyon sayılarının grafiksel gösterimi	129
Şekil 6.44 Thyroid veri kümesinde duyarlılık eşik ölçüsü parametresinin değişimine göre YBTS ve BK-YBTS'nin bellek hücreleri sayılarının grafiksel gösterimi	129
Şekil 6.45 Thyroid veri kümesinde duyarlılık eşik ölçüsü parametresinin değişimine göre YBTS ve BK-YBTS'nin sınıflandırma doğruluklarının grafiksel gösterimi	130
Şekil 6.46 Thyroid veri kümesinde k değeri parametresinin değişimine göre YBTS ve BK-YBTS'nin sınıflandırma doğruluklarının grafiksel gösterimi	131
Şekil 6.47 Thyroid veri kümesinde parça değeri parametresinin değişimine göre BK-YBTS'nin çalışma sürelerinin grafiksel gösterimi	133
Şekil 6.48 Thyroid veri kümesinde parça değeri parametresinin değişimine göre BK-YBTS'nin iterasyon sayılarının grafiksel gösterimi	133
Şekil 6.49 Thyroid veri kümesinde parça değeri parametresinin değişimine göre BK-YBTS'nin bellek hücreleri sayılarının grafiksel gösterimi	134
Şekil 6.50 Thyroid veri kümesinde parça değeri parametresinin değişimine göre BK-YBTS'nin sınıflandırma doğruluklarının grafiksel gösterimi	134
Şekil 6.51 Damar tıkanıklığı veri kümesinde çekirdek hücre sayısı parametresinin değişimine göre YBTS ve BK-YBTS'nin çalışma sürelerinin grafiksel gösterimi	139
Şekil 6.52 Damar tıkanıklığı veri kümesinde çekirdek hücre sayısı parametresinin değişimine göre YBTS ve BK-YBTS'nin iterasyon sayılarının grafiksel gösterimi	140
Şekil 6.53 Damar tıkanıklığı veri kümesinde çekirdek hücre sayısı parametresinin değişimine göre YBTS ve BK-YBTS'nin bellek hücreleri sayılarının grafiksel gösterimi	140
Şekil 6.54 Damar tıkanıklığı veri kümesinde çekirdek hücre sayısı parametresinin değişimine göre YBTS ve BK-YBTS'nin sınıflandırma doğruluklarının grafiksel gösterimi	141
Şekil 6.55 Damar tıkanıklığı veri kümesinde kaynak sayısı parametresinin değişimine göre YBTS ve BK-YBTS'nin çalışma sürelerinin grafiksel gösterimi	142
Şekil 6.56 Damar tıkanıklığı veri kümesinde kaynak sayısı parametresinin değişimine göre YBTS ve BK-YBTS'nin iterasyon sayılarının grafiksel gösterimi	143
Şekil 6.57 Damar tıkanıklığı veri kümesinde kaynak sayısı parametresinin değişimine göre YBTS ve BK-YBTS'nin bellek hücreleri sayılarının grafiksel gösterimi	143

Şekil 6.58 Damar tıkanıklığı veri kümesinde kaynak sayısı parametresinin değişimine göre YBTS ve BK-YBTS'nin sınıflandırma doğruluklarının grafiksel gösterimi	144
Şekil 6.59 Damar tıkanıklığı veri kümesinde uyarma eşiği parametresinin değişimine göre YBTS ve BK-YBTS'nin çalışma sürelerinin grafiksel gösterimi.....	145
Şekil 6.60 Damar tıkanıklığı veri kümesinde uyarma eşiği parametresinin değişimine göre YBTS ve BK-YBTS'nin iterasyon sayılarının grafiksel gösterimi.....	146
Şekil 6.61 Damar tıkanıklığı veri kümesinde uyarma eşiği parametresinin değişimine göre YBTS ve BK-YBTS'nin bellek hücreleri sayılarının grafiksel gösterimi .	146
Şekil 6.62 Damar tıkanıklığı veri kümesinde uyarma eşiği parametresinin değişimine göre YBTS ve BK-YBTS'nin sınıflandırma doğruluklarının grafiksel gösterimi	147
Şekil 6.63 Damar tıkanıklığı veri kümesinde mutasyon oranı parametresinin değişimine göre YBTS ve BK-YBTS'nin çalışma sürelerinin grafiksel gösterimi.....	148
Şekil 6.64 Damar tıkanıklığı veri kümesinde mutasyon oranı parametresinin değişimine göre YBTS ve BK-YBTS'nin iterasyon sayılarının grafiksel gösterimi.....	149
Şekil 6.65 Damar tıkanıklığı veri kümesinde mutasyon oranı parametresinin değişimine göre YBTS ve BK-YBTS'nin bellek hücreleri sayılarının grafiksel gösterimi .	149
Şekil 6.66 Damar tıkanıklığı veri kümesinde mutasyon oranı parametresinin değişimine göre YBTS ve BK-YBTS'nin sınıflandırma doğruluklarının grafiksel gösterimi	150
Şekil 6.67 Damar tıkanıklığı veri kümesinde duyarlılık eşik ölçüsü parametresinin değişimine göre YBTS ve BK-YBTS'nin çalışma sürelerinin grafiksel gösterimi	151
Şekil 6.68 Damar tıkanıklığı veri kümesinde duyarlılık eşik ölçüsü parametresinin değişimine göre YBTS ve BK-YBTS'nin iterasyon sayılarının grafiksel gösterimi	152
Şekil 6.69 Damar tıkanıklığı veri kümesinde duyarlılık eşik ölçüsü parametresinin değişimine göre YBTS ve BK-YBTS'nin bellek hücreleri sayılarının grafiksel gösterimi	152
Şekil 6.70 Damar tıkanıklığı veri kümesinde duyarlılık eşik ölçüsü parametresinin değişimine göre YBTS ve BK-YBTS'nin sınıflandırma doğruluklarının grafiksel gösterimi	153
Şekil 6.71 Damar tıkanıklığı veri kümesinde k değeri parametresinin değişimine göre YBTS ve BK-YBTS'nin sınıflandırma doğruluklarının grafiksel gösterimi	154
Şekil 6.72 Damar tıkanıklığı veri kümesinde parça değeri parametresinin değişimine göre BK-YBTS'nin çalışma sürelerinin grafiksel gösterimi.....	156
Şekil 6.73 Damar tıkanıklığı veri kümesinde parça değeri parametresinin değişimine göre BK-YBTS'nin iterasyon sayılarının grafiksel gösterimi	156
Şekil 6.74 Damar tıkanıklığı veri kümesinde parça değeri parametresinin değişimine göre BK-YBTS'nin bellek hücreleri sayılarının grafiksel gösterimi	157
Şekil 6.75 Damar tıkanıklığı veri kümesinde parça değeri parametresinin değişimine göre BK-YBTS'nin sınıflandırma doğruluklarının grafiksel gösterimi	157
Şekil 6.76 Bakteri veri kümesinde çekirdek hücre sayısı parametresinin değişimine göre YBTS ve BK-YBTS'nin çalışma sürelerinin grafiksel gösterimi.....	162
Şekil 6.77 Bakteri veri kümesinde çekirdek hücre sayısı parametresinin değişimine göre YBTS ve BK-YBTS'nin iterasyon sayılarının grafiksel gösterimi.....	162
Şekil 6.78 Bakteri veri kümesinde çekirdek hücre sayısı parametresinin değişimine göre YBTS ve BK-YBTS'nin bellek hücreleri sayılarının grafiksel gösterimi	163
Şekil 6.79 Bakteri veri kümesinde çekirdek hücre sayısı parametresinin değişimine göre YBTS ve BK-YBTS'nin sınıflandırma doğruluklarının grafiksel gösterimi	163
Şekil 6.80 Bakteri veri kümesinde kaynak sayısı parametresinin değişimine göre YBTS ve BK-YBTS'nin çalışma sürelerinin grafiksel gösterimi	165

Şekil 6.81 Bakteri veri kümesinde kaynak sayısı parametresinin değişimine göre YBTS ve BK-YBTS'nin iterasyon sayılarının grafiksel gösterimi.....	165
Şekil 6.82 Bakteri veri kümesinde kaynak sayısı parametresinin değişimine göre YBTS ve BK-YBTS'nin bellek hücreleri sayılarının grafiksel gösterimi.....	166
Şekil 6.83 Bakteri veri kümesinde kaynak sayısı parametresinin değişimine göre YBTS ve BK-YBTS'nin sınıflandırma doğruluklarının grafiksel gösterimi	166
Şekil 6.84 Bakteri veri kümesinde uyarma eşiği parametresinin değişimine göre YBTS ve BK-YBTS'nin çalışma sürelerinin grafiksel gösterimi	168
Şekil 6.85 Bakteri veri kümesinde uyarma eşiği parametresinin değişimine göre YBTS ve BK-YBTS'nin iterasyon sayılarının grafiksel gösterimi.....	168
Şekil 6.86 Bakteri veri kümesinde uyarma eşiği parametresinin değişimine göre YBTS ve BK-YBTS'nin bellek hücreleri sayılarının grafiksel gösterimi.....	169
Şekil 6.87 Bakteri veri kümesinde uyarma eşiği parametresinin değişimine göre YBTS ve BK-YBTS'nin sınıflandırma doğruluklarının grafiksel gösterimi	169
Şekil 6.88 Bakteri veri kümesinde mutasyon oranı parametresinin değişimine göre YBTS ve BK-YBTS'nin çalışma sürelerinin grafiksel gösterimi.....	171
Şekil 6.89 Bakteri veri kümesinde mutasyon oranı parametresinin değişimine göre YBTS ve BK-YBTS'nin iterasyon sayılarının grafiksel gösterimi.....	171
Şekil 6.90 Bakteri veri kümesinde mutasyon oranı parametresinin değişimine göre YBTS ve BK-YBTS'nin bellek hücreleri sayılarının grafiksel gösterimi	172
Şekil 6.91 Bakteri veri kümesinde mutasyon oranı parametresinin değişimine göre YBTS ve BK-YBTS'nin sınıflandırma doğruluklarının grafiksel gösterimi	172
Şekil 6.92 Bakteri veri kümesinde duyarlılık eşik ölçüsü parametresinin değişimine göre YBTS ve BK-YBTS'nin çalışma sürelerinin grafiksel gösterimi.....	174
Şekil 6.93 Bakteri veri kümesinde duyarlılık eşik ölçüsü parametresinin değişimine göre YBTS ve BK-YBTS'nin iterasyon sayılarının grafiksel gösterimi.....	174
Şekil 6.94 Bakteri veri kümesinde duyarlılık eşik ölçüsü parametresinin değişimine göre YBTS ve BK-YBTS'nin bellek hücreleri sayılarının grafiksel gösterimi .	175
Şekil 6.95 Bakteri veri kümesinde duyarlılık eşik ölçüsü parametresinin değişimine göre YBTS ve BK-YBTS'nin sınıflandırma doğruluklarının grafiksel gösterimi	175
Şekil 6.96 Bakteri veri kümesinde k değeri parametresinin değişimine göre YBTS ve BK-YBTS'nin sınıflandırma doğruluklarının grafiksel gösterimi	177
Şekil 6.97 Bakteri veri kümesinde parça değeri parametresinin değişimine göre BK-YBTS'nin çalışma sürelerinin grafiksel gösterimi.....	178
Şekil 6.98 Bakteri veri kümesinde parça değeri parametresinin değişimine göre BK-YBTS'nin iterasyon sayılarının grafiksel gösterimi	179
Şekil 6.99 Bakteri veri kümesinde parça değeri parametresinin değişimine göre BK-YBTS'nin bellek hücreleri sayılarının grafiksel gösterimi	179
Şekil 6.100 Bakteri veri kümesinde parça değeri parametresinin değişimine göre BK-YBTS'nin sınıflandırma doğruluklarının grafiksel gösterimi	180

TABLolar LİSTESİ

Tablo 4.1 Tenis örneği için veri özellikleri.....	44
Tablo 4.2 Tenis örneği için veri kümesi	45
Tablo 4.3 Örnek bir özelliğin eşit aralık genişliği yaklaşımı ile ayrık hale getirilmiş hali	48
Tablo 4.4 Örnek bir özelliğin eşit sıklık yaklaşımı ile ayrık hale getirilmiş hali	49
Tablo 4.5 Örnek bir özelliğin Holte'un 1R ayırıcısı yaklaşımı ile ayrık hale getirilmiş hali ..	50
Tablo 5.1 BK-YBTS'nin çalışma adımları	83
Tablo 5.2 Büyük O ifadeleri ve anlamları.....	85
Tablo 5.3 YBTS ve BK-YBTS için karmaşıklık analizi.....	85
Tablo 6.1 YBTS ve BK-YBTS parametreleri	87
Tablo 6.2 Iris veri kümesi için YBTS ve BK-YBTS'nin başlangıç parametre değerleri	93
Tablo 6.3 Iris veri kümesinde çekirdek hücre sayısı parametresi için YBTS ve BK-YBTS karşılaştırma sonuçları	93
Tablo 6.4 Iris veri kümesinde kaynak sayısı parametresi için YBTS ve BK-YBTS karşılaştırma sonuçları	96
Tablo 6.5 Iris veri kümesinde uyarma eşiği parametresi için YBTS ve BK-YBTS karşılaştırma sonuçları	99
Tablo 6.6 Iris veri kümesinde mutasyon oranı parametresi için YBTS ve BK-YBTS karşılaştırma sonuçları	102
Tablo 6.7 Iris veri kümesinde duyarlılık eşik ölçüsü parametresi için YBTS ve BK-YBTS karşılaştırma sonuçları	105
Tablo 6.8 Iris veri kümesinde k parametresi için YBTS ve BK-YBTS karşılaştırma sonuçları	108
Tablo 6.9 Iris veri kümesinde parça değeri parametresi için BK-YBTS sonuçları.....	110
Tablo 6.10 Iris veri kümesi için optimum BK-YBTS parametre değerleri.....	113
Tablo 6.11 Iris veri kümesi için BK-YBTS ve diğer sınıflandırıcıların sınıflandırma doğrulukları (Watkins 2005).....	113
Tablo 6.12 Thyroid veri kümesinin sınıf dağılımı	114
Tablo 6.13 Thyroid veri kümesi için YBTS ve BK-YBTS'nin başlangıç parametre değerleri	115
Tablo 6.14 Thyroid veri kümesinde çekirdek hücre sayısı parametresi için YBTS ve BK-YBTS karşılaştırma sonuçları	115
Tablo 6.15 Thyroid veri kümesinde kaynak sayısı parametresi için YBTS ve BK-YBTS karşılaştırma sonuçları	119
Tablo 6.16 Thyroid veri kümesinde uyarma eşiği parametresi için YBTS ve BK-YBTS karşılaştırma sonuçları	121
Tablo 6.17 Thyroid veri kümesinde mutasyon oranı parametresi için YBTS ve BK-YBTS karşılaştırma sonuçları	124
Tablo 6.18 Thyroid veri kümesinde duyarlılık eşik ölçüsü parametresi için YBTS ve BK-YBTS karşılaştırma sonuçları	128
Tablo 6.19 Thyroid veri kümesinde k değeri parametresi için YBTS ve BK-YBTS karşılaştırma sonuçları	131
Tablo 6.20 Thyroid veri kümesinde parça değeri parametresi için BK-YBTS sonuçları	132
Tablo 6.21 Thyroid veri kümesi için optimum BK-YBTS parametre değerleri	135
Tablo 6.22 Thyroid veri kümesi için BK-YBTS ve literatürdeki diğer sınıflandırıcılar tarafından elde edilen sınıflandırma doğrulukları	135

Tablo 6.23 Damar tıkanıklığı veri kümesi için YBTS ve BK-YBTS'nin başlangıç parametre değerleri.....	138
Tablo 6.24 Damar tıkanıklığı veri kümesinde çekirdek hücre sayısı parametresi için YBTS ve BK-YBTS karşılaştırma sonuçları.....	139
Tablo 6.25 Damar tıkanıklığı veri kümesinde kaynak sayısı parametresi için YBTS ve BK-YBTS karşılaştırma sonuçları.....	142
Tablo 6.26 Damar tıkanıklığı veri kümesinde uyarma eşiği parametresi için YBTS ve BK-YBTS karşılaştırma sonuçları.....	144
Tablo 6.27 Damar tıkanıklığı veri kümesinde mutasyon oranı parametresi için YBTS ve BK-YBTS karşılaştırma sonuçları.....	148
Tablo 6.28 Damar tıkanıklığı veri kümesinde duyarlılık eşik ölçüsü parametresi için YBTS ve BK-YBTS karşılaştırma sonuçları.....	151
Tablo 6.29 Damar tıkanıklığı veri kümesinde k parametresi için YBTS ve BK-YBTS karşılaştırma sonuçları.....	154
Tablo 6.30 Damar tıkanıklığı veri kümesinde parça değeri parametresi için BK-YBTS sonuçları	155
Tablo 6.31 Damar tıkanıklığı veri kümesi için BK-YBTS'nin 10 kat çapraz onaylama yöntemi ile ayrıntılı sınıflandırma doğrulukları	158
Tablo 6.32 Damar tıkanıklığı veri kümesi için BK-YBTS ve literatürdeki diğer sınıflandırıcılar tarafından elde edilen sınıflandırma doğrulukları.....	159
Tablo 6.33 Bakteri veri kümesi ayrıntıları	160
Tablo 6.34 Bakteri veri kümesi için YBTS ve BK-YBTS'nin başlangıç parametre değerleri.....	161
Tablo 6.35 Bakteri veri kümesinde çekirdek hücre sayısı parametresi için YBTS ve BK-YBTS karşılaştırma sonuçları.....	161
Tablo 6.36 Bakteri veri kümesinde kaynak sayısı parametresi için YBTS ve BK-YBTS karşılaştırma sonuçları.....	164
Tablo 6.37 Bakteri veri kümesinde uyarma eşiği parametresi için YBTS ve BK-YBTS karşılaştırma sonuçları.....	167
Tablo 6.38 Bakteri veri kümesinde mutasyon oranı parametresi için YBTS ve BK-YBTS karşılaştırma sonuçları.....	170
Tablo 6.39 Bakteri veri kümesinde duyarlılık eşik ölçüsü parametresi için YBTS ve BK-YBTS karşılaştırma sonuçları.....	173
Tablo 6.40 Bakteri veri kümesinde k parametresi için YBTS ve BK-YBTS karşılaştırma sonuçları	176
Tablo 6.41 Bakteri veri kümesinde parça değeri parametresi için BK-YBTS sonuçları	178
Tablo 6.42 Bakteri veri kümesi için BK-YBTS parametre değerleri.....	181
Tablo 6.43 Bakteri veri kümesi için BK-YBTS ve diğer sınıflandırıcılar tarafından elde edilen sınıflandırma doğrulukları	181
Tablo 6.44 BK-YBTS'nin bütün veri kümeleri için sınıflandırma doğrulukları ve parametre değerleri.....	182
Tablo 6.45 YBTS ve BK-YBTS'nin bütün veri kümeleri için sınıflandırma doğrulukları ..	182

1. GİRİŞ

Teknolojinin ilerlemesiyle beraber bilinen yöntemler karmaşık problemleri çözmede yetersiz kalmaya başlayınca, farklı çözüm yolları araştırılmaya başlanmıştır. Bunu takiben ortaya çıkan Yapay Sistemler, özellikle 1900'li yılların ortalarından itibaren adını sıkça duyurmaya başlamıştır. İnsanın biyolojik çalışmasından esinlenerek, yapay sinir ağları ve genetik algoritmalar gibi değişik yapay modeller geliştirilmiştir. Bilim adamları, mühendisler, matematikçiler, filozoflar ve diğer araştırmacılar bu sistemin yetenekleri ile özellikle ilgilenmiş ve *Yapay Bağışıklık Sistemi (YBS)* olarak adlandırılan yeni bir araştırma alanı geliştirilmiştir. Bu sistemin özellikleri arasında başlıcaları hatırlama, genelleme, sınıflandırma, tanıma, özellik çıkarımı, çeşitlilik, öğrenme, hafıza, gürültü toleransı vb. sayılabilir (De Castro ve Von Zuben 1999; 2000b, De Castro ve Timmis 2002a).

Bağışıklık sistemi, yapay bağışıklık sisteminin gelişmesi için bir esin kaynağı olarak bilgi işleme yetenekleri bakımından oldukça zengindir. Diğer Yapay Zeka tekniklerinin ortaya çıkışında olduğu gibi gelişimini, bağışıklık sisteminde gerçekleşen olayları kavramaya yönelik oluşturulan bağışıklık sistemi modellerine borçludur. Bu modellerden yola çıkılarak oluşturulan sistemler kısa zamanda araştırmacıların katkıları sayesinde pek çok alanda yapılan çalışmalar ile bu alandaki gelişmelere katkıda bulunmaktadır. Bu tez çalışması da bu amaca hizmet etmek için hazırlanmıştır.

Tez çalışmasında, Watkins (2001) tarafından geliştirilen danışmanlı bir sınıflandırma sistemi olan *Yapay Bağışıklık Tanıma Sistemi (Artificial Immune Recognition System- AIRS)* temel alınarak diğer YBS algoritmalarında da uygulanabilecek değişiklikler önerilmektedir.

YBS algoritmalarının çoğu hücreler arasındaki etkileşimi hesaplamak için uzaklık kriterini kullanmaktadır. Verilerin özellik sayısının arttığı durumlarda bu algoritmaların performansı düşmektedir. Tez çalışmasında ilk olarak verinin özelliklerine ağırlık verilmesi önerilmektedir. Ağırlıklandırma işlemi karar ağacı sınıflandırma metodunda kullanılan her bir özelliğin bilgi kazancı değeri

hesaplanarak yapılmaktadır. Bilgi kazancı değeri büyük olan özelliklere daha büyük ağırlık, bilgi kazancı değeri küçük olan özelliklere daha küçük ağırlık verilmektedir.

Yine aynı şekilde çoğu *YBS* algoritması mutasyon işlemini uygulamaktadır. Bu çalışmada mutasyon işlemi rasgele olmayıp, verinin duyarlılığı ile ters orantılı olarak olmaktadır. Yani duyarlılığı yüksek bir veri daha düşük bir oranda mutasyona uğramaktadır. Duyarlılığı düşük bir veri daha yüksek bir oranda mutasyona uğramaktadır.

YBTS'de kaynaklar antijenle aynı sınıftaki antikorlar için en fazla kullanıcı tarafından belirlenmiş olan toplam maksimum kaynağın yarısı olacak şekilde tahsis edilmektedir ve kalan yarısı diğer sınıflara ait antikorlar arasında eşit olarak paylaşılmaktadır. Geliştirilen sistemde bu mekanizma veri kümesindeki sınıfların oranı hesaba katılarak değiştirilmiştir.

Geliştirilen sistemde, bellek hücresi seçimine karar veren ölçüt, eğitim hücrelerinin tamamı arasındaki ortalama duyarlılık değeri yerine aynı sınıfa sahip eğitim hücrelerinin ortalama duyarlılık değerini temel alacak şekilde değiştirilmiştir.

Bir diğer öneri olarak ta sistem eğitildikten sonra test veri kümesindeki verilerin sınıfına karar verilirken bellek hücreleri üzerinde *k en yakın komşu* yönteminde çoğunluk oylaması yapılırken sınıf sayılarında eşitlik durumunda veri kümesinde o sınıflara ait verilerden çok olan sınıf temel alınmaktadır.

Tüm bu değişiklikleri içerecek şekilde geliştirilen sisteme *Bilgi Kazancı Tabanlı Yapay Bağışıklık Tanıma Sistemi (BK-YBTS)* adı verilmiştir. *BK-YBTS*'nin sınıflandırma performansı makine öğrenmesi standart veri kümeleri ve gerçek dünya veri kümeleri üzerinde değerlendirilmiştir.

1.1. Literatür Araştırması

YBS alanındaki çalışmalar 1990'lı yılların sonlarına doğru ivme kazanmaya başlamıştır. *YBS* alanında ilk olarak bağışıklık sistemlerini temel alan hesaplama modelleri inşa etme amacıyla teorik tartışmalar ve uygulamalar gerçekleştirilmiştir. Araştırmacıların bazıları tam olarak bağışıklık sistemini simule etmeye çalışmışlar bazıları da yaratılış modelinden esinlenen yapay sinir ağları ve genetik algoritmalar

alanları gibi bilgi işleme mantığıyla bağışıklık sistemini kullanmaya çalışmışlardır. *YBS*, anormallik tespiti, kümeleme, optimizasyon ve sınıflandırma gibi konularla ilgili bugüne kadar yapılmış çalışmalar söz konusudur. Aşağıda bu çalışmaların bazılarına kısaca yer verilmiştir.

Perelson ve Weisbuch (1997) tarafından doğal bağışıklık sisteminin matematiksel bir perspektiften derin bir inceleme yapılmıştır. Perelson ve Oster (1979) antijen ve antikor arasındaki etkileşimi açıklamak amacıyla “*Şekil Uzay*” kavramını sunmuşlardır. *Şekil Uzay* kavramı pek çok *YBS* çalışmasında kullanılmıştır. Antijenler ve antikorlar özellik sayıları olan L boyutlu uzayda temsil edilmiştir. *Öklid*, *Manhattan* ve *Hamming* uzaklık ölçütleri ile şekil uzayında çeşitli çalışmalar gerçekleştirilmiştir (Segel ve Perelson 1988; De Boer ve ark. 1992; Smith ve ark. 1997, Perelson ve ark. 1996; Oprea ve Forrest 1999).

Matzinger (1994), bağışıklık sisteminin saldırganları nasıl tanıdığını anlayabilmek için mevcut olan anormallikleri açıklamak için tehlike teorisini (*Danger Theory*) ilk kez önermiştir. Tehlike teorisi saldırganların hücre ölümü veya hücresel baskı başlatarak (tehlike sinyali) hücresel moleküllerin oluşumunu tetiklediğini ileri sürmektedir. Bu moleküller bağışıklık tepkisini başlatan antijen tanıyan hücreler olarak bilinmektedir.

Forrest ve ark. (1994), bağışıklık sisteminde T hücrelerinin oluşumunu temel alan bir değişikliği algılayarak doğal hücrelerimiz ve yabancı hücrelerin veya maddelerin (self/nonself) ayırımını modelleyen *Negatif Seçim Algoritmasını* sunmuşlardır. Hesaplama maliyeti açısından matematiksel bir analiz gerçekleştirilmiş ve bilgisayar virüsleri üzerinde algoritmanın nasıl uygulanabileceği gösterilmiştir. Dasgupta ve Forrest (1995) self/nonself ayırımını yapan *negatif seçim algoritmasını* kullanarak değişikliği algılama algoritması sunmuşlardır. Sinyal işleme uygulamalarında problem olarak karşımıza çıkan zaman serisi verilerindeki anormallikler tespit edilmeye çalışılmıştır. D’haeseleer ve ark. (1996) algoritmanın çalışma maliyetini düşürmeye çalışmışlardır.

Forrest ve Hofmeyr (1999), yapay bağışıklık sistemlerinin saldırılara karşı bilgisayar ağlarını korumada kullanılabilecek özelliklerini belirtmişlerdir. *YBS* bir sınıflandırıcı sistemle karşılaştırılmıştır. Sınıflandırıcı sisteme olan benzerliği ve farklılıkları tartışılmıştır.

Kim ve Bentley (1999), bağışıklık sistemi ve ağ saldırı tespit sistemleri arasındaki ilişkiyi incelemişlerdir. Çalışmada mevcut saldırı tespit sistemlerinden kısaca bahsedilmiştir. Bağışıklık sisteminin özelliklerinin ağ tabanlı saldırı tespit sistemi geliştirmeye olan katkıları sunulmuştur.

Watkins (2000), saldırı tespit sistemlerini incelemiştir. Bağışıklık sisteminden modellenen saldırı tespit yaklaşımının geleneksel yöntemlerden daha etkili olabileceğini tartışmıştır.

Hofmeyr ve Forrest (1999), negatif seçime dayalı *ARTIS (Artificial Immune System)* adını verdikleri sistemi tasarlamışlardır. Bağışıklık sistemindeki negatif-pozitif seçim mekanizmasının modellendiği *ARTIS* sistemi özellikle güvenlik uygulamaları için uygun bir sistemdir. Uygulama alanları arasında ağ güvenliği, bilgisayar güvenliği, hata tespit sistemleri gibi alanlar yer almaktadır. Güvenlik alanında ortaya atılmış çoğu *YBS* algoritmasının temelini *ARTIS* sistemi oluşturmaktadır.

Hofmeyr ve Forrest (2000), *LISYS (Lightweight Intrusion Detection System)* olarak adlandırdıkları bir ağ saldırı tespit sistemi geliştirmişlerdir. Yürütülen çalışmalarda bağışıklık sistemindeki üç temel özellikten yararlanılmaya çalışılmıştır: Dağıtılmış tanıma, duyarlılık olgunlaşması ve çapraz-reaktif hafıza. Tespit işlemini ağ üzerindeki *TCP* paketlerini izleyerek gerçekleştirmişlerdir. New Mexico Üniversitesi Bilgisayar Bilimi Bölümündeki 50 bilgisayardan toplanan veriler üzerinde sistem test edilmiştir. Ortalama olarak %90'lara varan bir başarıya ulaşılmıştır. Balthrop ve ark. (2002) başka bir veri kümesi üzerinde *LISYS* uygulaması gerçekleştirmişlerdir.

Dasgupta ve ark. (2004), uçaklardaki hata tespiti için *negatif seçim algoritmasını* araştırmışlardır. Algılama algoritması anormallikleri tespit edebilmek için normal uçuş davranışı örüntülerini kullanmaktadır. Veri kümesi olarak *C-17 uçuş simülatörü* kullanılmıştır. Kuyruk ve kanatlardaki hata tespit oranları belirtilmiştir.

Ji ve Dasgupta (2006), *negatif seçim algoritmasının* çeşitli uygulama alanlarını araştırmışlardır. Yazarlar *negatif seçim algoritmasının* uygulandığı alanların metodun kullanımı için uygun olmadığını belirtmişlerdir. Yapay veriler ve iyi bilinen gerçek dünya verileri çeşitli uzaklık ölçüleri ile algoritmanın etkili ve

güçlü bir algoritma olduğu gösterilmiştir. Yazarlar her metodun her veri kümesi için uygun olmadığını belirtmişlerdir.

Yapay sinir ağlarının insan beynini ve genetik algoritmaların doğadaki evrimi model aldığı gibi *YBS* bağışıklık sistemlerinden esinlenilerek hesaplama için bir model olarak ortaya çıkarılmıştır. Vücudun bağışıklık sistemi çok fazla komplekstir ve çoğu parçası tam olarak anlaşılamamıştır. 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* hücreleridir. *T* hücreleri olgunlaşma sürecinde timüste (*Thymus*) dolaştığı için, *B* hücreleri kemik iliğinde (*Bone marrow*) olgunlaştığı için bu adları almışlardır. *YBS* çalışmalarının çoğu *B* hücrelerine odaklanarak makine öğrenmesi için *YBS* kullanmayı araştırmaktadırlar. *B* hücreleri bağışıklık sistemindeki bellek hücresi gibi kullanılmaktadır. *YBS* dizaynında pek çok yararlı özellik barındıran bağışıklık ağ teorisi Jerne (1974) tarafından formüle edilmiş ve Perelson (1989) tarafından geliştirilmiştir. Bu teoriye göre bağışıklık tepkisi sadece *B* hücresi ile antijen arasındaki etkileşimi temel almamakta aynı zamanda *B* hücrelerinin kendi aralarındaki etkileşimi de temel almaktadır. Bu hücreler birbirlerini hem uyarmakta hem de baskılamaktadır. Hunt ve Cooke (1996); Timmis ve ark. (1999); (2000); Timmis ve Neal (2000); Timmis(2001)'deki çalışmalar bağışıklık ağ teorisinin uygulamaları olarak karşımıza çıkmaktadır.

Çok ilginç fikirlerden biri olan kaynak sınırlı yaklaşım kavramı, Timmis tarafından çalışılan *Yapay Bağışıklık Ağından* esinlenilerek geliştirilmiştir. Timmis (2000) danışmansız öğrenme algoritması olarak *AINE* (*Artificial Immune Network*) olarak adlandırdığı sistemi geliştirmiştir. Bu kavramın temelinde aynı özelliklere sahip *B* hücrelerinin bir topluluğunu temsil eden *Yapay Tanıma Topları* (*YTT*) kullanılmaktadır. Temel kavram oldukça basittir. Geliştirilen sistemde eğitim kümesi olarak antijen kümesi tanımlanmıştır ve amaç bu antijenlere uygun *B* hücreleri veya antikolar üretmektir. *AINE* eğitim kümesindeki verilerin anahtar özelliklerini açıklayan bir ağ oluşturmaktadır. Her bir *B* hücresi antijenleri tanıma yeteneğindedir. Benzer *B* hücreleri birbirlerine bağlanılarak *B* hücrelerinin bir ağını oluşturmaktadır.

Daha sonra B hücreleri klonlanıp mutasyona uğratarak veriyi daha iyi temsil edebilecek B hücreleri oluşturulmaya çalışılmaktadır. Belirli sayıda B hücrelerinin oluşturduğu yapıya yapay tanıma topu ismi verilmektedir. *AINE* bağışıklık ağı oluştururken tek tek B hücreleri yerine yapay tanıma toplarının popülasyonunu kullanmaktadır. Verilen bir antijen ile bağışıklık hücresinin her bir karşılaşmasında her bir hücre, uyarma seviyesini temel alarak kaynaklar elde etmektedir. Bununla beraber burada sistemin kaynakları sınırlıdır. Eğer sistemde izin verilenden daha fazla kaynak tüketilirse kaynakları kullanan bazı hücreler sistemden çıkartılır. Bu işleme en az uyarılan hücelere 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. *AINE*'de dört adet kontrol parametresi söz konusudur. Bunlar ağ duyarlılık eşiği, mutasyon oranı, yapay tanıma toplarının sayısı ve klonların sayısıdır. Bu parametrelerin etkisini daha sonra Knight ve Timmis (2001) araştırmışlardır.

Mevcut kaynak sınırlamalı *AINE* sisteminde hala bir takım eksikliklerin ve dezavantajların bulunması nedeniyle *AINE* sistemi Knight ve Timmis (2001) tarafından yeniden ele alınmış mevcut eksiklikler giderilmeye çalışılmıştır. Söz konusu yenilikler algoritmadaki kaynak paylaşımı mekanizması ve işlemlerin yapılış sırası ile ilgili olarak gerçekleştirilmiştir. Yenilenmiş *AINE* algoritmasının performansını eskisi ile karşılaştırmak amacıyla eski *AINE* algoritmasının uygulandığı problemlerin hepsine yeni *AINE* algoritması uygulanarak sonuçlar analiz edilmiştir ve yapılan yeniliklerin performansı artırdığı gözlemlenmiştir.

De Castro ve Von Zuben (2001), veri analizi, tanıma ve sınıflandırma problemlerini çözmek için *aiNet* (*Artificial Immune Network*) adını verdikleri yapay bağışıklık ağ modelini geliştirmişlerdir. Bu sistemde amaç B hücreleri veya yapay tanıma toplarının yerine antikor popülasyonunu geliştirmektedir. *AINE* antijenlerin rasgele alt kümesini kullanırken *aiNet* antikor popülasyonunu rasgele seçmektedir. Daha sonra klonsal seçim ve mutasyon işlemi uygulamaktadır. Algoritma ağın dinamiklerini klonsal seçim ile kontrol etmektedir. *aiNet*'in performansı *spir*, *chainlink* ve *5-NLSC* veri kümelerinde değerlendirmiştir.

Ağ tabanlı algoritmalar arasında dikkate değer bir diğer algoritma *AIN* (*Artificial Immune Network*) sistemidir. Wierzchon ve Kuzelevska (2001) tarafından oluşturulan *AIN* sisteminde de temel olarak *AINE* sistemi alınmış, *AINE* sistemindeki eşik seviyesi hesabı ve uyarım seviyesi hesabı değiştirilmiştir. Oluşturulan sistem standart verilerde oldukça iyi sonuçlar elde etmiştir. *AIN* sistemi ayrıca *AINE* ve *aiNet* sistemi ile de karşılaştırılmıştır.

Nasraoui ve ark. (2002) yapay tanıma topunu geliştirmek için eğitim veri kümesi üzerine bulanık küme uygulayarak bulanık tanıma topu kavramını önermişlerdir. Web kayıt dosyaları üzerinde kümeleme uygulamaları gerçekleştirmişlerdir. Klonlama aşamasını ve popülasyon boyunu kontrol ederek erken yakınsamadan kaçınmaya çalışmışlardır.

YBS optimizasyon alanında popüler bir araştırma konusudur. Optimizasyon problemlerini çözmek için çeşitli bağışıklık algoritmaları önerilmiştir. Bersini ve Varela (1990) nümerik optimizasyon, Mori ve ark. (1993) bağışıklık ağ teorisini kullanarak bir genetik algoritmanın global arama performansını geliştiren bir bağışıklık algoritması önermişlerdir. Chun ve ark. (1997) Mori'nin algoritmasını geliştirmişlerdir ve elektromagnetik cihazların şekil optimizasyonu için uygulamışlardır. Jiao ve Wang (2000) genetik algoritmanın yakınsama hızını arttırmak için bir bağışıklık operatörü kullanan bir bağışıklık algoritması önermişlerdir.

Yapay Bağışıklık Sistemleri arasında en önemli yeri tutan *Klonsal Seçim Algoritması*, De Castro ve Von Zuben (2000a) tarafından ortaya atılmıştır. Ortaya atılan algoritma antijene karşı bağışıklık tepkisinin temel özelliklerini tanımlamak için Bağışıklık Sistemindeki klonsal seçme mekanizması model alınarak oluşturulmuştur. Bu çalışmada temel düşünce antijeni tanıyan hücreleri çoğaltmaktır. Çoğaltılan hücreler duyarlılık olgunlaşması sürecine sokulmuştur. Örüntü tanıma ve optimizasyon gibi görevlerde algoritmanın performansı değerlendirilmiştir. Değerlendirilen görevler sırasıyla her bir karakterin boyutu 120 olan bir bit karakter dizisi ile ifade edildiği ikili karakter tanıma problemi, fonksiyon maksimizasyonu ve gezgin tüccar problemidir.

Ağ tabanlı *YBS* algoritmalarının çoğu danışmansız eğitime prosedürü uygulamışlardır. Danışmanlı *YBS* algoritması olarak iki sistem karşımıza

çıkmaktadır. Bunlar Carter (2000) tarafından ortaya atılan *Immunos_81* algoritması ve Watkins (2001) tarafından oluşturulan *YBTS`dir (Yapay Bağışıklık Tanıma Sistemi)*.

Carter (2000) tarafından geliştirilen *Immunos_81* algoritmasında bağışıklık sistemindeki *T* hücreleri, *B* hücreleri, Antikorlar ve bunlar arasındaki etkileşimler model alınmıştır. Bazı standart verilere uygulandığında *Immunos_81* algoritmasının performansının %63.2-%96 arasında değiştiği gözlemlenmiştir.

Watkins (2001) *Yapay Bağışıklık Tanıma Sistemi (YBTS)* adını verdiği bağışıklık sisteminden esinlenen danışmanlı bir öğrenme algoritması geliştirmiştir. Algoritma, Timmis'in ortaya attığı *AINE* sistemindeki Yapay Tanıma Topu kavramını baz alarak oluşturulmuştur. Watkins tez çalışmasında geliştirdiği sistemi bilinen veri kümelerini kullanarak literatürdeki diğer sınıflandırıcılarla performansını karşılaştırmıştır. Eğitim ve test kümeleri çapraz onaylama metoduyla belirlenmiştir. Watkins ve Boggess (2002) çeşitli veri kümeleri üzerinde *YBTS'nin* performansını test etmişlerdir. Marwah ve Boggess (2002) *Yeast* ve *E.Coli* veri kümeleri üzerinde *YBTS'nin* performansını değerlendirmişlerdir. Sistemin eğitim aşamasından sonra test verilerinin hangi sınıfta olduğu hesaplanırken *k en yakın komşu* yönteminde çoğunluk oylaması yapılırken eşitlik durumunda çözümler sunmuşlardır. Watkins ve Timmis (2002) *YBTS'nin* orijinal versiyonunda değişiklikler yapmışlardır. Yeni versiyon ile orijinal versiyonu karşılaştırmışlardır. Bellek hücresi sayısında azalma sağlanırken sınıflandırma doğruluğunda fazla bir iyileşme gerçekleşmemiştir.

Secker ve ark. (2003) elektronik postaların sınıflandırılması için *AISEC (Artificial Immune System for E-mail Classification)* adını verdikleri bağışıklıktan esinlenilmiş bir algoritma geliştirmişlerdir. Ekim 2002 ve Mart 2003 arasındaki 2268 adet elektronik postayı sınıflandırmışlar ve *Bayes* sınıflandırıcısı ile karşılaştırmışlardır.

Greensmith (2003) tezinde, çok sınıflı doküman sınıflandırma için *YBTS* kullanımını önermiştir. Doküman HTML belgelerinin ön işlemeden geçmiş halidir. *YBTS* çeşitli internet dokümanlarını sınıflandırmak için kullanılmış ve bir başka metin sınıflandırma paketi ile karşılaştırılmıştır.

Hamaker ve Boggess (2004) dört adet iyi bilinen sınıflandırma problemi için *Öklid Uzaklığı* dışındaki uzaklık ölçütlerini kullanarak *YBTS'nin* performansını

araştırmışlardır. Meng ve ark. (2005) *YBTS'nin* çeşitli değerlendirme verileri üzerindeki sonuçlarını sunmuşlardır.

Watkins (2005) doktora tezinde bağışıklık sisteminin öğrenme yeteneklerini detaylı olarak araştırmıştır. Ayrıca *YBTS* ve *klonsal seçim algoritmasının* paralel versiyonları geliştirilmiştir.

Stibor ve ark. (2005) yapay veri kümeleri üzerinde değişik boyutlu algılayıcılarla *negatif seçim algoritmasının* davranışını araştırmışlardır. Negatif seçim, pozitif seçim ve istatistiksel anormallik tespiti tekniklerinin yüksek boyutlu *KDD (Knowledge Discovery and Data Mining)* veri kümesi üzerindeki sınıflandırma performansları analiz edilmiştir. Değişken boyutlu algılayıcılarla negatif seçim *KDD* veri kümesinde istatistiksel anormallik tespiti teknikleri ile rekabet edememektedir. Ayrıca değişken boyutlu algılayıcıları olan negatif seçim birçok parametre için çok hassas bulunmuştur.

Xu ve ark. (2007) *YBTS'yi* enerji kesilmesine sebep olan üç faktör için (ağaç, hayvan ve yıldırım) Duke Enerji kesilmesi verisine uygulamışlardır. *YBTS'nin* performansını *YSA* ile karşılaştırmışlardır.

Igawa ve Ohashi (2007) sınıflandırıcı olarak *DAIS (Discrimination Based Artificial Immune System)* adını verdikleri bir sistem geliştirmişlerdir. Geliştirdikleri sistem bağışıklık sistemindeki *T* hücrelerinin self/nonself ayırımı prensibini temel almaktadır. *DAIS* çeşitli veri kümeleri üzerinde uygulanmış ve diğer sınıflandırıcılara göre makul bir performans göstermiştir.

Sathyanath ve Sahin (2001) görüntü sınıflandırma görevini uygulamak için genetik algoritma ile birlikte bağışıklık sisteminin self/nonself ayırımı ve negatif seçim gibi özelliklerini kullanarak bir sistem geliştirmişlerdir. Sınıflandırmada kullanılan görüntüler mutfak dolabı üretiminde kullanılan ahşap parçalarıdır. Görüntüleri Virginia Polytechnic Enstitüsünden elde etmişlerdir.

Şahan (2004) yüksek lisans tez çalışmasında *YBS* alanında yapılan çalışmalara katkıda bulunmak ve Yapay Zeka ile ilgili gelişmeleri hızlandırma amacıyla üç farklı *YBS* algoritması geliştirmiştir. Geliştirilen algoritmalarından ilki olan *YBS_VQ* algoritması *YBS'deki* süreçler ile vektör kuantizasyonu algoritmasındaki benzerlikler örnek alınarak oluşturulmuştur. Bağışıklık tabanlı bir vektör kuantizasyonu sistemi geliştirilmiştir. *YBS_VQ* algoritmasının parametreleri

doğrusal olarak ayrılabilir beş sınıflı veri kümesi ile analiz edilmiş ve yine aynı veri kümesinde sistemin simülasyonu gerçekleştirilmiştir. Uygun parametre değerleri ile istenen hata seviyesinde kuantizasyon işlemi gerçekleştirilmiştir. Geliştirilen bir diğer algoritma *AWAIS* (*Attribute Weighted Artificial Immune System*) algoritmasıdır. *YBS*'de kullanılan şekil uzay gösteriminden kaynaklanan bazı problemleri ortadan kaldırmaya yönelik olarak oluşturulmuştur. *AWAIS* veri üzerinde önem taşıyan özelliklere diğerlerine göre daha fazla ağırlık verilmesiyle oluşturulmuş danışmanlı bir *YBS* algoritmasıdır. Algoritmanın performans analizi *Two Spirals* ve *Chainlink* yapay veri kümeleri üzerinde yapılmış ve benzer modellemenin yapıldığı *aiNet* sistemi ile karşılaştırılmıştır. *AWAIS*, *aiNet*'ten daha az sistem birimi ile aynı sınıflandırma doğruluğuna ulaşmıştır. Son olarak geliştirilen *SAMA* (*Supervised Affinity Maturation*) algoritması, Bağışıklık Sitemindeki birimlerin farklı bir modellemesinin öne sürüldüğü ve oluşturulmuş duyarlılık olgunlaşması modellerinden daha farklı bir eğitime algoritmasının kullanıldığı bir *YBS* algoritmasıdır.

Şahan ve ark. (2004) *AWAIS* algoritmasının etkinliğini *Two Spirals* ve *Chainlink* yapay veri kümeleri ile gerçek zamanlı bir veri kümesi olan *Wine* veri kümesi üzerinde test etmişlerdir. Aynı algoritma UCI (University of California at Irvine) makine öğrenmesi veri ambarından alınan tıbbi veri kümeleri olan *Statlog Heart Disease* ve *Pima Indians Diabet* veri kümeleri üzerinde uygulanmıştır (Şahan ve ark. 2005).

Polat ve ark. (2005a) *YBTS*'ye ön işleme uygulayarak geliştirdikleri *FS-AIRS* (*Feature Selection-Artificial Immune Recognition System*) algoritmasının etkinliğini UCI makine öğrenmesi veri ambarından alınan *Breast Cancer* veri kümesi üzerinde değerlendirmişlerdir. Ayrıca *YBTS*'deki kaynak tahsisi mekanizmasının bulanık mantık kullanılarak geliştirildiği bir diğer sistem ile çeşitli veri kümeleri üzerinde çalışmalar gerçekleştirilmiştir (Polat ve ark. 2005b; Polat ve ark. 2007). Şahan ve ark. (2007) hem bulanık mantık hem de *k en yakın komşu* yöntemi kullanarak bir *YBS* geliştirmiş ve UCI makine öğrenmesi veri ambarından alınan *Breast Cancer* veri kümesi üzerinde performansı değerlendirmişlerdir.

YBS alanında çalışan araştırmacılardan bir kısmı yukarıda özetlenenler gibi etkin sonuçlar elde edecek yeni algoritmalar oluşturmaya yönelirken, bir kısmı da

YBS'den uygulama bağlamında yararlanmaya çalışmışlardır. Uygulamaya yönelik oluşturulan sistemlerde oluşturulan sistemin belirli bir sabit yapısı yoktur. Her probleme özgü ayrı bir gösterim tarzı, ayrı bir öğrenme stratejisi ve ayrı bir hafıza mekanizması modellenmiştir. *YBS'nin* uygulama alanları ve bu alanlarda oluşturulmuş sistemler De Castro ve Von Zuben (1999) tarafından yazılmış teknik raporda kısaca özetlenmiştir.

Tez çalışmasında, yapay bağışıklık sistemler ve ona bağlı çalışmalar üzerinde daha önceden yapılan bu çalışmalardan bazı özellikler kullanılmaktadır. Özellikle bağışıklık hücresi olarak antikorlar ve kaynak sınırlama fikri kullanılmaktadır.

1.2. Çalışmanın Temel Amacı ve Literatüre Katkıları

Bu çalışmanın temel amacı, *YBS* sınıflandırıcısı olan *YBTS*'yi temel alarak danışmanlı bir yapay bağışıklık sınıflandırma sistemi geliştirmek, geliştirilen sistemin performansını makine öğrenmesi için standart veri kümeleri ve gerçek dünya veri kümeleri üzerinde değerlendirmek ve elde edilen sonuçlarla geliştirilen sınıflandırıcının kullanılabilirliğini ve diğer *YBS* algoritmaları için yararlı mekanizmalarını göstermektir.

Şekil-Uzay gösteriminin kullanıldığı çoğu *YBS* algoritmaları uzaklık kriterini kullanmaktadır. Bu algoritmalarda verinin bütün özellikleri uzaklık hesabında aynı etkiye sahiptir. Veri kümesi üzerinde etkisi çok az olan özelliklerle veri kümesi üzerinde çok önemli özellikleri eşit ağırlıkta saymaktadır. Özellikle verilerin boyutunun büyük olduğunda bu durum sistemin sınıflandırma performansını olumsuz yönde etkilemektedir. Bu durum literatürde “*curse of dimensionality (boyutluluk problemi)*” olarak adlandırılmaktadır (Bellman 1961; Mitchell 1997; Beyer ve ark. 1999; Şahan ve ark. 2004; Şahan ve ark. 2005). Bu problemin üstesinden gelebilmek için veri kümesinde daha önemli olan özelliklere uzaklık hesabında daha yüksek ağırlık verilmektedir. Bu amaçla geliştirilmiş olan *BK-YBTS* bu problemin etkisini minimuma indirmek için özelliklerin bilgi kazancını temel almaktadır. Özelliklerin önem derecesi bilgi kazancı hesabıyla belirlenmektedir. Bu açıdan bakılırsa önerilen bilgi kazancı tabanlı özellik ağırlıklandırma yaklaşımı şu

temellere bağlıdır: eğer bir özelliğin bilgi kazancı diğer özelliklerin bilgi kazancından yüksekse bu özelliğe diğer özelliklerden daha çok ağırlık verilmelidir. Böylece bilgi kazancı tabanlı ağırlıklandırma kullanılarak sınıflandırma performansını arttırmak amaçlanmaktadır.

Performans artımına katkıda bulunmak amacıyla sistemdeki antikorlara mutasyon işlemi sırasında rasgele bir değer ataması yapan bir mutasyon mekanizması yerine duyarlılığı temel alan bir mutasyon mekanizmasını kullanılmaktadır. Yani duyarlılığı yüksek olan antikorlar düşük oranda, duyarlılığı düşük olan antikorlar yüksek oranda mutasyona uğramaktadırlar.

YBTS'de kaynaklar, antijenle aynı sınıftaki antikorlar için en fazla kullanıcı tarafından belirlenmiş olan toplam maksimum kaynağın yarısı olacak şekilde tahsis edilmektedir ve kalan yarısı, diğer sınıflara ait antikorlar arasında eşit olarak paylaşılmaktadır. *BK-YBTS'de* bu mekanizma, veri kümesindeki sınıfların oranı hesaba katılarak değiştirilmiştir. Bu metotla birlikte sistem kaynakları, eğitim kümesinde sınıf dağılımı ile orantılı olarak farklı sınıf antikorları arasında bölünmektedir.

BK-YBTS'de bellek hücresi seçimine karar veren ölçüt, eğitim hücrelerinin tamamı arasındaki ortalama duyarlılık değeri yerine aynı sınıfa sahip eğitim hücrelerinin ortalama duyarlılık değerini temel alacak şekilde değiştirilmiştir. Eğitim verilerinin tamamını kullanan geniş bir metottan ziyade aynı sınıfa sahip eğitim örneklerinin ortalama duyarlılığı temel alınarak duyarlılık eşiğine karar verilmiştir.

Test verilerinin hangi sınıfta olduğu bellek hücreleri kümesi üzerine *k en yakın komşu* yönteminin uygulanması ile belirlenmektedir. Bu aşamada çoğunluk oylaması yapılırken eşitlik olması durumunda test verisinin sınıfı olarak daha çok verisi bulunan sınıf atanmaktadır. Böylece sınıflandırma performansını arttırmanın yanı sıra diğer *YBS* algoritmalarında da bu değişikliklerin adapte edilmesi amaçlanmaktadır.

Bu çalışmada sadece bağışıklık sistemi prensiplerini temel alan bir sınıflandırma sistemini göstermek amaçlanmamış aynı zamanda geliştirilmiş olan sistemin doğru sınıflandırmayı sağladığını göstermek amaçlanmıştır. Bu amaçla standart makine öğrenmesi veri kümelerinden olan *Iris* ve *Thyroid* veri kümeleri, Erciyes Üniversitesi Tıp Fakültesinden alınan *Damar tıkanıklığı (Atherosclerosis)*

veri kümesi ve Erciyes Üniversitesi Eczacılık Fakültesinden alınan *Bakteri* veri kümesi üzerinde *BK-YBTS'nin* performansı, *YBTS* ve bu veri kümelerini sınıflandıran literatürdeki diğer sınıflandırıcılarla karşılaştırılarak değerlendirilmiştir. *YBTS'den* farklı olarak *BK-YBTS* ile beraber önerilen değişikliklerin etkisini araştırmak amacıyla *çekirdek hücre sayısı*, *kaynak sayısı*, *uyarma eşiği*, *mutasyon oranı*, *duyarlılık eşik ölçüsü* ve *k* parametreleri için *YBTS* ve *BK-YBTS* her bir veri kümesi için çalışma süreleri, iterasyon sayıları, oluşan bellek hücreleri sayısı ve sınıflandırma doğrulukları temel alınarak karşılaştırılmıştır.

Bağışıklık sistemi, yapay bağışıklık sisteminin gelişmesi için bir esin kaynağı olarak bilgi işleme yetenekleri bakımından oldukça zengindir. Bağışıklık sisteminde gerçekleşen olayları kavramaya yönelik oluşturulan bağışıklık sistemi modellerinden yola çıkılarak oluşturulan bu sistem, *YBS* alanında literatüre katkıda bulunmak amacıyla hazırlanmıştır.

1.3. Tezin Organizasyonu

Danışmanlı bir yapay bağışıklık sınıflandırıcısına odaklanmış olan bu tez çalışmasının bölümleri aşağıdaki gibi organize edilmiştir:

Birinci bölümde, tez çalışmasının konusu hakkında genel bir özet verilmektedir. Çalışmanın amacına, literatürdeki mevcut çalışmalara ve çalışmanın literatüre katacaklarına değinilmektedir.

İkinci bölümde, biyolojik olarak bağışıklık sisteminin genel bir özeti verilmektedir. Bağışıklık sisteminin temel bileşenleri, fizyolojisi ve mekanizmaları hakkında bilgi verilmektedir. Uygulamada kullanılan doğal ve adaptif bağışıklık sistemi açıklanmaktadır.

Üçüncü bölümde, bir önceki bölümde detaylı bir şekilde anlatılmış olan biyolojik bağışıklık sisteminin mekanizmaları kullanılarak meydana çıkmış olan yapay bağışıklık sistemlerinden bahsedilmektedir. Alt bölümlerde bağışıklık sisteminden modellenmiş yapılar ve mekanizmalardan bahsedilmektedir.

Dördüncü bölümde tezde kullanılan özellik ağırlıklandırma kavramının esin kaynağı olan bilgi kazancını kullanan karar ağacı sınıflandırma metodundan

bahsedilmektedir. Örnek bir veri kümesi üzerinde ağaç yapısı oluşturulmaktadır. Ayrıca sürekli değer içeren veri özelliklerinin ayrık hale dönüştürülmesinde kullanılan tekniklerden bahsedilmekte ve bu teknikler basit örneklerle açıklanmaya çalışılmaktadır.

Beşinci bölümde, çalışmada önerilen değişiklikleri uygulamak için temel alınan, Watkins tarafından geliştirilmiş olan *Yapay Bağışıklık Tanıma Sistemi (YBTS)* ve tez çalışmasında geliştirilmiş olan *Bilgi Kazancı Tabanlı Yapay Bağışıklık Tanıma Sistemi (BK-YBTS)* ayrıntılı bir şekilde açıklanmaktadır. *BK-YBTS'nin* eğitim sürecinde olan başlıca beş aşama belirtilmektedir. Bu aşamalardan ilki uzaklık hesabında kullanılmak üzere veri özelliklerinin bilgi kazançları değerlerinin hesaplanmasıdır. İkinci aşama veri normalizasyonu aşamasıdır. Üçüncü aşama antikoların oluşumu ve bellek hücresi belirlemedir. Dördüncü aşama aday bellek hücresinin gelişmesinde kaynaklar için olan rekabettir. Sistemin son aşaması, mevcut bellek hücrelerinin içerisine aday bellek hücresinin sunumudur. *BK-YBTS'nin* meydana getirilmesinde kullanılan tüm bu rutinler, metotlar ve eşitlikler bu bölümde detaylı bir şekilde sunulmaktadır.

Altıncı bölümde, öncelikle uygulamada kullanılan parametreler hakkında bilgi verilmektedir. Daha sonra makine öğrenmesi literatüründe standart olan veri kümeleri ve gerçek dünya veri kümelerinin her biri hakkında bilgi verildikten sonra bu veri kümeleri üzerinde *YBTS* ve *BK-YBTS'nin*; çekirdek hücre sayısı, kaynak sayısı, uyarma eşiği, mutasyon oranı, duyarlılık eşik ölçüsü ve k parametrelerinin değişik değerleri için her iki sistem, çalışma süreleri, iterasyon sayıları, oluşan bellek hücreleri sayısı ve sınıflandırma doğrulukları temel alınarak karşılaştırılmaktadır. *BK-YBTS'ye* özgü parametre olan *parça değerinin* aldığı değişik değerlerde *BK-YBTS'nin* çalışma süreleri, iterasyon sayıları, oluşan bellek hücreleri sayıları ve sınıflandırma doğrulukları araştırılmaktadır. Son olarak *BK-YBTS* ile diğer sınıflandırıcıların elde ettikleri sınıflandırma sonuçları sunulmaktadır.

Yedinci bölümde ise tez çalışması ile ilgili genel sonuçlara yer verilmektedir. Kaynaklar bölümünde ise bu tez çalışmasında faydalanan kaynaklara yer verilmektedir.

2. BAĞIŞIKLIK SİSTEMİ

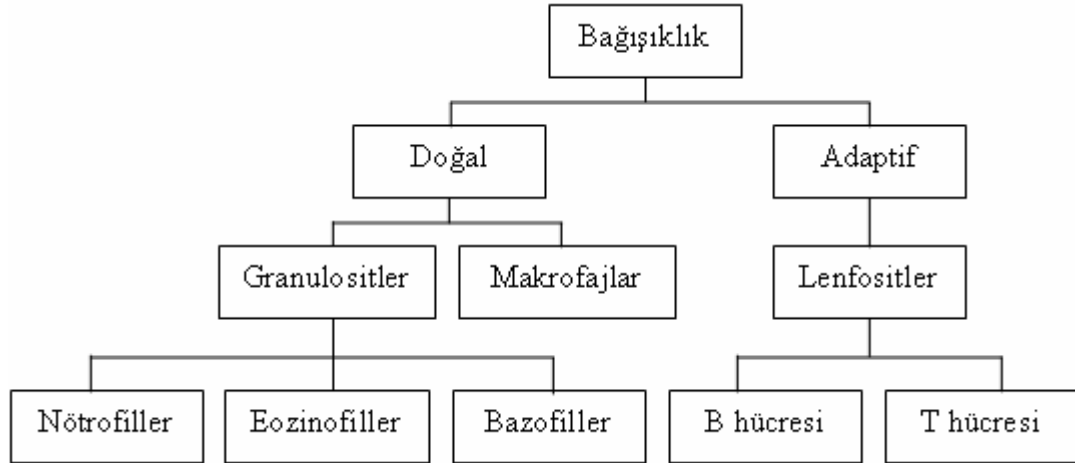
Bağışıklık sistemi, hücrelerimizden ve dışarıdan gelen bulaşıcı mikroorganizmaların hareketlerini algılama ve savaşma yeteneği olan hücreler, moleküller ve organların karışımından oluşan bir sistemdir. Bağışıklık sistemi ve diğer sistemler arasındaki etkileşim vücudun düzenlenmesine izin vermektedir (Jerne 1973; Janeway 1992). Bağışıklık sisteminde görev alan organlar hemen hemen sınırsız çeşitlilikteki bulaşıcı yabancı hücreleri ve maddeleri (*nonself*) tanıyabilmekte ve bulaşıcı olmayan doğal hücreleri (*self*) ayırt edebilmektedir. Bir patojen vücuda girdiği zaman yok edilmek için taranır ve yok edilmeye çalışılır. Bağışıklık sistemi, bu patojenlerin vücuda zarar vermesini engellemek ve bir an önce vücudu terk etmelerini sağlamak için gerekli işlemleri başlatmaktadır. Bu işlemlerin hepsine bağışıklık tepkisi adı verilmektedir. Sistemin her enfeksiyonu hatırlama yeteneği vardır. Aynı patojene ikinci kez bulaşınca daha etkili bir şekilde tepki verilmektedir.

2.1. Temel Bileşenler

Bağışıklık sistemi enfeksiyonlara karşı doğal, hızlı ve etkili mekanizmalar içermektedir. Bağışıklık sisteminde iki katlı bir savunma sistemi söz konusudur. Bunlar doğal ve adaptif bağışıklık sistemidir. Her iki sistem beyaz kan hücrelerinin (akyuvar) aktivitesine bağlıdır. Bu aktiviteler doğal bağışıklıkta granulositler ve makrofajlar yoluyla, adaptif bağışıklıkta ise lenfositler yoluyla olmaktadır. Şekil 2.1'de bağışıklık sistemleri bileşenleri gösterilmektedir (De Castro ve Timmis 2002a).

Makrofajlar ve nötrofiller, çeşitli mikroorganizmaları ve antijenik partikülleri yutma ve sindirme yeteneğine sahiptirler. Makrofajlar aynı zamanda Antijen Sunan Hücreler (Antigen Presenting Cell - APC) olarak ifade edilen hücrelere antijen sunma yeteneğine sahiptir. Granulositler enzimlerle doldurulmuş granüller içeren

çekirdekli beyaz kan hücrelerinin özel bir grubunu oluşturmaktadır. Nötrofiller en bol bulunan hücresel elemanlardır ve doğal bağışıklıkta en önemli görevleri görmektedir. Eozinofiller parazitler tarafından enfeksiyonlara karşı rol oynamaktadır. Bununla beraber bazofilin temel görevi henüz anlaşılamamıştır (Janeway ve ark. 1999).



Şekil 2.1 Bağışıklık sistemi bileşenleri (De Castro ve Timmis 2002a)

Adaptif bağışıklık sisteminde, antikorlar bulaşıcılara tepki olarak üretilmektedir. Bir bireydeki antikorların varlığı o bireyin patojenlere maruz kaldığını göstermektedir. Adaptif sistemin hücreleri bağışıklık belleği geliştirme yeteneğine sahiptir. Aynı organizma vücuda tekrar saldırdığı zaman aynı antijenik uyarıyı tanıma yeteneğindedir. Bu yetenek organizma içerisinde hastalığın tekrar bulunmasını engellemektedir. Böylece adaptif bağışıklık tepkisi, bağışıklık sisteminin belirli bir antijenle her karşılaşmada tepkinin gelişmesine imkân tanımaktadır.

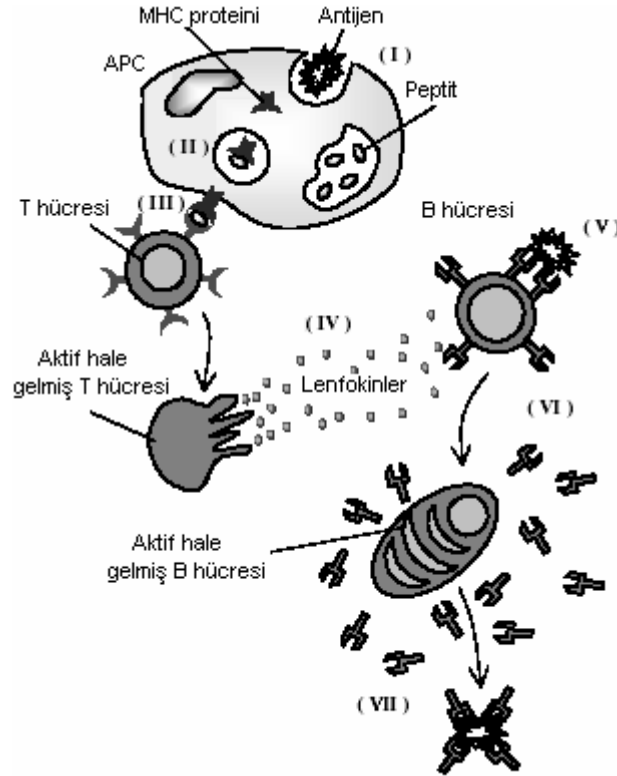
Lenfositler adaptif bağışıklık tepkisinde görev almaktadır ve patojenik ajanların tanınmasından ve yok edilmesinden sorumludur. Lenfositler bir antijenik uyarı ile etkileşime girdiği zaman aktif hale gelir ve çoğalırlar. Lenfositlerin iki temel tipi söz konusudur. Bunlar *B* lenfosit (*B* hücresi) ve *T* lenfositidir (*T* hücresi).

Adaptif bağışıklık tepkisi aynı bulaşıcı ajanın tekrar bulaşmasına karşı bağışıklık sağlarken, doğal bağışıklık tepkisi antijene maruz kalınmasa da bireyin yaşamı boyunca sabit kalmaktadır. Doğal ve adaptif bağışıklık sistemleri oldukça

etkili savunma mekanizması sağlamaktadır. Bu durumda potansiyel olarak patojenik bulaşıcılarla hayatımızı sürdürmemize rağmen nadiren hastalanmamızı garanti etmektedir.

Bağışıklık sistemi, enfeksiyona karşı vücudumuzu koruyan hücreler ve moleküllerin kompleks bir kümesinden meydana gelmektedir. Vücudumuz adaptif bağışıklık sistemimizi uyaran antijenler (*Ag*) tarafından devamlı saldırı altındadır. Patojenlerin üzerinde bulunan yüzey molekülleri olarak yabancı antijenler bulunduğu gibi vücudumuzun hücrelerinde veya moleküllerinde bulunan kendi antijenlerimizde (self-antigen) bulunmaktadır.

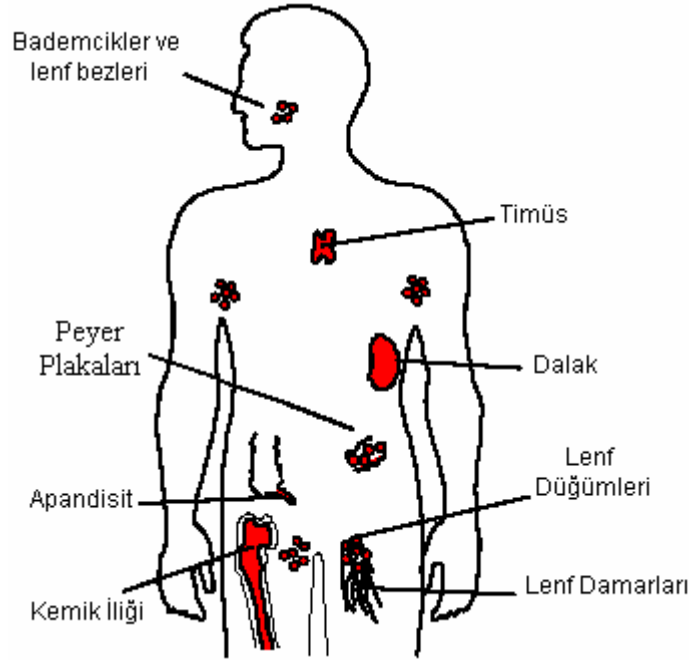
Şekil 2.2 temel bağışıklık tanıma ve aktivasyon mekanizmalarının basitleştirilmiş işleyişini göstermektedir. Şekil 2.2’de 1. aşamada vücudumuzda dolaşarak antijenleri yutan ve sindiren Antijen Sunan Hücreler (Antigen Presenting Cell - APC) olan makrofajların nasıl görev yaptığı gösterilmektedir. Bu hücreler antijeni antijenik peptitlere parçalamaktadır (Nossal 1993). Bu peptitlerin parçaları Major Histocompatibility Complex (MHC) moleküllerine bağlanmaktadır. Bağlanan MHC’ler MHC/peptit kompleksi olarak APC hücre yüzeyinde bulunur (2. aşama). *T* hücreleri farklı MHC/peptit komplekslerini taşıyabilen yüzey alıcıları taşımaktadır (3. aşama). MHC/peptit bir kez tanınarak aktif olunca, *T* hücreleri de aktif hale gelir, bölünür ve lenfokinler veya kimyasal sinyaller salgırlar. Bu sinyaller bağışıklık sisteminin diğer bileşenlerini harekete geçirmek için uyarırlar (4. aşama). *B* hücreleri MHC moleküllerinin yardımı olmaksızın antijenleri tanıma yeteneğine sahip alıcıları vardır (5. aşama). Bu *B* hücrelerinin yüzey alıcıları spesifik bir antijene tepki vermektedir. Bu *B* hücre alıcıları tarafından bir sinyal alındığı zaman, *B* hücresi aktif hale gelir ve plazma hücreleri olarak çoğalır ve farklılaşır. Plazma hücreleri yüksek miktarda antikor molekülleri salgılamaktadır (6. aşama). Bu ortaya çıkan antikorlar (*B* hücre alıcılarının değişik formları) patojeni etkisiz hale getirmede kullanılır (7. aşama). Bu aktif hale gelmiş *B* ve *T* hücrelerinin bazıları bellek hücrelerine ayrılır. Bu hücreler organizmanın hayatı boyunca vücutta dolaşmakta ve böylece aynı antijenin gelecekteki saldırısına karşı vücuda koruma sağlamaktadır.



Şekil 2.2 Bağışıklık tanıma ve aktivasyon mekanizmaların basitleştirilmiş gösterimi (De Castro ve Timmis 2002a)

2.1.1. Bağışıklık sisteminin fizyolojisi

Bağışıklık sistemini meydana getiren vücutta pek çok organ söz konusudur. Bu organlar lenfoid (lymphoid) organlar olarak adlandırılmaktadır. Bu organlar çoğunlukla *B* ve *T* hücreleri olan lenfositlerin üretiminden ve gelişmesinden sorumludur. Lenfoid organlarda lenfositler ya üretildikleri zaman ya da adaptif bağışıklık tepkisi başlangıcında çeşitli hücrelerle etkileşime girmektedirler. Lenfoid organlar birincil ve ikincil olmak üzere ikiye ayrılmaktadır. Birincil organlar lenfositlerin üretiminden ve olgunlaşmasından sorumludur. İkincil organlar lenfositleri antijenik uyarı ile etkileşime sokmaktadırlar. Böylece adaptif bağışıklık tepkisi başlamaktadır. Şekil 2.3'te vücuttaki lenfoid organlar gösterilmektedir (De Castro ve Timmis 2002a).



Şekil 2.3 Bağışıklık sisteminin anatomisi (De Castro ve Timmis 2002a)

Birincil Lenfoid organlar:

- **Kemik iliği (Bone marrow):** En uzun kemikler içerisindeki yumuşak dokudur. Burası kırmızı kan hücreleri (alyuvarlar), monositler, granulositler, lenfositler ve trombositlerin üretildiği en önemli yerdir. Aynı zamanda burada *B* hücrelerinin farklılaşması için çevre sağlanmaktadır. *B* hücresi adını bu organdan almaktadır.
- **Timüs (Thymus):** Göğüs kemiğinin arkasında bulunmaktadır. *T* hücrelerinin farklılaşması için çevre sağlamaktadır ve *T* hücresi olgunlaşması için önemli hormonları üretmektedir. Timüs, endokrin sistemle de etkileşim içerisinde. *T* hücresi adını bu organdan almaktadır.

İkincil Lenfoid organlar:

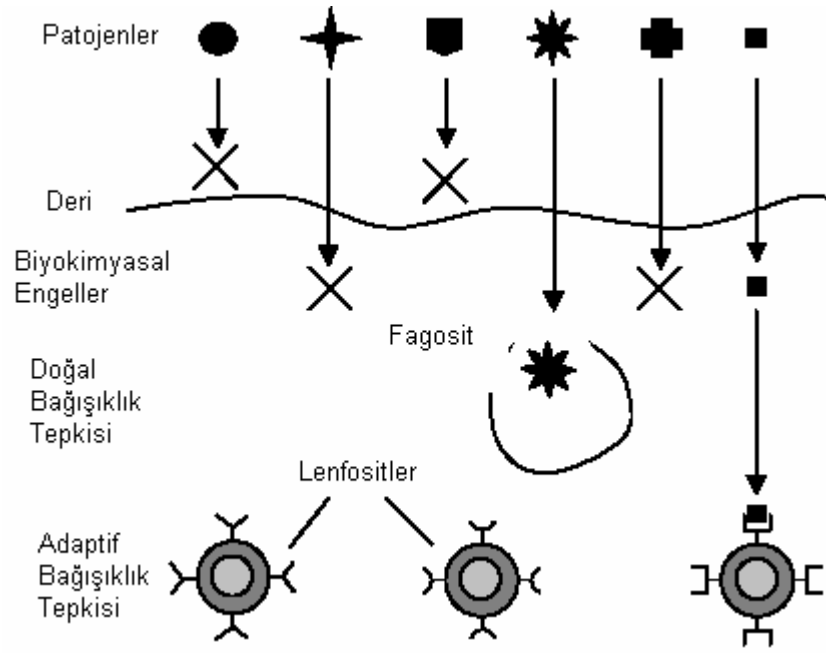
- **Bademcik (tonsil) ve lenf bezi (adenoid):** Bademcikler ağız ve yutak arasında yanlara doğru bulunmaktadır. Lenf bezleri boğazın üstünde burnun arkasında bulunmaktadır. Lenfosit hücrelerinin çok sayıda bulunduğu bu organlar solunum sisteminin korunmasından sorumludur.
- **Lenf düğümleri (lymph nodes):** Lenfatik sistem boyunca çeşitli noktalarda küçük yapı parçaları olarak bulunmaktadır. Bu düğümler kanı geriye taşırlar

ve dokulardan hücrel akışı toplarlar. Bu hücrel akış sürekli olarak kan filtre edilerek üretilir ve ak kan olarak adlandırılır. Lenf düğümleri aynı zamanda adaptif bağışıklık tepkisinin meydana geldiği yerlerdir.

- **Apandis (Appendix) ve Peyer plakaları:** Apandis kalın bağırsağın sonunda ve ince bağırsağın duvarında bulunmaktadır. Peyer plakaları ise ince bağırsakları çevreleyen lenf dokusu plakalarıdır. Bu organlar sindirim sistemini korumak için özelleşmiş bağışıklık hücreleri içeren lenf düğümleridir.
- **Dalak (Spleen):** Diyaframın altında vücudun solunda bulunan, geniş, kapsüllü ve fasulye şeklinde bir organdır. Dalak kan akışını sağlayan lenfoid organdır. Bu organ kan akışını engelleyen mikro organizmalara karşı savaşan lenfositleri bulundurmaktadır. Dalak eski kan hücrelerinin yok edilmesinden sorumlu olan kırmızı bir parça ve antijenlere tepki veren lenfoid hücrelerini kan yoluyla taşıyan beyaz bir parça içermektedir.
- **Lenf damarları (lymphatic vessels):** Bu damarlar kana ve lenfoid organlara akkan (lenf) taşımaktadır. Ak kan ise lenf damarlarıyla dokulardan lenf düğümlerine taşınmaktadır.

Bağışıklık sistemi çeşitli seviyelere yayılmış savunma mekanizmaları ile katmanlı bir mimariye sahiptir. Şekil 2.4'te bu katmanlar gösterilmektedir. Koruma katmanları şu şekilde bölünebilir (Janeway ve ark. 1999; Rensberger 1996; Hofmeyr 2000):

- **Fiziksel engeller:** Derimiz saldırganlara karşı bir kalkan hizmeti vermektedir. Solunum sistemi de aynı zamanda potansiyel enfeksiyonlardan korunmada yardımcı olmaktadır. Bu koruma burun kılları, mukus, kirpikler, öksürme, hapşıрма yoluyla gerçekleşmektedir.
- **Biyokimyasal engeller:** Tükürük, ter ve gözyaşı yok edici enzimler içermektedir. Mide asitleri sindirilen yiyecek ve içeceklerdeki pek çok zararlı mikro organizmayı öldürmektedir. Vücudun pH seviyesi ve sıcaklığı bazı saldırganlar için istenmeyen yaşama şartlarına yol açmaktadır.
- **Doğal ve adaptif bağışıklık sistemi:** Bağışıklık sisteminin iki temel yapısıdır.



Şekil 2.4 Bağışıklık sistemi koruma katmanları

2.1.2. Doğal bağışıklık sistemi

Doğal bağışıklık sistemi çeşitli mikro organizmalara karşı bağışıklığın ilk tepkisidir. Doğal bağışıklık sistemi, yok edici hücrelerden meydana gelmektedir. Bu hücrelere makrofajlar ve nötrofiller örnek olarak verilebilir. Doğal bağışıklık sisteminin hücreleri adaptif bağışıklık tepkisini başlatma ve kontrol etme sürecinde kritik bir rol oynamaktadır. Adaptif bağışıklık sistemi enfeksiyona cevap vermeden önce belirli bir zaman geçebilmektedir. Bu arada doğal bağışıklık sistemi enfeksiyonu kontrolden sorumludur.

Makrofajlar ve nötrofiller yüzey alıcılarına sahiptir. Bu yüzey alıcılar sadece mikro organizmada bulunan moleküler parçayı tanımak ve bağlanmak için gelişmiştir. Bakteriyel moleküllerine bağlanan bu alıcılar bakteriyi yutması için hücreleri uyarırlar ve aynı zamanda sitokin gibi aktif moleküllerin salgılanmasını tetiklerler. Sitokinler lenfosit hücreleri tarafından salgılanan proteinlerdir. Yüzey alıcıları, Örüntü Tanıma Alıcıları (Pattern Recognition Receptors- PRR) olarak bilinmektedir. PRR, Patojen Bulunduran Moleküler Örüntüler (Pathogen Associated Molecular Patterns - PAMP) olarak adlandırılan patojenleri tanıma yeteneğine

sahiptir. PAMP'lar sadece patojenler tarafından üretilir ve asla ev sahibi organizmada üretilmezler. Bu yüzden PRR tarafından tanınma gerçekleşmesi bir patojenin varlığına işaret etmektedir.

Doğal bağışıklık sistemi tarafından tanınan yapılar vücudun hücreleri ve dokularına zarar vermekten kaçınmak için self antijenlerden kesinlikle ayrı olmak zorundadır. Bu mekanizmanın sonucu olarak doğal bağışıklık sistemi self-nonsel self ayırımı sürecinde rol oynamaktadır. Bir PAMP tanındığı zaman doğal bağışıklık sisteminin yutma (fagositoz) mekanizmasını aktive eder.

2.1.3. Adaptif bağışıklık sistemi

Bütün hayvanlar patojenlere karşı direnme yeteneğine sahiptir. Bu direncin doğası organizmanın tipine göre farklıdır. Geleneksel olarak bağışıklık bilimi araştırmalarının hemen hemen hepsi omurgalı savunma reaksiyonları ve özellikle de fare ve insanlar gibi memelilerin bağışıklık sistemi üzerinde yapılmaktadır. Omurgalıların savunma sistemi gelişmiştir. Omurgalı bağışıklık sistemi pek çok antijen türlerine karşı enfeksiyondan vücudu korumaktadır.

Lenfositler adaptif bağışıklık sisteminin en önemli hücreleridir. Lenfositler sadece omurgalılarda bulunmaktadır. Omurgalılar gelecek enfeksiyonlara karşı daha etkili ve çok yönlü savunma mekanizmasına sahip bir sistem geliştirmişlerdir.

Adaptif bağışıklık tepkisinin davranışı için ilk açıklamalardan biri 1959 yılında McFarlane Burnet tarafından “*Seçici Teori*” adlı çalışma ile yapılmıştır (Burnet 1959). Bu çalışmada her bireyin sadece maruz kaldığı antijene karşı antikor üretildiği açıklanmıştır. Yazar potansiyel olarak farklı antikorlar üretebilen pek çok hücrenin varlığını kabul etmiştir. Bu hücrelerin her biri farklı özgünlükte antikoru sentezleyebilme yeteneğine sahiptir. Antikor antijene bağlandıktan sonra, hücre geniş bir klon oluşturacak şekilde çoğaltmak için aktif hale gelir. Bir klon, bir hücre ya da hücrelerin kümesidir ve tek bir ata hücreden çoğaltılır. Bu prensip “*klonsal seçim teorisi*” olarak adlandırılmıştır ve adaptif bağışıklık tepkisinin özünü oluşturmaktadır. Bu teorinin daha ileri uzantıları bağışıklık öğrenmesi ve bellektir.

Klonsal seçim teorisi, lenfositlerin doğal seleksiyona benzer bir süreçten geçtiğini temel almaktadır. Antijenler lenfositlerle karşılaşınca, lenfositler çoğalmak ve farklılaşmak üzere aktif hale gelirler. Bu bakış açısı, yapay bağışıklık sisteminin oluşumunda önemli rol oynamaktadır.

2.2. Bağışıklık Mekanizmaları

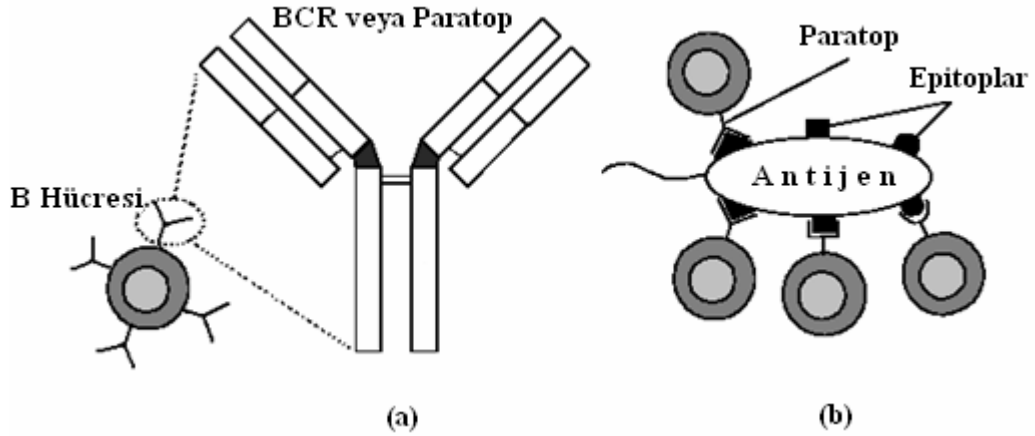
Bu bölümde Yapay Bağışıklık Sisteminde kullanılan Bağışıklık Sisteminin temel mekanizmalarından bahsedilmektedir. Bu mekanizmalar arasında antijen ve antikor moleküllerinin birbirlerini tanıma olayı, bu tanımaya tepki olarak *klonsal seçim prensibinin* gerçekleştirilmesi, antijene karşı verilen tepki sonucu bellek hücresinin oluşumu ve gelecekte karşılaşılabilecek benzer antijenlere karşı daha etkili tepki verecek şekilde duyarlılığın olgunlaşması sayılabilir. Alt bölümlerde bu mekanizmalardan bahsedilmektedir.

2.2.1. Örüntü tanıma

Bağışıklık sisteminin örüntü tanıma bakış açısından en önemli karakteristiği *B* ve *T* hücrelerinin antijenleri tanıyabilme yeteneğine sahip yüzey alıcı molekülleri taşımasıdır. *T* hücresi alıcısı (T cell receptor TCR) antijeni tanımakta ve *MHC* olarak adlandırılan yüzey molekülüne bağlanmaktadır (Şekil 2.2). *B* hücresi alıcısı (*B* cell receptor - BCR) antijen molekülleri ile etkileşime girmektedir.

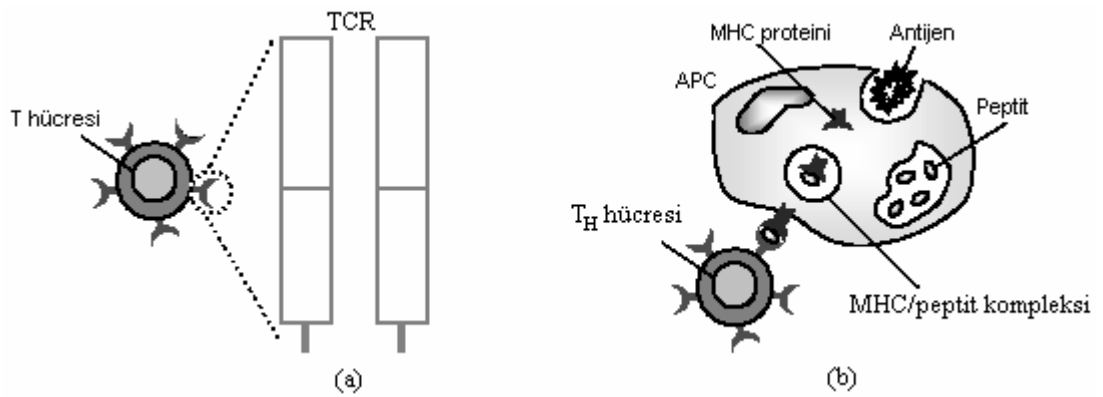
BCR aktif hale geldiği zaman antikorlar formunda salgılanır. *B* hücresinin temel rolü patojenik ajana tepkide antikorları üretmek ve salgılamaktır. Bu antikorlar proteini tanıma ve bağlanma yeteneğine sahiptir.

Şekil 2.5.a *B* hücresinin yüzeyindeki antikor molekülünün parçalarını göstermektedir. Bağışıklık tanınması moleküler seviyede olur. Tanıma olayı antikor alıcısının bağlandığı bölge (*paratop*) ile antijenin bir parçası olan *epitop* arasında tamamlayıcılığı temel almaktadır. Antikorlar tek tip alıcı sunarken, antijenler çok sayıda *epitop* sunabilir. Şekil 2.5.b de bu durum gösterilmektedir.



Şekil 2.5 (a) B hücresi alıcısı (b) Antikor-Antijen etkileşimi

B hücreleri, yüzeylerinde antikor molekülleri sunmaktadır. B hücreleri kemik iliğinde olgunlaşırken, T hücrelerinin olgunlaşma süreci timüste gerçekleşmektedir. T hücresi görevleri, vücutta enfeksiyona sebep olan hücelere direkt atak yapmak ve diğer hücreleri düzenlemektir. T hücreleri iki ana gruba ayrılabilir. Bunlar yardımcı T hücreleri (T_H) ve öldürücü T hücreleri (T_K). Şekil 2.6.a da gösterildiği gibi T hücresi antijenik alıcıları (T cell antigenic receptors TCR) B hücresi alıcıları ile karşılaştırıldığında yapısal olarak farklıdır. TCR, MHC tarafından sunulan antijenleri tanımaktadır. Şekil 2.6.b T hücresi alıcısını ve MHC/peptit kompleksini göstermektedir.



Şekil 2.6 (a) T hücresi alıcısı (b) Tanıma olayı

2.2.2. Klonsal seçim prensibi

Enfeksiyona karşı yeterince hücre üretmek için, aktif hale gelmiş lenfosit çoğaltılmalı ve farklılaşmalıdır. Klonsal genişleme olarak adlandırılan bu süreç lenf düğümlerinde meydana gelmektedir. Bütün adaptif bağışıklık sistemlerinde klonsal genişleme söz konusudur (Weissman ve Cooper 1993; Tarlinton 1998).

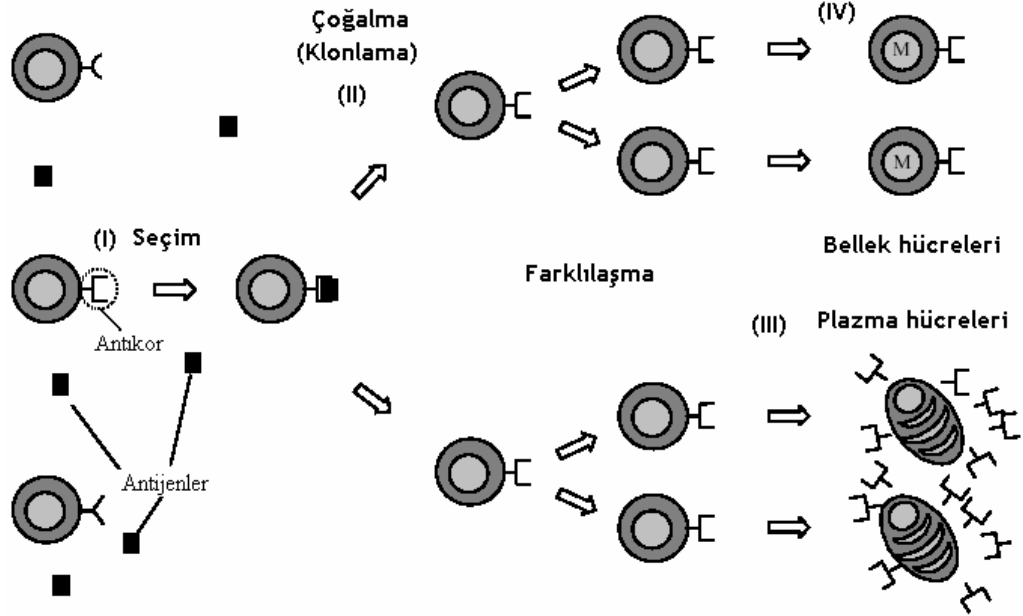
Klonsal seçim prensibi bir antijenik uyarıya adaptif bağışıklık tepkisinin temel özelliklerini açıklamada kullanılan bir prensiptir. Antijenik uyarıyı tanıyabilme yeteneğine sahip hücreler çoğaltılır ve farklılaştırılır. Çoğalma boyunca mutasyon ile farklılaşma söz konusudur. *B* hücresinin klonsal genişleme sürecinde mutasyon ve seçme olaylarının varlığı bu hücrelerin repertuar çeşitliliğini arttırmasına izin vermekte ve antijeni tanıma yeteneğini arttırmaktadır.

Vücut bir antijene maruz kaldığı zaman *B* hücreleri antikolar üreterek tepki göstermektedir. Her bir *B* hücresi antikorun sadece antijen için spesifik bir türünü salgılamaktadır. Şekil 2.7 *klonsal seçim prensibinin* basitleştirilmiş bir açıklamasını sunmaktadır. Bir *B* hücresi üzerindeki antijenik alıcılar bir antijene bağlanır (1) ve T_H hücreleri gibi yardımcı hücrelerden ikinci bir sinyal (uyarıcı sinyal) ile *B* hücresi antijene karşı uyarılmış olur. Bu uyarılma *B* hücresinin çoğalmasına (bölünme) sebep olur (2) ve klonlanan bu *B* hücreleri plazma hücreleri olarak adlandırılan antikor salgılayan hücrelerde olgunlaşır (3). Plazma hücreleri en aktif antikor salgılayıcı iken, hızla bölünen *B* hücreleri de az da olsa antikor salgırlar. Plazma hücrelerde çoğalan ve farklılaşan *B* hücreleri uzun ömürlü *B* bellek hücreleri olarak farklılaşabilir (4). Bellek hücreleri kanda, lenfte ve dokularda dolaşmaktadır. İkinci kez aynı antijenik uyarı ile karşılaşıldığında plazma hücrelerinde yüksek duyarlılıklı antikorlar mevcut olacaktır.

Özetle *klonsal seçim teorisinin* temel özellikleri şu şekildedir:

- **Negatif seçim:** Vücudun kendisine ait self antijen olarak adlandırılan antijenik örüntülere karşı lenfositlerin aktif olması engellenmektedir.
- **Klonsal genişleme:** Vücuttaki yabancı antijenlerle lenfositlerin temas kurarak lenfositlerin çoğalması ve farklılaşması.
- **Mutasyon:** Antikor örüntülerinin çeşitliliği olarak ifade edilen yeni rasgele genetik değişikliklerin somatik mutasyon şeklinde gerçekleşmesidir. Yani

duyarlılığı yüksek antikorların daha az oranda mutasyona uğraması düşük duyarlılıklı antikorların daha yüksek bir oranda mutasyona uğramasıdır.



Şekil 2.7 Klonal seçim prensibinin basitleştirilmiş açıklaması

2.2.3. Bağışıklık öğrenimi ve bellek

Bağışıklık sisteminin uzun zaman sürelerince koruyucu olması için antijen tanıma yeterli değildir. Bağışıklık sistemi daha sonra tekrar karşılaştığı patojene karşı etkili bir tepki verebilmesi için yeterli kaynaklara da sahip olmalıdır. Tipik av-avcı durumlarında olduğu gibi (Taylor 1984) patojen popülasyonunun boyutu ile o patojeni yok etmek için klonlanmış lenfosit popülasyonunun boyutu arasındaki oran enfeksiyonun üstesinden gelmede çok önemlidir. Bağışıklık sisteminde *öğrenme* antijen tanıma aşaması boyunca lenfositlerin popülasyon boyutunun ve duyarlılıklarının artışı içermektedir. Böylece bağışıklık repertuarı rasgele bir repertuardan ziyade antijenik ortamı daha açık bir şekilde yansıtmaya eğilimlidir.

Bağışıklık sisteminde lenfositlerin toplam sayısı düzenlenirken bazı klonların sayısının artması diğer klonların sayısının azalması anlamına gelmektedir. Bununla beraber lenfositlerin toplam sayısı tamamen sabit tutulmamaktadır. Eğer bağışıklık sistemi sadece spesifik lenfositlerin popülasyon boyutunu arttırmayı öğrenirse, ya

boyutu artmış olan önceden öğrendiği antijenleri unutacak ya da rasgele oluşturulan yeni antijenlere tepkiden sorumlu repertuarın oranı azalacaktır (Perelson ve Weisbuch 1997).

Bağışıklık sisteminin normal gelişiminde organizmanın ömrü boyunca defalarca belli bir antijenle karşılaşacağı beklenmektedir. Bir antijen ilk kez organizmaya uğradığı zaman adaptif bağışıklık tepkisi, her biri farklı duyarlılıklara sahip antikorlar üreten *B* hücrelerinin küçük bir miktarı ile yürütülür. İlk karşılaşmada yüksek duyarlılıkta antikor üreten hücreler saklanır. Saklanan bu hücreler *bellek hücreleri* olarak adlandırılır. Bu hücreler antijen ile bir sonraki karşılaşmada daha etkili bağışıklık tepkisi vermek için kullanılır. Böylece antijenle her karşılaşmada sıfırdan tepki vermektense her karşılaşmadan sonra daha hızlı ve güçlü tepki verilmektedir. Bu durum takviyeli öğrenme stratejisinin esasıdır. Bu stratejide sistem çevre ile her etkileşiminde sürekli öğrenme işlemi gerçekleştirmektedir. *B* hücreleri giderek antijen tanımada daha da özelleşmektedir.

Şekil 2.8’de bağışıklık tepkisi gösterilmektedir. Bir *A* antijeni ilk kez vücuda bulaşmış olsun. *A* antijeni için az sayıda antikor olacaktır ve belli bir gecikme fazından sonra *A* antijenine karşı antikorların belli bir seviyede yoğunlaşma ve duyarlılık artacaktır. Enfeksiyon yok edildiği zaman yoğunlaşma azalmaya başlayacaktır. Bu ilk aşama *birincil tepki* olarak adlandırılmaktadır. *A* antijeninden farklı olarak *B* antijeni vücuda bulaştığı zaman, aynı şekilde tepki gösterilir. Fakat tanınmış olan *A* antijeni için gösterilenden farklı antikorlar üretilir. Diğer taraftan bağışıklık belleğinin önemli bir karakteristiği birliktelik özelliğidir. *A* antijenine adapte olan antikorlar ikincil tepkide sadece *A* antijenine değil aynı zamanda yapısal olarak alakalı *A'* antijenine daha hızlı ve daha etkili tepki gösterilir.

Önceden görülmüş olan antijenle yapısal olarak ilgili antijene tepki, çapraz tepki olarak adlandırılmaktadır. İkincil tepki, birincil tepki ile karşılaştırıldığında daha kısa bir bekleme fazı, daha yüksek oranda antikor üretimi ve daha kalıcı antikorlar söz konusudur.

Özet olarak bağışıklık öğrenimi ve bellek aşağıdaki durumlardan çıkarılmaktadır:

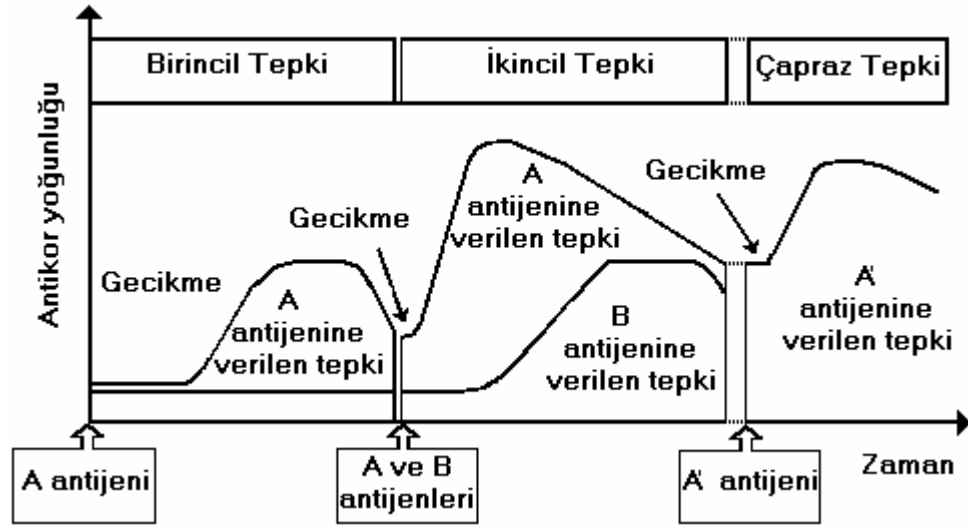
Öğrenme:

- Antijenik uyarıya defalarca maruz kalma;

- Spesifik *B* hücreleri alt popülasyonlarının (klonlar) boyutunda artma ve
- Antikoru duyarlılık olgunlaşması.

Bellek:

- Antijen ile bir sonraki karşılaşmaya kadar lenfositlerin dinlenme durumunda kalması;
- Çapraz tepki.



Şekil 2.8 Birincil, ikincil ve çapraz bağışıklık tepkisi

2.2.4. Duyarlılık olgunlaşması

T hücresine bağımlı bağışıklık tepkisinde, aktif hale gelmiş *B* hücrelerinin repertuarı mutasyon olayı ile değişmektedir. Bu değişikliklerden sonra sadece yüksek duyarlılıklı hücreler bellek hücresi olarak seçilmektedir (Tarlinton 1998).

İkincil tepkide ortaya çıkan antikolar ortalama olarak birincil tepkide ortaya çıkan antikorlardan daha yüksek bir duyarlılığa sahiptir. Bu durum bağışıklık tepkisinin olgunlaşması olarak adlandırılmaktadır. Bu olgunlaşma, antikor moleküllerinin antijene bağlandığı yerlerin yapısal olarak birincil tepkideki yerlerden farklı olmasını gerektirmektedir (Allen ve ark. 1987).

Klonal genişleme boyunca, rasgele değişikliklerin bazıları antikoru duyarlılığını arttırmaktadır. Bu yüksek duyarlılıklı antikolar bellek hücresi havuzuna

seçilmektedir. Mutasyon sürecinin doğası gereği mutasyona uğrayan bazı genlerle düşük duyarlılıklı hücrelerde ortaya çıkabilir. Bu hücreler daha sonra yok edilir. Şekil 2.8’de gösterildiği gibi birincil tepkiden ikincil tepkiye antikor duyarlılığının artışı, bağışıklık tepkisinin olgunlaşmasının sürekli bir süreç olduğunu göstermektedir.

3. YAPAY BAĞIŞIKLIK SİSTEMLERİ

1900'lü yılların ikinci yarısından itibaren gelişmeye başlayan yapay zeka teknikleri, insan gibi düşünen ve davranan makinelerin ortaya çıkarılması gibi amaçlara yönelik çalışmalarda büyük ölçüde insanın biyolojik sisteminin işleyişini referans almıştır. Bağışıklık sistemindeki olayların aydınlatılmaya başlanması ve Yapay Zeka'ya duyulan ilginin artması, Yapay Bağışıklık Sistemlerinin ortaya çıkışının ardında yatan temel nedenler arasında yer almaktadır. Bağışıklık Sistemi Modelleri, Yapay Zeka araştırmacılarının ilgisini çekmiş ve oluşturulan yeni algoritma ve uygulamalarla Yapay Bağışıklık Sistemleri yavaş yavaş adından söz ettirmeye başlamıştır. Özellikle 1990'lı yılların sonlarından itibaren Yapay Zeka alanında çalışmakta olan çoğu araştırmacı, dikkatini Yapay Bağışıklık Sistemlerine yöneltmiş ve bu alanda etkin yeni algoritmalar oluşturulmaya başlanmıştır.

Bu bölümde, ikinci bölümde detaylı bir şekilde anlatılmış olan biyolojik bağışıklık sisteminin mekanizmaları kullanılarak meydana çıkmış olan yapay bağışıklık sistemlerinden bahsedilmiştir. Alt bölümlerde bağışıklık sisteminden modellenmiş yapılar ve mekanizmalardan bahsedilmiştir. Bunlar optimizasyon, örüntü tanıma, sınıflandırma, ağ güvenliği ve veri madenciliği gibi problemlerin çözümünde kullanılmaktadır.

3.1. B Hücrelerinin Modellenmesi

YBS'de en önemli prensip *B* hücrelerinin temsilidir. *B* hücrelerinin temel fonksiyonu, bakteri, virüs ve tümör hücreleri gibi dış kaynaklı proteinlere tepki vermek için antikor üretmek ve salgılamaktır. Her *B* hücresi belirli bir antikor üretmek için programlanmıştır. Antikorlar antijeni tanıyan ve ona bağlanan proteinlerdir. Antikor üretimi ve bağlanması genellikle maddeyi kuşatıp sindirme veya yok etme diğer hücreleri öldürmek için sinyal verme şeklinde olmaktadır.

Bağıışıklık sistemi adaptif bağıışıklık tepkileri için mekanizma sağılayan B hücrelerinin geniş bir popölasyonunu içermektedir. Bu tepki tanıma, uyarma ve klonsal çoğıalma süreçleri ile başarılr. Her bir B hücresi yabancı hücrelerin bağılandığı yerleri bulunan antikorların bir popölasyonudur. B hücresi saldırgan hücrenin antijenlerinin *epitoplari* ve antikorun *paratoplari* arasındaki bağılanma derecesine bağılı olarak saldırgan hücre ile eşleşir. Antikor-antijen etkileşimi B hücresinin tanıma yeteneğini ortaya çıkarmaktadır. Bununla beraber B hücresi antijene verilen bağıışıklık tepkisinde önemli bir rol oynamaktadır. Antikorlar ve antijenler arasındaki *bağılanma (duyarlılık)* derecesini temel alan bir B hücresi uyarılmış olur. Uyarılmış B hücreleri sonra klonlanır ve mutasyon işlemine tabi tutularak yeni ürünler oluşturulur. Bu ürünler yabancı maddeye atak yapmada ve yok etmede kullanılırlar. Bağıışıklık sisteminde mevcut olan tanıma ve tepki verme mekanizmaları arasındaki yakın bir ilişki söz konusudur.

Yapay bağıışıklık sistemlerinde B hücresi örüntü tanıma ve makine öğrenmesi için temel kavramdır. Özellikle başlıca tanıma mekanizmaları sık sık antikor-antijen bağılanımını modellemektedir. Uyarılmış bir B hücresinin özellik vektörü ve bir eğitim verisinin özellik vektörü arasındaki *Öklid uzaklığını* kullanarak antikorlar ve antijenler arasındaki bağılanmanın derecesine karar verilen çalışmalar söz konusudur (Timmis ve ark. 1999, 2000; Timmis ve Neal 2000; Timmis 2001; Knight ve Timmis 2001; De Castro ve Von Zuben 2001). Antikorlar ve antijenleri *ikili (binary) karakter dizileri* olarak ifade edildiğı ve birbirleriyle eşleşen bitlerin sayısını iki hücre arasındaki duyarlılık ölçüsü (*Hamming uzaklığı*) olarak kullanıldığı çalışmalar yapılmıştır (Hunt ve Cooke 1996; De Castro ve Von Zuben 2002b). İki hücre arasındaki duyarlılığa karar vermek için birden fazla metot kullanıldığı çalışmalar da söz konusudur (Hunt ve ark. 1998).

Bağılanma mekanizmalarına aldırmaksızın bu yaklaşımların tamamı sistemde mevcut hücreleri (antikorlar) ve eğitim verileri (antijenler) arasındaki etkileşimleri simule etmek için hesaplama açısından tutarlı bir mekanizma sağılamaya çalışmışlardır. Yapay bağıışıklık sistemlerinin neredeyse tamamı adaptif öğrenme mekanizmaları sağılamak için B hücresi uyarılması kavramını kullanmıştır.

3.2. Yapay Tanıma Toplari ve Kaynak Rekabeti

YBTS'de bir *B* hücre si tek bir antikoru temsil edecek şekilde basitleştirilmiştir (Watkins 2001). Bu antikor her bir elemanın farklı bir özelliği temsil ettiğı bir karakter dizisi olarak düşünölebilir. Bundan dolayı bir *B* hücre si tüm özellik uzayında belirli bir noktayı ifade etmektedir. Bu anlamda Timmis ve Neal (2000) tarafından, *Yapay Tanıma Toplari (YTT)* kavramı ilk kez önerilmiş, Timmis (2001) ve Knight ve Timmis (2001) çalışmalarında daha da geliştirmişlerdir.

Her bir *YTT* aynı sınıftaki antikorları temsil etmektedir. *Timmis ve Neal (2000)* sistemin kaynakları olarak *YTT*'ye ait antikorların uyarılma seviyelerinin toplamını kabul etmişlerdir. Yapay bağışıklık sistemi, sistemin temel bağışıklık bellek mekanizmasını oluşturmak için çok sayıda *YTT* içermektedir. Kaynakların sayısı belirli bir sayı ile sınırlıdır. Öğrenme olayı kaynakların rekabeti ile meydana gelmektedir. *YTT'ler* antijene tepki verecek şekilde uyarılmaktadır. Belirli bir antijen sisteme geldikten sonra her bir *YTT* uyarma seviyesini temel alarak kaynakları tüketmeye teşebbüs etmektedir. Bununla beraber kaynakların sayısı sınırlı olduğu için sadece en çok uyarılan *YTT'ler* kaynakları tüketecektir. Kalan *YTT'ler* (örneğin herhangi bir kaynağı olmayanlar) sistemden çıkartılmaktadır. Kaynaklar için bu rekabet sistemde sadece en güçlü *YTT'lerin* kalmasını garantilemektedir.

3.3. Bellek Hücreleri, Mutasyon ve Klonsal Seçim

Antikor popölasyonu bilinmeyen bir antijenle karşılaştığı zaman, en fazla uyarılan hücreler saldırganla savaşan yeni ürünler oluşturmak için klonlanacak ve mutasyona uğratılacaktır. Bu bağışıklık sisteminin ilk tepkisi olarak düşünölebilir. Bu ilk tepki boyunca belirli antijenleri veya yapısal olarak benzer antijenleri önleyen antikorların gelecekte antijenle karşılaşma olasılığına karşı sistemde kalmasına izin verilir. Bu kalıcı hücreler bundan sonra bellek hücre si (De Castro ve Von Zuben 2002b) olarak adlandırılır ve bağışıklık sisteminin ikincil tepkisi için temel teşkil ederler. Bağışıklık sistemi daha önceden karşılaşmış antijenlerle veya 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.

De Castro ve Von Zuben (2001); (2002b) tarafından *YBS* için bellek hücrelerinin gelişiminde, klonsal seçim prensibinin ve duyarlılık olgunlaşmasının rolü tartışılmıştır. *YBS*'de klonsal seçim, temel olarak bir antijen için en büyük duyarlılığa sahip veya en çok uyarılan hücrelerin yeni ürün oluşturmak için seçilmesidir. *YBS*'de mutasyon, rasgele özellik mutasyonu şeklindedir. Eğer bir antikoru belirli bir özelliği mutasyon için seçilirse, o özellik belirli bir aralık içerisinde herhangi bir rasgele değer ile mutasyona uğratılır. *YBS*'de duyarlılık olgunlaşması antijen tarafından en fazla uyarılan ürünleri üreten hücrelerin hayatta kalması ile sağlanır. Bu işlem kaynaklar için rekabeti temel almaktadır. Sadece başarılı olarak elde edilmiş kaynaklara klonlama için izin verilir. Sonuçta klonsal seçim, özellik mutasyonu ve duyarlılık olgunlaşması sürecinde belirli bir antijen tarafından en çok uyarılan tek bir 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.

3.4. Bağışıklık Ağları ve Hücreler Arası Etkileşimler

YBS'de rol oynanan bağışıklık sistem prensiplerinden biri bağışıklık ağ teorisidir. Jerne (1974) tarafından sunulan ve Perelson (1989) tarafından geliştirilen bağışıklık ağ teorisinde bağışıklık tepkileri sadece *B* hücreleri ile patojenlerin etkileşimini temel almaz aynı zamanda *B* hücrelerinin diğer *B* hücreleri ile etkileşimini de temel almaktadır. *B* hücreleri birbirlerini hem uyarırlar hem de bastırırlar ve bu etkileşim sonucunda bellek bağışıklık sisteminde alıkonulur. Timmis'in *bağışıklık ağ teorisinin* önemli noktalarından biri hücreler arası duyarlılık

ve bağlantı kuvvetlerinin düzenlenmesi fikridir. Timmis'in çalışmasında antikorların bir ağı, antikorlar arasındaki benzerliği veya duyarlılığı temel almaktadır.

Hücreler arası duyarlılık *YBS*'de önemli bir rol oynamaktadır. *YBS*'nin özelliklerinden biri veri azaltılmasını uygulama yeteneğidir. Eğitilmiş *YBS* sınıflandırıcısı tipik olarak orijinal eğitim örneklerinden çok daha az bellek hücresine sahiptir. Bu durum bellek hücresi yer değişimi ile başarılı ve bellek hücresi yer değişimi için anahtar kriter yeni geliştirilmiş bellek hücresi ve mevcut bellek hücresi arasındaki hücre etkileşiminin temel alınmasıdır.

3.5. Klonsal Seçim Algoritması

Klonsal seçim algoritması antijenik uyarıcı için bağışıklık tepkisinin temel özelliklerini açıklamak için bağışıklık sistemini kullanan bir algoritmadır. Bu prensip sadece antikorların hızlıca çoğalmasını sağlayacak hücrelere uygulanır. *Klonsal seçim algoritması* 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 baz alarak *Klonsal Seçim Algoritmasını* ortaya atmış ve oluşturdukları algoritmayı karakter tanıma, optimizasyon gibi problemlere uygulayarak performansını analiz etmişlerdir (De Castro ve Von Zuben (1999); (2000a)).

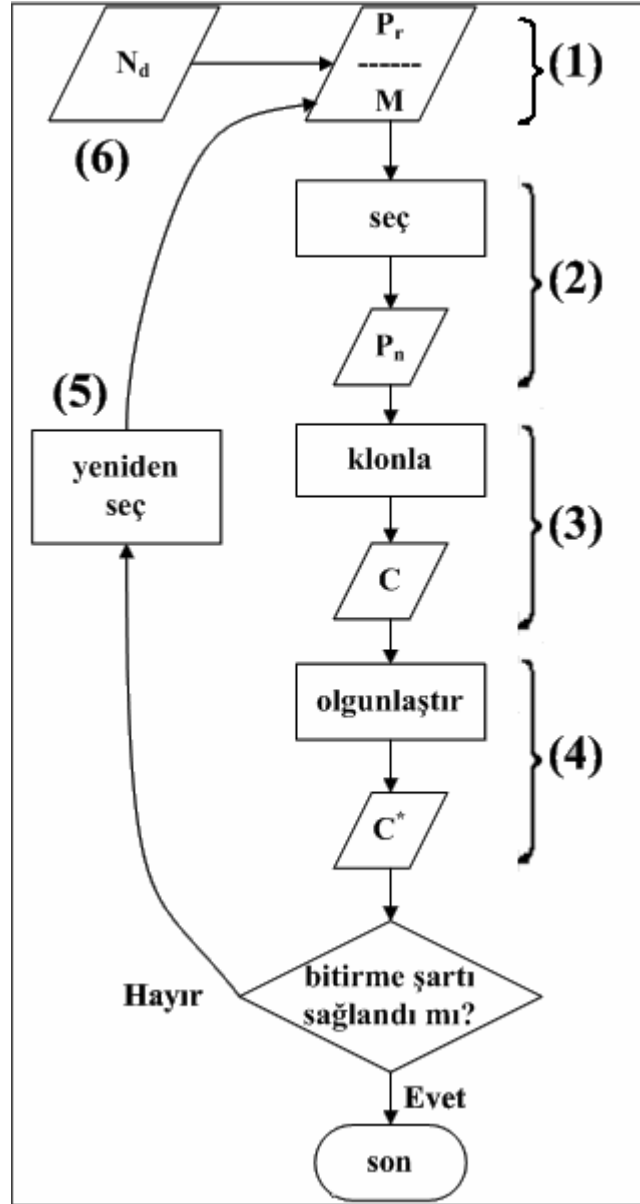
Klonsal Seçim Algoritması iki temel esas üzerine şekillendirilmiştir: Birincisi sadece antijeni tanıyan hücreler çoğalma için seçilirler, ikincisi seçilen ve çoğalan hücreler duyarlılık olgunlaşması işlemine tabi tutularak, antijene olan duyarlılıkları artırılır. *Klonsal Seçim Algoritmasının* blok diyagramı ise Şekil 3.1'de verilmiştir.

Algoritmanın çalışması şu şekildedir:

- 1- Aday çözümlerden oluşan bir popülasyon oluşturulur (*P*). Bu popülasyon, bellek hücreleri (*M*) ve geri kalan popülasyondan (*P_r*) oluşmaktadır ($P=P_r+M$).

- 2- Duyarlılık ölçütüne bağlı olarak popülasyondaki en iyi n tane eleman seçilir ve P_n popülasyonu oluşturulur.
- 3- 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.
- 4- Klonlardan oluşan C kümesi bir mutasyon işlemine tabi tutulur. Mutasyon işlemi de duyarlılık ile ters orantı söz konusudur. Mutasyondan sonra C^* kümesi oluşmaktadır.
- 5- Klonlama ve mutasyondan sonra gelişmiş bireyleri popülasyona eklemek için tekrar bir seçme işlemi yapılır ve M bellek kümesi oluşturulur. Bu seçme işleminden sonra P popülasyonundaki bazı hücrelerin yerlerini C^* kümesindeki bazı hücrelere bırakmaları olasıdır.
- 6- 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, *Klonsal Seçim Algoritması* standart bir *Genetik Algoritma* ile karşılaştırıldığında lokal optimumlardan oluşan çok çeşitli bir çözüm kümesi üretebilmektedir. *Genetik Algoritmalar* ise tüm popülasyonu en iyi bireye benzetmeye çalışmaktadır. 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 etmektedir. Her iki yöntemde de çözüme ulaşmada popülasyon büyüklüğü önemli bir parametredir. *Genetik Algoritmalar*da popülasyon içerisindeki her bir birey kromozom olarak adlandırılırken *Klonsal Seçim Algoritması*nda antikor olarak adlandırılmaktadır. Genetik algoritmalarda optimal ya da optimale yakın çözüme ulaşmada kullanılan en önemli operatör çaprazlama operatörüdür. *Klonsal Seçim Algoritması*nda çaprazlama operatörü kullanılmamaktadır. Mutasyon her iki algoritmada da kullanılan bir operatördür. Mutasyon operatörü, algoritmaların yeni çözümlere ulaşma sürecinde ince ayar yapmasını sağlamaktadır.



Şekil 3.1 Klonsal Seçim Algoritmasının blok diyagramı (De Castro ve Von Zuben 2000a)

3.6. Takviyeli Öğrenme 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 patojene tepki veren *lenfositin* popülasyonunun boyutu enfeksiyonun sonucuna karar vermede 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 rasgele oluşturulmuş repertuar sürekli olarak azalır (Perelson ve Weisbuch 1997).

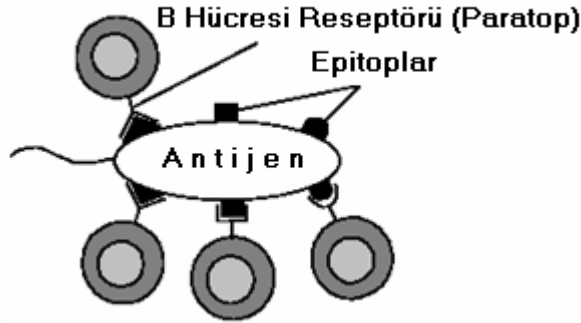
Bağışıklık sisteminin gelişiminde bir organizmanın hayatı boyunca defalarca belirli bir antijenle karşılaşması beklenebilir. Adaptif bağışıklık tepkisini uyaran bir antijen ilk açığa çıktığında, her bir farklı eğilim için antikor üreten *B* hücrelerinin küçük klonları tarafından ele alınır. İkinci kez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

Bağışıklık belleğini göstermek için düşünelim ki bir *A* antijeni ortaya çıksın ve organizma içerisinde birkaç adet antikor bulsun. Bir gecikme 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 gösterir (Janeway ve ark. 1999). Bir başka deyişle bağışıklık belleğinin önemli bir karakteristiği ilişkili olmasıdır. *B* hücreleri sadece *A* antijeni değil aynı zamanda yapısal olarak *B* antijeniyle ilgili daha hızlı ve etkili ikinci tepkiyi vermek için uyarlanır. Bu durum *çapraz reaksiyon* veya *çapraz tepki* olarak adlandırılır (Hoffman 1986; Ada ve Nossal 1987; Sprent 1994; Smith ve ark. 1997; Hodgkin 1998; Mason 1998) (Bu durum Şekil 2.8’de gösterilmiştir.).

Bazı araştırmacılar (Allen 1987; Coutinho 1989) önceki antijene tepki vermiş uzun ömürlü *B* hücrelerinin durumunu koruduğunu ve ikinci antikor tepkisinde önemli bir rol oynadıklarını ileri sürmüştür. Burada önemli bir nokta en yüksek eğilimli hücrelerin en yüksek kalitedeki aday çözümleri olarak korunmalarının gerekliliğidir.

3.7. Örüntü Tanıma

Bağışıklık sisteminde örüntü tanımanın bakış açısından hem *B* hem de *T* hücrelerinin en önemli özelliği antijenleri tanıyabilen ara yüzlerindeki reseptör moleküllerine sahip olmasıdır (Şekil 3.2).



Şekil 3.2 Epitop olarak adlandırılan ve antikor tarafından tanınan antijen parçası

Bağışıklık 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 *epitoplarla* birlikte birbirlerini etkiler. Antijen moleküllerini eritebilir. *T* hücre reseptörleri sadece hücre arayüz molekülleri ile etkileşimdedir. *T* hücreleri bağışıklık tepkisinin düzenlenmesinde çok önemli bir rol oynayan diğer hücreleri öldürebilen veya onların gelişmesini ilerleten kimyasal maddeler salgılamaktadır.

3.8. Şekil Uzay Modeli

Antijen tanıma için gerekli olan tüm hücrelerin oluşturduğu kümeye, *bağışıklık repertuarı* adı verilmektedir. Vücudumuzun 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.

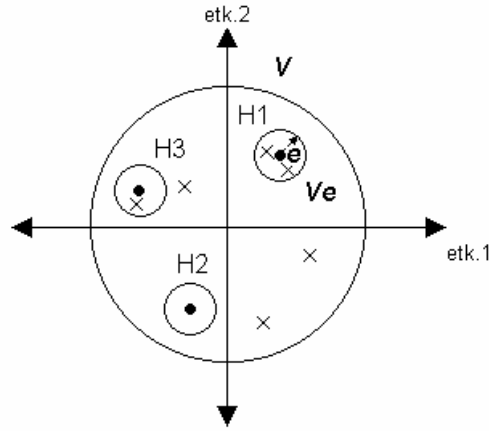
Bağışıklık hücreleri kümesi zararlı saldırganlardan vücudu uygun şekilde korumak için bağışıklık tepkisi olarak adlandırılan antijenleri tanımaya çalışırlar. Nicel olarak bağışıklık hücresi molekülleri ve antijenler arasındaki etkileşimleri açıklamak için Perelson ve Oster (1979) “*Şekil Uzayı (Shape Space)*” kavramını ortaya atmışlardır.

Şekil uzayı modelinde bir reseptör ve bir molekül arasında bağlayıcılığın derecesi temel alınır. L adet parametreden oluşan antikor bağlanma yerinin (*paratop*) olduğu kabul edilir. Parametrelerin kesin sayısı veya onların değerleri burada önemli değildir. Şekil uzayı 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 3.3’te tanıma çemberi kavramı ve şekil uzayında repertuar tamlığı gösterilmektedir.

Şekil 3.3’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. Şekil 3.3’te $H1$ 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 3.3’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 antikoron 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

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



Şekil 3.3 Tanıma topu kavramı ve repertuar gösterimi

3.9. Antijen-Antikor Gösterimleri ve Duyarlılık

Antijen-Antikor gösterimi, kısmen etkileşim derecesini (tamamlayıcılık) hesaplamada kullanılan uzaklık ölçüsüne karar vermektedir. 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*, uzaklık ölçüsü yoluyla tahmin edilebilen uzaklık ile ilişkilidir. *Öklid uzaklığı* veya *Manhattan uzaklığı* bu ölçülerden bazılarıdır. *Öklid Uzaklığı* durumunda eğer bir antikorun koordinatları $\langle ab_1, ab_2, \dots, ab_L \rangle$ ile ve antijenin koordinatları $\langle ag_1, ag_2, \dots, ag_L \rangle$ ile verilirse antikor ve antijen arasındaki *Öklid Uzaklığı* Eşitlik 3.1’de gösterilmiştir. Eşitlik 3.2’de Manhattan uzaklığı gösterilmiştir.

$$\text{Öklid_Uzaklığı}(ab, ag) = \sqrt{\sum_{i=1}^L (ab_i - ag_i)^2} \quad (3.1)$$

$$Manhattan_Uzaklığı(ab, ag) = \sqrt{\sum_{i=1}^L |ab_i - ag_i|} \quad (3.2)$$

Gerçel (reel) değerli koordinatlar kullanan ve Eşitlik 3.1'deki gibi uzaklığı ölçen şekil-uzayları *Öklid Şekil Uzayı* olarak adlandırılmaktadır (Segel ve Perelson 1988; De Boer ve ark. 1992; Smith ve ark. 1997). Gerçel değerli koordinatlı fakat Öklid uzaklığı yerine Manhattan uzaklığı kullanan şekil-uzayı *Manhattan Şekil Uzayı* olarak adlandırılır.

Antijenler ve antikorlar gerçel değerli vektörler yerine *ikili* sembollerle ifade edilirlse 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* Eşitlik 3.3'te gösterilmiştir.

$$Hamming_Uzaklığı(ab, ag) = \sum_{i=1}^L \delta_i, \begin{cases} \delta_i = 1 & \text{eğer } ab_i \neq ag_i \\ \delta_i = 0 & \text{aksi durumda} \end{cases} \quad (3.3)$$

4. KARAR AĞAÇLARI VE BİLGİ KAZANCI KAVRAMI

Karar ağaçları geniş olarak kullanılan bir sınıflandırma metodudur. Bir karar ağacı karar düğümleri, dallar ve yaprak düğümlerinden oluşmaktadır (Mitchel 1997; Kamber 2000). Bu metotta kök düğümünden yaprak düğüme kadar aşağıya doğru ilerleyen bir ağaç yapısı oluşturulmaktadır. Ağaçtaki her karar düğümü verinin bir özelliğini test etmektedir ve her dal bu özelliğin olabilecek değerlerinden birine göre uygun düğümü gösterecek şekilde ağacın kök seviyesinden alt seviyelerine doğru inmektedir. Bu süreç yeni düğümü kök kabul eden alt ağaç içinde tekrar edilmektedir. Ağacın her bir dalı sınıflandırma işlemini tamamlamaya adaydır. Eğer bir dalın ucunda sınıflandırma işlemi gerçekleşmiyorsa yani farklı sınıf değerleri söz konusu ise, o dalın sonucunda bir karar düğümü oluşmaktadır. Ancak dalın sonundaki verilerin hepsi belirli bir sınıfta ise, o dalın sonunda yaprak düğümü vardır. Bu yaprak düğüm, veri üzerinde belirlenmek istenen sınıflardan birisidir.

Karar ağacı tabanlı algoritmalar Böl ve Fethet öğrenme metoduna göre işlem yaparak örnekleri alt setlere bölmektedir. Böl ve Fethet öğrenme metodunun ana fikri, bir örnek kümesini, her bir alt kümeye sadece tek bir sınıf kalıncaya kadar alt kümelere bölmektir. Bu metodun sonucu olarak bir karar ağacı oluşmaktadır.

Karar ağacı algoritmasını kullanarak verinin sınıflandırılması iki basamaklı bir süreci içermektedir (Kamber 2000). İlk basamak öğrenme basamağıdır. Öğrenme basamağında önceden bilinen bir eğitim verisi, model oluşturmak amacıyla sınıflandırma algoritması tarafından analiz edilmektedir. Öğrenilen model, sınıflandırma kuralları veya karar ağacı olarak gösterilmektedir. İkinci basamak ise sınıflandırma basamağıdır. Sınıflandırma basamağında test verisi, sınıflandırma kurallarının veya karar ağacının doğruluğunu belirlemek amacıyla kullanılmaktadır.

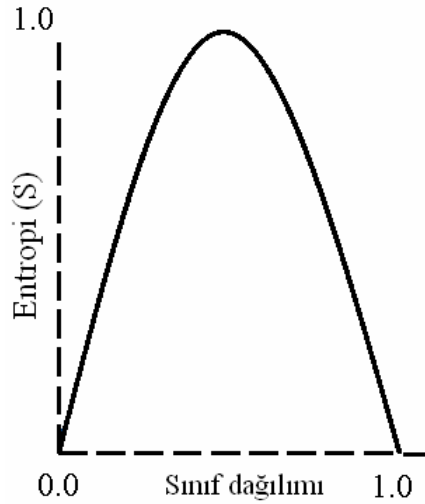
Test verisine uygulanan bir modelin doğruluğu, yaptığı doğru sınıflandırmanın test verisindeki tüm sınıflara oranıdır. Her test örneğinde bilinen sınıf, model tarafından tahmin edilen sınıf ile karşılaştırılmaktadır.

Karar ağacı algoritmasında en önemli aşama ağaçtaki her bir düğümde hangi özelliğin seçileceğine karar vermektir. Karar ağacı tabanlı algoritmalar karar ağacını

oluşturmak için en büyük bilgi kazancını veren özelliklerden arama yapmak için bilginin Entropi (Entropy) ölçüsünü kullanmaktadır (Quinlan 1986; Blahut 1987; Cendrowska 1990; Michalski ve Kodratoff 1990; Mitchell 1997). Özelliğin değerinin ölçüsü *bilgi kazancı (information gain)* adı verilen istatistik bir değer ile belirlenmektedir. *Entropi*, örneklerin belirsizliğini (gürültüsünü) karakterize etmektedir. Eşitlik 4.1’de S kümesinin *Entropi* değeri gösterilmiştir.

$$Entropi(S) = - \sum_{i=1}^c p_i \log_2 p_i \quad (4.1)$$

burada p_i , S kümesindeki i . sınıfa ait örneklerin sayısının S kümesindeki bütün örneklerin sayısına oranıdır. c , sınıf sayısıdır. S kümesindeki bütün örnekler aynı sınıfta ise *Entropi* değeri sıfırdır yani belirsizlik yoktur. S kümesindeki örneklerin sınıf sayısı eşit ise *Entropi* değeri maksimum değeri olan bir değerini almaktadır yani belirsizlik maksimumdur. 2 sınıflı bir veri kümesi için Şekil 4.1’de bu durum gösterilmektedir. c adet sınıflı bir veri kümesi için *Entropi* değeri en fazla $\log_2 c$ ’dir.



Şekil 4.1 Sınıf dağılımına göre belirsizlik değişimi

Bir özelliğin etkinlik ölçüsü bilgi kazancı terimi ile kullanılmaktadır. Eşitlik 4.2’de A özelliğinin bilgi kazancı $Kazanç(S,A)$ tanımlanmaktadır.

$$Kazanç(S, A) = Entropi(S) - \sum_{v \in Değerler(A)} \frac{|S_v|}{|S|} Entropi(S_v) \quad (4.2)$$

Burada $Değerler(A)$, A özelliğinin bütün olası değerlerinin kümesidir; S_v ise S kümesindeki A özelliğinin v değerine sahip örneklerinin kümesidir.

4.1. Standart Bir Veri kümesi Üzerinde Karar Ağacı Uygulaması

Karar ağacı algoritmasının anlaşılabilmesi için 14 günlük hava koşullarına göre tenis oynama durumu örneği standart bir örnek olarak literatürde verilmektedir (Mitchell 1997). Bu örnekte tenis oynamak isteyen kişiler için hava koşulları kayıtları yardımıyla ağaç yapısı oluşturulmaktadır. Hedef özellik olarak tenis oynama durumu iki değer alabilmektedir. Bunlar oynanabilir ve oynanamaz şeklindedir. Tablo 4.1’de bu örnek için verilmiş özellikler ve bunların alabileceği olası değerler verilmiştir. Tablo 4.2’de bu özellikleri içeren öğrenme verisi verilmiştir.

Tablo 4.1 Tenis örneği için veri özellikleri

Özellik adı	Olası değerler
<i>Hava durumu</i>	<i>Güneşli, Bulutlu, Yağmurlu</i>
<i>Sıcaklık</i>	<i>Sıcak, Ilık, Soğuk</i>
<i>Nem</i>	<i>Yüksek, Normal</i>
<i>Rüzgar</i>	<i>Zayıf, Güçlü</i>

Verilen tenis örneği için,

$$Entropi(S) = -(9/14)\log_2(9/14) - (5/14)\log_2(5/14) = 0.94$$

olarak hesaplanmaktadır. Bu aşamadan sonra kök düğüm olarak hangi özelliğin seçileceğine karar verebilmek için her bir özelliğin bilgi kazancı hesaplanmaktadır. Bilgi kazancı en yüksek olan özellik kök düğüm olarak belirlenmektedir.

“*Hava durumu*” özelliği için,

$$Değerler(Hava durumu) = \{Güneşli, Bulutlu, Yağmurlu\}$$

$$S = [9+, 5-]$$

$$S_{Güneşli} = [2+, 3-]$$

$$S_{Bulutlu} = [4+, 0-]$$

$$S_{Yağmurlu} = [3+, 2-]$$

$$\begin{aligned} \text{Kazanç}(S, \text{Hava durumu}) &= \text{Entropi}(S) - \sum_{v \in \{\text{Güneşli}, \text{Bulutlu}, \text{Yağmurlu}\}} \frac{|S_v|}{|S|} \text{Entropi}(S_v) \\ &= 0.94 - (5/14)\text{Entropi}(S_{\text{Güneşli}}) - (4/14)\text{Entropi}(S_{\text{Bulutlu}}) - (5/14)\text{Entropi}(S_{\text{Yağmurlu}}) \\ &= 0.94 - (5/14)0.971 - (4/14)0.0 - (5/14)0.971 \\ &= 0.247 \text{ olarak hesaplanmaktadır.} \end{aligned}$$

Tablo 4.2 Tenis örneği için veri kümesi

Gün	Hava durumu	Sıcaklık	Nem	Rüzgar	Tenis Oynar
1	Güneşli	Sıcak	Yüksek	Zayıf	Hayır
2	Güneşli	Sıcak	Yüksek	Güçlü	Hayır
3	Bulutlu	Sıcak	Yüksek	Zayıf	Evet
4	Yağmurlu	Ilık	Yüksek	Zayıf	Evet
5	Yağmurlu	Soğuk	Normal	Zayıf	Evet
6	Yağmurlu	Soğuk	Normal	Güçlü	Hayır
7	Bulutlu	Soğuk	Normal	Güçlü	Evet
8	Güneşli	Ilık	Yüksek	Zayıf	Hayır
9	Güneşli	Soğuk	Normal	Zayıf	Evet
10	Yağmurlu	Ilık	Normal	Zayıf	Evet
11	Güneşli	Ilık	Normal	Güçlü	Evet
12	Bulutlu	Ilık	Yüksek	Güçlü	Evet
13	Bulutlu	Sıcak	Normal	Zayıf	Evet
14	Yağmurlu	Ilık	Yüksek	Güçlü	Hayır

“Sıcaklık” özelliği için,

Değerler(Sıcaklık)= {Sıcak, Ilık, Soğuk}

$$S = [9+, 5-]$$

$$S_{Sıcak} = [2+, 2-]$$

$$S_{Ilık} = [4+, 2-]$$

$$S_{Soğuk} = [3+, 1-]$$

$$\begin{aligned}
Kazanç(S, Sıcaklık) &= Entropi(S) - \sum_{v \in \{Sıcak, Ilık, Soğuk\}} \frac{|S_v|}{|S|} Entropi(S_v) \\
&= 0.94 - (4/14)Entropi(S_{Sıcak}) - (6/14)Entropi(S_{Ilık}) - (4/14)Entropi(S_{Soğuk}) \\
&= 0.94 - (4/14)1.0 - (6/14)0.91 - (4/14)0.81 \\
&= 0.029 \text{ olarak hesaplanmaktadır.}
\end{aligned}$$

“Rüzgar” özelliği için,

Değerler(Rüzgar)= {Zayıf, Güçlü}

$S = [9+, 5-]$

$S_{Zayıf} = [6+, 2-]$

$S_{Güçlü} = [3+, 3-]$

$$\begin{aligned}
Kazanç(S, Rüzgar) &= Entropi(S) - \sum_{v \in \{Zayıf, Güçlü\}} \frac{|S_v|}{|S|} Entropi(S_v) \\
&= 0.94 - (8/14)Entropi(S_{Zayıf}) - (6/14)Entropi(S_{Güçlü}) \\
&= 0.94 - (8/14)0.811 - (6/14)1.0 \\
&= 0.048 \text{ olarak hesaplanmaktadır.}
\end{aligned}$$

“Nem” özelliği için,

Değerler(Nem)= {Yüksek, Normal}

$S = [9+, 5-]$

$S_{Yüksek} = [3+, 4-]$

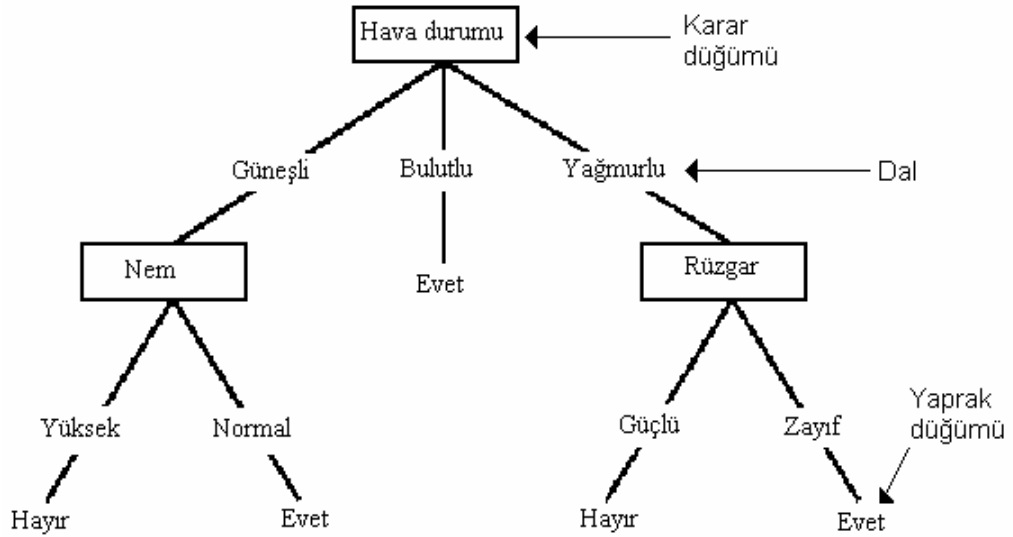
$S_{Normal} = [6+, 1-]$

$$\begin{aligned}
Kazanç(S, Nem) &= Entropi(S) - \sum_{v \in \{Yüksek, Zayıf\}} \frac{|S_v|}{|S|} Entropi(S_v) \\
&= 0.94 - (7/14)Entropi(S_{Yüksek}) - (7/14)Entropi(S_{Normal}) \\
&= 0.94 - (7/14)0.985 - (7/14)0.592 \\
&= 0.151 \text{ olarak hesaplanmaktadır.}
\end{aligned}$$

Bu bilgi kazancı değerlerine göre en yüksek bilgi kazancına sahip özellik olan “Hava durumu” özelliği kök düğüm olarak belirlenmektedir. Daha sonra “Hava durumu” özelliği için olası değerlere uygun olarak gerekli ise alt ağaç için kök düğüm ya da yaprak düğüm belirlenecektir. Ağaçta her bir düğüm, kökten başlayarak

yol üzerinde henüz dikkate alınmamış nitelikler arasından en büyük kazancı veren özellikler seçilerek belirlenmektedir.

Tablo 4.1'deki özellikler ve Tablo 4.2'deki öğrenme verileri kullanıldığında ortaya Şekil 4.2'deki karar ağacı çıkmaktadır.



Şekil 4.2 Tenis örneği için oluşan karar ağacı

4.2. Özelliklerin Değerlerinin Ayırık Hale Getirilmesi

Karar ağacı metodunda özelliklerin bilgi kazancı hesaplanırken özellik değerlerinin ayırık değerler olması gerekmektedir. Bu sebepten dolayı sürekli değerli özelliklerin ayırık hale dönüştürülmesi gerekmektedir. Bu amaçla çeşitli yöntemler söz konusudur.

4.2.1. Danışmansız ayırım

Sürekli özelliklerin ayırık hale dönüştürülmesinde en basit yaklaşımlardandır. Bu yaklaşımlarda kullanıcıdan aralık sayısı veya bir aralığa düşecek eleman sayısı alınmaktadır.

4.2.1.1. Eşit aralık genişliği ayırım yaklaşımı

Bu yaklaşımda verinin sürekli olan her bir özelliği kendi içerisinde sıralanmaktadır. Daha sonra özellik değerleri eşit veri aralıklarına bölünmektedir. Kullanıcı tarafından k aralık sayısı belirlenmektedir. Verinin özellik değerinin maksimum ve minimum değerleri arasındaki fark, k değerine bölünerek her bir aralığın genişliği belirlenmektedir. Eşitlik 4.3'te bu durum gösterilmiştir. Tez çalışmasında sürekli değerli özelliklerin ayırık hale getirilmesinde bu yaklaşım kullanılmıştır.

$$aralık = \frac{\max(X) - \min(X)}{k} \quad (4.3)$$

Basit bir örnekle açıklamak gerekirse veri özellik değerleri $X = \{40, 45, 55, 60, 64, 67, 78, 89, 110, 140, 154, 167\}$ için, k değeri 4 olsun. Bu durumda

$$aralık = \frac{167 - 40}{4} = \frac{127}{4} = 31.75 \text{ olarak hesaplanmaktadır. Veri özelliklerinin}$$

ayırık hali Tablo 4.3'teki gibi olacaktır.

Tablo 4.3 Örnek bir özelliğin eşit aralık genişliği yaklaşımı ile ayırık hale getirilmiş hali

Özellik değeri	40	45	55	60	64	67	78	89	110	140	154	167
Ayrık değeri	A1	A1	A1	A1	A1	A1	A2	A2	A3	A4	A4	A4

4.2.1.2. Eşit sıklık ayırım yaklaşımı

Bu yaklaşımda da verinin sürekli olan her bir özelliği kendi içerisinde sıralanmaktadır. Daha sonra özellik değerleri her bir veri aralığına eşit sayıda özellik değeri düşecek şekilde bölünmektedir. Kullanıcı tarafından k sayısı belirlenmektedir. Veri sayısı, k değerine bölünerek her bir aralığın eleman sayısı belirlenmektedir. Eşitlik 4.4'te bu durum gösterilmiştir.

$$adet = \frac{\text{veri sayısı}}{k} \quad (4.4)$$

Aynı örnekle bu yaklaşımı açıklamak gerekirse veri özellik değerleri $X=\{40, 45, 55, 60, 64, 67, 78, 89, 110, 140, 154, 167\}$ için, k değeri 3 olsun. Bu durumda

$$adet = \frac{12}{3} = 4 \text{ olarak hesaplanmaktadır. Tablo 4.4'te veri özelliklerinin ayrık}$$

hali gösterilmiştir.

Tablo 4.4 Örnek bir özelliğin eşit sıklık yaklaşımı ile ayrık hale getirilmiş hali

Özellik değeri	40	45	55	60	64	67	78	89	110	140	154	167
Ayrık değeri	A1	A1	A1	A2	A2	A2	A3	A3	A3	A4	A4	A4

4.2.2. Danışmanlı ayırım

Danışmanlı ayırım yaklaşımları ayırım süreci boyunca veri değerleri ve bu değerlerin ait olduğu sınıf değerini kullanmaktadır.

4.2.2.1. Holte'un 1R ayırıcısı

Bu metot özellik değerlerinin aralıklara bölmeye karar verirken karşılaşılan hatalarla birlikte kullanıldığı için hata tabanlı bir yaklaşım kullanmaktadır. Özellik değerleri artan sırada sıralanmaktadır ve her bir aralığa aynı sınıfa sahip veri özellikleri gelecek şekilde özellik değerleri aralıklara bölünmektedir. Burada her bir özellik için bir aralık kullanımı söz konusu olabilmektedir. Bu problemin üstesinden gelebilmek için kullanıcı tarafından belirtilen bir değer, belirli sınıftaki özelliklerin sayısının minimum değeri olarak belirlenmektedir. Bu durumda aralık, içerisinde birden fazla sınıf içerebilmektedir. Örnek olarak bir sınıftan en az 6 örnek olana kadar aralık soldan sağa doğru kaymaktadır. Deneysel analizler sonucunda aralık içerisinde 6 değeri önerilmektedir (Holte 1993).

Tablo 4.5'te basit bir örnekle metot açıklanmaktadır. Burada bir sınıf için minimum veri sayısı 6 alınmıştır. Öncelikle veri artan sırada sıralanmıştır. İlk veriden

İtibaren herhangi bir sınıftaki veri sayısı 6 olana kadar aralığını sınırı sağa doğru kaymaktadır. Verilen örnek için 9. veri ile beraber bir sınıf için minimum veri sayısı olan 6 değerine ulaşılmaktadır. Bu durumda 9. veriden itibaren sınıf değeri değişimi olana kadar sağa doğru ilerlenmektedir. 10. veri de aynı sınıfta olduğu için sınır değeri belirli değildir. 11. verinin sınıf değeri farklı olduğu için aralığın üst sınırı belirlenebilmektedir. Aralığın üst sınırı 10. ve 11. verinin aritmetik ortalaması olarak belirlenmektedir. Yani 75 ve 80 değerinin aritmetik ortalaması olan 77.5'tir. 64 ile 77.5 arasındaki veri değerleri bir aralıktadır, 77.5'ten büyük değer diğer bir aralıktadır.

Tablo 4.5 Örnek bir özelliğin Holte'un 1R ayırıcısı yaklaşımı ile ayırık hale getirilmiş hali

Veri No	1	2	3	4	5	6	7	8	9	10	11	12	13	14
Özellik Değeri	64	65	68	69	70	71	72	72	75	75	80	81	83	85
Sınıf Değeri	1	2	1	1	1	2	2	1	1	1	2	1	1	2
Ayrık Değeri	A1	A1	A1	A1	A1	A1	A1	A1	A1	A1	A2	A2	A2	A2

4.2.2.2. C4.5 ayırıcısı

C4.5 algoritması sürekli özellikleri bir eşik değeri kullanarak her bir düğümde ikiye ayırarak karar ağacı oluşturmaktadır (Pazzani 1995). Bilgi kazancı teorisi, veriyi bölmek için en büyük kazanç oranını veren eşik değerini belirlemede kullanılmaktadır. Böl ve Fethet algoritması her iterasyonda aralığın daha küçük alt kümelerle bölünüp bölünmeyeceğine karar vermek için uygulanmaktadır.

4.2.2.3. Fayyad ve Irani'nin entropi tabanlı MDL metodu

Bu metod ikili aralıklardan ziyade çoklu aralıklar tercih eden bir yaklaşımdır ve ayırık değerler üretmek için aynı düğümde sürekli değerli özelliklerin 2'den fazla ayrılmasıyla ağaç yapısı oluşturmaktadır. Aralık değerlerinin başlangıç ve bitiş değerlerine karar vermede yanlış karar verme olasılığını minimuma indiren Minimum Tanımlama Boyut Prensipleri (Minimum Description Length Principle-MDLP) temel alınmaktadır. Bilgi entropi minimizasyonu, bütün olası eşikler

arasında entropi fonksiyonunu minimize eden eřiđi bulan eřik sınırlarını seřmede kullanılmaktadır (Fayyad ve Irani 1993).

5. GELİŞTİRİLEN SİSTEM

Bu bölümde geliştirilen sistemde kullanılan tanımlamalar verildikten sonra tez çalışmasında temel alınan ve Watkins (2001) tarafından geliştirilen *Yapay Bağışıklık Tanıma Sistemi (YBTS)* incelenmektedir. *YBTS*'de tespit edilen bazı eksiklikler alt bölüm olarak sunulduktan sonra bu eksiklikleri gidermek amacıyla tez çalışmasında geliştirilmiş olan *Bilgi Kazancı Tabanlı Yapay Bağışıklık Tanıma Sistemi (BK-YBTS)* ayrıntılı olarak açıklanmaktadır.

5.1. Tanımlamalar

Bu bölümde, *BK-YBTS*'de kullanılan önemli terimlerin ve kavramların tanımları sunulmaktadır.

- **Antijen:** Özellik vektörü ve sınıf değerinden oluşan eğitim verisidir.
- **Antikor:** Antijen ile aynı şekilde temsil edilmektedir. Sınıf değeri ile bir özellik vektörünün birleştirilmiş halidir. Eğitim aşamasında oluşmakta ve antijene karşı tepki vermektedir.
- **Duyarlılık:** İki antikor veya antijen arasındaki yakınlığın veya benzerliğin bir ölçüsüdür. Uygulamada bu değer 0 ile 1 aralığını kapsayacak şekilde kalması garantilenmiştir ve bu değer iki özellik vektörü arasında *bilgi kazancı tabanlı uzaklık* ile hesaplanmaktadır. Küçük uzaklık değerleri güçlü duyarlılığa işaret etmektedir.
- **Duyarlılık Eşiği:** Geliştirilen sistemde eğitim kümesindeki her bir sınıftaki antijenlerin aralarındaki ortalama duyarlılık değeridir.
- **Duyarlılık Eşik Ölçüsü:** Bu değer, duyarlılık eşiği ile çarpılarak sistemin eğitim rutinde bellek hücresi yer değişimi için bir eşik değeri sağlamaktadır.

- **Bilgi Kazancı Tabanlı Yapay Bağışıklık Tanıma Sistemi (BK-YBTS):** Bağışıklık sistemlerinden esinlenilen bir sınıflandırma sistemidir.
- **Entropi:** Verideki rasgeleliği, belirsizliği, kirliliği, beklenmeyen durumun ortaya çıkma olasılığını karakterize eden bir ölçüdür.
- **Bilgi Kazancı:** Verideki özelliklerin etkinliğinin ölçüsünü hesaplamada kullanılmaktadır
- **B Hücresi:** Antikor olarak kullanılmaktadır.
- **Aday Bellek Hücresi:** Bir antijen sisteme sunulduktan sonra en çok uyarılan ve eğitim antijeni ile aynı sınıftaki antikordur.
- **Sınıf:** Verilmiş olan bir özellik vektörünün kategorisini belirtmektedir. Bu aynı zamanda bir hücrenin çıkışı olarak da kullanılmaktadır.
- **Klonlama Oranı:** Belirli bir antikorun üretmeye izin verilen kadar mutasyona uğramış klonlarının sayısına karar vermek için kullanılan bir tamsayı değeridir. Uygulamada seçilmiş bir antikorun belirli bir antijene tepki verdikten sonra *klonlama oranı x uyarma değeri* kadar mutasyona uğramış klonların üretilmesine izin verilmektedir. Bu üretim, aynı zamanda bir antikor için kaynak tahsisinde kullanılmaktadır. Bundan dolayı *klonlama oranı* hücre popülasyonu için *kaynak tahsisi çarpanı* ve *klonsal mutasyon çarpanı* olarak iki rolde hizmet sunmaktadır.
- **Yerleşmiş Bellek Hücresi:** Kaynakların rekabeti ile hayatta kalan ve belirli bir eğitim antijeni için en çok uyarılan bir antikordur ve bellek hücrelerinin kümesine eklenmektedir.
- **Özellik Vektörü:** Tek bir veriye ait değerlerin bir sırası olarak ifade edilmektedir. Her bir değer veri içerisindeki farklı bir özelliği ifade etmektedir.
- **Klon Sayısı:** Belirli bir hücre popülasyonu içerisine eklenilmesine izin verilen bellek hücresinin veya antikorların mutasyona uğrayacak klonlarının sayısını ifade etmektedir. Uygulamada bu değer seçilmiş bellek hücresi klonlanırken *klonlama_oranı x uyarma değeri*, antikor hücreleri klonlanırken *ab.uyarım x klonlama_oranı* olarak hesaplanmaktadır.
- **k En Yakın Komşu:** Daha önceden gözlemlenmemiş veriyi k adet en yakın veri noktası arasında bir çoğunluk oylaması ile karar vererek sınıflandıran bir

sınıflandırma yöntemidir. *BK-YBTS* 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:*** Antijenin sisteme sunulması sonunda eğitim antijeni tarafından en fazla uyarılan bir antikordur. Bir *bellek hücresi* yenisi ile değiştirilebilir. Bu değişiklik, sadece bir aday *bellek hücresi*, eğitim antijenini en çok uyararak *yerleşmiş bellek hücresinden* daha fazla uyardığı ve *yerleşmiş bellek hücresi* ve *aday bellek hücresi* arasındaki duyarlılık, *duyarlılık eşiği* değeri ve *duyarlılık eşik ölçüsü* değeri çarpımından daha az olduğu zaman meydana gelmektedir.
- ***Mutasyon oranı:*** Bir antikorun herhangi bir özelliğinin mutasyona uğrama olasılığını işaret eden 0 ile 1 arasında bir parametredir.
- ***Çıkış:*** Bir hücre ile birleştirilmiş sınıf değeridir. Hücreye uygun özellik vektörünün sınıfı ile aynıdır.
- ***Kaynaklar:*** Sistemde izin verilen antikorların sayısını sınırlayan bir parametredir. Her bir antikora, uyarma değerini ve klonlama oranını temel olarak kaynak ayrılmaktadır. Sistemdeki kaynakların toplam sayısı belirli bir sınırla belirlenmiştir. Eğer sistemde bulunması için izin verilenden daha fazla kaynak tüketilirse, sistemde izin verilen kaynak sayısına ulaşana kadar en az uyarılan başlayarak antikorlar yok edilir.
- ***Çekirdek hücre:*** Eğitim başlangıcında bellek hücresi ve antikor popülasyonlarını başlatmak için kullanılan ve eğitim kümesinden alınan bir antikordur.
- ***Uyarma fonksiyonu:*** Bir antikorun bir antijene veya bir başka antikora tepkisinin ölçüsü için kullanılmaktadır. Bu çalışmada sunulan *BK-YBTS* uygulaması için uyarma fonksiyonu antikorun özellik vektörleri ve antijen arasındaki *bilgi kazancı tabanlı uzaklık* ile ters orantılıdır. Yani uzaklık ne kadar küçük ise uyarma o derece büyüktür.
- ***Uyarma değeri:*** Uyarma fonksiyonu tarafından döndürülen değerdir.
- ***Uyarma eşiği:*** Spesifik bir antijenin eğitimi için bir durma kriteri olarak kullanılan 0 ile 1 arasında bir parametredir. Uygulamada her bir sınıf ait

antikorların ortalama uyarma değeri, uyarma eşliğinden daha yüksek olduğu zaman o antijenin eğitimi durdurulmaktadır.

- **Test kümesi:** Eğitilmiş sistemin sınıflandırma performansını değerlendirmek için kullanılan antijenlerin kümesidir.
- **Eğitim kümesi:** Sistemi eğitmek için kullanılan antijenlerin kümesidir.

5.2. Yapay Bağışıklık Tanıma Sistemi (YBTS)

Bu bölümde Watkins (2001) tarafından geliştirilen *YBTS'nin* meydana getirilmesinde kullanılan başlıca rutinler, metotlar ve eşitlikler gözden geçirilmiştir. *YBTS'nin* başlıca beş aşaması söz konusudur. İlk aşama veri normalizasyonu aşamasıdır. İkinci aşama antikorların oluşumu ve bellek hücresi belirlemedir. Üçüncü aşama aday bellek hücresinin gelişmesinde kaynaklar için olan rekabettir. Dördüncü aşaması, mevcut bellek hücrelerinin içerisine aday bellek hücresinin sunumudur. Sistemin son aşaması, eğitim sonunda oluşan bellek hücrelerini kullanarak test kümesini *k en yakın komşu* yöntemiyle sınıflandırmaktır (Şekil 5.1). Tüm bu rutinler aşağıdaki alt başlıklarda detaylı bir şekilde anlatılmaktadır.



Şekil 5.1 YBTS'de kullanılan temel rutinler

5.2.1. YBTS için normalizasyon

Sistemin ilk aşaması aslında bir veri ön işleme aşaması olarak düşünülebilir. Bu aşamada öncelikle veri kümesindeki herhangi iki veri arasındaki bütün özellikler arasındaki Eşitlik 3.1’de belirtilen *Öklid Uzaklığı* $[0,1]$ aralığında olacak şekilde *normalize* edilmektedir. *Öklid Uzaklık kriteri*, hem *duyarlılık* hem de *uyarmayı* hesaplarken başlıca kriterdir. Normalizasyondan sonra *duyarlılık eşiği* hesaplanmaktadır. YBTS’de bellek hücresi yenileme için eşiğe karar veren ölçüt eğitim hücrelerinin tamamı arasındaki ortalama duyarlılık değerini temel alınmaktadır. *Duyarlılık_Eşiği* Eşitlik 5.1’de açıklandığı gibi hesaplanmaktadır.

$$Duyarlılık_Eşiği = \frac{\sum_{i=1}^{n-1} \sum_{j=i+1}^n Duyarlılık(ag_i, ag_j)}{\frac{n \cdot (n-1)}{2}} \quad (5.1)$$

Burada n, eğitim antijenlerinin sayısıdır, ag_i ve ag_j sırasıyla i. ve j. eğitim antijenleridir, $Duyarlılık(ag_i, ag_j)$ özellik vektöründeki iki antijen arasındaki *Öklid Uzaklığıdır*. *Duyarlılık Eşiği* bellek hücresinin değerlendirilmesinde kullanılmaktadır.

Başlangıç aşamasında son adım bellek hücrelerinin ve başlangıç antikor popülasyonunun oluşturulmasıdır. Bu rasgele olarak bir veya daha fazla antijenin bellek hücrelerine ve antikorların kümesine eklenmesiyle yapılmaktadır. Oluşturulan bu başlangıç hücresine *çekirdek hücre* adı verilmektedir.

5.2.2. YBTS için antikor oluşumu ve bellek hücresi belirleme

YBTS’de bu aşamadan itibaren eğitim döngüsel olarak devam etmektedir. Bu aşamada ilk adım bellek hücresinin belirlenmesi ve antikorların oluşumudur. ag eğitim antijenine tepki olarak Eşitlik 5.2’deki gibi hesaplanan başlangıç bellek hücresi mc_{ilk} oluşmaktadır. Bu mc_{ilk} bellek hücresi mevcut antijene ilk tepkide

oluşmaktadır. Bu bellek hücresi, mevcut antikorlar içerisinde ag eğitim antijenine en yüksek uyarım gösteren antikordur. Eşitlik 5.3'te iki özellik vektörü arasındaki uyarma değeri hesabı gösterilmektedir.

$$mc_{ilk} = \operatorname{argmax}_{mc \in MC_{ag,c}} uyarma(ag, mc) \quad (5.2)$$

$$uyarma(x, y) = 1 - \text{Oklid_Uzaklığı}(x, y) \quad (5.3)$$

Eğer antijen ile aynı sınıftaki bellek hücrelerinin kümesi boş ise o antijen bellek hücresinin kümesine eklenmektedir.

Bir kez mc_{ilk} tanımlandınca, bu bellek hücresi, önceden mevcut antikorların popülasyonuna (önceki antijenlere maruz kalan sistemdeki antikorlar) yeni antikorlar eklemek için kullanılmaktadır. Bu işlem Şekil 5.2'de gösterilen metotla gerçekleştirilmektedir. Burada *Antikor_Oluştur*(x) fonksiyonu x vektörünün sınıfında olan ve en yüksek uyarımı veren antikoru döndürmektedir. x vektörünün sınıfına ait antikör yoksa x vektörü antikör olarak döndürülmektedir. Burada seçilen bellek hücresi mc_{ilk} , antijene olan uyarım değeri ile doğru orantılı olarak klonlanmakta ve klonlanan bu antikorlar mutasyon işlemine maruz kalmaktadırlar.

```

MU ← ∅
MU ← MU ∪ Antikor_Oluştur ( mcilk )
uyarma_değeri ← uyarma ( ag , mcilk )
klon_sayısı ← klonlama_oranı * uyarma_değeri
while ( |MU| < klon_sayısı )
do
    mut ← false
    mcklon ← mcilk
    mcklon ← mutasyon ( mcklon, mut )
    if (mut = true)
    {
        MU ← MU ∪ Antikor_Oluştur ( mcklon )
    }
done
AB ← AB ∪ MU

```

Şekil 5.2 YBTS'de antikör oluşumu için mutasyon

$Mutasyon(x,b)$ Şekil 5.3'te tanımlanmıştır. Şekil 5.3'teki $drandom()$ fonksiyonu $[0,1]$ aralığında rasgele bir değer döndürmektedir. YBTS'de antikorlara mutasyon işlemi sırasında rasgele bir değer atayan mutasyon mekanizması kullanılmaktadır.

```

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

```

Şekil 5.3 YBTS'de mutasyon rutini

5.2.3. YBTS için bir aday bellek hücresinin gelişimi ve kaynaklar için rekabet

Bu noktada antikor kümesinin içerisinde mc_{ilk} , mc_{ilk} 'ten mutasyonla çoğaltılmış antikorlar ve önceki karşılaşılan antijenlere olan tepkiden oluşan kalan antikorlar bulunmaktadır. Sistem ayrı aşamalara bölünmüştür ve her bir antijen kesin olarak en az bir kez bu sürece katılmaktadır. Sistemin bir sonraki parçasının amacı bir aday bellek hücresi geliştirmektir. Bu hücre belirli bir ag antijenini doğru olarak sınıflandıran en başarılı hücredir ve başlıca üç mekanizma ile gerçekleştirilir. Birinci mekanizma sistemin geniş kaynakları için olan rekabettir. Antikorların ag antijenine olan uyarma değeri *normalize* edilerek kaynaklar antikorlara tahsis edilir. İkinci

mekanizma çeşitlilik ve şekil-uzayını araştırmak için mutasyonun kullanılmasıdır. Üçüncü mekanizma *ag* antijeni için eğitimi durdurma zamanına karar vermek için bir kriter olarak *ortalama uyarma eşiğinin* kullanılmasıdır.

Genetik algoritmaların prensiplerine benzer olarak geliştirilen sistem antikor popülasyonu içerisinde bireylerin hayatta kalması için uygunluk kavramını kullanmaktadır. Bir antikorum hayatta kalmasına birbiri ile bağlantılı olarak iki aşamada karar verilir. Önce *AB* popülasyonunda her bir antikor, *uyarma* seviyesine karar vermek için *ag* antijeni ile etkileşime girer. Bu *uyarma* hem ham olarak uyarma seviyesini hem de antikorum sınıfını (*ab.c*) temel alarak antikor popülasyonunda normalize edilir. Bu normalize edilmiş *uyarma* değerini temel alan her bir $ab \in AB$ olan antikor kaynakların sınırlı sayısında yerleştirilir. Eğer kaynakların bu yerleşimi popülasyon için izin verilen kaynaklardan fazla olarak sonuçlanırsa, sistemdeki kaynakların sayısı izin verilen kaynak sayısına ulaşınca kadar kaynaklardan en zayıf antikorlar (en az uyarılan) silinir. Sıfır kaynağa sahip antikorlar antikor popülasyonundan silinir. Bu süreç Şekil 5.4'te gösterilmiştir.

Kaynak tahsisi rutinin iki önemli bakış açısı söz konusudur. Birincisi bir antikorum *uyarma* değerine sadece Eşitlik 5.4'teki uyarma fonksiyonu tarafından karar verilmez aynı zamanda antikorum sınıfı temel alınarak karar verilir. Şekil 5.4'te gösterilen *uyarma* hesaplama metodu hem *ag* antijeni ile aynı sınıfta olan ve yüksek derecede uyarılan antikorlar, hem de *ag* antijeninden farklı sınıfta olan ve güçlü reaksiyon sergilemeyen antikorlar ile takviye edilir. İkincisi kaynakların dağılımının antikorların sınıfını temel almasıdır. *YBTS'de* kaynaklar antijenle aynı sınıftaki antikorlar için en fazla kullanıcı tarafından belirlenmiş olan toplam maksimum kaynağın yarısı olacak şekilde tahsis edilmektedir ve kalan yarısı diğer sınıflara ait antikorlar arasında eşit olarak paylaşılmaktadır.

AB kümesi sadece *ag* antijeni tarafından en çok uyarılan antikorları içermektedir veya daha spesifik olarak *AB*, kaynaklar için başarılı olarak rekabet edebilen antikorlar içermektedir. Bu noktada *AB* kümesindeki antikorlar *ag* antijeninin eğitimi durdurmak için yeterince uyarılıp uyarılmadığı kontrol edilerek devam edilmektedir. Bu kontrol *AB* kümesinin her bir alt sınıfı için ortalama uyarma değerini içeren *nc* boyutlu bir *s* vektörü tanımlanarak yapılmaktadır (Eşitlik 5.4).

```

minUyarma ← 1.0
maxUyarma ← 0.0
foreach(ab ∈ AB)
do
    uyarma_değeri ← uyarma(ag, ab)
    if (stim < minStim)
        minUyarma ← uyarma_değeri
    if (stim > maxUyarma)
        maxUyarma ← uyarma_değeri
    ab.uyarım ← uyarma_değeri
done
foreach(ab ∈ AB)
do
    if (ab.c = ag.c)
        ab.uyarım ←  $\frac{ab.uyarım - minUyarma}{maxUyarma - minUyarma}$ 
    else
        ab.uyarım ←  $1 - \frac{ab.uyarım - minUyarma}{maxUyarma - minUyarma}$ 
    ab.kaynak ← ab.uyarım * klonlama_oranı
done
i ← 1
while(i ≤ nc)
do
    ayrılan_kaynak ←  $\frac{|AB_i|}{\sum_{j=1}^n ab_j.kaynak}$ ,  $ab_j \in AB_i$ 
    if (i = ag.c)
        izin_verilen_kaynak ←  $\frac{kaynak\_sayısı}{2}$ 
    else
        izin_verilen_kaynak ←  $\frac{kaynak\_sayısı}{2 * (nc - 1)}$ 
    while(ayrılan_kaynak > izin_verilen_kaynak)
    do
        silinecek_kaynak ← ayrılan_kaynak - izin_verilen_kaynak
         $ab_{sil} \leftarrow \underset{ab \in AB_i}{\text{argmin}} (ab.uyarım)$ 
        if ( $ab_{sil}.kaynak \leq silinecek\_kaynak$ )
             $AB_i \leftarrow AB_i - ab_{sil}$ 
            ayrılan_kaynak ← ayrılan_kaynak -  $ab_{sil}.kaynak$ 
        else
             $ab_{sil}.kaynak \leftarrow ab_{sil}.kaynak - silinecek\_kaynak$ 
            ayrılan_kaynak ← ayrılan_kaynak - silinecek_kaynak
        done
        i = i + 1
    done
done

```

Şekil 5.4 YBTS’de uyarma, kaynak tahsisi ve antikor çıkarılması

$$s_i \leftarrow \frac{\sum_{j=1}^{|AB_i|} ab_j \cdot uyarım}{|AB_i|}, ab_j \in AB_i \quad (5.4)$$

burada her bir sınıfa ait uyarma seviyeleri hesaplanmaktadır ($i=1,2,\dots,nc$). Durma kriteri $s=\{s_1,s_2,\dots,s_{nc}\}$ deki bütün elemanlar için Eşitlik 5.1’de tanımlanan *Duyarlılık Eşiğinin* geçilmesidir.

Durma kriterine ulaşıp ulaşılmadığına bakmaksızın algoritma mutasyona uğrayarak oluşturulmuş ürünlerle oluşturulan AB ’deki her bir antikör hesaba katılarak ilerlemektedir. Burada Şekil 5.2’de gösterilen metot ile benzer mutasyona uğramış ürünler oluşturulurken Şekil 5.2’deki metottan birkaç farklılık söz konusudur. Bu değişime uğramış mutasyon oluşturma rutini Şekil 5.5’te sunulmuştur.

```

MU ← ∅
Foreach ( ab ∈ AB )
do
    rd ← drandom()
    if ( ab.uyarım > rd )
    {
        klon_sayısı ← ab.uyarım * klonlama_oranı
        i ← 1
        while ( i ≤ klon_sayısı )
        do
            mut ← false
            ab_klon ← ab
            ab_klon ← mutasyon ( ab_klon , mut )
            if ( mut = true )
                MU ← MU ∪ ab_klon
            i ← i + 1
        done
    }
done

```

Şekil 5.5 YBTS’de hayatta kalan antikörlerin mutasyonu

Hayatta kalan her bir antikörden mutasyona uğramış ürünler üretildikten sonra durma kriteri kontrol edilmektedir. Eğer durma kriteri ile karşılaşılmışsa bu antijen için eğitim durmaktadır. Eğer durma kriteri sağlanamamışsa bu sürecin

tamamı Şekil 5.4'te gösterilen metottan başlayarak durma kriteri sağlanan kadar tekrar edilmektedir. Bu tekrar işleminde her tekrarda tek istisna *uyarma* ve *kaynak tahsisi* aşamasından sonra eğer durma kriteri ile karşılaşırsa mutasyona uğramış antikör üretimi uygulanmamaktadır. Çünkü durma kriterine ulaşılmıştır. Durma kriteri ile bir kez karşılaşıncı *aday bellek hücresi* seçilmektedir. *Aday bellek hücresi* mc_{aday} , *ag* eğitim antijenin sınıfı ile aynı sınıftaki en fazla uyarılan antikör olan mutasyondan önce sistemde bulunan antikörün sınıfına sahiptir ve özellik vektörüdür.

5.2.4. YBTS için bellek hücresinin değerlendirilmesi

Eğitim rutininde son aşamada mevcut *MC* bellek hücreleri kümesi içerisindeki en çok geliştirilmiş aday bellek hücresi mc_{aday} belirlenmekte ve daha sonra antijen için eğitimin başında belirlenmiş olan mc_{ilk} ile mc_{aday} 'ın yer değiştirilip değiştirilmeyeceğine başlangıçta hesaplanan *Duyarlılık Eşiği* ile karar verilmektedir. mc_{aday} *aday bellek hücresi*, *ag* eğitim antijeni tarafından mc_{ilk} 'ten daha fazla uyarılırsa ve eğer mc_{aday} ve mc_{ilk} arasındaki duyarlılık değeri, duyarlılık eşiği ve kullanıcı tarafından belirtilen duyarlılık eşik ölçüsünün (*Affinity Threshold Scalar-ATS*) çarpımından daha az ise mc_{ilk} bellek hücresi kümesinden çıkartılarak yerine mc_{aday} bellek hücresi kümesine eklenmektedir. Bir diğer ihtimal olarak mc_{aday} *aday bellek hücresi*, *ag* eğitim antijeni tarafından mc_{ilk} 'ten daha fazla uyarılırsa ve eğer mc_{aday} ve mc_{ilk} arasındaki duyarlılık, duyarlılık eşiği ve kullanıcı tarafından belirtilen duyarlılık eşik ölçüsünün çarpımından daha büyükse mc_{ilk} bellek hücresi kümesinde muhafaza edilmekte ve mc_{aday} bellek hücresi kümesine eklenmektedir. Son bir ihtimal olarak *Aday bellek hücresi*, *ag* eğitim antijeni tarafından mc_{ilk} 'ten daha az uyarılırsa mc_{aday} bellek hücresine eklenmemektedir. Buradaki *uyarma işlemi* Eşitlik 5.3'te tanımlanmaktadır. Bu süreç Şekil 5.6'da gösterilmektedir.

Aday bellek hücresi, mevcut bellek hücreleri içerisinde değerlendirilince bu antijen için eğitim tamamlanmaktadır. Bu süreç, bütün antijenler sisteme sunulana kadar devam etmektedir.


```

AdayUyarı ← uyarma ( ag , mcaday )
İlkUyarı ← uyarma ( ag , mcilk )
HücrelerArasıDuyarlılık ← duyarlılık ( mcaday , mcilk )
if ( AdayUyarı > İlkUyarı )
{
    if ( HücrelerArasıDuyarlılık < Duyarlılık_Eşiği * ATS )
    {
        MC ← MC - mcilk
    }
    MC ← MC ∪ mcaday
}

```

Şekil 5.6 YBTS’de bellek hücresi seçimi

5.2.5. YBTS için sınıflandırma

Eğitim aşaması tamamlandıktan sonra geliştirilen bellek hücreleri sınıflandırmada kullanılmak için uygun duruma gelmektedir. Sistemin son aşamasında *k en yakın komşu* yöntemi uygulanmaktadır. Test verilerinin hangi sınıfta olduğu bellek hücreleri kümesi üzerine *k en yakın komşu* yönteminin uygulanması ile belirlenmektedir. Bir verinin sistem tarafından sınıflandırılmasına test verisini en fazla uyarı *k* adet bellek hücrelerinin çıkışlarının bir *çoğunluk oylaması* kullanılarak karar verilmektedir.

5.3. YBTS’deki Bazı Eksiklikler

Geliştirilen sistemden bahsetmeden önce *YBTS’de* tespit edilen bazı eksiklikler bu bölümde özetlenmektedir. Tez çalışmasında bu eksiklikler giderilerek diğer *YBS* algoritmaları için faydalı olabilecek yaklaşımlar önerilmektedir.

YBTS şekil uzayında uzaklık kriteri olarak Eşitlik 3.1’deki *Öklid uzaklık kriterini* kullanmaktadır. Bu eşitliğe göre verinin bütün özellikleri uzaklık hesabında aynı etkiye sahiptir. Veri kümesi üzerinde etkisi çok az olan özelliklerle veri kümesi üzerinde çok önemli özellikleri eşit ağırlıkta saymaktadır. Özellikle verilerin

boyutunun büyük olduğunda bu durum sistemin performansı olumsuz yönde etkilemektedir. Bu durum literatürde “*curse of dimensionality (boyutluluk problemi)*” olarak adlandırılmaktadır (Bellman 1961; Mitchell 1997; Beyer ve ark. 1999; Şahan ve ark. 2004; Şahan ve ark. 2005). Verinin bir özelliği, aynı sınıfa ait iki verinin şekil uzayında birbirlerinden uzak konumlarda yer almasına sebep olabilmektedir. Bu durumda iki verinin farklı sınıflara ait sistem birimleri tarafından sınıflandırılması söz konusu olmaktadır. Örneğin üç adet özelliğe sahip bir veri kümesinde aynı sınıfa ait iki veri için ilk özellik değerleri farklı, diğer iki özellik değerleri aynı olsun. Bu durumda bu iki veri üç boyutlu şekil uzayında birbirlerinden çok farklı noktalarda yerleşmiş gibi görülebilmektedir. Uzaklık tabanlı sınıflandırma sistemlerinde bu iki veri, farklı sınıflara veriler olarak sınıflandırılabilir. Şekil uzayında sınıf dağılımlarının doğrusal olarak ayrılabilirdiği durumlarda *YBTS*’de kullanılan bu uzaklık kriterinin kullanımı sorun oluşturmamaktadır. Fakat gerçek dünya problemlerinin çoğunda doğrusal olmayan bir yapı söz konusudur. *YBTS*’deki gibi tüm özelliklerin eşit ölçüde kullanıldığı bir uzaklık hesabı, yanlış sınıflamaların çoğunlukta olmasına ve sistem performansının hissedilir ölçüde düşmesine yol açacaktır. Bu tarz verilerde her bir özelliğin sınıflama üzerindeki etkisi bulunmalı ve bu özelliklere bu etkiye dayalı olarak bir ağırlık verilmelidir. Bu amaçla *BK-YBTS*’de veri özelliklerinin bilgi kazancı değerleri hesaplandıktan sonra bu değerler uzaklık hesabında ağırlık olarak kullanılarak, bu problemin üstesinden gelmeye çalışılmaktadır.

YBTS’de bellek hücresi yenilenirken eşik değeri olarak eğitim antijenlerinin tamamı arasındaki ortalama duyarlılık değeri kullanılmaktadır. Geliştirilen sistemde bellek hücresi yenilerken eşik değeri olarak aynı sınıfa sahip eğitim antijenlerinin ortalama duyarlılık değeri temel alınmaktadır.

YBTS’de antikorlar mutasyon işlemine maruz kalırken antikor veya bellek hücresinin değeri yerine rasgele bir değer atanmaktadır. Bunun yerine mutasyona uğrayacak antikoru veya bellek hücresinin duyarlılığını temel alan bir mutasyon mekanizması kullanılmaktadır. Yani duyarlılığı yüksek olan antikorlar düşük oranda, duyarlılığı düşük olan antikorlar yüksek oranda mutasyon işlemine maruz kalmaktadır. Bu durum yüksek duyarlılıklı hücrelerin hemen çevresindeki arama alanının araştırılmasına izin vermektedir. Diğer taraftan daha düşük duyarlılıklı

hücreler için daha yüksek mutasyon oranı, daha az verimli alanlardan uzaklaşmayı kolaylaştırmak için arama uzayında daha büyük atlamalara izin vermektedir.

Kaynak tahsisi rutininde *YBTS'de* kaynaklar antijenle aynı sınıftaki antikorlar için en fazla kullanıcı tarafından belirlenmiş olan toplam maksimum kaynağın yarısı olacak şekilde tahsis edilmektedir ve kalan yarısı diğer sınıflara ait antikorlar arasında eşit olarak paylaşılmaktadır. Bunun yerine kaynakların sınıf dağılımına göre yapılması daha faydalı olacağı öngörülmektedir.

YBTS'de tek bir eğitim antijeni için, Eşitlik 5.1'de hesaplanan s kümesindeki bütün değerlerin, eğitim antijenlerinin tamamı arasındaki ortalama duyarlılık değerinden büyük olması durumunda eğitim durmaktadır. Geliştirilen sistemde tek bir eğitim antijeni için eğitimi durdurma kriteri, s kümesindeki bütün değerlerin, eğitim antijeni ile aynı sınıfa sahip tüm eğitim antijenlerin kendi aralarındaki ortalama duyarlılık değerinden büyük olması olarak değiştirilmiştir.

Tek bir antijen için eğitim tamamlandıktan sonra *YBTS'de* aday bellek hücresi mc_{aday} belirlenmektedir ve bu aşama boyunca önceden belirlenmiş olan mc_{ilk} ile eğitim sonucu oluşan mc_{aday} hücresinin yer değiştirilip değiştirilmeyeceğine Eşitlik 5.11'de başlangıçta hesaplanan *Sınıf Tabanlı Duyarlılık Eşiği* ile karar verilmektedir. Bu süreç Şekil 5.16'da gösterilmiştir.

Bütün antijenler eğitildikten sonra geliştirilen bellek hücreleri sınıflandırmada kullanılmak için uygun duruma gelmektedir. *YBTS'nin* son aşamasında k en yakın komşu yöntemi uygulanmaktadır. Bir verinin sistem tarafından sınıflandırılmasına test verisini en fazla uyaran k adet bellek hücrelerinin çıkışlarının bir *çoğunluk oylaması* kullanılarak karar verilmektedir. Bu aşamada eşitlik olması durumunda *YBTS* test verisinin sınıf değerini tahmin ederken sınıf değeri küçük olan sınıf değerini atamaktadır. Geliştirilen sistemde çoğunluk oylaması yapılırken eşitlik olması durumunda test verisinin sınıfı olarak, tüm veri kümesinde daha çok verisi bulunan sınıf atanmaktadır.

Bölüm 5.4'te *YBS* algoritmaları için faydalı olabilecek yaklaşımlar içeren *BK-YBTS* açıklanırken tüm bu yaklaşımlar hakkında ayrıntılı bilgi verilmektedir.

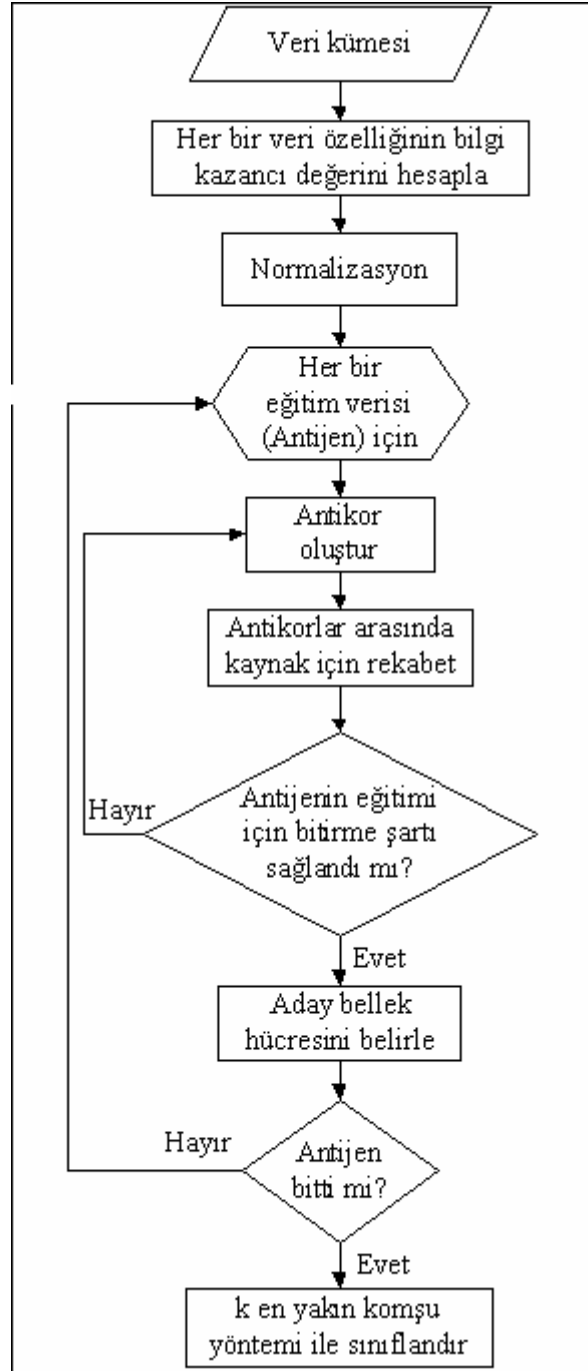
5.4. Bilgi Kazancı Tabanlı Yapay Bağışıklık Tanıma Sistemi (BK-YBTS)

Bu bölümde *YBTS*'yi temel alan ve diğer *YBS* algoritmaları için faydalı olabilecek yaklaşımlar içeren *BK-YBTS* açıklanmaktadır. *BK-YBTS*'de kullanılan bağışıklık sistemi mekanizmaları *kaynak rekabeti*, *mutasyon*, *klonsal seçim*, *duyarlılık olgunlaşması* ve *bellek hücreleri oluşumudur*. Bölüm 2'de bu mekanizmaların biyolojik olarak bağışıklık sistemindeki yeri hakkında bilgiler verilmektedir. Aynı şekilde Bölüm 3'te ise bu mekanizmaların *YBS*'de kullanımı hakkında bilgiler verilmektedir.

BK-YBTS beş aşamadan oluşmaktadır. İlk aşama uzaklık hesabında kullanılmak üzere veri özelliklerinin bilgi kazançları değerlerinin hesaplanması aşamasıdır. İkinci aşama veri normalizasyonu aşamasıdır. Üçüncü aşama antikorların oluşumu ve bellek hücresi belirlemedir. Dördüncü aşama aday bellek hücresinin gelişmesinde kaynaklar için olan rekabettir. Sistemin son aşaması, mevcut bellek hücrelerinin içerisine aday bellek hücresinin sunumudur. Şekil 5.7'de *BK-YBTS*'nin temel işleyiş adımları gösterilmektedir.

Temel olarak geliştirilen sistemde, antijen olarak kabul edilen eğitim verilerini tanıyan bellek hücreleri oluşturulmaktadır. Antijene (eğitim verisi) uygun tepki veren antikorlar arasından bellek hücreleri seçilmektedir. Bu seçim sürecinde antikorlara kaynak tahsis edilirken antijene yüksek uyarım veren antikorlara daha büyük kaynak ayrılmaktadır. Bu mekanizma ile antikorlar arasında rekabet oluşturulmaktadır. Güçlü olan (kaynak sayısı yüksek) antikorlar hayatta kalmakta zayıf olan (kaynak sayısı düşük) antikorlar sistemden silinmektedir. Bu sürecin sonunda eğitim antijenine en güçlü uyarımı veren antikor, bellek hücresi olarak belirlenmektedir.

Her bir antijen için yukarıda bahsedilen süreçler tamamlandıktan sonra sistemin eğitim aşaması sona ermektedir. Bu aşamadan sonra test verileri, eğitim sonucunda oluşan bellek hücreleri üzerinde *k en yakın komşu* yöntemiyle sınıflandırılmaktadır. Alt bölümlerde *BK-YBTS*'nin çalışma aşamaları ayrıntılı olarak açıklanmaktadır.



Şekil 5.7 BK-YBTS'nin temel işleyiş adımları

5.4.1. BK-YBTS için veri özelliklerinin bilgi kazancı değerlerinin hesaplanması

Şekil-uzay gösteriminin kullanıldığı çoğu YBS algoritmaları Eşitlik 3.1'deki *Öklid uzaklık kriterini* kullanmaktadır. Bu eşitliğe göre verinin bütün özellikleri

uzaklık hesabında aynı etkiye sahiptir. Veri kümesi üzerinde etkisi çok az olan özelliklerle veri kümesi üzerinde çok önemli özellikleri eşit ağırlıkta saymaktadır. Özellikle verilerin boyutunun büyük olduğunda bu durum sistemin sınıflandırma performansını olumsuz yönde etkilemektedir. Bu durum literatürde “*curse of dimensionality (boyutluluk problemi)*” olarak adlandırılmaktadır (Bellman 1961; Mitchell 1997; Beyer ve ark. 1999; Şahan ve ark. 2004; 2005). Bu problemin üstesinden gelebilmek için veri kümesinde daha önemli olan özelliklere uzaklık hesabında daha yüksek ağırlık vermek ön görülmüştür. Bu amaçla geliştirilmiş olan *BK-YBTS* bu problemin etkisini minimuma indirmek için özelliklerin bilgi kazancını temel almaktadır. Özelliklerin önem derecesi bilgi kazancı hesabıyla belirlenmektedir. Bu açıdan bakılırsa önerilen bilgi kazancı tabanlı özellik ağırlıklandırma yaklaşımı şu temellere bağlıdır: eğer bir özelliğin bilgi kazancı diğer özelliklerin bilgi kazancından yüksekse bu özelliğe diğer özelliklerden daha çok ağırlık verilmelidir.

BK-YBTS’nin özellik ağırlıklandırma işleminde öncelikle verinin özellikleri her bir özelliğin bilgi kazancı değeri hesaplanması için sürekli değerli veri özellikleri ayrık forma dönüştürülmektedir. Dönüştürme işleminde Bölüm 4.2.1.1’de açıklanan eşit aralık genişliği ayırım yaklaşımı kullanılmıştır. Bu yaklaşımda öncelikle sürekli değerli her bir özellik için maksimum ve minimum değerler belirlenmektedir. Eşitlik 5.5’e göre veri aralıkları hesaplanmaktadır. Daha sonra Eşitlik 5.6’daki formüle göre ayrık değerli veriler elde edilmektedir. Bu ayrık değerli veriler bilgi kazancı değerlerinin hesaplanmasında kullanılmaktadır.

$$aralık(j) = \frac{\max(veri(:, j)) - \min(veri(:, j))}{parça} \quad (5.5)$$

$$ayrik_veri(i, j) = round \left[\frac{veri(i, j)}{aralık(j)} \right] \quad (5.6)$$

burada *parça* aralıkların sayısıdır, *j* verinin özellik numarasıdır ($j=1, 2, \dots, L$; *L* özelliklerin boyutudur), *veri(:, j)* veri kümesindeki *j*. özellik değerlerinin kümesidir,

$veri(i,j)$ i . verinin j . özelliğinin değeridir. Ayrım süreci tamamlandıktan sonra, özelliklerin bilgi kazancı değerlerinin hesaplanmasına geçilmektedir.

Bilgi kazancı hesaplanırken bilgi teorisinde kullanılan *Entropi* (*Entropy*) olarak adlandırılan bir ölçü kullanılmaktadır (Quinlan 1986; Blahut 1987; Mitchell 1997). Eşitlik 5.7’de belirtilmiş olan bu ölçü verideki rasgeleliği, belirsizliği, kirliliği beklenmeyen durumun ortaya çıkma olasılığını karakterize etmektedir. Sınıflandırmada örneklerin tümü aynı sınıfta ise *Entropi* değeri sıfırdır. Yani belirsizlik yoktur. Örnekler sınıflar arasında eşit dağılmışsa *Entropi* değeri birdir. Yani belirsizlik maksimum seviyededir. Örnekler sınıflar arasında rasgele dağılmışsa *Entropi* değeri sıfır ile bir arasındadır.

$$Entropi(S) \equiv \sum_{i=1}^c -p_i \log_2 p_i \quad (5.7)$$

burada S veri kümesinin tamamıdır, p_i i . sınıfa ait verilerin sayısının S kümesindeki verilerin sayısına oranıdır, c sınıf sayısıdır.

Verideki kirliliğin bir ölçüsü olarak *Entropi* aynı zamanda veriyi sınıflandırmada bir özelliğin etkinliğinin ölçüsünü hesaplamada kullanılmaktadır. Bu ölçüye *bilgi kazancı* adı verilmektedir (Eşitlik 5.8).

$$Kazanç(S,A) \equiv Entropi(S) - \sum_{v \in Değerler(A)} \frac{|S_v|}{S} Entropi(S_v) \quad (5.8)$$

burada $Değerler(A)$, S veri kümesindeki A özelliğinin alabileceği bütün değerlerinin kümesidir; S_v , S veri kümesindeki A özelliğinin v değerine sahip elemanlarından oluşan kümedir. ($S_v = \{s \in S \mid A(s) = v\}$). Hesaplanan bu kazanç değerleri daha sonra Eşitlik 5.9’daki gibi normalize edilmektedir.

$$NormKazanç_i = \left(\frac{Kazanç_i}{\max(Kazanç)} \right), i = (1, 2, \dots, L) \quad (5.9)$$

burada $Kazanç_i$, veri kümesindeki i . özelliğin kazanç değeridir; $\max(Kazanç)$, özelliklere ait kazanç değerleri arasında maksimum kazanç değeridir; L , özellik sayısıdır. Bu süreçlerden sonra *BK-YBTS*'de kullanılacak uzaklık kriteri Eşitlik 5.10'daki gibi olmaktadır.

$$bilgi\ kazancı\ tabanlı\ uzaklık(ab, ag) = \sqrt{\sum_{i=1}^L NormKazanç_i \cdot (ab_i - ag_i)^2} \quad (5.10)$$

5.4.2. BK-YBTS için normalizasyon

Bu aşamada *YBTS*'den farklı olarak *Öklid Uzaklık kriteri* yerine Eşitlik 5.10'da belirtilen *bilgi kazancı tabanlı uzaklık* kullanılmaktadır. Veri kümesindeki herhangi iki veri arasındaki bütün özellikler arasındaki *bilgi kazancı tabanlı uzaklık* $[0,1]$ aralığında olacak şekilde *normalize* edilmektedir. *Bilgi kazancı tabanlı uzaklık* hem *duyarlılık* hem de *uyarmayı* hesaplarken başlıca ölçek olarak kullanılmaktadır. Normalizasyondan sonra *duyarlılık eşiği* hesaplanmaktadır.

BK-YBTS'de bellek hücresi yenileme için eşik değerine karar verilirken *YBTS*'de olduğu gibi eğitim antijenlerinin tamamı arasındaki ortalama duyarlılık değeri yerine aynı sınıfa sahip eğitim antijenlerinin ortalama duyarlılığı temel alınmaktadır. *Sınıf Tabanlı Duyarlılık Eşiği*, Eşitlik 5.11'de açıklandığı gibi hesaplanmaktadır.

$$Sınıf\ Tabanlı\ Duyarlılık\ Eşiği_i = \frac{\sum_{j=1}^{|sınıf_i|-1} \sum_{k=j+1}^{|sınıf_i|} duyarlılık(ag_j, ag_k)}{\frac{|sınıf_i| \cdot (|sınıf_i| - 1)}{2}} \quad (5.11)$$

Burada $|sınıf_i|$, i . sınıfa ait eğitim antijenlerinin sayısıdır; ag_j ve ag_k , sırasıyla i . sınıfa ait j . ve k . eğitim antijenleridir ($ag_{j_c} \in sınıf_i$, $ag_{k_c} \in sınıf_i$) ve $i=\{1,2,...,nc\}$ 'dir. $duyarlılık(ag_j, ag_k)$, özellik vektöründeki iki antijen arasındaki *bilgi*

kazancı tabanlı uzaklıktır. Sınıf Tabanlı Duyarlılık Eşiği bölüm 5.4.5'te bellek hücresi değerlendirilmesinde kullanılmaktadır.

Başlangıç aşamasında son adım bellek hücrelerinin ve başlangıç antikor popülasyonunun oluşturulmasıdır. Bu rasgele olarak bir veya daha fazla antijenin bellek hücrelerine ve antikorların kümesine eklenmesiyle yapılmaktadır. Oluşturulan bu başlangıç hücresine *çekirdek hücre* adı verilmektedir.

5.4.3. BK-YBTS için antikor oluşumu ve bellek hücresi belirleme

Bu aşamadan itibaren eğitim *YBTS'de* olduğu gibi döngüsel olarak devam etmektedir. Bu aşamada ilk adım bellek hücresinin belirlenmesi ve antikor oluşumudur. *ag* eğitim antijenine tepki olarak Eşitlik 5.12'deki gibi hesaplanan başlangıç bellek hücresi mc_{ilk} oluşmaktadır. Bu mc_{ilk} bellek hücresi mevcut antijene ilk tepkide oluşmaktadır. Eşitlik 5.13'te iki özellik vektörü arasındaki uyarma değeri hesaplanmaktadır.

$$mc_{ilk} = \underset{mc \in MC_{ag,c}}{\operatorname{argmax}} bilgi_kazançlı_uyarma(ag, mc) \quad (5.12)$$

$$bilgi_kazançlı_uyarma(x, y) = 1 - bilgi_kazancı_tabanlı_uzaklık(x, y) \quad (5.13)$$

Eğer antijen ile aynı sınıftaki bellek hücrelerinin kümesi boş ise antijen bellek hücresinin kümesine eklenmektedir. Bir kez mc_{ilk} tanımlandıncaya, bu bellek hücresi, önceden mevcut antikorların popülasyonuna (önceki antijenlere maruz kalan sistemdeki antikorlar) yeni antikorlar eklemek için kullanılmaktadır. Antijen ile mc_{ilk} arasındaki bilgi kazançlı uyarma değeri ile doğru orantılı olarak mc_{ilk} klonlanmakta ve oluşan klonlar duyarlılık tabanlı mutasyon işlemine maruz kalmaktadır. Bu işlem Şekil 5.8'de gösterilen metotla gerçekleştirilmektedir.

```

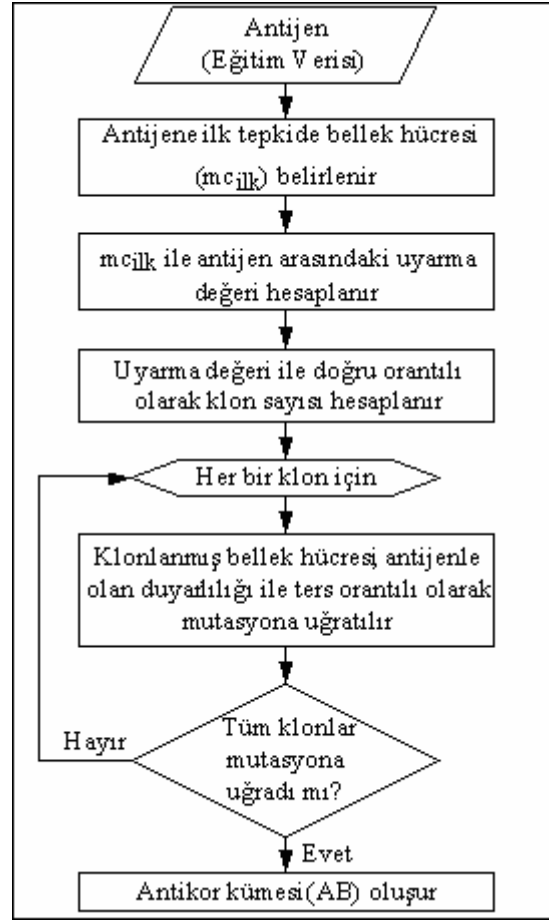
MU ← ∅
MU ← MU ∪ Antikor_Oluştur ( mcilk )
uyarma_değeri ← bilgi_kazançlı_uyarma ( ag , mcilk )
klon_sayısı ← klonlama_oranı * uyarma_değeri
while ( |MU| < klon_sayısı )
do
    mut ← false
    mcklon ← mcilk
    mcklon ← duyarlılık_tabanlı_mutasyon ( mcklon , mut )
    if ( mut = true )
    {
        MU ← MU ∪ Antikor_Oluştur ( mcklon )
    }
done
AB ← AB ∪ MU

```

Şekil 5.8 BK-YBTS’de antikor oluşumu için mutasyon

Antikor_Oluştur(x) fonksiyonu x vektörünün sınıfında olan ve en yüksek uyarıyı veren antikoru döndürmektedir. x vektörünün sınıfına ait antikor yoksa x vektörü antikor olarak döndürülmektedir. Antikor oluşumunun akış diyagramı şekil 5.9’da verilmektedir.

Şekil 5.8’de belirtilen *duyarlılık_tabanlı_mutasyon(x,b)*, Şekil 5.10’da tanımlanmıştır. Şekil 5.10’daki *drandom()* fonksiyonu $[0,1]$ aralığında rasgele bir değer döndürmektedir. Burada performans artımına katkıda bulunmak amacıyla sistemdeki antikorlara mutasyon işlemi sırasında antikor veya bellek hücresinin kalitesini arttırmak için *YBTS’de* olduğu gibi rasgele bir değer atamak yerine mutasyona uğrayacak antikorum veya bellek hücresinin duyarlılığını temel alan bir mutasyon mekanizması kullanılmaktadır. Yani duyarlılığı yüksek olan antikorlar düşük oranda, duyarlılığı düşük olan antikorlar yüksek oranda mutasyon işlemine maruz kalmaktadır. Bu durum yüksek duyarlılıklı hücrelerin hemen çevresindeki arama alanının araştırılmasına izin vermektedir. Diğer taraftan daha düşük duyarlılıklı hücreler için daha yüksek mutasyon oranı, daha az verimli alanlardan uzaklaşmayı kolaylaştırmak için arama uzayında daha büyük atlamalara izin vermektedir. Mutasyon rutininin akış diyagramı şekil 5.11’de verilmektedir.



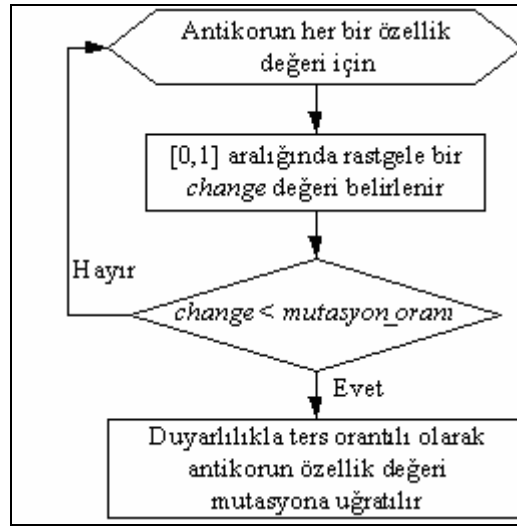
Şekil 5.9 BK-YBTS’de antikor oluşumu için mutasyon mekanizmasının akış diyagramı

```

duyarlılık_tabanlı_mutasyon ( x , b )
{
    foreach ( x.fi in x.f )
    do
        change ← drandom()
        change_to ← drandom()
        if ( change < mutasyon_oranı )
        {
            mutasyon_aralığı = 1 - x.uyarım
            değer ← x.fi ± x.fi * mutasyon_aralığı * change_to
            if ( ( değer > 0 ) and ( değer < 1 ) )
            {
                x.fi ← değer
                b ← true
            }
        }
    }
done
}

```

Şekil 5.10 BK-YBTS’de mutasyon rutini



Şekil 5.11 BK-YBTS’de mutasyon rutininin akış diyagramı

5.4.4. BK-YBTS için aday bellek hücresinin gelişimi ve kaynaklar için rekabet

Bu noktada sistemin amacı bir aday bellek hücresi geliştirmektir. Bu hücre belirli bir *ag* antijenini doğru olarak sınıflandıran en başarılı hücredir ve başlıca üç mekanizma ile ortaya çıkmaktadır. Birinci mekanizma sistemin geniş kaynakları için olan rekabettir. Antikorların *ag* antijenine olan uyarma değerleri *normalize* edilmektedir. Normalize edilmiş olan uyarma değerleri ile doğru orantılı olarak antikorlara kaynaklar tahsis edilmektedir. İkinci mekanizma çeşitlilik ve şekil-uzayını araştırmak için mutasyonun kullanılmasıdır. Üçüncü mekanizma *ag* antijeni için eğitimi durdurma zamanına karar vermek için bir kriter olarak *ortalama uyarma eşiğinin* kullanılmasıdır.

Kaynak tahsisi rutininde *YBTS’de* kaynaklar antijenle aynı sınıftaki antikorlar için en fazla kullanıcı tarafından belirlenmiş olan toplam maksimum kaynağın yarısı olacak şekilde tahsis edilmektedir ve kalan yarısı diğer sınıflara ait antikorlar arasında eşit olarak paylaşılmaktadır. *BK-YBTS’de* ise bu mekanizma veri kümesindeki sınıfların oranı hesaba katılarak değiştirilmektedir. Bu metotla birlikte sistem kaynakları, eğitim kümesinde sınıf dağılımı ile orantılı olarak farklı sınıf antikorları arasında bölünmektedir. Şekil 5.12’de geliştirilen sistem için kaynak

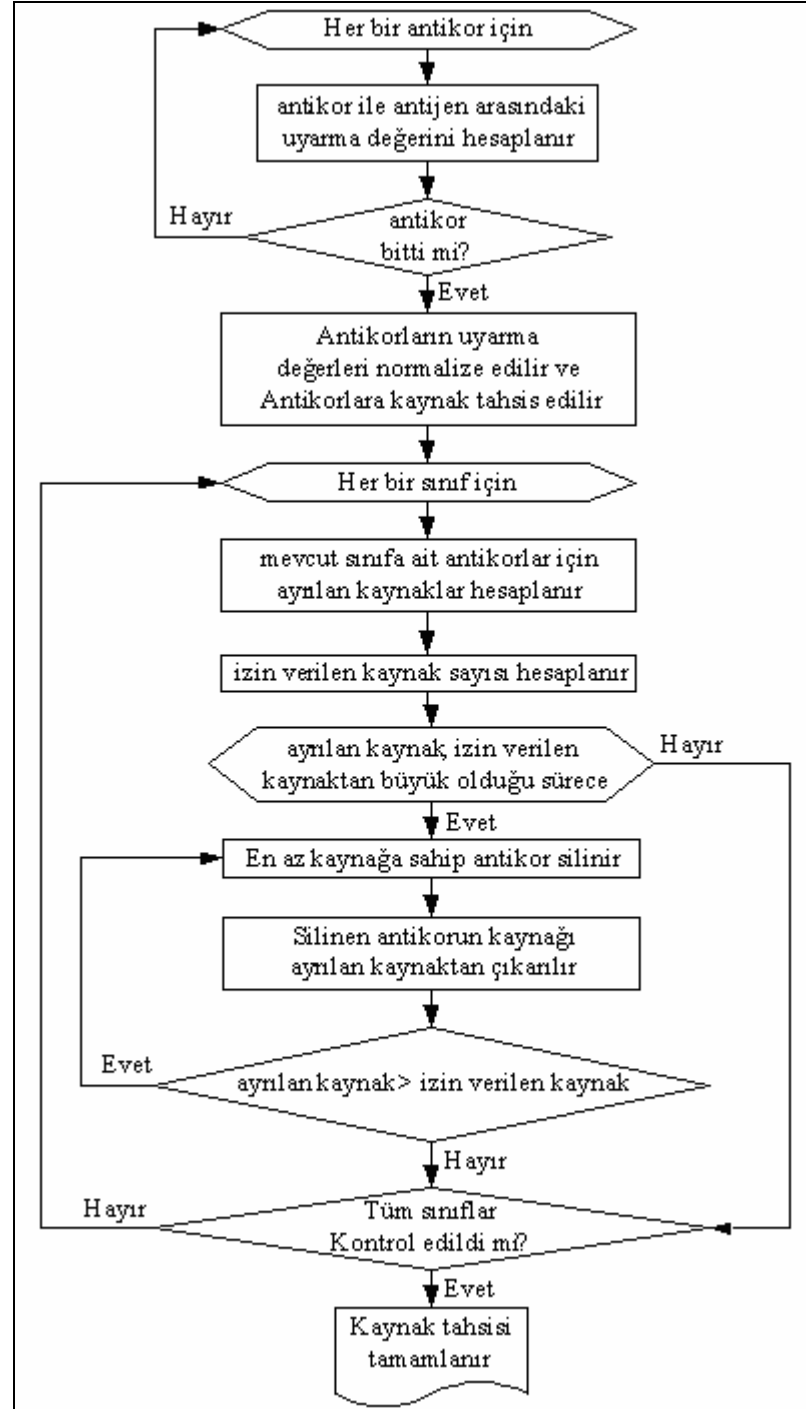
tahsisi yaklaşımı gösterilmektedir. Şekil 5.13'te kaynak tahsisi yaklaşımının akış diyagramı verilmektedir.

```

minUyarma ← 1.0
maxUyarma ← 0.0
foreach(ab ∈ AB)
do
    uyarma_değeri ← uyarma_bilgi_kazançlı(ag, ab)
    if (stim < minStim)
        minUyarma ← uyarma_değeri
    if (stim > maxUyarma)
        maxUyarma ← uyarma_değeri
    ab.uyarım ← uyarma_değeri
done
foreach(ab ∈ AB)
do
    if (ab.c = ag.c)
        ab.uyarım ←  $\frac{ab.uyarım - minUyarma}{maxUyarma - minUyarma}$ 
    else
        ab.uyarım ←  $1 - \frac{ab.uyarım - minUyarma}{maxUyarma - minUyarma}$ 
    ab.kaynak ← ab.uyarım * klonlama_oranı
done
i ← 1
while(i ≤ nc)
do
     $ayrılan\_kaynak ← \sum_{j=1}^{|AB_i|} ab_j.kaynak, ab_j \in AB_i$ 
    if (i = ag.c)
        izin_verilen_kaynak ← kaynak_sayısı *  $\left( \frac{|sınıf(ag_c)|}{veri\_sayısı} \right)$ 
    else
        izin_verilen_kaynak ← kaynak_sayısı *  $\left( \frac{\left( 1 - \frac{|sınıf(ag_c)|}{veri\_sayısı} \right)}{2} \right)$ 
    while(ayrılan_kaynak > izin_verilen_kaynak)
    do
        silinecek_kaynak ← ayrılan_kaynak - izin_verilen_kaynak
         $ab_{sil} \leftarrow \underset{ab \in AB_i}{\text{argmin}} (ab.uyarım)$ 
        if (  $ab_{sil}.kaynak \leq silinecek\_kaynak$  )
             $AB_i \leftarrow AB_i - ab_{sil}$ 
            ayrılan_kaynak ← ayrılan_kaynak -  $ab_{sil}.kaynak$ 
        else
             $ab_{sil}.kaynak \leftarrow ab_{sil}.kaynak - silinecek\_kaynak$ 
            ayrılan_kaynak ← ayrılan_kaynak - silinecek_
        done
        i = i + 1
    done
done

```

Şekil 5.12 BK-YBTS’de uyarma, kaynak tahsisi ve antikor çıkarılması



Şekil 5.13 BK-YBTS’de uyarma, kaynak tahsisi ve antikor çıkarılması sürecinin akış diyagramı

Şekil 5.12’de belirtilen AB antikor kümesi sadece ag antijeni tarafından en çok uyarılan antikorları içermektedir veya daha spesifik olarak AB, kaynaklar için başarılı olarak rekabet edebilen antikorlar içermektedir. Kaynak sayısı düşük olan antikorlar sistemden çıkarılmaktadır. Bu noktada AB kümesindeki antikorlar ag antijeninin eğitimini durdurmak için yeterince uyarılıp uyarılmadığı kontrol edilerek

devam edilmektedir. Bu kontrol AB kümesinin her bir alt sınıfı için ortalama uyarma değerini içeren nc boyutlu bir s vektörü tanımlanarak yapılmaktadır (Eşitlik 5.4).

$YBTS$ 'de bir antijenin eğitimi için durma kriteri, s kümesindeki bütün değerlerin Eşitlik 5.1'de belirtilen eğitim antijenlerinin tamamı arasındaki ortalama duyarlılık değerinden büyük olması durumudur. BK - $YBTS$ 'de ise bir antijenin eğitimi için durma kriteri, s kümesindeki bütün değerlerin Eşitlik 5.11'de belirtilen ve aynı sınıfa sahip eğitim antijenlerinin kendi aralarındaki ortalama duyarlılık değerinden büyük olması durumudur.

Durma kriterine ulaşıp ulaşılmadığına bakmaksızın algoritma mutasyona uğrayarak oluşturulmuş ürünlerle oluşturulan AB kümesindeki her bir antikor hesaba katılarak ilerlemektedir. Bu aşamada mutasyon oluşturma rutini Şekil 5.14'te sunulmaktadır. Şekil 5.15'te ise bu rutinin akış diyagramı verilmektedir. Şekil 5.14'ten de görülebileceği gibi AB antikor kümesindeki antikorlar antijenle uyarmı seviyeleri ile doğru orantılı olarak klonlanmaktadır. Daha sonra klonlanan bu antikorlar duyarlılık tabanlı mutasyon işlemine maruz kalmaktadır.

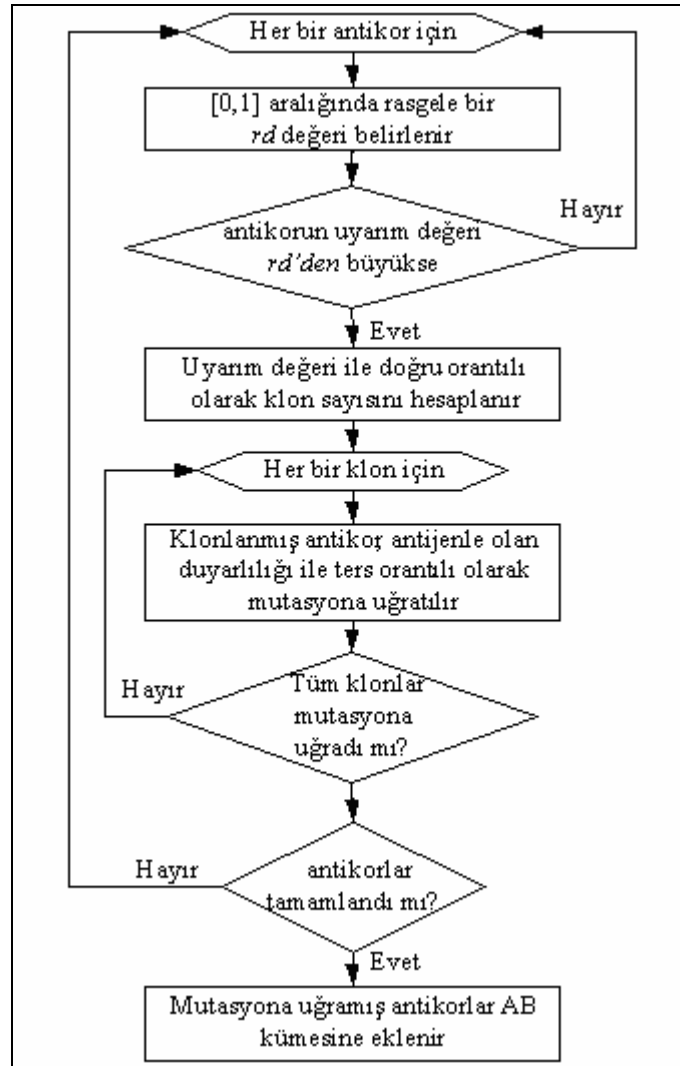
```

MU ← ∅
foreach ( ab ∈ AB )
do
    rd ← drandom()
    if ( ab.uyarım > rd )
    {
        klon_sayısı ← ab.uyarım * klonlama_oranı
        i ← 1
        while ( i ≤ klon_sayısı )
        do
            mut ← false
            ab_klon ← ab
            ab_klon ← duyarlılık_tabanlı_mutasyon ( ab_klon, mut )
            if ( mut = true )
                MU ← MU ∪ ab_klon
            i ← i + 1
        done
    }
done
AB ← AB ∪ MU

```

Şekil 5.14 BK-YBTS'de hayatta kalan antikorların mutasyonu

Hayatta kalan her bir antikordan mutasyona uğramış ürünler üretildikten sonra durma kriteri kontrol edilmektedir. Eğer durma kriteri ile karşılaşılmışsa bu antijen için eğitim durmaktadır. Eğer durma kriteri sağlanamamışsa bu sürecin tamamı Şekil 5.12’de gösterilen metottan başlayarak durma kriteri sağlanan kadar tekrar edilmektedir. Bu tekrar işleminde her tekrarda tek istisna *uyarma* ve *kaynak tahsisi* aşamasından sonra eğer durma kriteri ile karşılaşılsa mutasyona uğramış antikör üretimi uygulanmamaktadır. Çünkü durma kriterine ulaşılmıştır. Durma kriteri ile bir kez karşılaşılnca *aday bellek hücresi* seçilmektedir. *Aday bellek hücresi* mc_{aday} , ag eğitim antijeninin sınıfı ile aynı sınıftaki en fazla uyarılan antikör olan mutasyondan önce sistemde bulunan antikörün sınıfına sahiptir ve özellik vektörüdür.



Şekil 5.15 BK-YBTS’de hayatta kalan antikörlerin mutasyonunun akış diyagramı

5.4.5. BK-YBTS için bellek hücresinin değerlendirilmesi

Eğitim rutininde son aşamada mevcut MC bellek hücreleri kümesi içerisindeki en çok geliştirilmiş aday bellek hücresi mc_{aday} belirlenmektedir ve bu aşama boyunca önceden belirlenmiş olan mc_{ilk} ile eğitim sonucu oluşan mc_{aday} hücresinin yer değiştirilip değiştirilmeyeceğine Eşitlik 5.11’de başlangıçta hesaplanan *Sınıf Tabanlı Duyarlılık Eşiği* ile karar verilmektedir. Bu süreç Şekil 5.16’da gösterilmektedir. Bellek hücresi seçiminin akış diyagramı Şekil 5.17’de verilmektedir.

```

AdayUyarı ← bilgi_kazançlı_uyarma ( ag , mcaday )
İlkUyarı ← bilgi_kazançlı_uyarma ( ag , mcilk )
HücrelerArasıDuyarlılık ← bilgi_kazancı_tabanlı_uzaklık ( mcaday , mcilk )
if ( AdayUyarı > İlkUyarı )
{
    if ( HücrelerArasıDuyarlılık < SınıfTabanlıDuyarlılıkEşiğiag.c * ATS )
        MC ← MC - mcilk
        MC ← MC ∪ mcaday
}

```

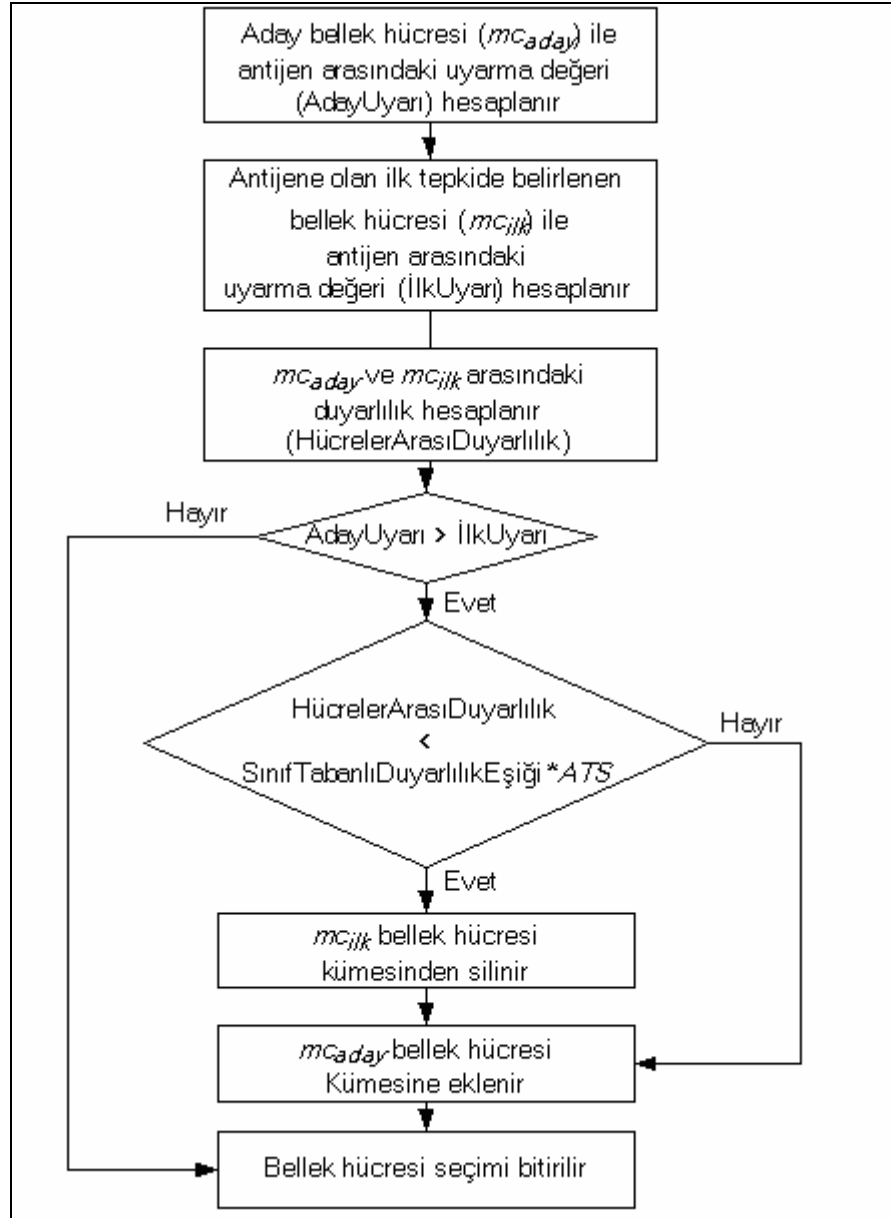
Şekil 5.16 BK-YBTS’de bellek hücresi seçimi

Aday bellek hücresi mevcut bellek hücreleri içerisinde değerlendirilince bu antijen için eğitim tamamlanmaktadır. Bu süreç bütün antijenler sisteme sunulana kadar devam etmektedir.

5.4.6. BK-YBTS için sınıflandırma

Eğitim aşaması tamamlandıktan sonra geliştirilen bellek hücreleri sınıflandırmada kullanılmak için uygun duruma gelmektedir. Sistemin son aşamasında *k en yakın komşu* yöntemi uygulanmaktadır. Bir test verisi sistem tarafından sınıflandırılırken, test verisini en fazla uyaran *k* adet bellek hücrelerinin çıkışları arasında bir *çoğunluk oylaması* yapılarak test verisinin sınıfına karar verilmektedir. Bu aşamada, *YBTS’den* farklı olarak *çoğunluk oylaması* yapılırken

eşitlik olması durumunda test verisinin sınıf değeri olarak, tüm veri kümesinde daha çok verisi bulunan sınıfın değeri atanmaktadır.



Şekil 5.17 BK-YBTS’de bellek hücresi seçiminin akış diyagramı

5.5. BK-YBTS Ayrıntıları

Bölüm 5.4’te *BK-YBTS*’nin bir özeti sunulurken, bu bölümde bu sınıflandırma sisteminin önemli elemanlarından bazıları kavramsal olarak

tartışılmıştır. Şekil-uzay gösteriminin kullanıldığı çoğu *YBS* algoritmaları *Öklid* uzaklık kriterini kullanmaktadır (Eşitlik 3.1). Bu uzaklık kriterine göre verinin bütün özellikleri uzaklık hesabında aynı etkiye sahiptir. Veri kümesi üzerinde etkisi çok az olan özelliklerle, veri kümesi üzerinde çok etkili özellikleri eşit ağırlıkta saymaktadır. Özellikle verilerin boyutunun büyük olduğunda bu durum sistemin sınıflandırma performansını olumsuz yönde etkilemektedir. Bu problemin üstesinden gelebilmek için veri kümesinde daha önemli olan özelliklere uzaklık hesabında daha yüksek ağırlık vermek ön görülmüştür. Bu amaçla geliştirilmiş olan *BK-YBTS* bu problemin etkisini minimuma indirmek için özelliklerin bilgi kazancını temel almaktadır. Özelliklerin önem derecesi bilgi kazancı hesabıyla belirlenmektedir. Bir özelliğin bilgi kazancı diğer özelliklerin bilgi kazancından yüksekse bu özelliğe diğer özelliklerden daha çok ağırlık verilmektedir.

Sistemin amacı sınıflandırma verisi olarak kullanılabilen bellek hücrelerinin bir kümesini geliştirmektir. Bu bellek hücreleri, doğal bağışıklık sistemlerinde gözlemlenen çeşitli karakteristikleri kapsamaktadır. Esasen bellek hücreleri vücuttaki olgunlaşma sürecinden geçmiş *B* hücrelerini temel almaktadır. Memeli bağışıklık sistemlerinde *B* hücreleri saldırgan antijenler tarafından uyarılmakta ve saldırgan patojenleri tanımaya tepki olarak mutasyon sürecinden geçmektedir.

BK-YBTS'nin en önemli aşaması antikor popülasyonundan bellek hücrelerinin geliştirilmesi sürecidir. Bu evrimsel süreç burada bahsedilen çeşitli anahtar kavramlara sahiptir. Bellek hücrelerinin geliştirilmesinde antikor popülasyonu için evrimsel basınç sağlamak için başlıca mekanizma sistemin geniş kaynakları için yapılan rekabettir. *Timmis ve Neal (2000)*; *Knight ve Timmis (2001)*; *Watkins (2001)* tarafından yapılan çalışmalardan esinlenen bu kavram, bellek hücresinin hayatta kalmasına karar vermede ve sınıflandırma kalitesini arttırmada kullanılmaktadır. Genetik algoritmalara benzer olarak kaynak rekabetinin amacı en uygun bireyleri geliştirmektir. *BK-YBTS* uygunluğa başlangıçta bir antijen ile ayrı ayrı antikorların uyarma tepkisi ile karar vermektedir. *Timmis ve ark. (2000)* danışmansız öğrenme çalışmalarında bir hücrenin uyarma değeri kaynak tahsisi için kullanmışlardır. 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 *BK-YBTS*'de antijenle aynı sınıftaki hücreler yüksek uyarmayla ve

antijenle aynı sınıfa sahip olmayan hücreler düşük uyarmayla ödüllendirilmektedir. Ödül daha fazla sistem kaynakları tahsis edilmesine izin verilmesidir. 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 hücrelerdir ve onlar böylece daha az ödüllendirilir. Elde edilmiş kaynaklar için en az yetenekli antikorlar sistemden çıkarıldığı için burada en çok ödül sağlayacak arama uzayının olduğu yere doğru gelişme olacaktır. Uygulamada antijenle aynı sınıfa sahip antikorların kaynakları antijenle farklı sınıfta olan antikorların kaynaklarından daha fazla olması sağlanmıştır. Bu eğitim antijeni civarındaki uzay araştırıldığı zaman daha büyük potansiyel çeşitliliğe izin verilerek yapılmaktadır.

Kaynakların rekabetinden hayatta kalan antikorlara mutasyona uğramış ürünler üretme fırsatı verilerek ödüllendirilmektedir. Tekrar genetik algoritmalarda olduğu gibi hayatta kalmak için bu rekabet aynı şekilde devam etmektedir. Belirli bir antijenin etrafındaki en uygun bireyler gerçekten bir bellek hücresi olarak hayatta kalmayabilirken onların mutasyona uğramış ürünleri hayatta kalabilir. Böylece sistemin desteklediği hücreler hayatta kalmaktadır. Bu, eğitim rutininde özelliklerin mutasyona uğratılması ile başarılmaktadır. Antikor popülasyonu içerisindeki mutasyona uğramış ürünleri oluşturmak, arama uzayının daha fazla araştırılmasını sağlamaktadır. Performans artımına katkıda bulunmak amacıyla sistemdeki antikorlara mutasyon işlemi sırasında duyarlılığı temel alan bir mutasyon mekanizmasını kullanılmıştır. Yani duyarlılığı yüksek olan antikorlar düşük oranda, duyarlılığı düşük olan antikorlar yüksek oranda mutasyon işlemine maruz kalmıştır. Böylece bellek hücresi için arzu edilebilir nitelikler sergileyen popülasyon içerisinde hücreler geliştirmek için evrimsel basınç arttırılmıştır.

Bir antikor, antikor popülasyonu arasındaki kaynaklar için başarılı olarak rekabet ettikten sonra eğer o antikor eğitim antijeni için tepkide en çok uyarılan olursa ve eğitim antijeni ile aynı sınıfta ise antikor, bellek hücresi havuzuna eklenmektedir. Bir antijen ilk kez sunulduğu zaman antijenle aynı sınıfta ve en çok uyarılan bellek hücresinin antikor popülasyonu içerisine mutasyona uğramış ürünlerini eklemesine izin verilmektedir. Bununla beraber bu antikor popülasyonunun evrimsel süreci sonunda bu orijinal bellek hücresi potansiyel olarak

geliştirilmiş bellek hücresi ile yer değiştirilmektedir. 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 *BK-YBTS'nin* veri azaltma yeteneğini sağladığını göstermektedir. Mevcut bellek hücresi ile şekil uzayında yakın yerlerde bulunan daha iyi sınıflandıran bellek hücrelerinin yer değişimine izin verilerek *BK-YBTS*, gerekli olan hücrelerin sayısını azaltmaktadır. Algoritma aynı zamanda genelleme yapmaktadır çünkü sistemde geliştirilmiş bellek hücrelerinin herhangi eğitim örneği ile aynı olması zorunlu değildir. *BK-YBTS'nin* çalışma adımları Tablo 5.1'de verilmektedir.

Tablo 5.1 BK-YBTS'nin çalışma adımları

1.	Verinin her bir özelliğinin bilgi kazancı değeri hesaplanır.
2.	Veri normalize edilir.
3.	Her bir eğitim örneği (<i>antijen</i>) için - Eğer bellek kümesi (<i>MC</i>) boş ise antijen bellek kümesine eklenir. <i>MC</i> kümesi içerisinde antijen ile en yüksek duyarlılığa sahip bellek hücresi (<i>mc</i>) seçilir. <i>mc</i> antijenle olan duyarlılığı ile doğru orantılı olarak klonlanır. - Her bir klon duyarlılıkla ters orantılı olarak mutasyona uğratılır ve <i>antikor</i> havuzuna (<i>AB</i>) eklenir. <i>AB</i> kümesi için kaynak tahsil edilir ve zayıf hücreler kümeden çıkarılır. <i>Antijen</i> ile <i>AB</i> kümesindeki antikorların ortalama uyarma değeri hesaplanır. Eğer bitirme şartı sağlanmışsa <i>i</i> adımına atlanır. <i>Antijen</i> ile olan uyarma seviyelerini temel alarak <i>AB</i> kümesindeki <i>antikorlar</i> klonlanır ve mutasyona uğratılır.

- h. e adımına geri dönülür.
 - i. AB kümesi içerisinde *antijen* ile duyarlılığı en yüksek olan *antikor*, aday bellek hücresi olarak belirlenir. Eğer *antijenin* aday bellek hücresine olan duyarlılığı, eğitimin başında seçilen bellek hücresine (mc_{ilk}) olan duyarlılıktan yüksekse aday bellek hücresi MC kümesine eklenir. Eğer mc_{ilk} bellek hücresi ve aday bellek hücresi benzer ise mc_{ilk} bellek hücresi, MC kümesinden çıkarılır.
4. Test verisi MC kümesi üzerinde k en yakın komşu yöntemiyle sınıflandırılır.

5.6. BK-YBTS'nin Karmaşıklık Analizi

Bir algoritmanın çalışma hızı karmaşıklıkla ifade edilmektedir. Bu kavram zaman birimiyle ifade edilmeyip doğrudan işlem adedi veya döngü sayısı ile ifade edilmektedir (Knuth 1968; Aho ve ark. 1975). Programın çalışma hızında zaman miktarı programın üzerinde çalıştığı donanıma bağlıdır; dolayısıyla algoritmanın çalışma hızını hesaplarken zaman miktarını kullanmak gerçekçi olmayıp yanılgılara neden olmaktadır. Bunun yerine, ilgili algoritmanın bilgisayar donanımından bağımsız olarak kaç adet işlem veya döngüyle gerçekleştirilebileceği hesaplanmaktadır.

Karmaşıklık, bir algoritmanın maliyet davranışındaki değişikliği gösteren kavramsal bir ifadedir. Karmaşıklık, eleman sayısı n 'nin sonsuza gitmesi durumunda algoritmanın maliyet hesabının davranışını görmek için kullanılmaktadır. Karmaşıklığı ifade edebilmek için genel olarak büyük O notasyonu (*big O notation*) daha çok kullanılmaktadır. Tablo 5.2'de büyük O notasyonuna dayanılarak verilen bazı temel fonksiyonlar verilmiştir. Örneğin bir algoritmanın karmaşıklığı $O(n^2)$ ise, bunun anlamı, n çok büyük değerlere giderken algoritmanın zaman maliyeti karesel olarak artar şeklindedir. Örneğin karmaşıklığı $O(n \log_2 n)$ olan bir sıralama algoritması $O(n^2)$ olana göre daha iyidir; n çok büyük değerlere giderken karmaşıklık

doğrusal çarpanlı logaritmik olarak artmaktadır. Karmaşıklık hesabında tüm sabit çarpanlar ve düşük dereceli terimler atılmaktadır.

Tablo 5.2 Büyük O ifadeleri ve anlamları

Büyük O	Değişim Şekli
$O(1)$	Sabit
$O(\log n)$	Logaritmik
$O(n)$	Doğrusal
$O(n \log n)$	Doğrusal çarpanlı logaritmik
$O(n^2)$	Karesel
$O(n^3)$	Kübik
$O(2^n)$	İki tabanından üssel
$O(10^n)$	On tabanından üssel
$O(n!)$	Faktöriyel

Tablo 5.3'te YBTS ve BK-YBTS için karmaşıklık analizi verilmiştir. Burada n , eğitim antijenlerinin sayısını; k , k en yakın komşu yönteminde kullanılan sabiti; a , eğitimi bitirme şartı sağlanana kadar gerçekleşen döngü sayısını ifade eden sabiti; L , kullanılan verinin boyutunu belirten sabiti ifade etmektedir.

Çalışmada geliştirilen sistemde duyarlılık eşiği hesaplanırken eğitim verileri (antijenler) arasındaki ortalama duyarlılık hesabında $\frac{n * (n-1)}{2}$ adet döngü söz konusudur ($O(n^2)$). Normalizasyon aşaması, bilgi kazancı hesabı, antikor oluşumu ve k en yakın komşu yöntemi uygulaması doğrusal olarak çalışmaktadır ($O(n)$). Bellek hücresi belirleme aşaması sabit karmaşıklıkta çalışmaktadır ($O(1)$). Geliştirilen toplam olarak $\frac{n^2}{2} + ((k+1) * L + a + \frac{1}{2}) * n + 1$ adımda çalışmaktadır. Bu durumda geliştirilen sistemin karmaşıklığı tüm sabit çarpanlar ve düşük dereceli terimler atılınca $O(n^2)$ olarak hesaplanmaktadır.

Tablo 5.3 YBTS ve BK-YBTS için karmaşıklık analizi

Sistem Aşamaları	YBTS	BK-YBTS
Bilgi kazancı hesabı	-	$L * n$
Normalizasyon	n	n
Duyarlılık Eşiği Hesabı	$\frac{n * (n - 1)}{2}$	$\frac{n * (n - 1)}{2}$
Antikor oluşumu	$a * n$	$a * n$
Bellek hücresi belirleme	l	l
k en yakın komşu yöntemi	$k * L * n$	$k * L * n$
Toplam	$\frac{n^2}{2} + (k * L + a + \frac{l}{2}) * n + l$	$\frac{n^2}{2} + ((k + 1) * L + a + \frac{l}{2}) * n + l$
Sonuç	$O(n^2)$	$O(n^2)$

6. UYGULAMALAR

Bu bölümde standart makine öğrenmesi ve gerçek dünya veri kümeleri üzerinde *YBTS* ve *BK-YBTS'nin* sınıflandırma performansı değerlendirilmiştir. Bu bölüm şu şekilde organize edilmiştir. Öncelikle *YBTS* ve *BK-YBTS'de* kullanılan parametreler hakkında bilgi verilmiştir. Çalışmada kullanılan dört adet veri kümesi *YBTS* ve *BK-YBTS* için sınıflandırma performansı değerlendirilmesinde kullanılmıştır. Değerlendirmelerde veri kümeleri hakkında kısaca bilgi verildikten sonra *BK-YBTS* ile beraber önerilen değişikliklerin etkisini araştırmak amacıyla her bir parametrenin değişik değerleri için *YBTS* ve *BK-YBTS'nin* çalışma zamanı, iterasyon sayısı, bellek hücresi sayısı ve sınıflandırma doğrulukları karşılaştırılmıştır. Daha sonra en iyi sınıflandırma doğruluğunu veren parametre değerleri ve *BK-YBTS'nin* sınıflandırma doğrulukları diğer sınıflandırıcıların sınıflandırma doğrulukları ile birlikte sunulmuştur. Son olarak dört adet veri kümesi için *YBTS* ve *BK-YBTS'de* elde edilen sonuçlar topluca verilmiştir.

6.1. BK-YBTS'de Kullanılan Parametreler

YBTS ve *BK-YBTS'de* kullanılan parametreler Tablo 6.1'de belirtilmiştir. Alt bölümlerde ise bu parametreler kısaca açıklanmıştır.

Tablo 6.1 YBTS ve BK-YBTS parametreleri

Parametre Adı	Açıklama
Çekirdek hücre sayısı	Eğitim başlangıcında kullanılan çekirdek hücrelerin sayısı
Kaynak sayısı	Sistemde izin verilen kaynakların sayısı
Uyarma eşiği	Eğitimi durdurmada kullanılan uyarma eşiği
Mutasyon oranı	Özellik mutasyonu için mutasyon oranı
Duyarlılık eşik ölçüsü	Bellek hücresi seçimi için kullanılan duyarlılık eşik ölçüsü
k değeri	Çoğunluk oylaması için kullanılan k değeri
Parça değeri	Sürekli değerli veri özelliğini ayrık forma dönüştürmede kullanılan parça değeri (sadece <i>BK-YBTS'de</i> kullanılmaktadır)

6.1.1. Çekirdek hücre sayısı

Çekirdek hücreleri, sistemdeki başlangıç bellek hücrelerini ve antikorları oluşturmaktadır. Bu hücreler eğitim antijenleri arasından rasgele seçilmektedir. *Çekirdek hücre sayısı* arttığı zaman bellek hücreleri ve antikorlar eğitim antijenine benzer yapıda oluşmaktadır. Aynı zamanda *çekirdek hücre sayısı* arttığı zaman sistemdeki bellek hücresi sayısı da artmaktadır. Bu durum test verilerinin sınıflandırma doğruluğunu olumsuz yönde etkilemektedir. Denemelerde *çekirdek hücre sayısının YBTS ve BK-YBTS üzerindeki etkisi*, $\{1, 10, 20, 30, 40, 50, 60, 70, 80, 90\}$ değerleri için analiz edilmiştir.

6.1.2. Kaynak sayısı

Kaynak sayısı parametresi, sistemde bulunacak antikor sayısını direkt olarak etkilemektedir. *Kaynak sayısı* parametre değeri arttıkça, sistemde bulunan bellek hücreleri ve antikorların çeşitliliği artmaktadır. Çeşitlilikteki bu artış sınıflandırma doğruluğunda artışı garanti etmemektedir. Eğitim sürecinde aynı sınıftaki antikorlar için tahsis edilen kaynak sayısı, izin verilen kaynak sayısından fazla olduğu durumlarda, en az kaynağa sahip antikorlardan başlanarak sistemden antikorlar çıkarılmaktadır. Bölüm 5.2.3 ve Bölüm 5.4.4'te kaynak tahsisi mekanizması ayrıntılı olarak anlatılmıştır. Denemelerde kaynak sayısının *YBTS ve BK-YBTS üzerindeki etkisi*, $\{50, 100, 150, 200, 250, 300, 350, 400, 450, 500\}$ değerleri için analiz edilmiştir.

6.1.3. Uyarma eşiği

Uyarma eşiği parametresi, belirli bir eğitim antijeninin eğitimini durdurma kriteri olarak kullanılmaktadır. Eğitimin durdurulup durdurulmayacağı kontrol edilirken Eşitlik 5.4'te belirtilen bir s vektörü hesaplanmaktadır. Bu vektör her bir sınıfa ait antikorların ortalama uyarma değerlerini içermektedir. s vektörünün her bir

elemanı, uyarma eşiği parametresinden büyük olduğu durumda eğitim durmaktadır ($s_i \geq \text{uyarma eşiği}$).

Uyarma eşiği parametresi, antikor popülasyonunun ne kadar uyarılacağını belirlemektedir. Denemelerde *uyarma eşiğinin YBTS ve BK-YBTS üzerindeki etkisi*, $\{0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9\}$ değerleri için analiz edilmiştir. *Uyarma eşiğinin* kullanımı Bölüm 5.2.3'te ayrıntılı olarak açıklanmıştır.

6.1.4. Mutasyon oranı

Mutasyon oranı parametresi, antikor popülasyonunun çeşitliliğini etkilemektedir. *Mutasyon oranı* parametre değeri büyük olduğu durumlarda, sistemin sınıflandırma doğruluğu ve çalışma zamanı negatif yönde etkilenmektedir. Özellikle *YBTS'de* mutasyon işleminin rasgele yapılması ile hedef arama uzayından uzaklaşma olduğu için sınıflandırma doğruluğu negatif yönde etkilenmektedir. *BK-YBTS'de* duyarlılık tabanlı mutasyon mekanizması kullanıldığı için *YBTS'deki* kadar sınıflandırma doğruluğuna negatif yönde etki gözlemlenmemiştir. Bölüm 5.2.2'de *YBTS'nin* mutasyon mekanizması, Bölüm 5.4.3'te ise *BK-YBTS'nin* mutasyon mekanizması ayrıntılı olarak açıklanmıştır.

Mutasyon oranı arttıkça eğitimi durdurma kriteri olan uyarma eşiği değerinin, antikorlar arasındaki ortalama uyarma değerinden küçük olma ihtimali azaldığı için çalışma süresi artmaktadır. Denemelerde mutasyon oranı parametresinin *YBTS ve BK-YBTS üzerindeki etkisi*, $\{0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9, 1.0\}$ değerleri için analiz edilmiştir.

6.1.5. Duyarlılık eşik ölçüsü

Duyarlılık eşik ölçüsü, bellek hücresi seçimini direkt olarak etkileyen bir parametredir. *YBTS'de ortalama duyarlılık eşiği*, eğitim antijenleri arasındaki ortalama duyarlılığı; *BK-YBTS'de sınıf tabanlı duyarlılık eşiği*, aynı sınıfa sahip eğitim antijenlerinin ortalama duyarlılığını ifade etmektedir. *Duyarlılık eşik ölçüsü*

değeri, *YBTS'de ortalama duyarlılık eşiği* değeri; *BK-YBTS'de sınıf tabanlı duyarlılık eşiği* değeri ile çarpılarak bellek seçimi gerçekleştirilmektedir. Tek bir antijen için başlangıçta belirlenen mc_{ilk} ve eğitimin sonunda oluşan mc_{aday} arasındaki duyarlılığa göre bellek hücresi seçilmektedir. *YBTS* için duyarlılık eşik ölçüsü kullanımı Bölüm 5.2.4'te, *BK-YBTS* için duyarlılık eşik ölçüsü kullanımı, Bölüm 5.4.5'te açıklanmıştır.

Duyarlılık eşik ölçüsü parametre değeri arttıkça, sistemde kalan bellek hücresi sayısı azalmaktadır. Ayrıca *duyarlılık eşik ölçüsü* küçük değerler aldığı zaman sınıflandırma doğruluğunda büyük değişiklikler olmamıştır; *duyarlılık eşik ölçüsü*, büyük değerler aldığı zaman sınıflandırma doğruluğunda büyük azalmalar gözlemlenmiştir. Denemelerde *duyarlılık eşik ölçüsü* parametresinin *YBTS* ve *BK-YBTS* üzerindeki etkisi, $\{0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9, 1.0\}$ değerleri için analiz edilmiştir.

6.1.6. *k* değeri

Eğitim aşaması tamamlandıktan sonra geliştirilen bellek hücreleri sınıflandırmada kullanılmak için uygun duruma gelmektedir. Sistemin son aşamasında *k en yakın komşu* yöntemi uygulanmaktadır. Test verilerinin hangi sınıfta olduğu bellek hücreleri kümesi üzerine *k en yakın komşu* yönteminin uygulanması ile belirlenmektedir. Bir verinin sistem tarafından sınıflandırılmasına test verisini en fazla uyaran *k* adet bellek hücrelerinin çıkışlarının bir *çoğunluk oylaması* kullanılarak karar verilmektedir.

YBTS'de çoğunluk oylaması yapılırken eşitlik olması durumunda test verisinin sınıfı olarak, en küçük indisli sınıf değeri atanmaktadır. *BK-YBTS'de* çoğunluk oylaması yapılırken eşitlik olması durumunda test verisinin sınıfı olarak, tüm veri kümesinde daha çok verisi bulunan sınıf değeri atanmaktadır. Denemelerde *k değeri* parametresinin *YBTS* ve *BK-YBTS* üzerindeki etkisi, $\{1, 2, 3, 4, 5, 6, 7, 8, 9, 10\}$ değerleri için analiz edilmiştir.

6.1.7. Parça değeri

BK-YBTS'ye özgü bir parametre olan *parça değeri*, veri özelliklerinin bilgi kazancını hesaplamak için sürekli değerli özelliklerin ayırık hale getirilmesinde kullanılmaktadır. Hesaplanan bilgi kazancı değerleri, uzaklık hesabında katsayı olarak kullanılmaktadır (Bölüm 5.4.1). Sürekli değerli özellikleri ayırık hale getirilmesi işleminde Bölüm 4.2.1.1'de açıklanan eşit aralık genişliği ayırım yaklaşımı kullanılmıştır. Denemelerde *parça değeri* parametresinin *YBTS* ve *BK-YBTS* üzerindeki etkisi, $\{2, 4, 6, 8, 10, 12, 14, 16, 18, 20\}$ değerleri için analiz edilmiştir.

6.2. YBTS ve BK-YBTS'nin Sınıflandırma Performanslarının Analizi

YBTS ve *BK-YBTS*'nin performansı standart makine öğrenmesi veri kümelerinden olan *Iris*, ve *Thyroid* veri kümeleri ile Erciyes Üniversitesi Tıp Fakültesinden alınan *Damar tıkanıklığı (Atherosclerosis)* veri kümesi ve Erciyes Üniversitesi Eczacılık Fakültesinden alınan *Bakteri* veri kümesi üzerinde değerlendirilmiştir.

Her bir veri kümesi için *BK-YBTS* ile beraber önerilen değişikliklerin etkisini araştırmak amacıyla *YBTS* ve *BK-YBTS* parametreleri olan *çekirdek hücre sayısı*, *kaynak sayısı*, *uyarma eşiği*, *mutasyon oranı*, *duyarlılık eşik ölçüsü* ve *k* parametrelerinin her biri aldığı değişik değerler için *YBTS* ve *BK-YBTS*'nin sınıflandırma performansı analizleri yapılmıştır. Sınıflandırma performansı analizinde her bir parametrenin değişik değerleri için iki sistemin üçer kez çalıştırıldıktan sonra ortalama olarak; *çalışma süreleri*, *iterasyon sayıları*, *oluşan bellek hücreleri sayıları* ve *sınıflandırma doğrulukları* kullanılmıştır. Sadece *BK-YBTS*'ye özgü bir parametre olan *parça değeri* parametresi, değişik değerler için *BK-YBTS*'de çalışma zamanı, iterasyon sayısı, oluşan bellek hücreleri sayısı ve sınıflandırma doğrulukları açısından analiz edilmiştir. Bir parametrenin değerinin değişiminin, *YBTS* ve *BK-YBTS*'nin çalışma sürelerine, iterasyon sayılarına, bellek

hücreleri sayılarına ve sınıflandırma doğruluklarına etkisi araştırılırken; sadece etkisi araştırılan parametre değeri değiştirilmiş ve diğer parametrelerin değerleri sabit alınmıştır.

Son olarak her bir veri kümesi için *BK-YBTS'de* deneme yanılma yoluyla belirlenen ve en yüksek sınıflandırma doğruluklarını veren parametre değerleri verildikten sonra bu parametre değerleri ile elde edilen elde edilen sınıflandırma doğrulukları, diğer sınıflandırıcıların sınıflandırma doğrulukları ile birlikte sunulmuştur.

6.2.1. Iris veri kümesi uygulamaları

Iris veri kümesi, makine öğrenmesi literatüründe çok iyi bilinen veri kümelerinden birisidir (<http://archive.ics.uci.edu/ml/>). Bu veri kümesi dört adet özelliğe ve üç adet sınıfa sahiptir. Burada her bir sınıfta 50 adet veri içeren 150 adet veri söz konusudur. Bu sınıflar *Iris setosa*, *Iris versicolor* ve *Iris virginica*dır. Dört adet özellik, *Iris (süsen)* bitkisinin farklı nitelik ölçülerini ifade etmektedir. Bu dört özellik *çanak yaprağı uzunluğu*, *çanak yaprağı genişliği*, *taç yaprağı uzunluğu* ve *taç yaprağı genişliğidir*.

Uygulama için veri kümesi rasgele olarak beş ayrı kümeye bölünmüştür. Bu kümeler *5 kat çapraz onaylama (five-fold cross validation)* yöntemiyle çalıştırılarak kullanılmıştır. 5 kat çapraz onaylama yönteminde veri kümesi beş parçaya bölünmektedir. Bu parçalardan dördü eğitim verisi olarak ve beşinci parça test verisi olarak kullanılmaktadır. Bu şekilde her bir parça, bir kez test verisi olarak kullanılmak üzere beş adet eğitim-test verisi kombinasyonu oluşturulmaktadır. Sonuçlar beş eğitim-test kümesi kombinasyonlarının her birinin *YBTS'de* ve *BK-YBTS'de* üçer kez çalıştırılması sonucu elde edilen değerlerin ortalamasıdır. Literatürdeki sınıflandırıcılar *Iris* veri kümesini sınıflandırırken eğitim ve test kümelerini *5 kat çapraz onaylama yöntemi* ile belirlediklerinden dolayı geliştirilen sistemde de bu yöntem kullanılarak eğitim ve test kümeleri belirlenmiştir.

YBTS'den farklı olarak *BK-YBTS* ile beraber önerilen değişikliklerin etkisini araştırmak amacıyla *çekirdek hücre sayısı*, *kaynak sayısı*, *uyarma eşiği*, *mutasyon*

oranı, duyarlılık eşik ölçüsü ve k parametreleri için Tablo 6.2’de belirtilen parametre değerleri temel alınarak *YBTS* ve *BK-YBTS* karşılaştırılmıştır.

Tablo 6.2 Iris veri kümesi için *YBTS* ve *BK-YBTS*’nin başlangıç parametre değerleri

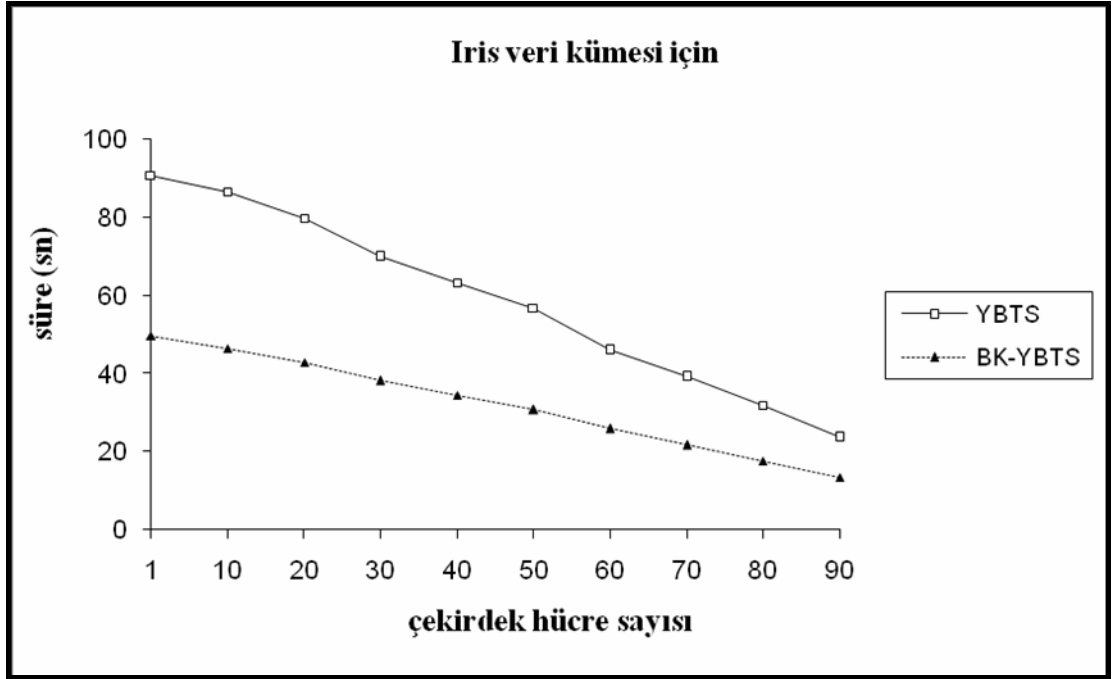
Parametre Adı	Parametre Değeri
Çekirdek hücre sayısı	1
Kaynak sayısı	200
Uyarma eşiği	0.9
Mutasyon oranı	0.1
Duyarlılık eşik ölçüsü	0.2
Çoğunluk oylaması için kullanılan k değeri	1
Parça değeri (<i>BK-YBTS</i> için)	2

Çekirdek hücre sayısı parametresinin aldığı değişik değerler için *YBTS* ve *BK-YBTS* karşılaştırma sonuçları Tablo 6.3’te verilmiştir.

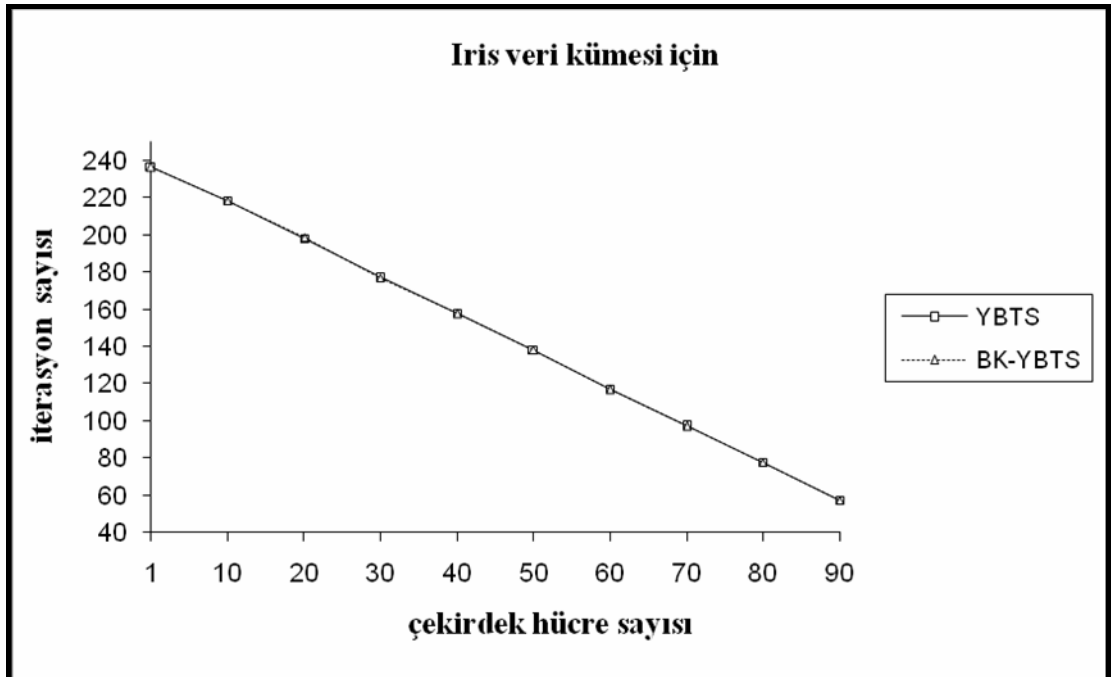
Tablo 6.3 Iris veri kümesinde çekirdek hücre sayısı parametresi için *YBTS* ve *BK-YBTS* karşılaştırma sonuçları

Çekirdek Hücre Sayısı	Süre (sn)		İterasyon Sayısı		Bellek Hücresi Sayısı		Sınıflandırma Doğruluğu (%)	
	YBTS	BK-YBTS	YBTS	BK-YBTS	YBTS	BK-YBTS	YBTS	BK-YBTS
1	90.6	49.6	236.5	236.3	42.8	57.9	93.8	96.0
10	86.5	46.5	218.1	218.2	40.9	57.9	95.6	94.9
20	79.6	42.8	198.0	198.3	44.3	63.5	95.6	94.4
30	69.8	38.1	177.3	177.0	47.1	68.7	94.9	96.0
40	63.3	34.4	157.7	157.4	54.8	72.2	94.2	95.3
50	56.7	30.7	138.0	138.1	60.2	77.9	95.6	95.6
60	45.9	25.8	117.0	116.7	68.3	82.9	95.6	95.8
70	39.1	21.6	97.0	97.0	76.8	88.1	95.8	95.6
80	31.8	17.7	77.4	77.4	84.9	94.5	95.1	96.0
90	23.6	13.3	57.4	57.4	93.5	101.4	95.3	95.6

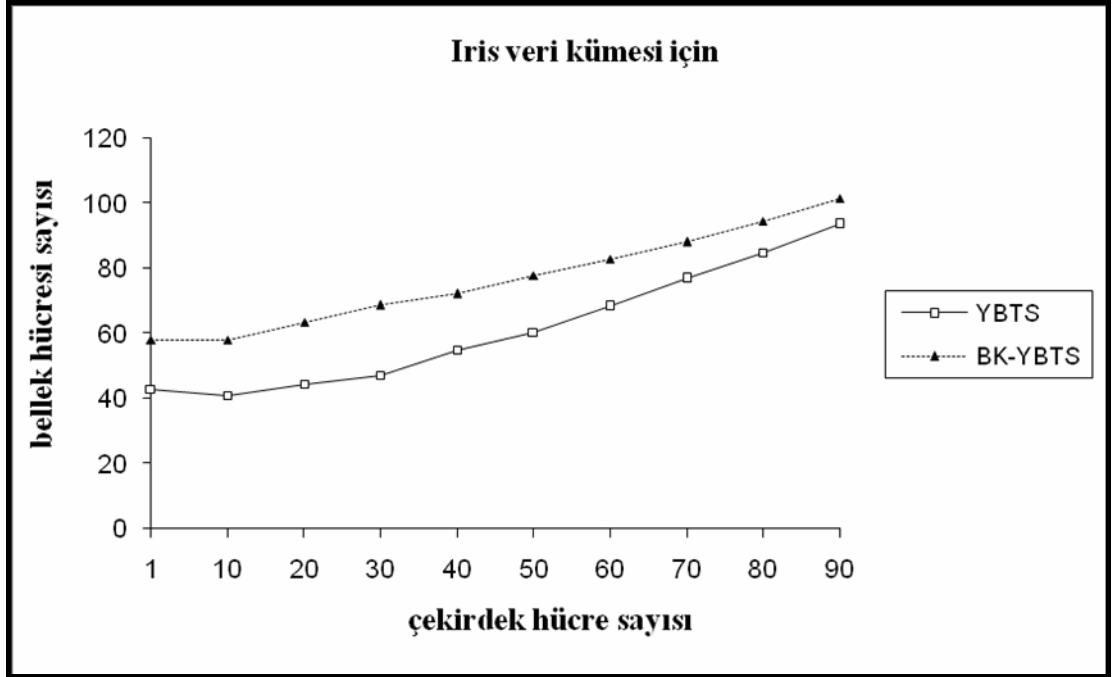
Çekirdek hücre sayısı parametresinin aldığı değişik değerler için iki sistemin çalışma süreleri açısından karşılaştırılmasının grafiksel olarak gösterimi Şekil 6.1’de, iterasyon sayıları açısından karşılaştırılmasının grafiksel olarak gösterimi Şekil 6.2’de, bellek hücreleri sayıları açısından karşılaştırılmasının grafiksel olarak gösterimi Şekil 6.3’te ve sınıflandırma doğrulukları açısından karşılaştırılmasının grafiksel olarak gösterimi Şekil 6.4’te verilmiştir.



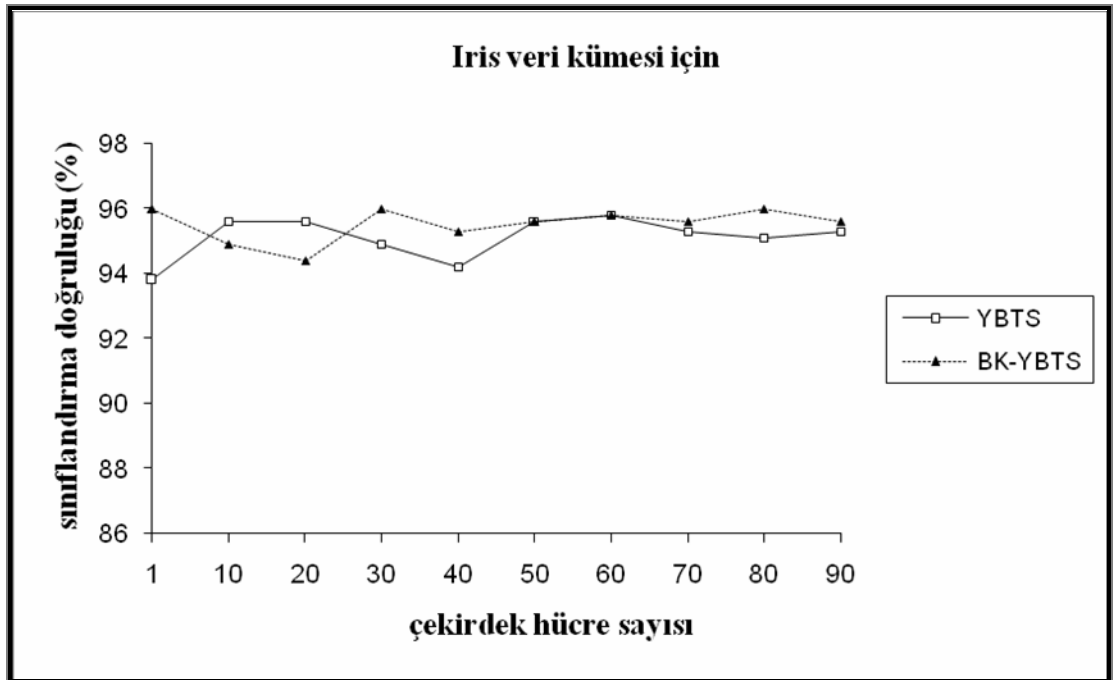
Şekil 6.1 Iris veri kümesinde çekirdek hücre sayısı parametresinin değişimine göre YBTS ve BK-YBTS'nin çalışma sürelerinin grafiksel gösterimi



Şekil 6.2 Iris veri kümesinde çekirdek hücre sayısı parametresinin değişimine göre YBTS ve BK-YBTS'nin iterasyon sayılarının grafiksel gösterimi



Şekil 6.3 Iris veri kümesinde çekirdek hücre sayısı parametresinin değişimine göre YBTS ve BK-YBTS'nin bellek hücreleri sayılarının grafiksel gösterimi



Şekil 6.4 Iris veri kümesinde çekirdek hücre sayısı parametresinin değişimine göre YBTS ve BK-YBTS'nin sınıflandırma doğruluklarının grafiksel gösterimi

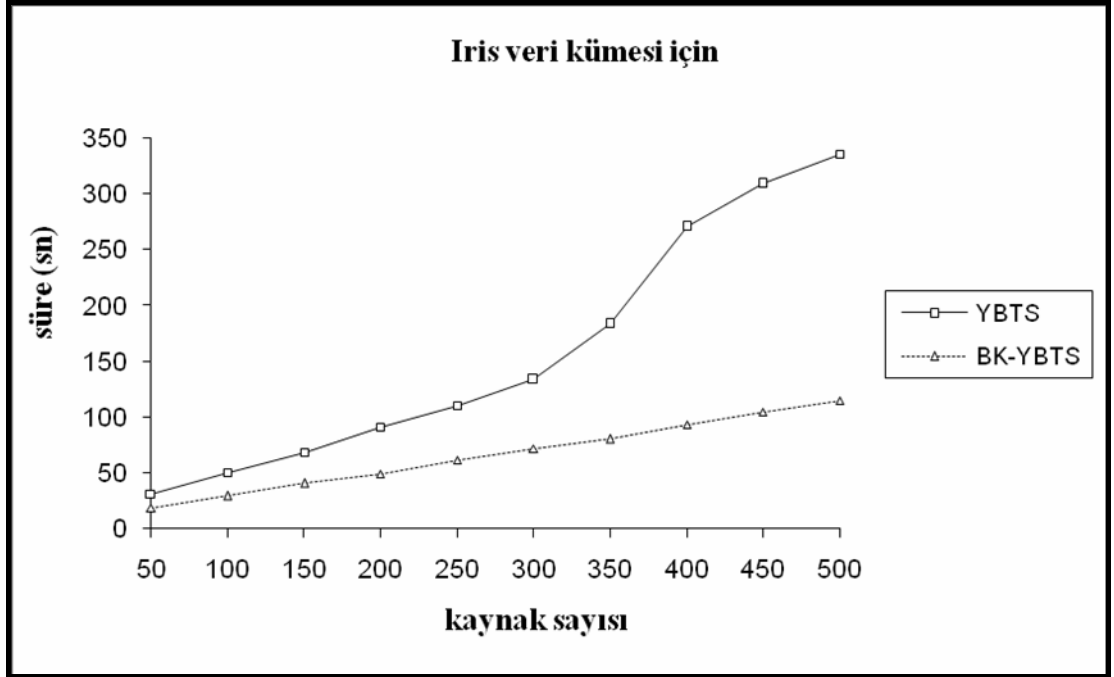
Çekirdek hücre sayısı arttığı zaman, eğitimin başlangıcında antijenler bellek hücresi ve antikör olarak belirlendiği için doğal olarak *YBTS* ve *BK-YBTS'nin* çalışma zamanı ve iterasyon sayısı azalmış, sistemdeki bellek hücresi sayısı ise artmıştır. *BK-YBTS* ve *YBTS* iterasyon sayısı olarak birbirlerine yakın çalışmakla beraber antikörlere kaynak tahsisi rutininde *YBTS* daha çok zaman harcadığı için *YBTS'nin* çalışma süresi, *BK-YBTS'nin* çalışma süresinden fazla ölçülmüştür. Bellek hücreleri sayısı *BK-YBTS* için *YBTS'den* daha fazla tespit edilmiştir. *Çekirdek hücre sayısı* için *YBTS* ve *BK-YBTS* yakın sınıflandırma doğruluklarına ulaşmışlardır. *Çekirdek hücre sayısı*, 1, 30, 40, 60, 80 ve 90 değerlerinde *BK-YBTS*, *YBTS'den* daha yüksek sınıflandırma doğruluklarına ulaşmıştır.

Çekirdek hücre sayısı parametresi için iki sistemin sınıflandırma performansı analiz edildikten sonra *kaynak sayısı* parametresinin değişik değerleri için karşılaştırma sonuçları Tablo 6.4'te verilmiştir.

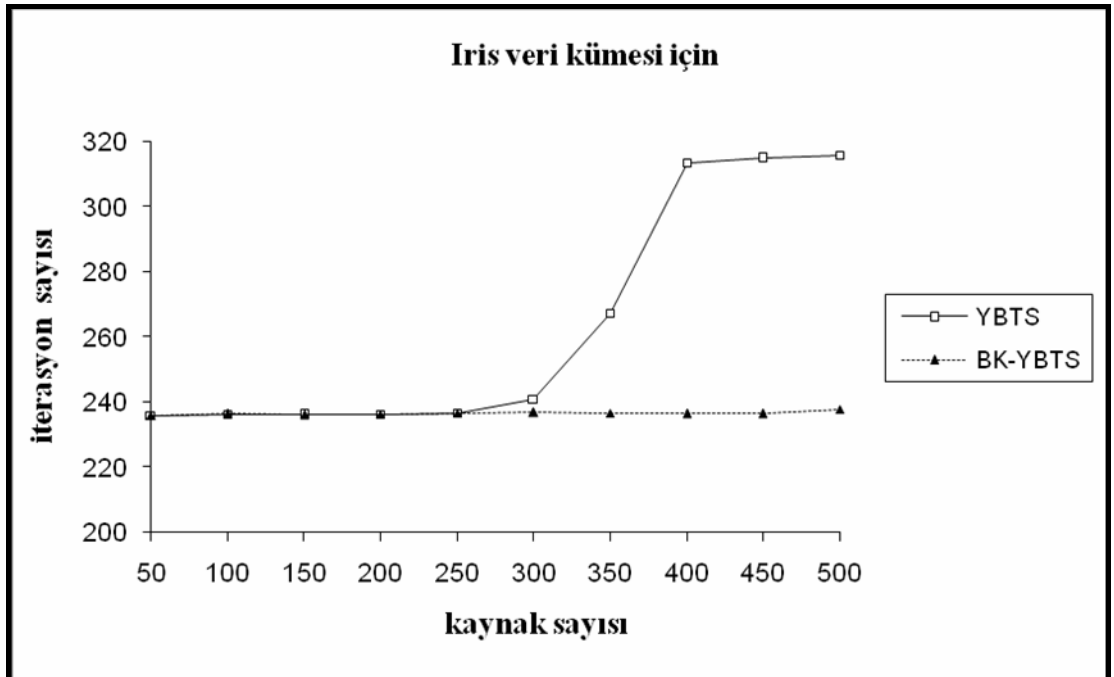
Tablo 6.4 Iris veri kümesinde kaynak sayısı parametresi için *YBTS* ve *BK-YBTS* karşılaştırma sonuçları

Kaynak Sayısı	Süre (sn)		İterasyon Sayısı		Bellek Hücresi Sayısı		Sınıflandırma Doğruluğu (%)	
	YBTS	BK-YBTS	YBTS	BK-YBTS	YBTS	BK-YBTS	YBTS	BK-YBTS
50	14.3	21.1	236.0	236.0	37.0	60.4	93.1	94.9
100	28.9	40.9	236.3	236.4	40.7	56.5	95.1	96.0
150	49.5	65.4	236.1	236.1	43.1	59.5	94.0	96.2
200	83.5	89.8	236.3	236.3	43.1	59.8	93.8	96.0
250	119.9	128.4	236.7	236.5	42.3	58.4	94.2	96.0
300	177.2	170.0	241.0	236.8	42.8	59.1	93.1	95.8
350	287.4	207.6	267.3	236.5	44.2	61.3	95.8	95.8
400	521.6	264.5	313.6	236.4	46.2	65.2	94.4	95.8
450	629.7	334.0	314.9	236.4	45.2	71.8	93.8	95.1
500	768.5	393.2	315.9	237.9	45.9	74.2	93.8	95.8

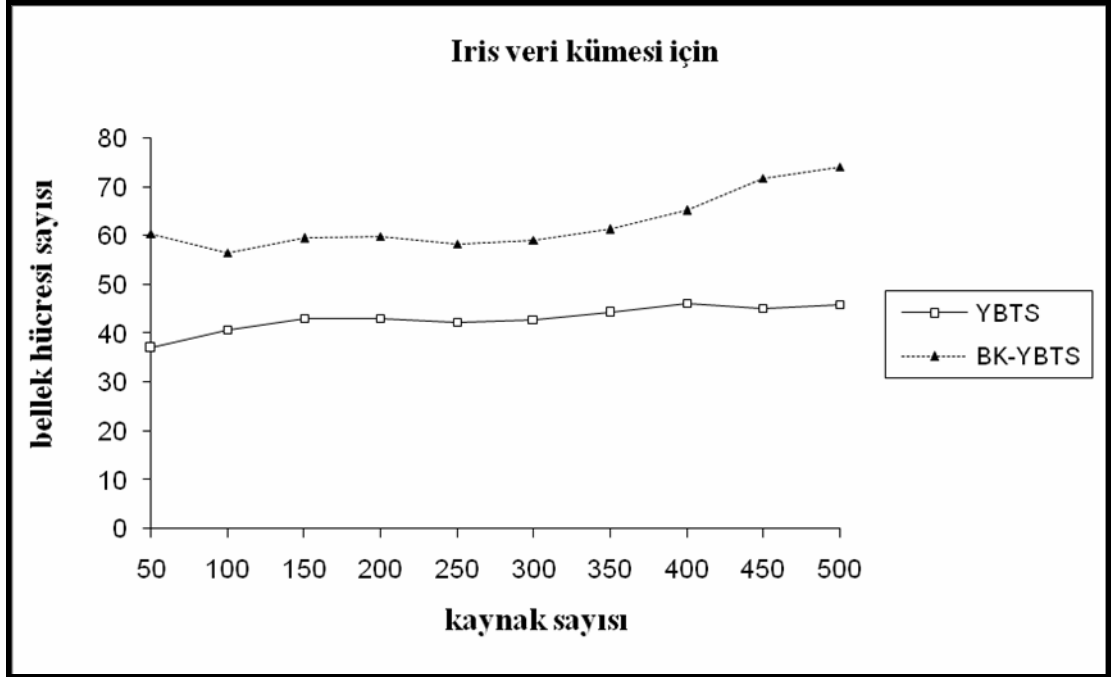
Kaynak sayısı parametresinin değişik değerleri için iki sistemin çalışma süreleri, iterasyon sayıları, bellek hücreleri sayıları ve sınıflandırma doğrulukları açısından karşılaştırılmasının grafiksel olarak gösterimleri sırasıyla Şekil 6.5 ile Şekil 6.8 arasında verilmiştir.



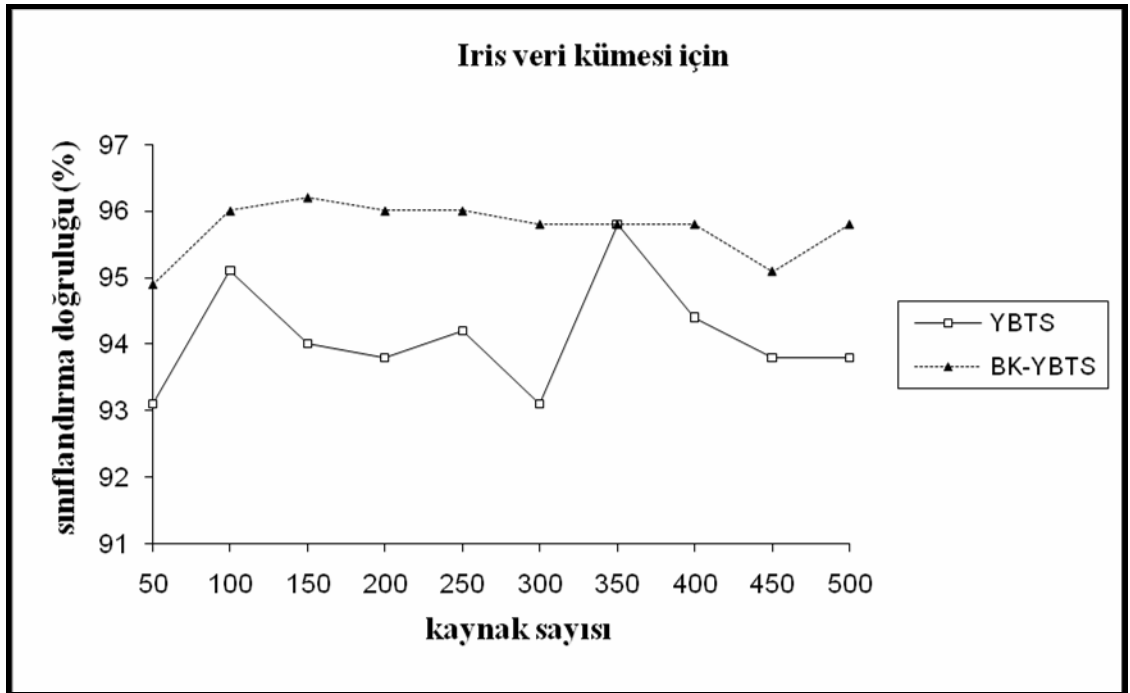
Şekil 6.5 Iris veri kümesinde kaynak sayısı parametresinin değişimine göre YBTS ve BK-YBTS'nin çalışma sürelerinin grafiksel gösterimi



Şekil 6.6 Iris veri kümesinde kaynak sayısı parametresinin değişimine göre YBTS ve BK-YBTS'nin iterasyon sayılarının grafiksel gösterimi



Şekil 6.7 Iris veri kümesinde kaynak sayısı parametresinin değişimine göre YBTS ve BK-YBTS'nin bellek hücreleri sayılarının grafiksel gösterimi



Şekil 6.8 Iris veri kümesinde kaynak sayısı parametresinin değişimine göre YBTS ve BK-YBTS'nin sınıflandırma doğruluklarının grafiksel gösterimi

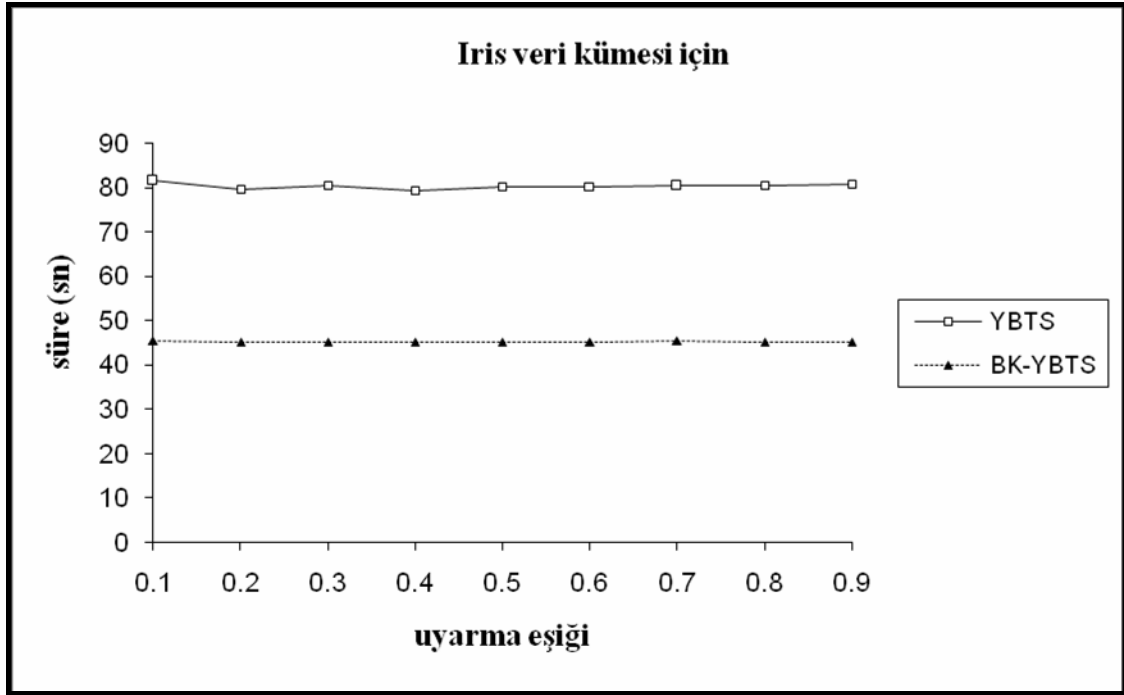
Kaynak sayısı parametresi değeri arttığı zaman, *YBTS* ve *BK-YBTS'nin* çalışma zamanı, iterasyon sayısı ve bellek hücresi sayısı artmıştır. *BK-YBTS'nin* iterasyon sayısı ve çalışma süresi *YBTS'nin* iterasyon sayısı ve çalışma süresinden daha az tespit edilirken, bellek hücreleri sayısı açısından *BK-YBTS'de YBTS'den* daha fazla bellek hücresi oluşmuştur. *Kaynak sayısı* parametresinin bir değeri hariç *BK-YBTS, YBTS'den* daha yüksek sınıflandırma doğruluğuna ulaşmıştır. Sadece *kaynak sayısı* parametresi 350 değerinde, *BK-YBTS* ve *YBTS* aynı sınıflandırma doğruluğuna ulaşmıştır.

Kaynak sayısı parametresi için iki sistemin sınıflandırma performansı analiz edildikten sonra *Uyarma eşiği* parametresinin değişik değerleri için karşılaştırma sonuçları Tablo 6.5'te verilmiştir.

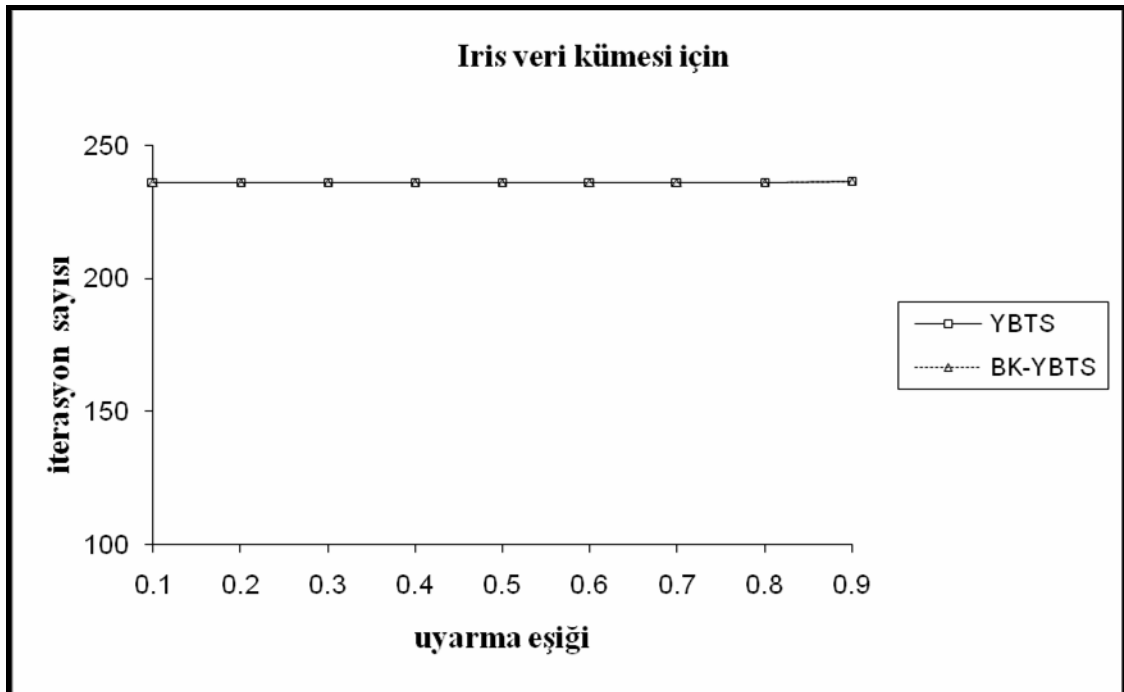
Tablo 6.5 Iris veri kümesinde uyarma eşiği parametresi için YBTS ve BK-YBTS karşılaştırma sonuçları

Uyarma Eşiği	Süre (sn)		İterasyon Sayısı		Bellek Hücresi Sayısı		Sınıflandırma Doğruluğu (%)	
	YBTS	BK-YBTS	YBTS	BK-YBTS	YBTS	BK-YBTS	YBTS	BK-YBTS
0.1	81.4	45.4	236.0	236.0	43.6	59.5	93.8	96.0
0.2	79.6	45.0	236.0	236.0	43.3	59.8	95.6	96.4
0.3	80.4	45.1	236.0	236.0	43.6	56.1	94.4	96.2
0.4	79.2	45.0	236.0	236.0	42.3	57.2	92.7	96.2
0.5	80.1	45.1	236.0	236.0	45.1	59.7	93.6	95.8
0.6	80.1	45.0	236.0	236.0	41.9	58.3	94.9	96.7
0.7	80.3	45.4	236.0	236.0	42.5	61.3	95.8	96.7
0.8	80.4	45.0	236.1	236.0	40.9	57.3	94.2	97.1
0.9	80.7	45.2	236.3	236.4	43.1	58.9	93.8	96.0

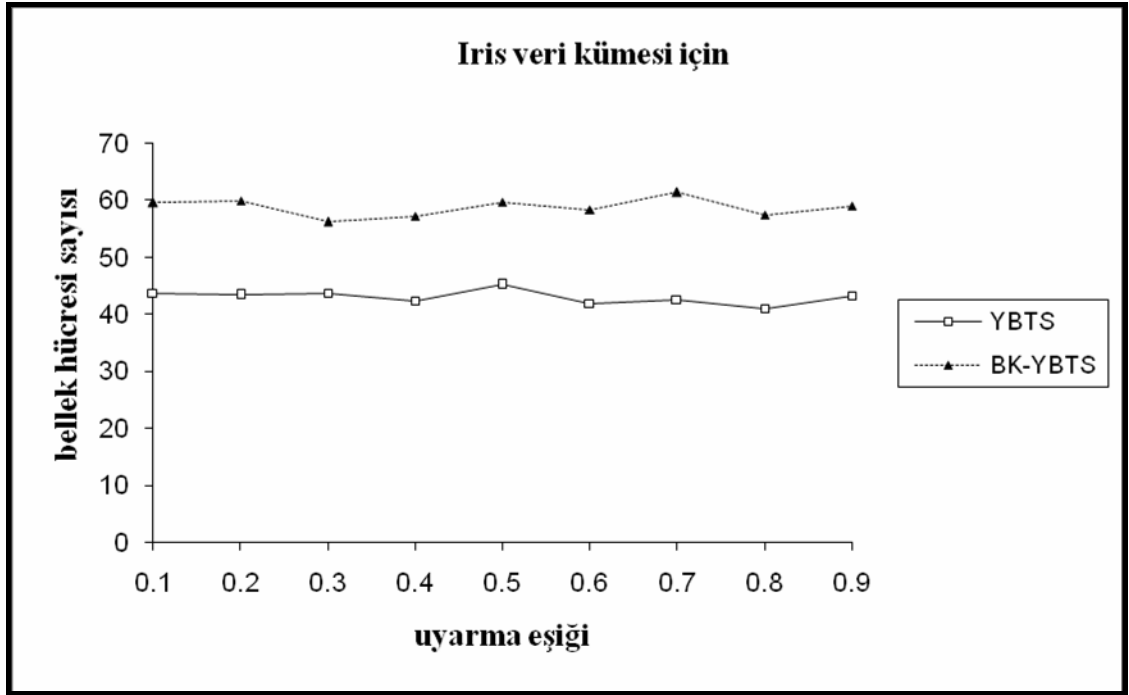
Uyarma eşiği parametresinin değişik değerleri için iki sistemin çalışma süreleri, iterasyon sayıları, bellek hücreleri sayıları ve sınıflandırma doğrulukları açısından karşılaştırılmasının grafiksel olarak gösterimleri sırasıyla Şekil 6.9 ile Şekil 6.12 arasında verilmiştir.



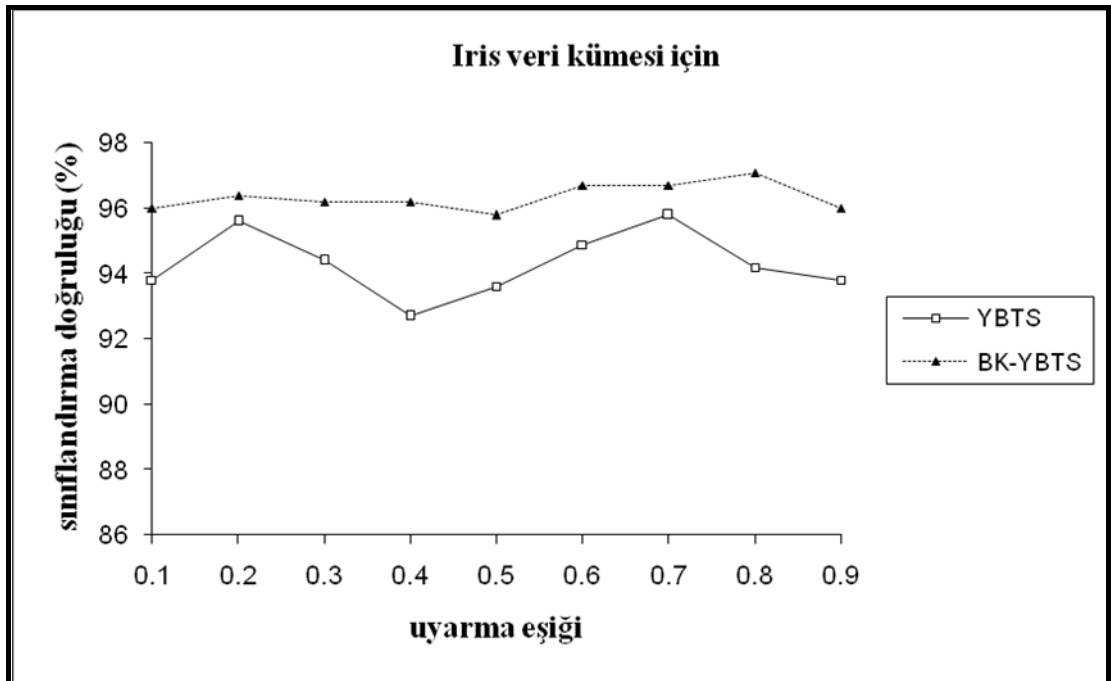
Şekil 6.9 Iris veri kümesinde uyarma eşiği parametresinin değişimine göre YBTS ve BK-YBTS'nin çalışma sürelerinin grafiksel gösterimi



Şekil 6.10 Iris veri kümesinde uyarma eşiği parametresinin değişimine göre YBTS ve BK-YBTS'nin iterasyon sayılarının grafiksel gösterimi



Şekil 6.11 Iris veri kümesinde uyarma eşiği parametresinin değişimine göre YBTS ve BK-YBTS'nin bellek hücreleri sayılarının grafiksel gösterimi



Şekil 6.12 Iris veri kümesinde uyarma eşiği parametresinin değişimine göre YBTS ve BK-YBTS'nin sınıflandırma doğruluklarının grafiksel gösterimi

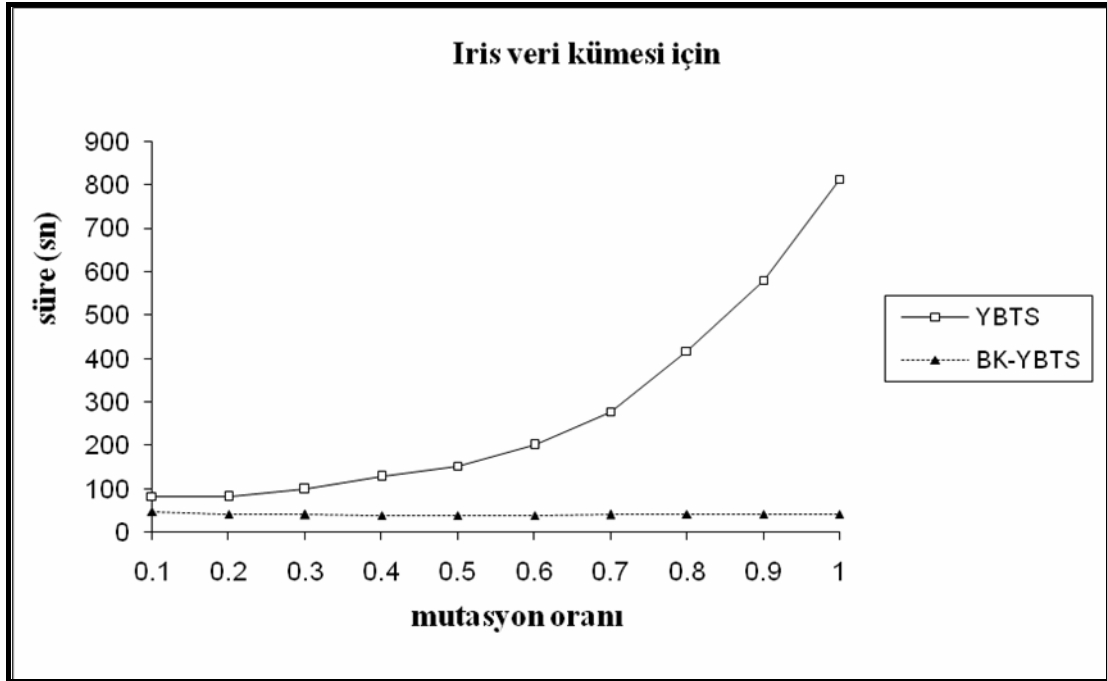
Uyarma eşiği parametresinin bütün değerleri için *BK-YBTS*, *YBTS'den* daha yüksek sınıflandırma doğruluğuna ulaşmıştır. *Uyarma eşiği* parametresi değişimi, her iki sistemde de çalışma zamanını, iterasyon sayısını ve bellek hücresi sayısını çok fazla etkilememiştir. *BK-YBTS'nin* çalışma süresi *YBTS'nin* çalışma süresinden daha az ölçülmüştür. Bellek hücreleri sayısı açısından *BK-YBTS'de YBTS'den* daha çok bellek hücresi oluşmuştur.

Uyarma eşiği parametresi için iki sistemin sınıflandırma performansı analiz edildikten sonra *Mutasyon oranı* parametresinin değişik değerleri için karşılaştırma sonuçları Tablo 6.6'da verilmiştir.

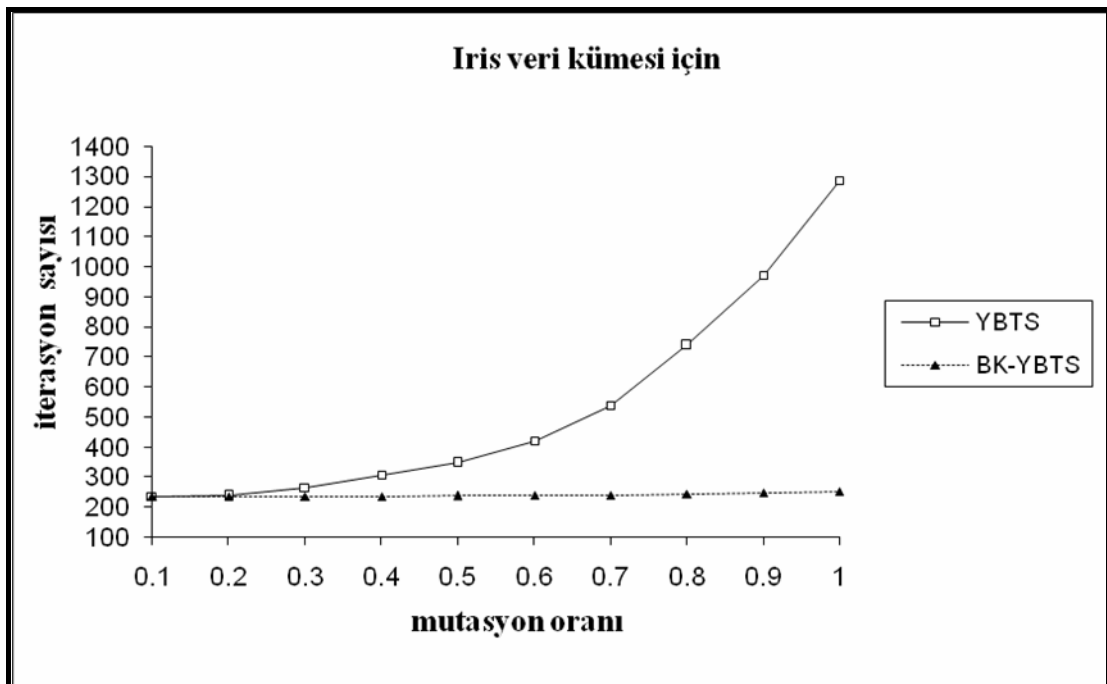
Tablo 6.6 Iris veri kümesinde mutasyon oranı parametresi için YBTS ve BK-YBTS karşılaştırma sonuçları

Mutasyon Oranı	Süre (sn)		İterasyon Sayısı		Bellek Hücresi Sayısı		Sınıflandırma Doğruluğu (%)	
	YBTS	BK-YBTS	YBTS	BK-YBTS	YBTS	BK-YBTS	YBTS	BK-YBTS
0.1	82.1	46.3	236.4	236.3	42.5	57.9	93.8	96.0
0.2	80.7	41.7	238.1	236.5	42.2	73.8	95.1	96.2
0.3	98.3	40.2	263.9	236.5	42.4	83.9	93.1	96.0
0.4	127.3	39.2	305.7	236.4	41.3	89.5	93.8	96.0
0.5	152.7	39.0	346.8	237.4	40.8	95.1	95.1	95.6
0.6	199.9	39.3	421.7	238.7	37.1	95.3	93.8	96.0
0.7	278.0	39.8	537.1	240.9	33.2	97.1	92.9	95.8
0.8	417.5	41.1	737.3	244.7	32.5	95.6	93.8	94.9
0.9	578.9	41.3	969.4	247.3	26.1	88.9	91.6	95.8
1.0	813.0	42.2	1286.6	251.1	26.3	85.3	93.3	95.1

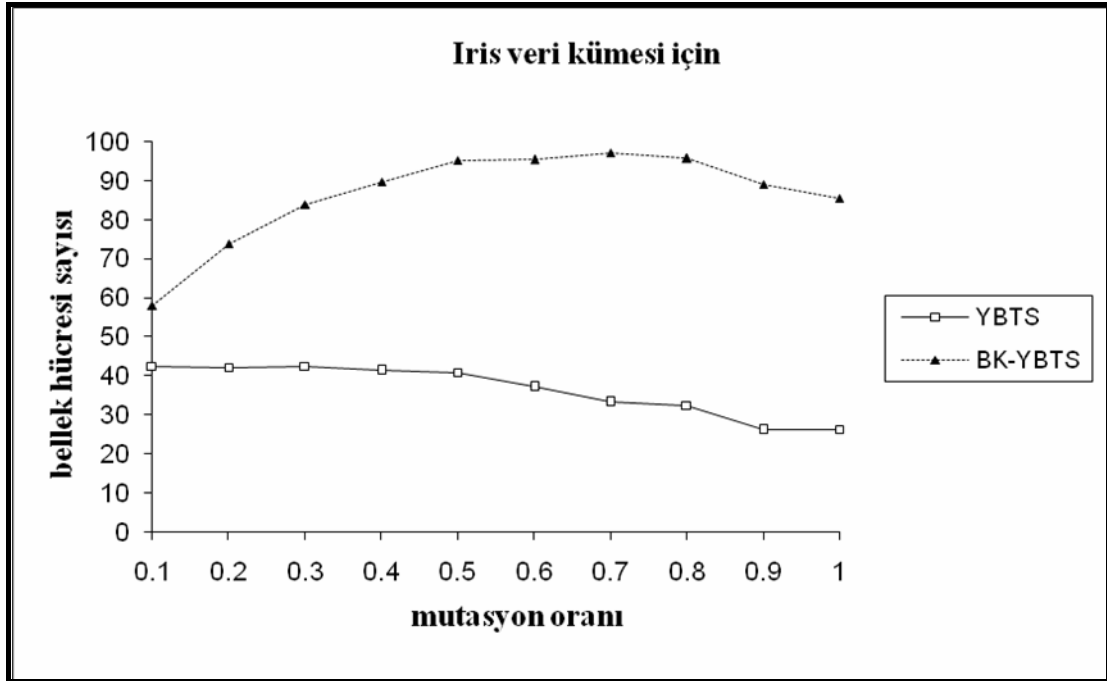
Mutasyon oranı parametresinin değişik değerleri için iki sistemin çalışma süreleri, iterasyon sayıları, bellek hücreleri sayıları ve sınıflandırma doğrulukları açısından karşılaştırılmasının grafiksel olarak gösterimleri sırasıyla Şekil 6.13 ile Şekil 6.16 arasında verilmiştir.



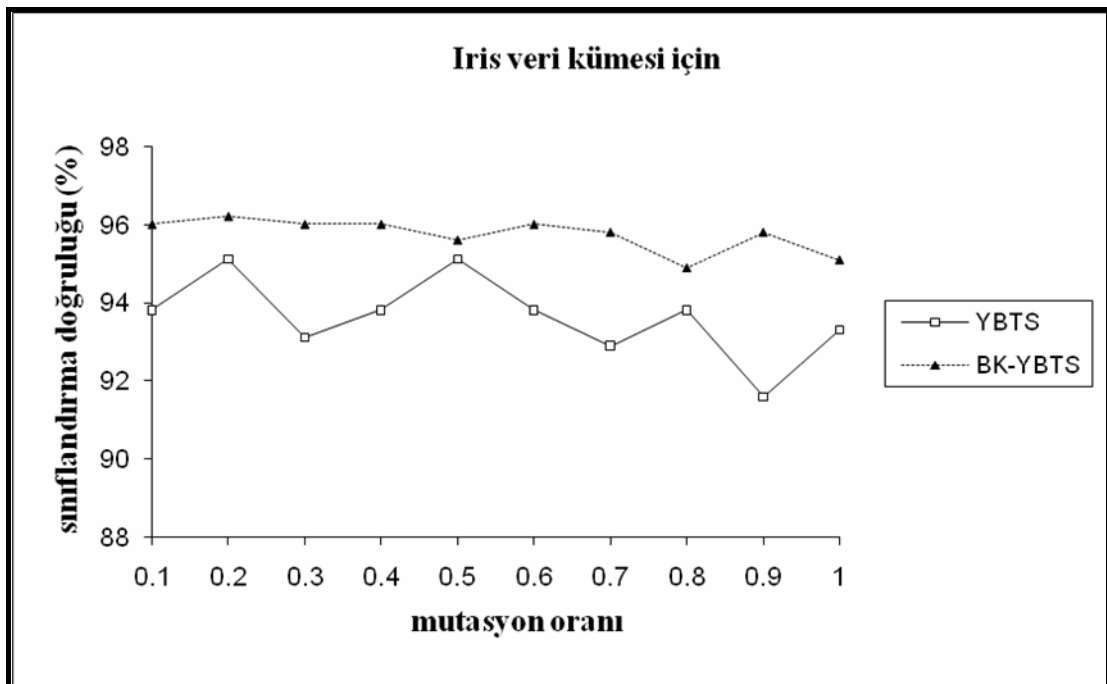
Şekil 6.13 Iris veri kümesinde mutasyon oranı parametresinin değişimine göre YBTS ve BK-YBTS'nin çalışma sürelerinin grafiksel gösterimi



Şekil 6.14 Iris veri kümesinde mutasyon oranı parametresinin değişimine göre YBTS ve BK-YBTS'nin iterasyon sayılarının grafiksel gösterimi



Şekil 6.15 Iris veri kümesinde mutasyon oranı parametresinin değişimine göre YBTS ve BK-YBTS'nin bellek hücreleri sayılarının grafiksel gösterimi



Şekil 6.16 Iris veri kümesinde mutasyon oranı parametresinin değişimine göre YBTS ve BK-YBTS'nin sınıflandırma doğruluklarının grafiksel gösterimi

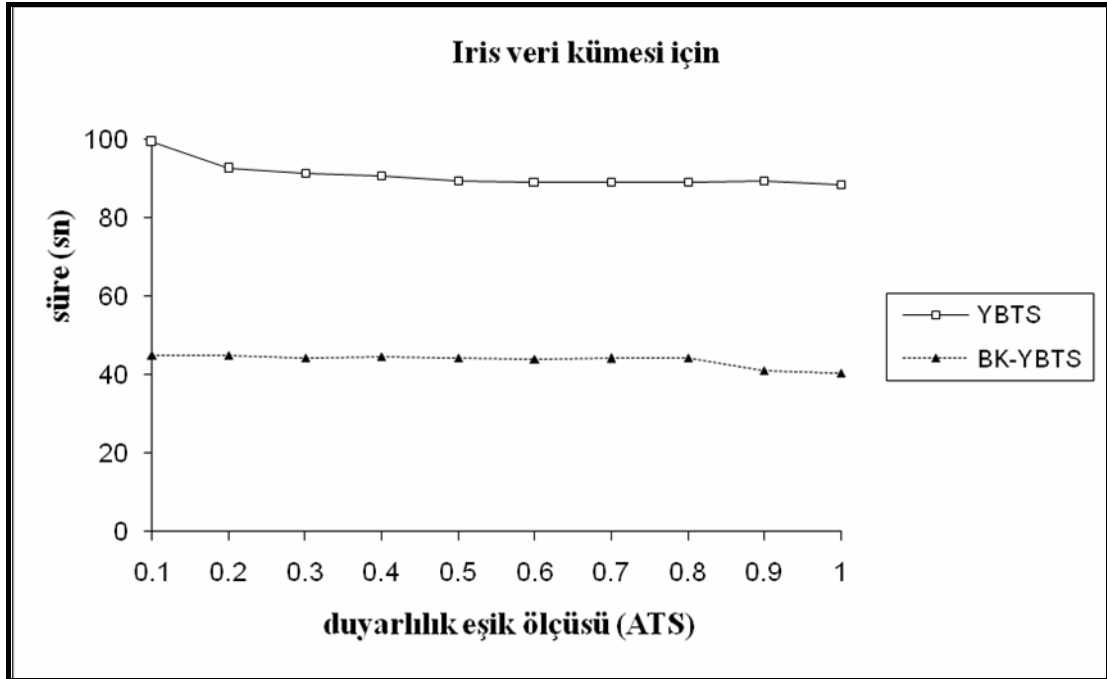
Mutasyon oranı parametresinin bütün değerleri için *BK-YBTS*, *YBTS*'den daha yüksek sınıflandırma doğruluğuna ulaşmıştır. *Mutasyon oranı* arttıkça *BK-YBTS*'de çalışma zamanında ve iterasyon sayısında önemli bir değişiklik gözlemlenmezken *YBTS*'de çalışma zamanında ve iterasyon sayısında ciddi artışlar gözlemlenmiştir. Bu durumun sebebi *YBTS*'de özellik mutasyonunun rasgele değer atanarak yapılmasıdır. *BK-YBTS*'de duyarlılık tabanlı mutasyon mekanizması kullanılmıştır. Bölüm 5.2.2'de *YBTS*'nin mutasyon mekanizması, Bölüm 5.4.3'te ise *BK-YBTS*'nin mutasyon mekanizması ayrıntılı olarak açıklanmıştır.

Mutasyon oranı parametresi için iki sistemin sınıflandırma performansı analiz edildikten sonra *Duyarlılık eşik ölçüsü* parametresinin değişik değerleri için karşılaştırma sonuçları Tablo 6.7'de verilmiştir.

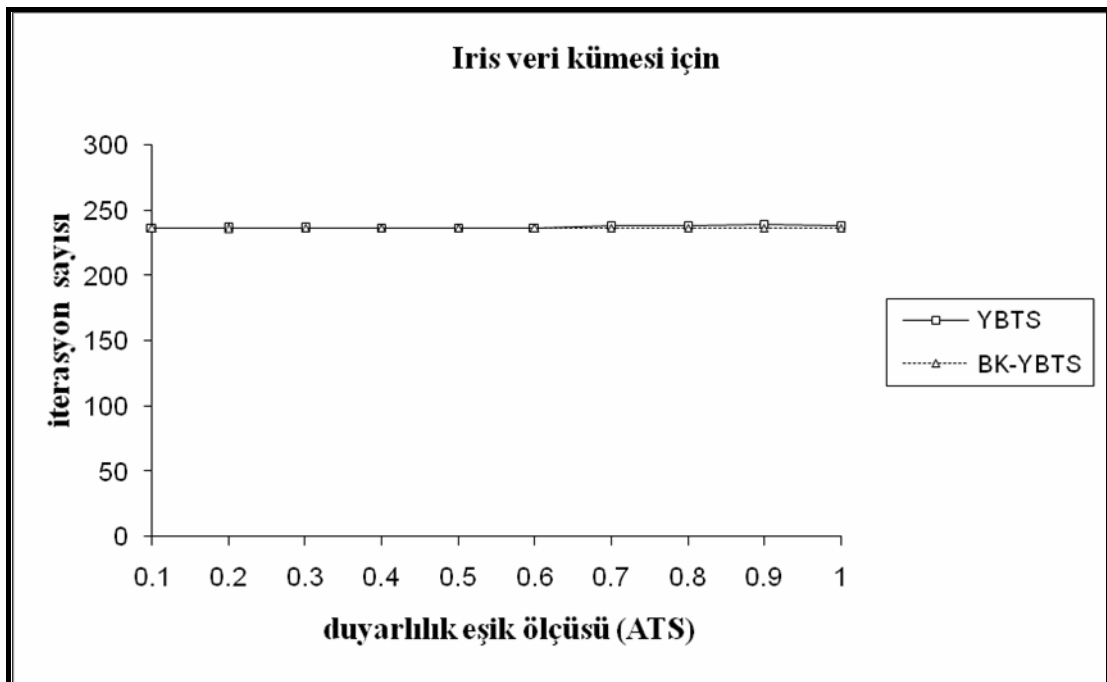
Tablo 6.7 Iris veri kümesinde duyarlılık eşik ölçüsü parametresi için YBTS ve BK-YBTS karşılaştırma sonuçları

Duyarlılık Eşik Ölçüsü (ATS)	Süre (sn)		İterasyon Sayısı		Bellek Hücresi Sayısı		Sınıflandırma Doğruluğu (%)	
	YBTS	BK-YBTS	YBTS	BK-YBTS	YBTS	BK-YBTS	YBTS	BK-YBTS
0.1	99.4	45.0	236.6	236.4	74.1	76.5	95.3	96.0
0.2	92.6	45.0	236.3	236.3	44.1	60.7	93.8	96.0
0.3	91.4	44.5	236.2	236.4	27.4	48.0	92.4	95.3
0.4	90.8	44.7	236.9	236.6	18.1	38.9	90.9	95.6
0.5	89.5	44.3	236.7	236.5	13.0	32.5	91.3	95.6
0.6	89.3	44.1	237.0	236.5	10.2	28.1	90.2	95.8
0.7	89.2	44.2	238.3	236.5	8.5	23.9	89.6	94.4
0.8	89.1	44.3	238.6	236.9	6.1	20.9	90.9	95.8
0.9	89.5	41.1	239.4	236.6	5.5	19.7	90.0	94.7
1.0	88.6	40.6	238.5	236.7	4.5	17.0	90.4	95.3

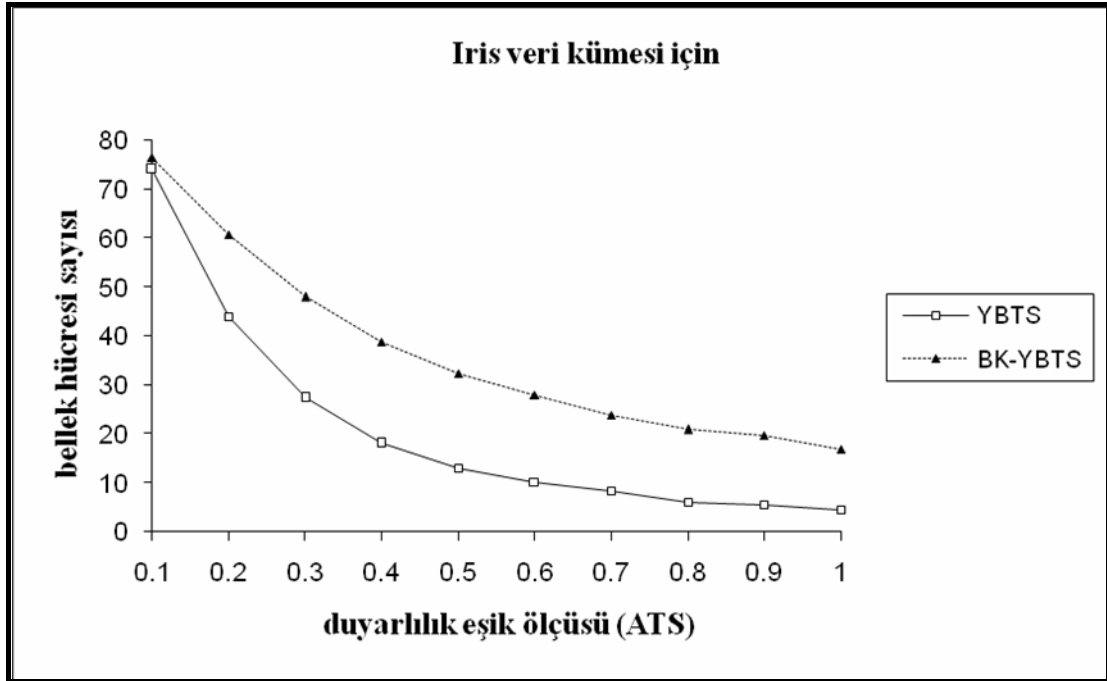
Duyarlılık eşik ölçüsü parametresinin değişik değerleri için iki sistemin çalışma süreleri, iterasyon sayıları, bellek hücreleri sayıları ve sınıflandırma doğrulukları açısından karşılaştırılmasının grafiksel olarak gösterimleri sırasıyla Şekil 6.17 ile Şekil 6.20 arasında verilmiştir.



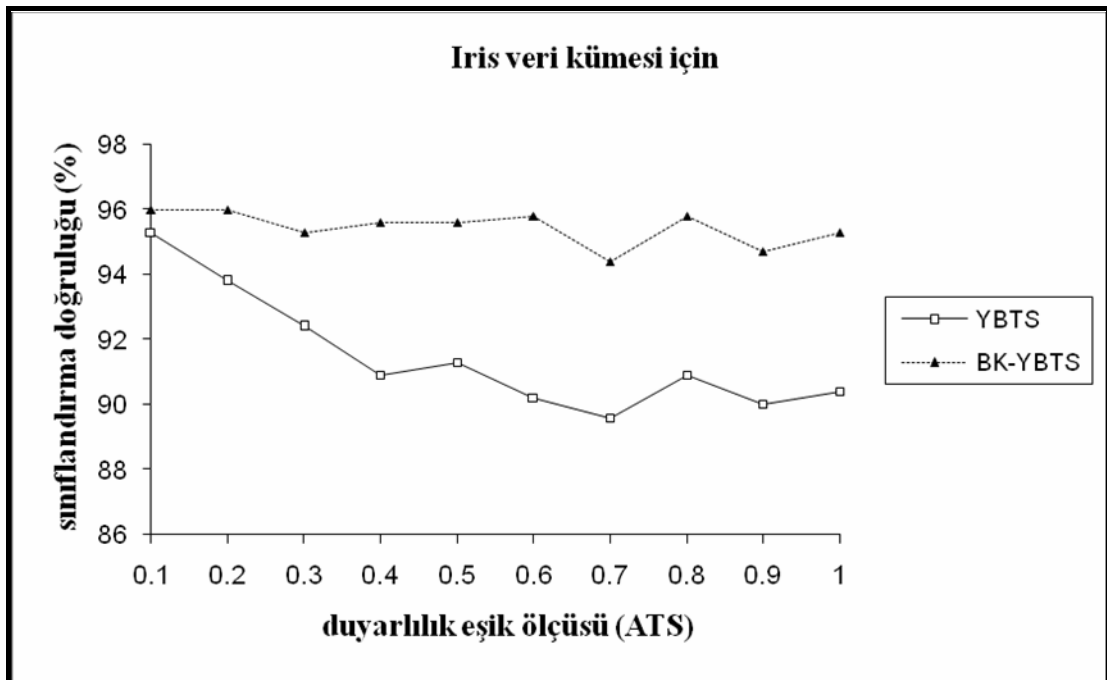
Şekil 6.17 Iris veri kümesinde duyarlılık eşik ölçüsü parametresinin değişimine göre YBTS ve BK-YBTS'nin çalışma sürelerinin grafiksel gösterimi



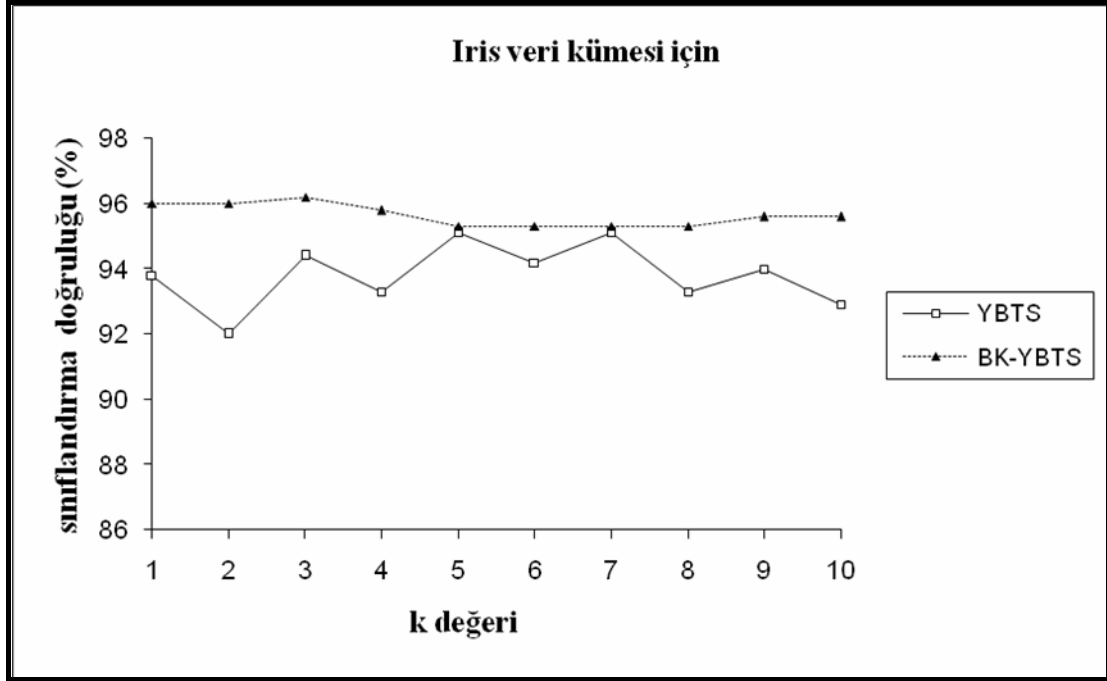
Şekil 6.18 Iris veri kümesinde duyarlılık eşik ölçüsü parametresinin değişimine göre YBTS ve BK-YBTS'nin iterasyon sayılarının grafiksel gösterimi



Şekil 6.19 Iris veri kümesinde duyarlılık eşik ölçüsü parametresinin değişimine göre YBTS ve BK-YBTS'nin bellek hücreleri sayılarının grafiksel gösterimi



Şekil 6.20 Iris veri kümesinde duyarlılık eşik ölçüsü parametresinin değişimine göre YBTS ve BK-YBTS'nin sınıflandırma doğruluklarının grafiksel gösterimi



Şekil 6.21 Iris veri kümesinde k değeri parametresinin değişimine göre YBTS ve BK-YBTS'nin sınıflandırma doğruluklarının grafiksel gösterimi

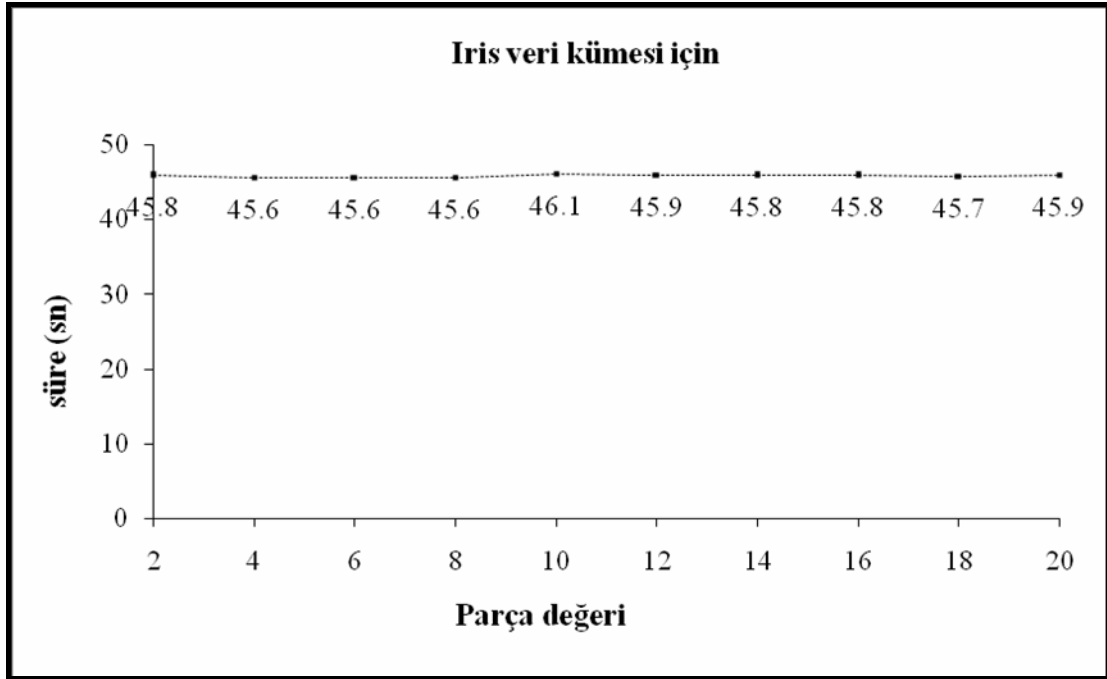
k değeri parametresinin aldığı bütün değerler için BK-YBTS, YBTS'den daha yüksek sınıflandırma doğruluklarına ulaşmıştır. Bununla beraber BK-YBTS'de ortalama olarak 57.9 bellek hücresi oluşurken, YBTS'de ortalama olarak 43.8 bellek hücresi oluşmuştur. Çalışma süreleri bakımından BK-YBTS ortalama olarak 51.9 saniye, YBTS ortalama olarak 89.3 saniye çalışmıştır.

Buraya kadar YBTS ve BK-YBTS için ortak parametreler olan çekirdek hücre sayısı, kaynak sayısı, uyarma eşiği, mutasyon oranı, duyarlılık eşik ölçüsü ve k parametrelerinin sınıflandırma performansına olan etkileri araştırılmıştır. Bu aşamadan sonra sadece BK-YBTS'ye özgü olan parça değeri parametresinin değişik değerleri için çalışma süreleri, iterasyon sayıları, bellek hücresi sayıları ve sınıflandırma doğrulukları Tablo 6.9'da verilmiştir.

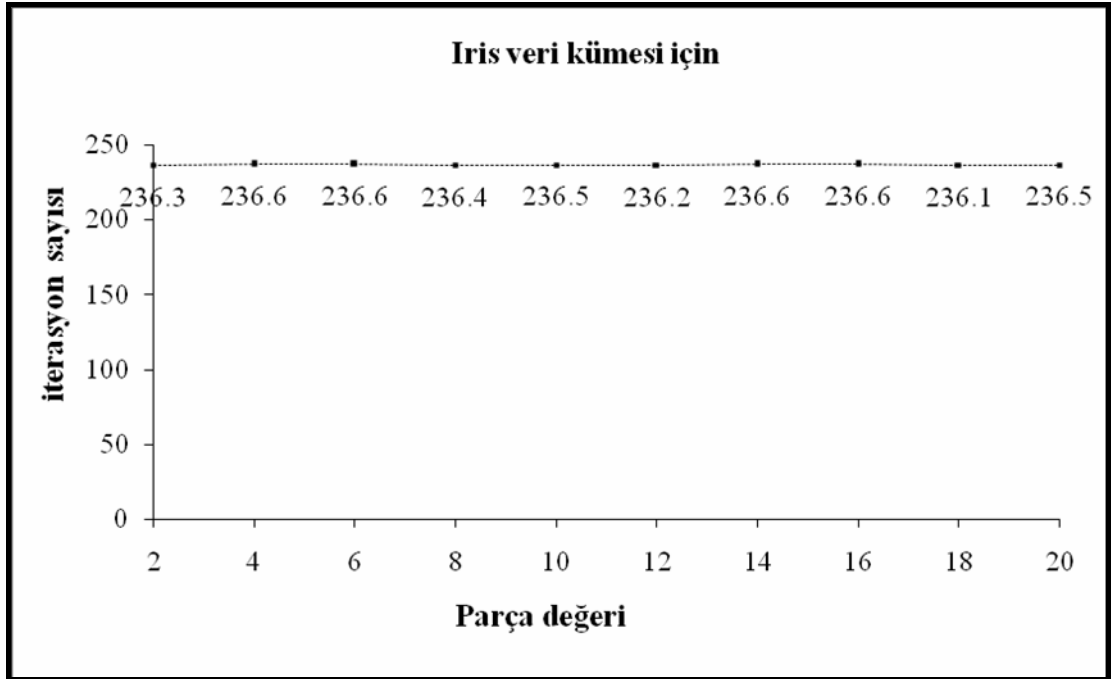
Parça değeri parametresinin aldığı değişik değerler için BK-YBTS'nin çalışma süreleri açısından grafiksel olarak gösterimi Şekil 6.22'de, iterasyon sayıları açısından grafiksel olarak gösterimi Şekil 6.23'te, bellek hücreleri sayıları açısından grafiksel olarak gösterimi Şekil 6.24'te ve sınıflandırma doğrulukları açısından grafiksel olarak gösterimi Şekil 6.25'te verilmiştir.

Tablo 6.9 Iris veri kümesinde parça değeri parametresi için BK-YBTS sonuçları

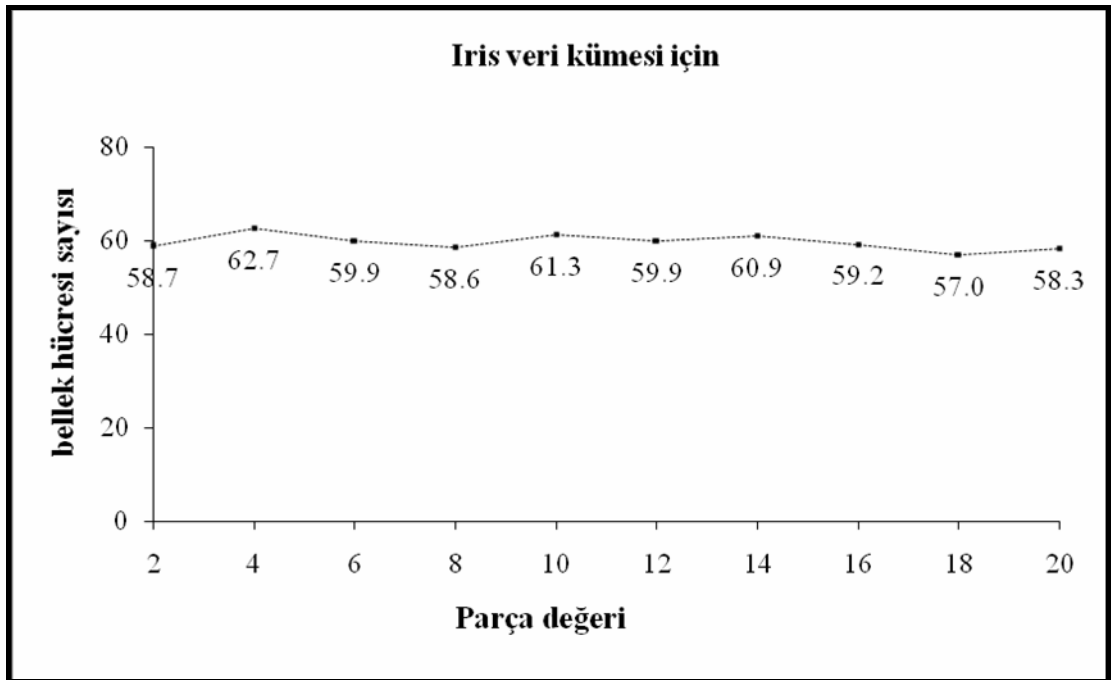
Parça Değeri	Süre (sn)	İterasyon Sayısı	Bellek Hücresi Sayısı	Sınıflandırma Doğruluğu (%)
2	45.8	236.3	58.7	96.0
4	45.6	236.6	62.7	96.2
6	45.6	236.6	59.9	95.6
8	45.6	236.4	58.6	96.2
10	46.1	236.5	61.3	96.4
12	45.9	236.2	59.9	96.2
14	45.8	236.6	60.9	94.9
16	45.8	236.6	59.2	95.3
18	45.7	236.1	57.0	95.1
20	45.9	236.5	58.3	95.1



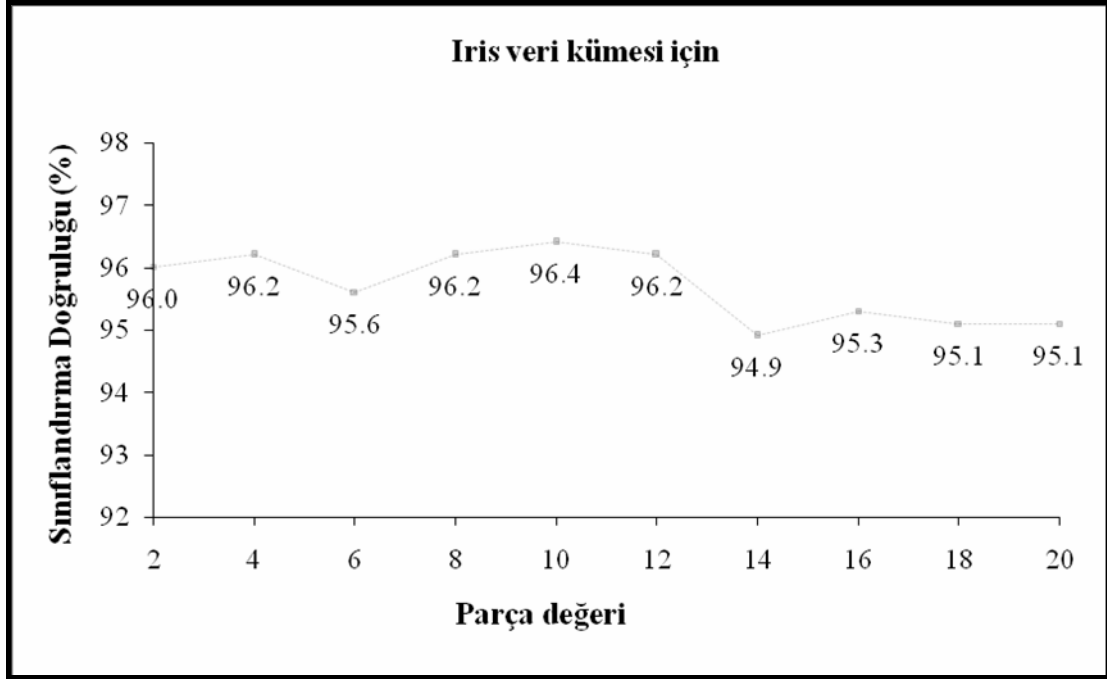
Şekil 6.22 Iris veri kümesinde parça parametresinin değişimine göre BK-YBTS'nin çalışma sürelerinin grafiksel gösterimi



Şekil 6.23 Iris veri kümesinde parça parametresinin değişimine göre BK-YBTS'nin iterasyon sayılarının grafiksel gösterimi



Şekil 6.24 Iris veri kümesinde parça parametresinin değişimine göre BK-YBTS'nin bellek hücreleri sayılarının grafiksel gösterimi



Şekil 6.25 Iris veri kümesinde parça parametresinin değişimine göre BK-YBTS'nin sınıflandırma doğruluklarının grafiksel gösterimi

Tablo 6.9'dan da anlaşılabileceği gibi *parça değeri* parametresi 10 olduğu durumda *BK-YBTS* en yüksek sınıflandırma doğruluğu olan % 96.4'e ulaşmıştır. *Parça değeri* parametresinin değişimi *BK-YBTS'nin* çalışma süresini, iterasyon sayısını ve bellek hücreleri sayısını çok fazla etkilememiştir.

Iris veri kümesinde tüm parametreler için genel olarak *BK-YBTS*, *YBTS'den* daha yüksek sınıflandırma doğruluğuna ulaşmıştır. Bu durum *BK-YBTS* ile beraber önerilen değişikliklerin, *YBTS'ye* göre sınıflandırma performansını arttırdığını göstermiştir.

Tablo 6.10'da *Iris* veri kümesi için deneme yanılma yoluyla elde edilen en iyi sınıflandırma doğruluğunu veren parametre değerleri verilmiştir. Tablo 6.11'de ise *Iris* veri kümesi için *BK-YBTS'nin* ve literatürdeki diğer sınıflandırıcıların sınıflandırma doğrulukları belirtilmiştir. *Parça değeri* 6 olduğu durumda, geliştirilen sistemin en yüksek sınıflandırma doğruluğu olan %97.3 değerine ulaşmıştır. Tablo 6.11'den de anlaşılaacağı üzere *BK-YBTS'nin* sınıflandırma doğruluğu *Iris* veri kümesi için *YBTS'den* daha yüksek elde edilmiştir. Ayrıca literatürde belirtilen diğer sınıflandırıcıların sınıflandırma doğruluğu ile kıyaslandığı kabul edilebilir sınıflandırma doğruluklarına ulaşmıştır.

Tablo 6.10 Iris veri kümesi için optimum BK-YBTS parametre değerleri

Parametre Adı	Değeri
<i>Çekirdek hücre sayısı</i>	1
<i>Kaynak sayısı</i>	200
<i>Uyarma eşiği</i>	0.9
<i>Mutasyon oranı</i>	0.2
<i>Duyarlılık eşik ölçüsü (ATS)</i>	0.3
<i>Çoğunluk oylaması için kullanılan k değeri</i>	3
<i>Parça değeri</i>	6

Tablo 6.11 Iris veri kümesi için BK-YBTS ve diğer sınıflandırıcıların sınıflandırma doğrulukları (Watkins 2005)

Metot	Doğruluk (%)
<i>Grobian (rough)</i>	100
<i>SSV</i>	98.0
<i>C-MLP2LN</i>	98.0
<i>PVM 2 rules</i>	98.0
<i>PVM 1 rule</i>	97.3
BK-YBTS	97.3
<i>YBTS</i>	96.7
<i>FuNe-I</i>	96.7
<i>NEFCLASS</i>	96.7
<i>CART</i>	96.0
<i>FUNN</i>	95.7

6.2.2. Thyroid veri kümesi uygulamaları

Tiroit bezi boyundaki en büyük bezdir. Nefes borusunu saran sol ve sağ tiroit bezleri ile ifade edilen iki kanat ile kelebek şeklindedir ve boynun ön kısmına yerleşmiştir. Tiroidin tek görevi tiroit hormonu salgılamaktır. Bu hormon hücrel aktivitenin arttığı vücudun neredeyse bütün dokularını etkilemektedir. Bu yüzden tiroit vücudun metabolizmasını düzenlemektedir (<http://www.endocrineweb.com/thyroid.html>).

Thyroid verisi tiroit bezinin ölçümünden elde edilmektedir. Tiroit iki aktif tiroit hormonu üretmektedir. Bu hormonlar *levothyroxine (T4)* ve *tri-iodothyronine'dir (T3)*. Bu hormonlar protein üretiminde, vücut sıcaklığının düzenlenmesinde ve enerji üretiminde önemli rol oynamaktadır.

Tiroit bezi oldukça genel olan bazı problemlerle sık sık karşılaşmaktadır. Çok az tiroit hormonu üretimi *hypo-thyroidism*'e sebep olmaktadır veya çok fazla tiroit hormonu üretimi *hyper-thyroidism*'e sebep olmaktadır. *Hypothyroidism*'in yani tiroit hormonu azlığının pek çok sebebi söz konusudur. Bu sebeplerden bazıları daha önceden bir tiroit operasyonu geçirilmesi, radyasyona maruz kalma, tiroidin iltihaplanması, iyot eksikliği, tiroit hormonu üreten enzimlerin eksikliği olabilir. *Hyperthyroidism* yani tiroit hormonu fazlalığının sebepleri arasında tiroidin iltihaplanması, çeşitli hastalıklar ve tiroit hormonu üretimi kontrolünün eksikliği sayılabilir. En genel sebeplerden biri ise *Graves* hastalığıdır. *Graves* hastalığı vücudun tiroit'te daha fazla tiroit hormonu üretmek için sürekli protein ürettiği zaman meydana gelmektedir (Özyılmaz ve Yıldırım 2002). Tiroit bozukluğu ciddiye alınmazsa tiroit fırtınası (ciddi *hyperthyroidism*) ve miksödem koma (*myxedema coma- hypothyroidism*'in son aşaması) gibi durumlarda ölüme sebep olabilir (Zhang ve Berardi 1998).

Thyroid verisi, Coomans ve ark. (1983) tarafından elde edilmiş ve UCI veri ambarında bulunmaktadır (<http://archive.ics.uci.edu/ml/>). Veri 215 adet hastadan elde edilmiştir. Bu kişiler üç grupta sınıflandırılmıştır. 150 adet kişi "*normal*" olarak adlandırılan sağlıklı kişilerdir. 35 adet kişi "*hyper*" olarak adlandırılan *hyperthyroidism* hastasıdır. 30 adet kişi ise "*hypo*" olarak adlandırılan *hypothyroidism* hastasıdır. Tablo 6.12'de *Thyroid* veri kümesi için sınıf dağılımı verilmektedir.

Veriler tiroid bezinin ölçüm değerleridir. Her bir özellik beş ayrı laboratuvar testi sonucundan oluşmaktadır. Bu test sonuçları:

- T3-resin uptake test (yüzde olarak)
- İzotopik yer değiştirme metodu ile ölçülen toplam serum thyroxin
- Radioimmuno testi ile ölçülen toplam serum tri-iodothyronine
- Radioimmuno testi ile ölçülen Basal thyroid-stimulating hormonu (TSH)
- Basal değeri karşılaştırmak için thyrotropin üreten hormonun 200 mikrogramının enjeksiyonundan sonra TSH değerinin maksimum mutlak farkıdır.

Tablo 6.12 Thyroid veri kümesinin sınıf dağılımı

Sınıf Adı	Veri Sayısı	Sınıf Dağılımı (%)
<i>Normal</i>	150	69.77
<i>Hyper</i>	35	16.28
<i>Hypo</i>	30	13.95

Yapılan uygulamalarda, bu veri kümesi için literatürdeki sınıflandırıcıların çoğunun kullandığı *10-kat çapraz onaylama yöntemi* kullanılmıştır. Her bir küme bir kez test kümesi olacak şekilde uygulanmıştır. Rapor edilen sonuç *10 eğitim-test* kombinasyonunun her birinin üç kez çalıştırılması sonucu elde edilen değerlerin ortalamasıdır.

Thyroid veri kümesi için *YBTS*'den farklı olarak *BK-YBTS* ile beraber önerilen değişikliklerin etkisini araştırmak amacıyla *çekirdek hücre sayısı*, *kaynak sayısı*, *uyarma eşiği*, *mutasyon oranı*, *duyarlılık eşik ölçüsü* ve *k* parametreleri için Tablo 6.13'de belirtilen parametre değerleri temel alınarak *YBTS* ve *BK-YBTS* karşılaştırılmıştır.

Tablo 6.13 Thyroid veri kümesi için YBTS ve BK-YBTS'nin başlangıç parametre değerleri

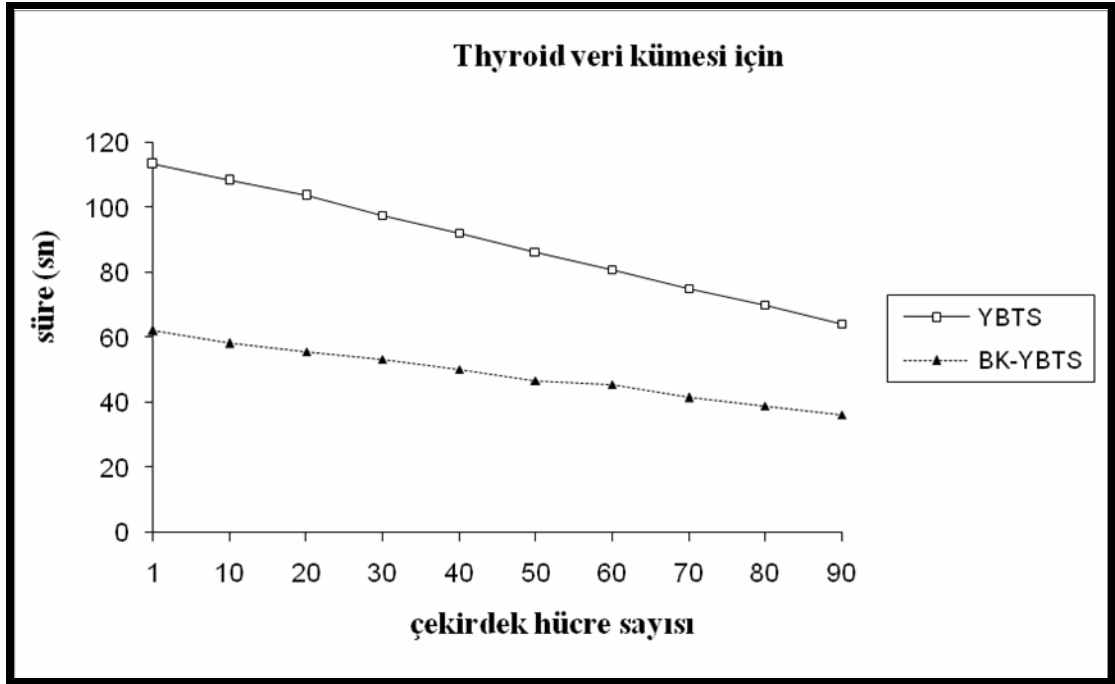
Parametre Adı	Parametre Değeri
<i>Çekirdek hücre sayısı</i>	1
<i>Kaynak sayısı</i>	200
<i>Uyarma eşiği</i>	0.9
<i>Mutasyon oranı</i>	0.1
<i>Duyarlılık eşik ölçüsü</i>	0.2
<i>Çoğunluk oylaması için kullanılan k değeri</i>	1
<i>Parça değeri (BK-YBTS için)</i>	2

Çekirdek hücre sayısı parametresinin aldığı değişik değerler için *YBTS* ve *BK-YBTS* arasındaki karşılaştırma sonuçları Tablo 6.14'te verilmiştir.

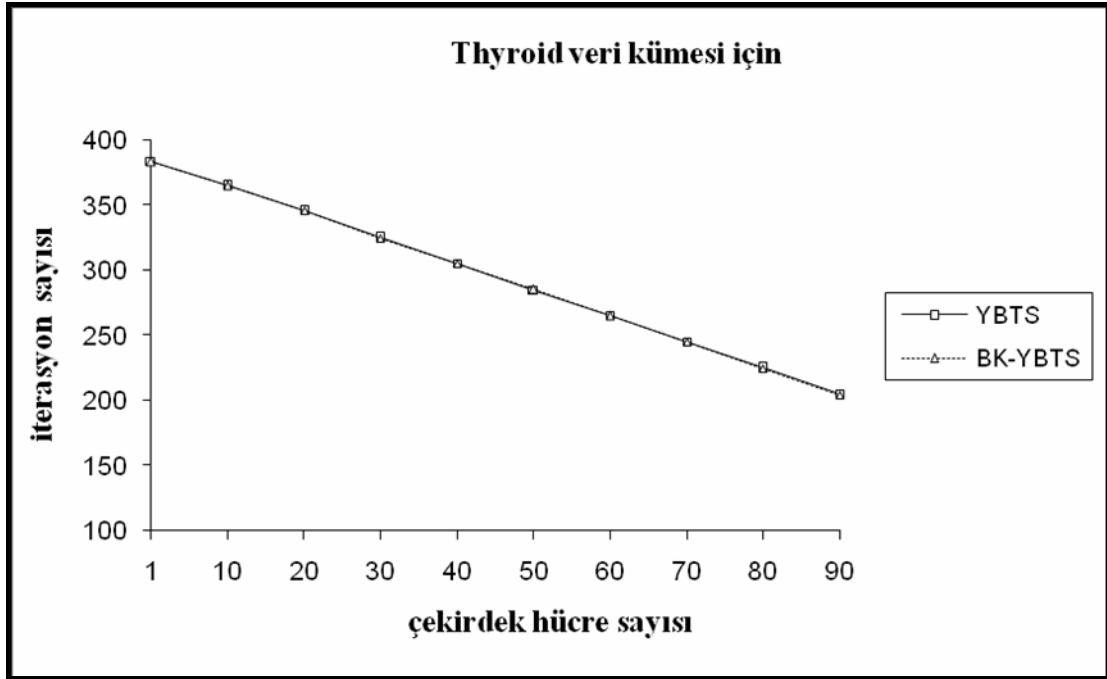
Çekirdek hücre sayısı parametresinin aldığı değişik değerler için iki sistemin çalışma süreleri açısından karşılaştırılmasının grafiksel olarak gösterimi Şekil 6.26'da, iterasyon sayıları açısından karşılaştırılmasının grafiksel olarak gösterimi Şekil 6.27'de, bellek hücreleri sayıları açısından karşılaştırılmasının grafiksel olarak gösterimi Şekil 6.28'de ve sınıflandırma doğrulukları açısından karşılaştırılmasının grafiksel olarak gösterimi Şekil 6.29'da verilmiştir.

Tablo 6.14 Thyroid veri kümesinde çekirdek hücre sayısı parametresi için YBTS ve BK-YBTS karşılaştırma sonuçları

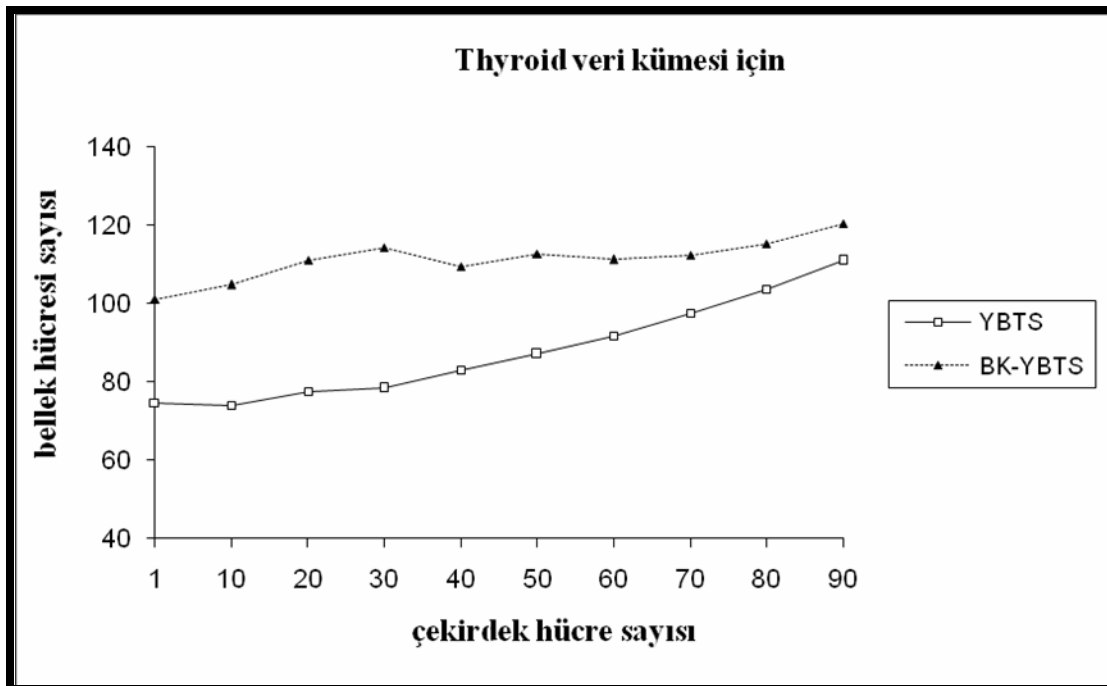
Çekirdek Hücre Sayısı	Süre (sn)		İterasyon Sayısı		Bellek Hücresi Sayısı		Sınıflandırma Doğruluğu (%)	
	YBTS	BK-YBTS	YBTS	BK-YBTS	YBTS	BK-YBTS	YBTS	BK-YBTS
1	113.3	62.1	383.4	383.2	74.7	101.1	94.2	94.4
10	108.3	58.2	364.8	364.9	74.1	104.9	94.0	94.2
20	103.6	55.7	345.5	345.7	77.5	111.1	93.8	93.8
30	97.6	53.3	325.0	324.8	78.4	114.4	93.1	93.4
40	92.2	50.1	305.1	304.8	83.2	109.6	93.5	93.2
50	86.4	46.8	285.0	285.2	87.1	112.7	93.4	93.6
60	80.9	45.5	265.0	264.8	91.7	111.3	93.0	93.2
70	75.2	41.5	245.0	244.7	97.7	112.5	93.8	93.5
80	70.0	39.0	225.0	224.4	103.7	115.4	94.3	94.1
90	64.2	36.2	205.0	204.2	111.0	120.4	94.5	94.3



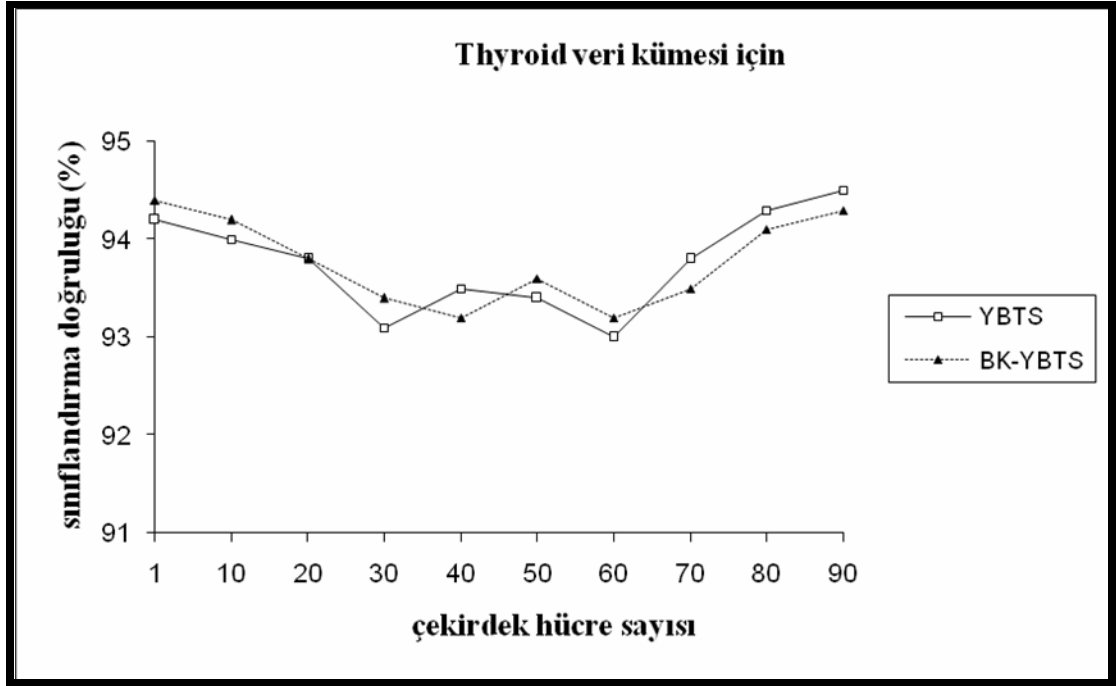
Şekil 6.26 Thyroid veri kümesinde çekirdek hücre sayısı parametresinin değişimine göre YBTS ve BK-YBTS'nin çalışma sürelerinin grafiksel gösterimi



Şekil 6.27 Thyroid veri kümesinde çekirdek hücre sayısı parametresinin değişimine göre YBTS ve BK-YBTS'nin iterasyon sayılarının grafiksel gösterimi



Şekil 6.28 Thyroid veri kümesinde çekirdek hücre sayısı parametresinin değişimine göre YBTS ve BK-YBTS'nin bellek hücreleri sayılarının grafiksel gösterimi



Şekil 6.29 Thyroid veri kümesinde çekirdek hücre sayısı parametresinin değişimine göre YBTS ve BK-YBTS'nin sınıflandırma doğruluklarının grafiksel gösterimi

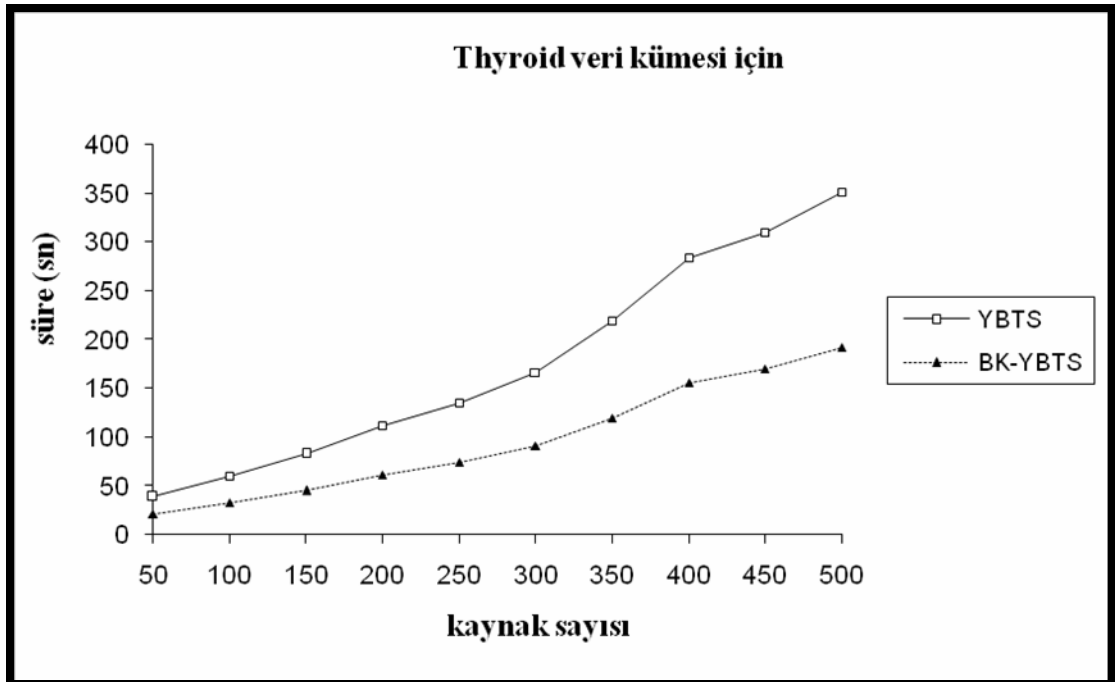
Çekirdek hücre sayısı arttığı zaman, eğitimin başlangıcında antijenler bellek hücresi ve antikör olarak belirlendiği için doğal olarak *YBTS* ve *BK-YBTS'nin* çalışma zamanı ve iterasyon sayısı azalmış, sistemdeki bellek hücresi sayısı ise artmıştır. *BK-YBTS* ve *YBTS* iterasyon sayısı olarak birbirlerine yakın çalışmakla beraber antikörlere kaynak tahsisi rutininde *YBTS* daha çok zaman harcadığı için *YBTS'nin* çalışma süresi, *YBTS'den* fazla ölçülmüştür. Bellek hücreleri sayısı *BK-YBTS* için *YBTS'den* daha fazla tespit edilmiştir. *Çekirdek hücre sayısı* için *YBTS* ve *BK-YBTS* yakın sınıflandırma doğruluklarına ulaşmışlardır. *Çekirdek hücre sayısı*, 1, 10, 30, 50 ve 60 değerlerinde *BK-YBTS*, *YBTS'den* daha yüksek sınıflandırma doğruluklarına ulaşmıştır.

Çekirdek hücre sayısı parametresi için iki sistemin sınıflandırma performansı analiz edildikten sonra *kaynak sayısı* parametresinin değişik değerleri için karşılaştırma sonuçları Tablo 6.15'te verilmiştir.

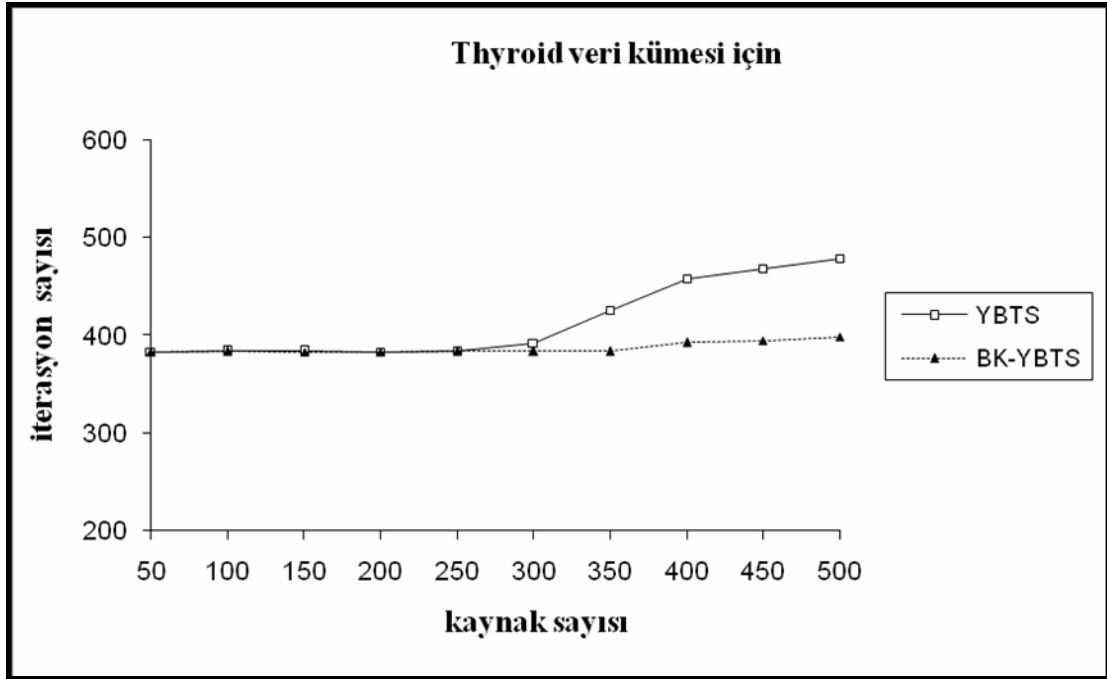
Kaynak sayısı parametresinin değişik değerleri için iki sistemin çalışma süreleri, iterasyon sayıları, bellek hücreleri sayıları ve sınıflandırma doğrulukları açısından karşılaştırılmasının grafiksel olarak gösterimleri sırasıyla Şekil 6.30 ile Şekil 6.33 arasında verilmiştir.

Tablo 6.15 Thyroid veri kümesinde kaynak sayısı parametresi için YBTS ve BK-YBTS karşılaştırma sonuçları

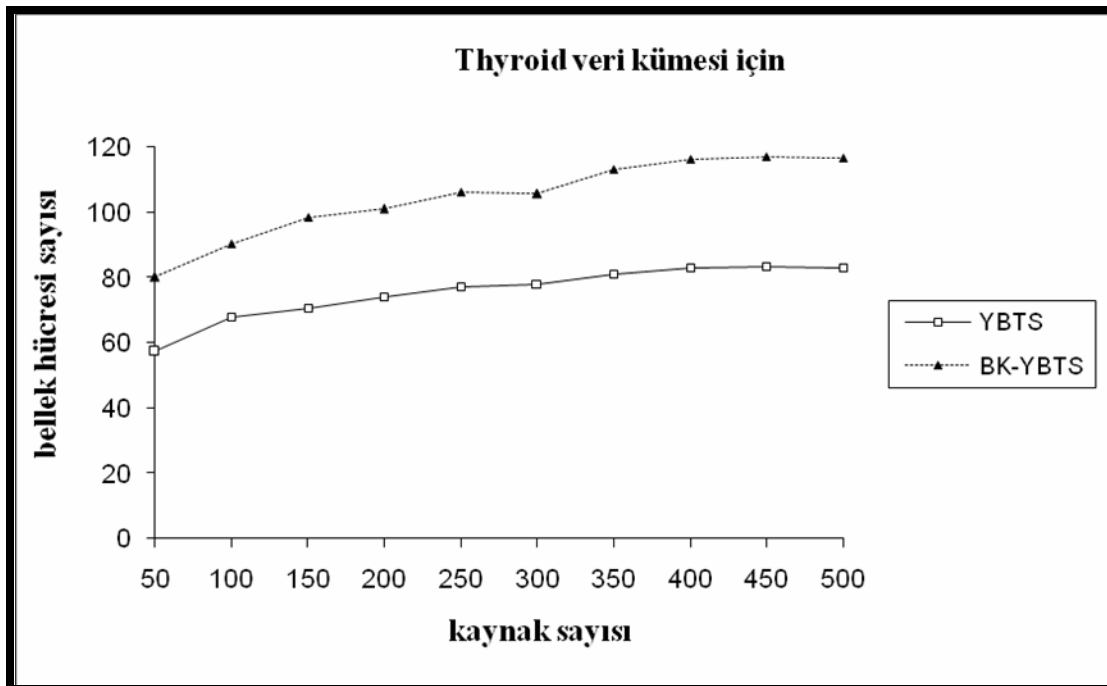
Kaynak Sayısı	Süre (sn)		İterasyon Sayısı		Bellek Hücresi Sayısı		Sınıflandırma Doğruluğu (%)	
	YBTS	BK-YBTS	YBTS	BK-YBTS	YBTS	BK-YBTS	YBTS	BK-YBTS
50	38.8	21.2	383.4	383.4	57.3	80.2	93.4	93.1
100	59.8	32.7	383.7	383.6	68.0	90.3	94.1	93.9
150	82.6	45.2	383.5	383.4	70.8	98.4	94.3	94.4
200	111.9	61.3	383.4	383.2	74.1	101.1	94.2	94.4
250	134.9	73.9	384.7	383.5	77.2	106.3	94.1	94.2
300	166.0	90.9	391.9	383.5	77.9	105.7	94.2	94.8
350	219.1	120.0	425.9	383.7	81.2	113.2	94.0	94.6
400	284.0	155.5	457.8	392.8	83.1	116.4	94.3	94.5
450	310.0	169.8	468.1	394.2	83.5	117.1	94.5	94.8
500	350.9	192.	478.6	398.5	83.1	116.8	94.3	94.4



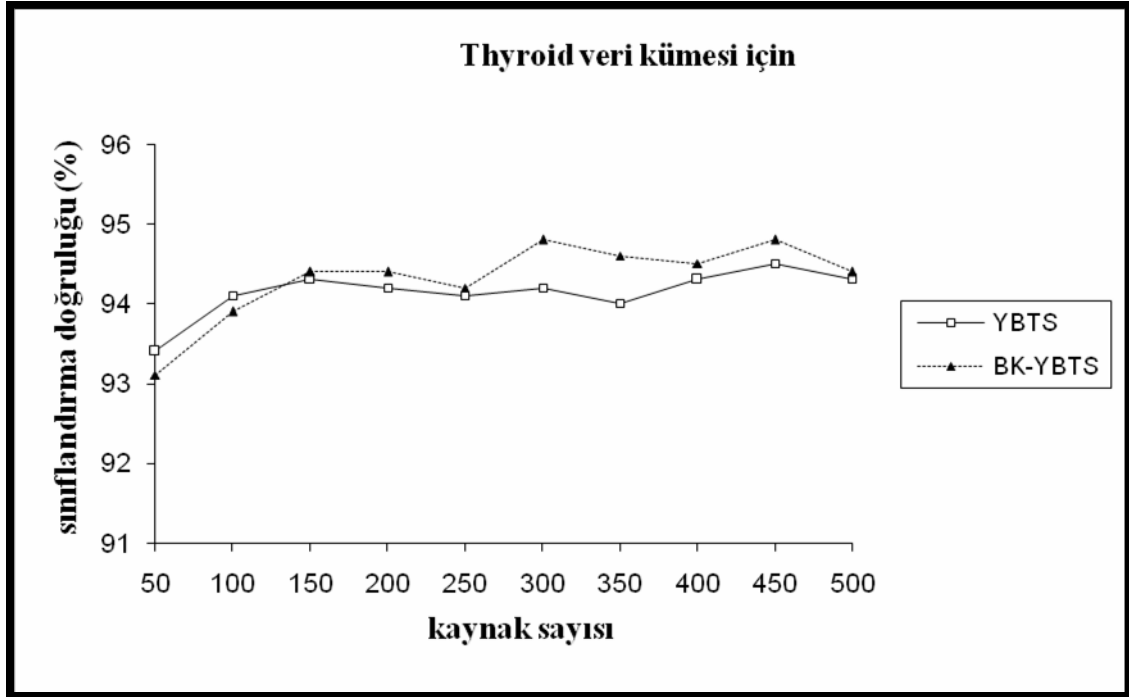
Şekil 6.30 Thyroid veri kümesinde kaynak sayısı parametresinin değişimine göre YBTS ve BK-YBTS'nin çalışma sürelerinin grafiksel gösterimi



Şekil 6.31 Thyroid veri kümesinde kaynak sayısı parametresinin değişimine göre YBTS ve BK-YBTS'nin iterasyon sayılarının grafiksel gösterimi



Şekil 6.32 Thyroid veri kümesinde kaynak sayısı parametresinin değişimine göre YBTS ve BK-YBTS'nin bellek hücreleri sayılarının grafiksel gösterimi



Şekil 6.33 Thyroid veri kümesinde kaynak sayısı parametresinin değişimine göre YBTS ve BK-YBTS'nin sınıflandırma doğruluklarının grafiksel gösterimi

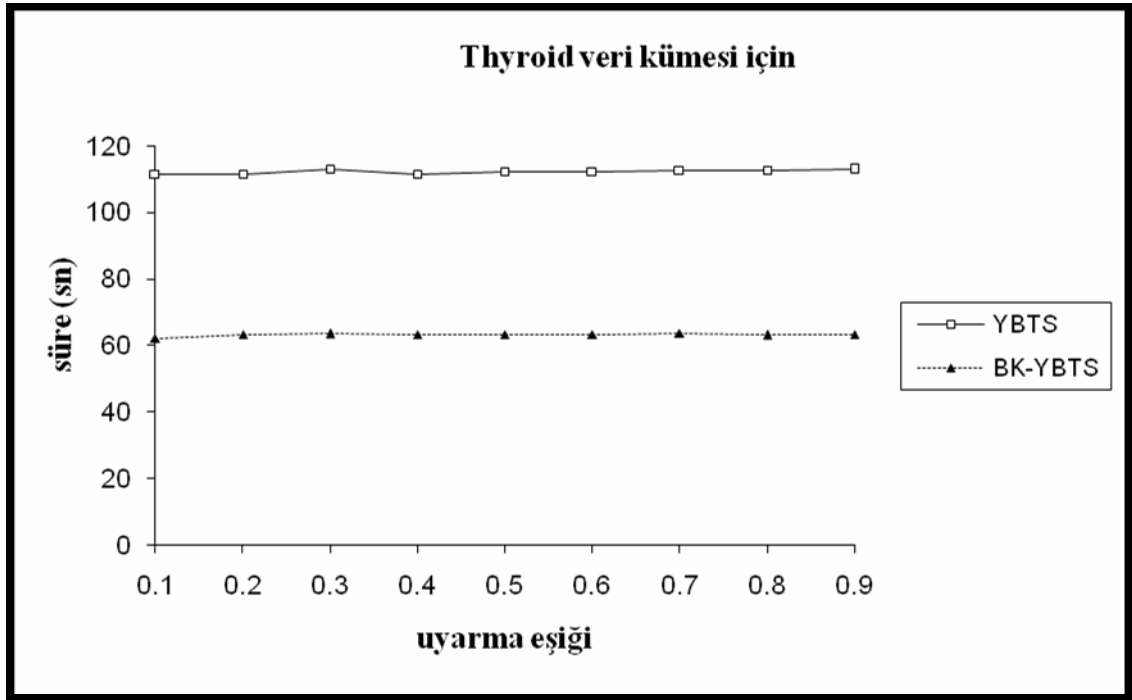
Kaynak sayısı parametresi değeri arttığı zaman, *YBTS* ve *BK-YBTS*'nin çalışma zamanı, iterasyon sayısı ve bellek hücresi sayısı artmıştır. *BK-YBTS*'nin iterasyon sayısı ve çalışma süresi *YBTS*'nin iterasyon sayısı ve çalışma süresinden daha az tespit edilirken, bellek hücreleri sayısı açısından *BK-YBTS*'de *YBTS*'den daha fazla bellek hücresi oluşmuştur. *Kaynak sayısı* parametresinin iki değeri hariç *BK-YBTS*, *YBTS*'den daha yüksek sınıflandırma doğruluğuna ulaşmıştır. Sadece *kaynak sayısı* parametresi 50 ve 100 değerlerinde, *BK-YBTS*, *YBTS*'den daha düşük sınıflandırma doğruluklarına ulaşmıştır.

Kaynak sayısı parametresi için iki sistemin sınıflandırma performansı analiz edildikten sonra *uyarma eşiği* parametresinin değişik değerleri için karşılaştırma sonuçları Tablo 6.16'da verilmiştir.

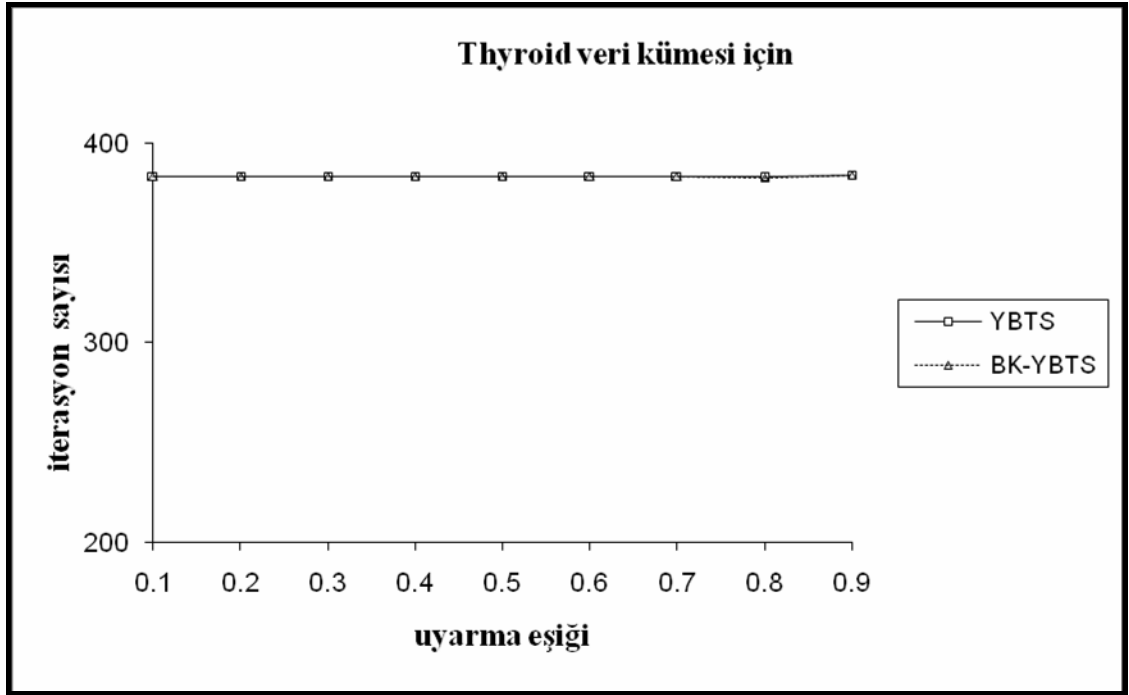
Uyarma eşiği parametresinin değişik değerleri için iki sistemin çalışma süreleri, iterasyon sayıları, bellek hücreleri sayıları ve sınıflandırma doğrulukları açısından karşılaştırılmasının grafiksel olarak gösterimleri sırasıyla Şekil 6.34 ile Şekil 6.37 arasında verilmiştir.

Tablo 6.16 Thyroid veri kümesinde uyarma eşiği parametresi için YBTS ve BK-YBTS karşılaştırma sonuçları

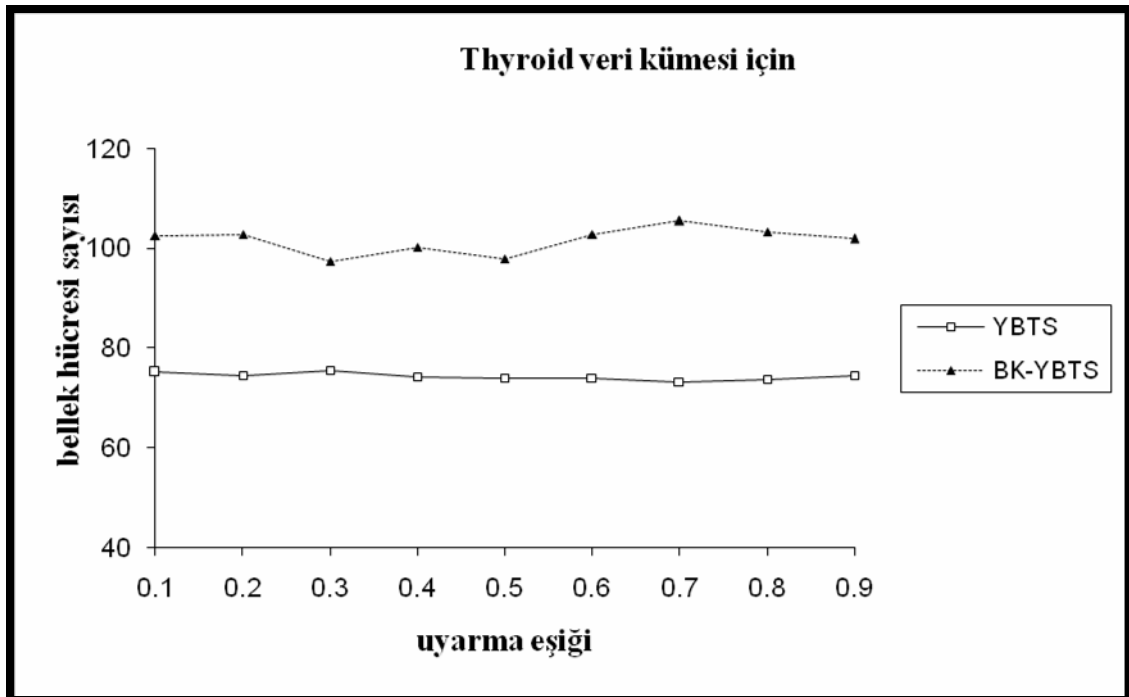
Uyarma Eşiği	Süre (sn)		İterasyon Sayısı		Bellek Hücresi Sayısı		Sınıflandırma Doğruluğu (%)	
	YBTS	BK-YBTS	YBTS	BK-YBTS	YBTS	BK-YBTS	YBTS	BK-YBTS
0.1	111.5	62.2	383.0	383.0	75.1	102.5	92.7	92.9
0.2	111.6	63.1	383.0	383.0	74.4	102.8	92.8	93.1
0.3	113.0	63.4	383.0	383.0	75.4	97.3	93.1	93.5
0.4	111.4	63.3	383.0	383.0	74.1	100.2	93.6	93.7
0.5	112.1	63.1	383.0	383.0	73.9	97.8	93.6	93.8
0.6	112.3	63.1	383.0	383.0	73.9	102.8	93.3	94.0
0.7	112.6	63.7	383.0	383.0	73.1	105.4	94.3	94.3
0.8	112.6	63.0	383.0	382.7	73.7	103.3	93.8	94.2
0.9	112.8	63.2	383.6	383.8	74.5	101.8	94.2	94.4



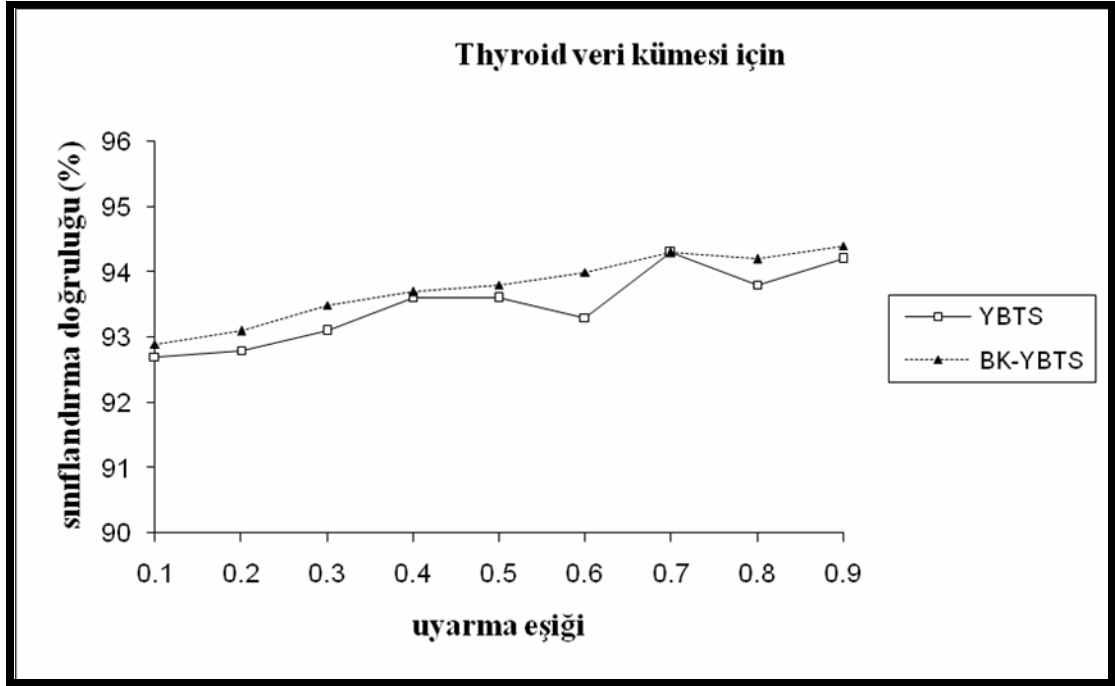
Şekil 6.34 Thyroid veri kümesinde uyarma eşiği parametresinin değişimine göre YBTS ve BK-YBTS'nin çalışma sürelerinin grafiksel gösterimi



Şekil 6.35 Thyroid veri kümesinde uyarma eşiği parametresinin değişimine göre YBTS ve BK-YBTS'nin iterasyon sayılarının grafiksel gösterimi



Şekil 6.36 Thyroid veri kümesinde uyarma eşiği parametresinin değişimine göre YBTS ve BK-YBTS'nin bellek hücreleri sayılarının grafiksel gösterimi



Şekil 6.37 Thyroid veri kümesinde uyarma eşiği parametresinin değişimine göre YBTS ve BK-YBTS'nin sınıflandırma doğruluklarının grafiksel gösterimi

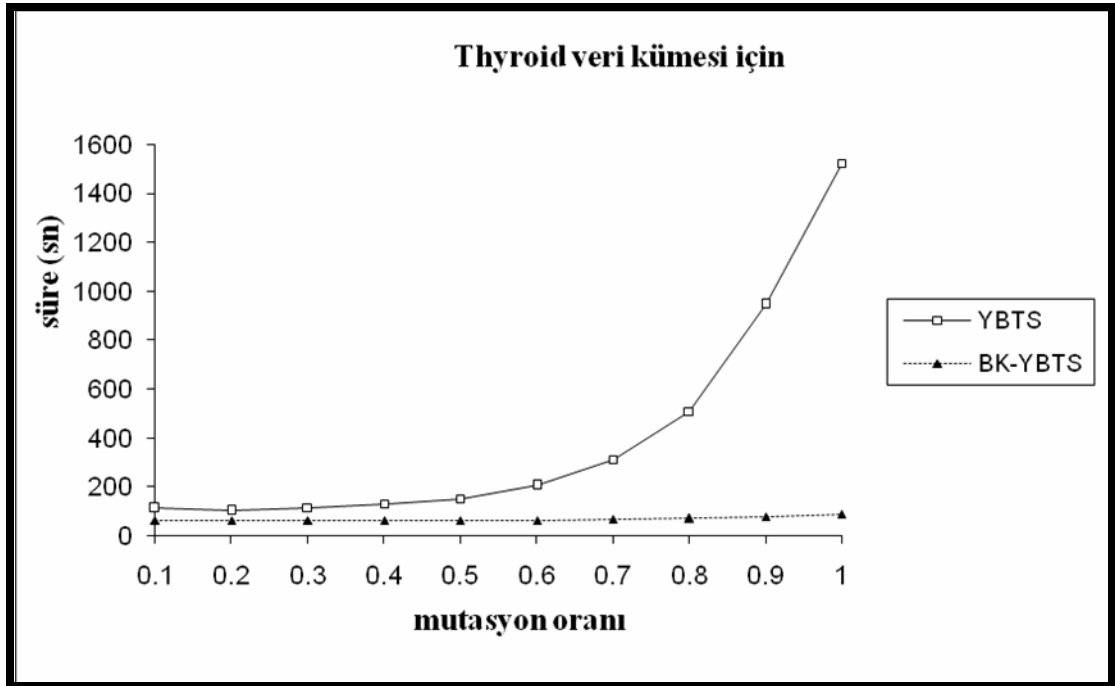
Uyarma eşiği parametresinin bir değeri hariç *BK-YBTS*, *YBTS*'den daha yüksek sınıflandırma doğruluğuna ulaşmıştır. *Uyarma eşiği* parametresi 0.7 değerinde iki sistem aynı sınıflandırma doğruluklarına ulaşmıştır. *Uyarma eşiği* parametresinin değerinin değişimi, her iki sistemde de çalışma zamanını, iterasyon sayısını ve bellek hücresi sayısını çok fazla etkilememiştir. *BK-YBTS*'nin çalışma süresi *YBTS*'nin çalışma süresinden daha az ölçülmüştür. Bellek hücreleri sayısı açısından *BK-YBTS*'de *YBTS*'den daha çok bellek hücresi oluşmuştur.

Uyarma eşiği parametresi için iki sistemin sınıflandırma performansı analiz edildikten sonra *mutasyon oranı* parametresinin değişik değerleri için karşılaştırma sonuçları Tablo 6.17'de verilmiştir.

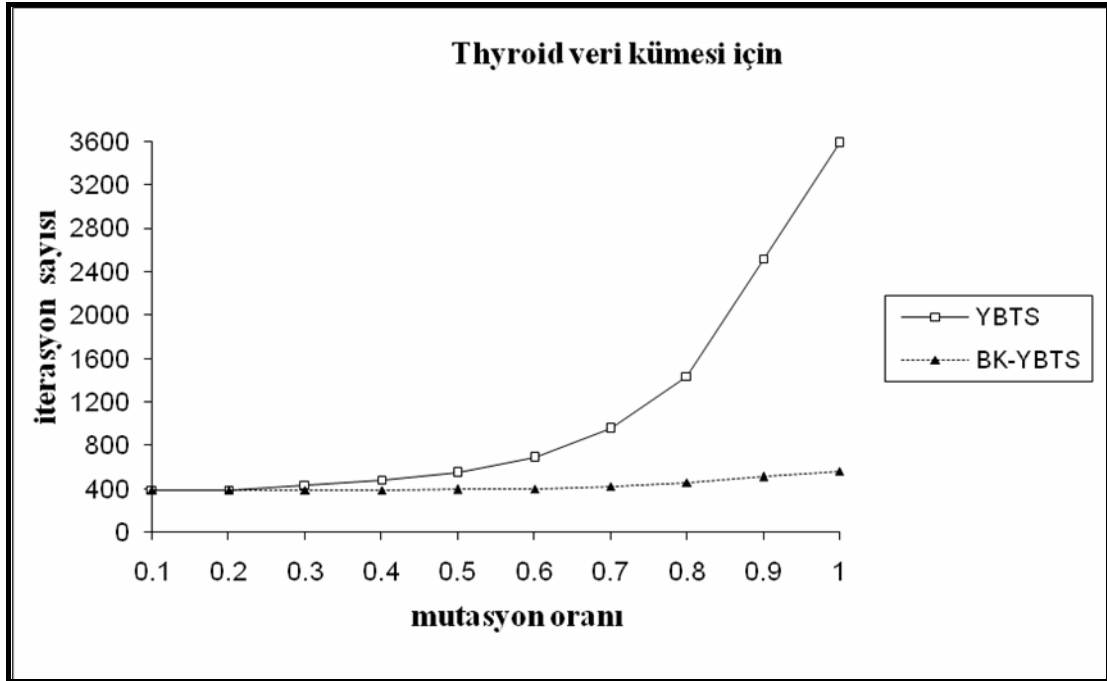
Mutasyon oranı parametresinin değişik değerleri için iki sistemin çalışma süreleri, iterasyon sayıları, bellek hücreleri sayıları ve sınıflandırma doğrulukları açısından karşılaştırılmasının grafiksel olarak gösterimleri sırasıyla Şekil 6.38 ile Şekil 6.41 arasında verilmiştir.

Tablo 6.17 Thyroid veri kümesinde mutasyon oranı parametresi için YBTS ve BK-YBTS karşılaştırma sonuçları

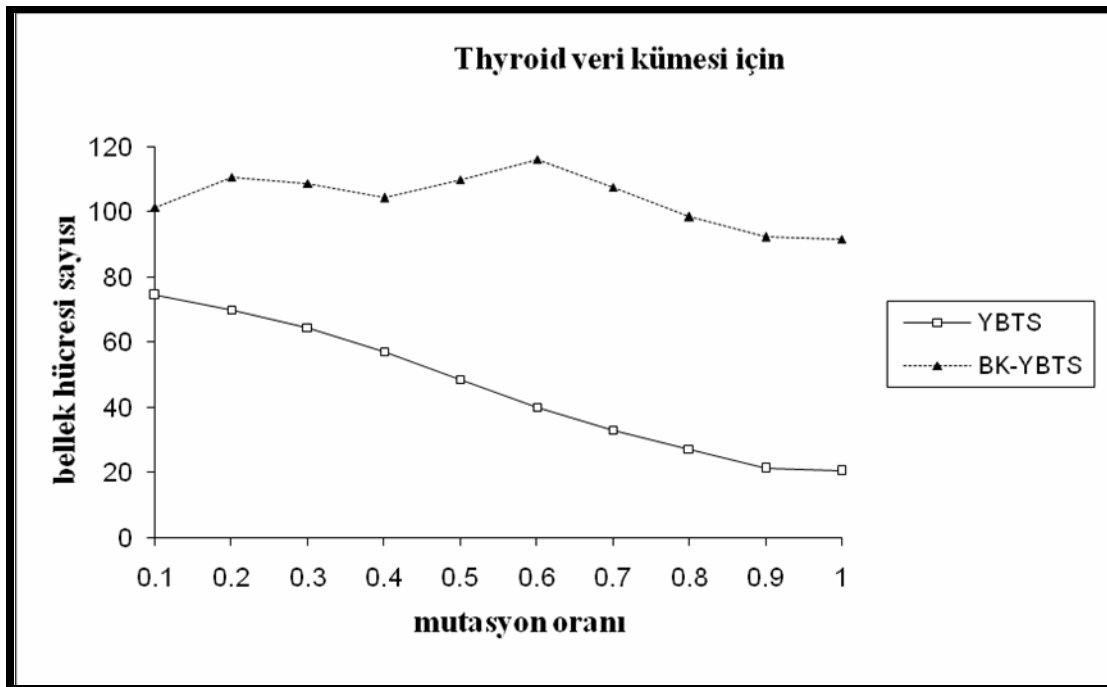
Mutasyon Oranı	Süre (sn)		İterasyon Sayısı		Bellek Hücresi Sayısı		Sınıflandırma Doğruluğu (%)	
	YBTS	BK-YBTS	YBTS	BK-YBTS	YBTS	BK-YBTS	YBTS	BK-YBTS
0.1	112.6	62.1	383.8	383.2	74.4	101.3	94.2	94.4
0.2	101.5	62.0	390.0	383.2	69.8	110.5	93.0	94.8
0.3	113.5	62.1	431.0	383.2	64.6	108.6	94.2	94.2
0.4	129.0	62.2	476.0	383.2	57.2	104.2	93.8	94.0
0.5	151.8	62.3	545.0	392.4	48.5	109.8	91.0	93.4
0.6	205.5	62.8	684.2	398.3	40.1	116.1	91.5	93.2
0.7	313.4	67.9	949.9	419.8	33.0	107.4	91.4	92.7
0.8	508.5	71.2	1437.0	452.5	27.2	98.4	86.7	92.8
0.9	944.6	78.5	2513.3	507.9	21.2	92.2	84.1	88.7
1.0	1520.4	87.9	3582.0	557.6	20.5	91.7	87.7	88.2



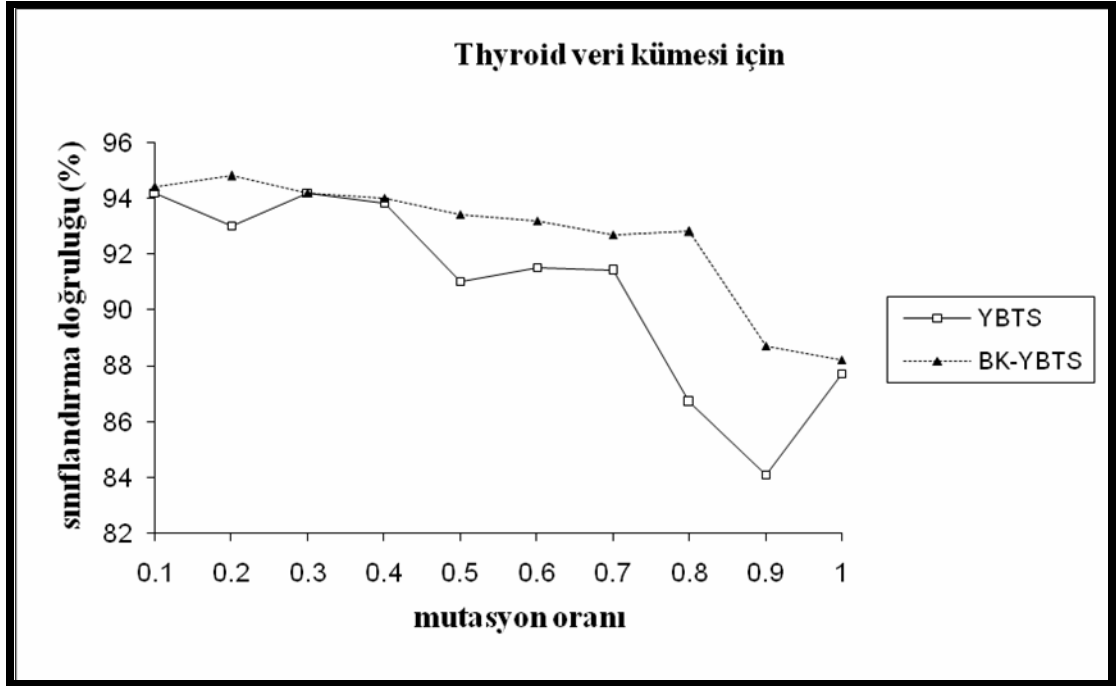
Şekil 6.38 Thyroid veri kümesinde mutasyon oranı parametresinin değişimine göre YBTS ve BK-YBTS'nin çalışma sürelerinin grafiksel gösterimi



Şekil 6.39 Thyroid veri kümesinde mutasyon oranı parametresinin değişimine göre YBTS ve BK-YBTS'nin iterasyon sayılarının grafiksel gösterimi



Şekil 6.40 Thyroid veri kümesinde mutasyon oranı parametresinin değişimine göre YBTS ve BK-YBTS'nin bellek hücreleri sayılarının grafiksel gösterimi



Şekil 6.41 Thyroid veri kümesinde mutasyon oranı parametresinin değişimine göre YBTS ve BK-YBTS'nin sınıflandırma doğruluklarının grafiksel gösterimi

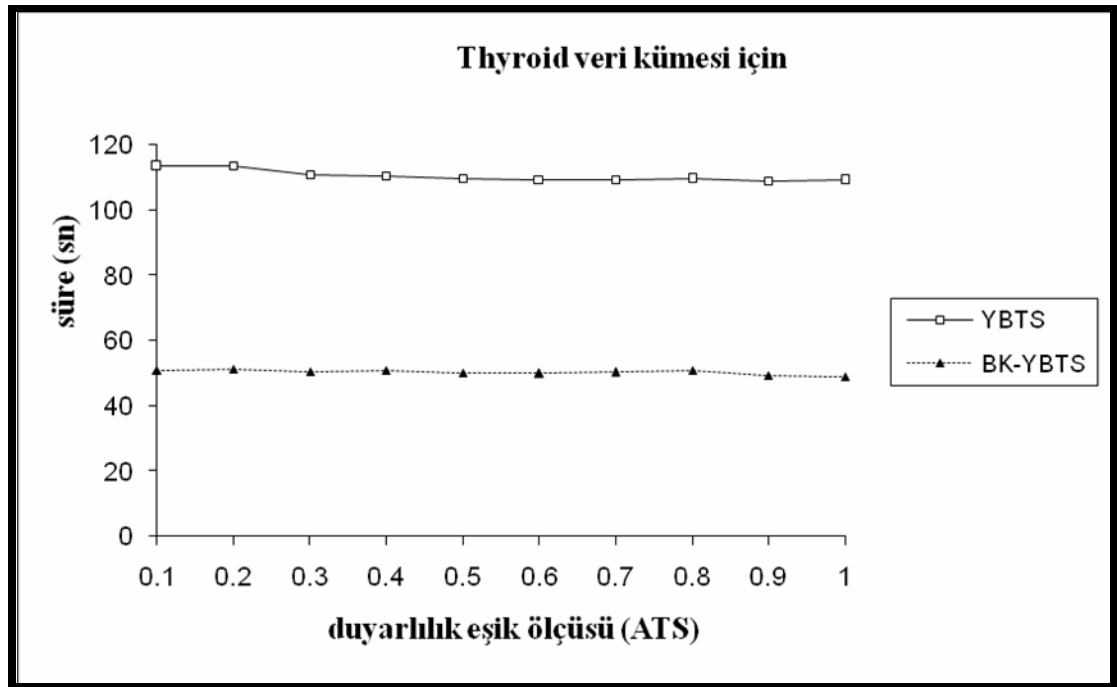
Mutasyon oranı parametresinin bir değeri hariç *BK-YBTS*, *YBTS*'den daha yüksek sınıflandırma doğruluğuna ulaşmıştır. *Mutasyon oranı* değeri 0.3 olduğunda her iki sistemde %94.2 sınıflandırma doğruluğuna ulaşmıştır. *Mutasyon oranı* arttıkça *BK-YBTS*'de çalışma zamanında ve iterasyon sayısında önemli bir değişiklik gözlemlenmezken *YBTS*'de çalışma zamanında ve iterasyon sayısında ciddi artışlar gözlemlenmiştir. Bu durumun sebebi *YBTS*'de özellik mutasyonunun rasgele değer atanarak yapılmasıdır. *BK-YBTS*'de ise duyarlılık tabanlı mutasyon mekanizması kullanılmıştır. Bölüm 5.2.2'de *YBTS*'nin mutasyon mekanizması, Bölüm 5.4.3'te ise *BK-YBTS*'nin mutasyon mekanizması ayrıntılı olarak açıklanmıştır.

Mutasyon oranı parametresi için iki sistemin sınıflandırma performansı analiz edildikten sonra *duyarlılık eşik ölçüsü* parametresinin değişik değerleri için karşılaştırma sonuçları Tablo 6.18'de verilmiştir.

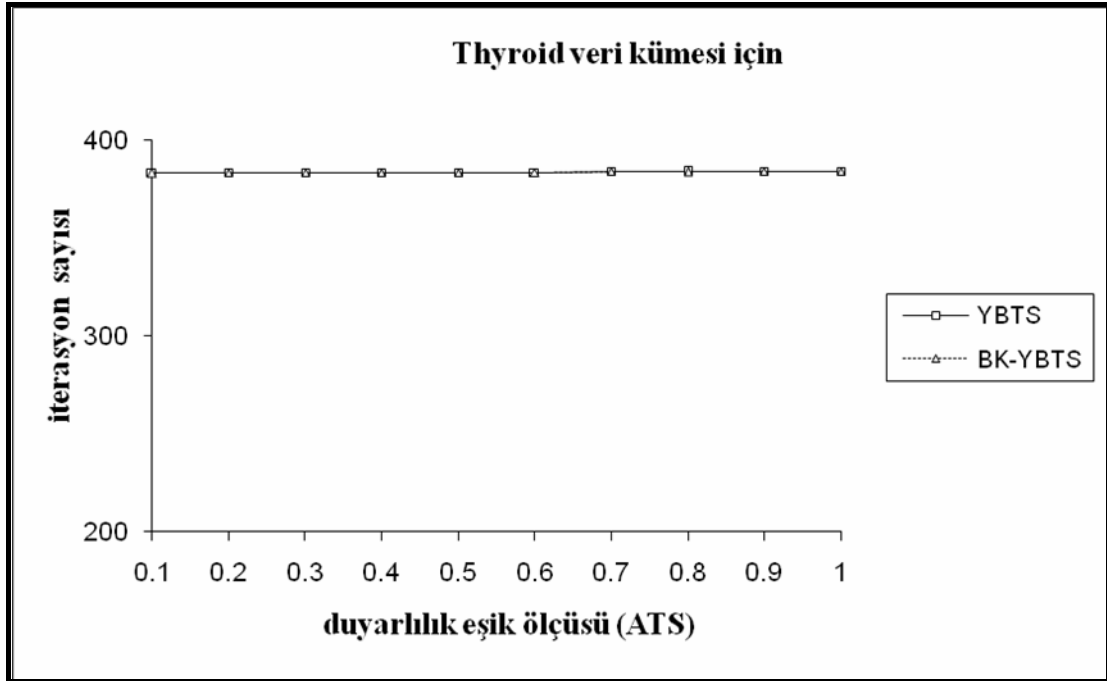
Duyarlılık eşik ölçüsü parametresinin değişik değerleri için iki sistemin çalışma süreleri, iterasyon sayıları, bellek hücreleri sayıları ve sınıflandırma doğrulukları açısından karşılaştırılmasının grafiksel olarak gösterimleri sırasıyla Şekil 6.42 ile Şekil 6.45 arasında verilmiştir.

Tablo 6.18 Thyroid veri kümesinde duyarlılık eşik ölçüsü parametresi için YBTS ve BK-YBTS karşılaştırma sonuçları

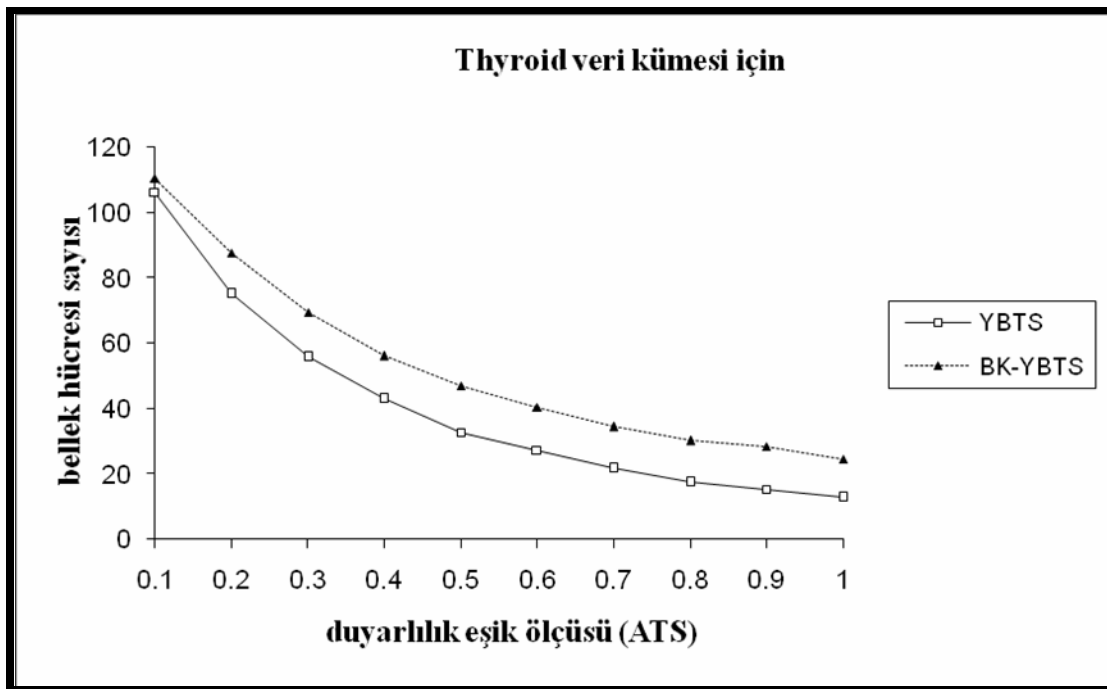
Duyarlılık Eşik Ölçüsü (ATS)	Süre (sn)		İterasyon Sayısı		Bellek Hücresi Sayısı		Sınıflandırma Doğruluğu (%)	
	YBTS	BK-YBTS	YBTS	BK-YBTS	YBTS	BK-YBTS	YBTS	BK-YBTS
0.1	113.4	51.0	383.7	383.2	106.2	110.6	94.0	94.0
0.2	113.5	51.2	383.5	383.5	75.1	87.8	94.2	94.4
0.3	111.0	50.5	383.5	383.3	55.8	69.4	93.8	93.9
0.4	110.5	50.8	383.8	383.5	43.0	56.2	91.3	93.4
0.5	109.7	50.1	383.7	383.5	32.7	47.0	91.3	92.8
0.6	109.3	50.0	383.8	383.6	27.2	40.6	87.3	90.9
0.7	109.4	50.4	384.2	383.5	21.7	34.5	89.1	91.4
0.8	109.6	50.9	383.9	383.9	17.5	30.2	86.7	89.8
0.9	109.1	49.3	384.0	383.8	15.2	28.5	85.7	89.4
1.0	109.2	48.9	384.3	383.5	12.8	24.6	84.1	88.9



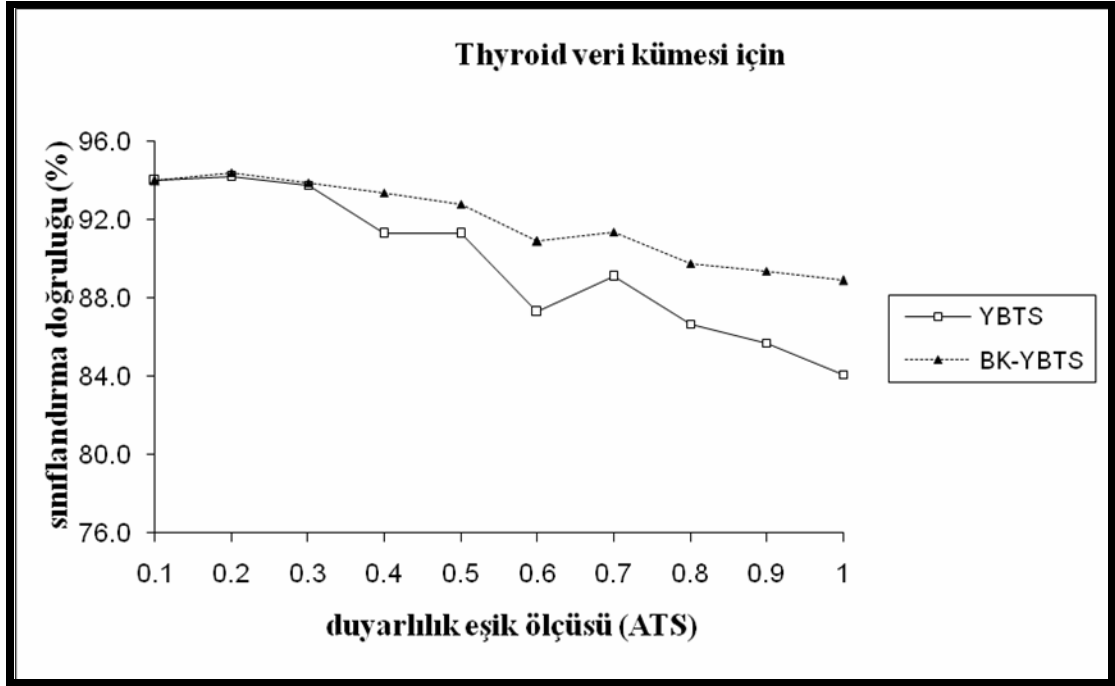
Şekil 6.42 Thyroid veri kümesinde duyarlılık eşik ölçüsü parametresinin değişimine göre YBTS ve BK-YBTS'nin çalışma sürelerinin grafiksel gösterimi



Şekil 6.43 Thyroid veri kümesinde duyarlılık eşik ölçüsü parametresinin değişimine göre YBTS ve BK-YBTS'nin iterasyon sayılarının grafiksel gösterimi



Şekil 6.44 Thyroid veri kümesinde duyarlılık eşik ölçüsü parametresinin değişimine göre YBTS ve BK-YBTS'nin bellek hücreleri sayılarının grafiksel gösterimi



Şekil 6.45 Thyroid veri kümesinde duyarlılık eşik ölçüsü parametresinin değişimine göre YBTS ve BK-YBTS'nin sınıflandırma doğruluklarının grafiksel gösterimi

Duyarlılık eşik ölçüsü parametresi için bir değer hariç *BK-YBTS*, *YBTS*'den daha yüksek sınıflandırma doğruluklarına ulaşmıştır. *Duyarlılık eşik ölçüsü* değeri 0.1 olduğunda her iki sistemde %94.0 sınıflandırma doğruluğuna ulaşmıştır. *Duyarlılık eşik ölçüsü* değeri arttıkça her iki sistemde de çalışma zamanında ve iterasyon sayısında çok az bir oranda azalma gözlemlenmiştir. Tam tersine *duyarlılık eşik ölçüsü* parametresi değeri arttığı zaman her iki sistemde de bellek hücresi sayısında önemli düşüşler gözlemlenmiştir. *Duyarlılık eşik ölçüsü* parametresi bellek hücresi seçiminde aktif rol oynadığı için bu parametrenin değişimi, bellek hücresi sayısını oldukça etkilemiştir. *YBTS*'de duyarlılık eşik ölçüsü kullanımı Bölüm 5.2.4'te, *BK-YBTS* için duyarlılık eşik ölçüsü kullanımı Bölüm 5.4.5'te açıklanmıştır.

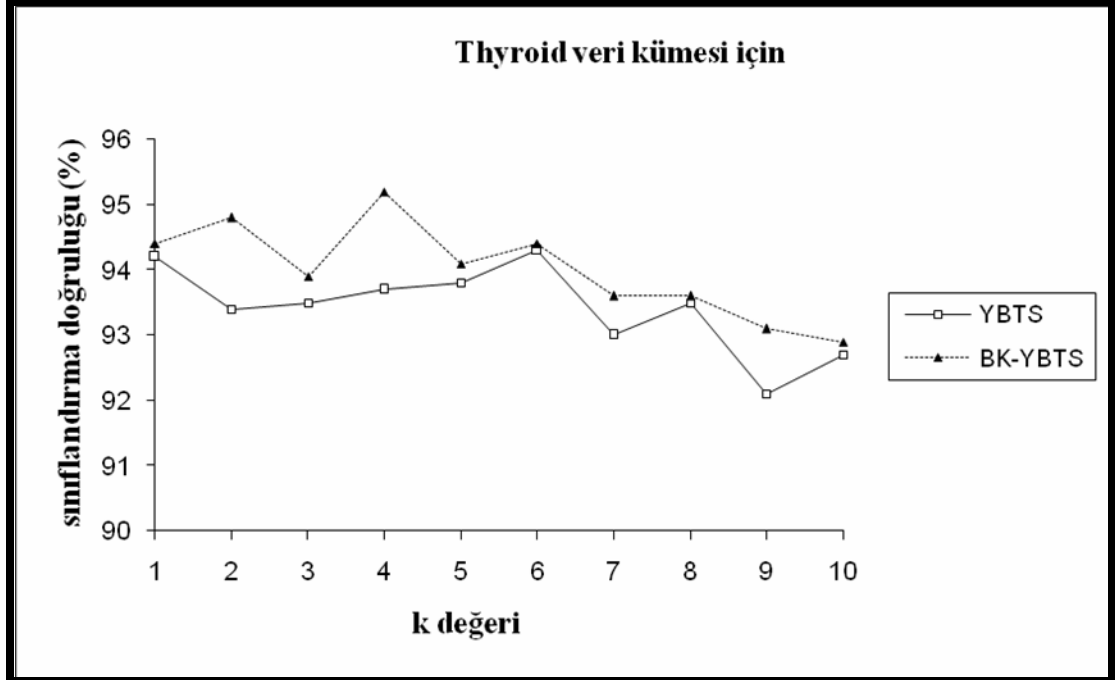
Duyarlılık eşik ölçüsü parametresi için iki sistemin sınıflandırma performansı analiz edildikten sonra *k değeri* parametresinin değişik değerleri için karşılaştırma sonuçları Tablo 6.19'da verilmiştir.

Test veri kümesini bellek hücreleri üzerinde çoğunluk oylaması yaparak sınıflandırma da kullanılan *k değeri* parametresinin değişik değerleri için iki sistemin sınıflandırma doğrulukları grafiksel olarak Şekil 6.46'da gösterilmiştir. *YBTS* ve *BK-YBTS* eğitimlerini tamamladıktan sonra oluşan bellek hücrelerini temel alarak *k en*

yakın komşu yöntemi kullanılmaktadır. Farklı k değerleri için iki sistem ayrı ayrı çalıştırılmak yerine bir kez çalıştırılmış ve oluşan bellek hücreleri farklı k değerleri ile test verisini sınıflandırmıştır. Bundan dolayı farklı k değerleri için iki sistemin çalışma süreleri, iterasyon sayıları ve bellek hücreleri sayıları ortak alınmıştır.

Tablo 6.19 Thyroid veri kümesinde k değeri parametresi için YBTS ve BK-YBTS karşılaştırma sonuçları

k değeri	Süre (sn)		İterasyon Sayısı		Bellek Hücresi Sayısı		Sınıflandırma Doğruluğu (%)	
	YBTS	BK-YBTS	YBTS	BK-YBTS	YBTS	BK-YBTS	YBTS	BK-YBTS
1	113.1	63.4	383.5	383.4	73.8	97.6	94.2	94.4
2							93.4	94.8
3							93.5	93.9
4							93.7	95.2
5							93.8	94.1
6							94.3	94.4
7							93.0	93.6
8							93.5	93.6
9							92.1	93.1
10							92.7	92.9



Şekil 6.46 Thyroid veri kümesinde k değeri parametresinin değişimine göre YBTS ve BK-YBTS'nin sınıflandırma doğruluklarının grafiksel gösterimi

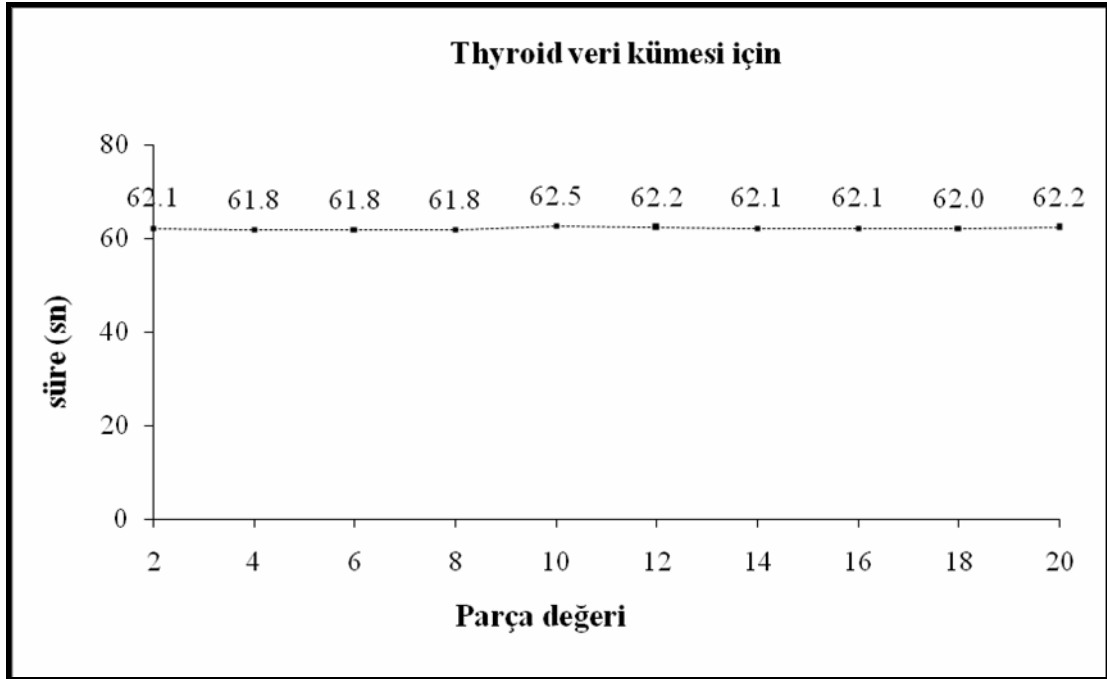
k değeri parametresinin aldığı bütün değerler için *BK-YBTS*, *YBTS'den* daha yüksek sınıflandırma doğruluklarına ulaşmıştır. Bununla beraber *BK-YBTS'de* ortalama olarak 97.6 adet bellek hücresi oluşurken, *YBTS'de* ortalama olarak 73.8 adet bellek hücresi oluşmuştur.

Buraya kadar thyroid veri kümesinde *YBTS* ve *BK-YBTS* için ortak parametreler olan çekirdek hücre sayısı, kaynak sayısı, uyarma eşiği, mutasyon oranı, duyarlılık eşik ölçüsü ve k değeri parametrelerinin, sınıflandırma performansına olan etkileri araştırıldıktan sonra sadece *BK-YBTS'ye* özgü olan *parça değeri* parametresinin değişik değerleri için çalışma süreleri, iterasyon sayıları, bellek hücresi sayıları ve sınıflandırma doğrulukları Tablo 6.20’de verilmiştir.

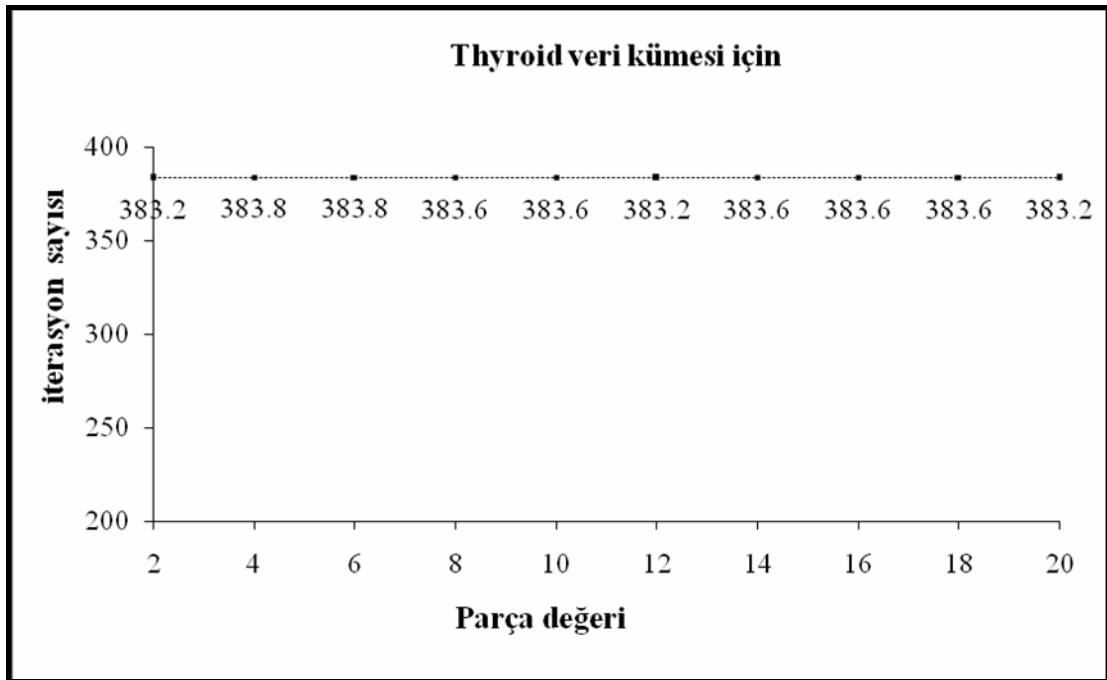
Parça değeri parametresinin aldığı değişik değerler için *BK-YBTS'nin* çalışma süreleri açısından grafiksel olarak gösterimi Şekil 6.47’de, iterasyon sayıları açısından grafiksel olarak gösterimi Şekil 6.48’de, bellek hücreleri sayıları açısından grafiksel olarak gösterimi Şekil 6.49’da ve sınıflandırma doğrulukları açısından grafiksel olarak gösterimi Şekil 6.50’de verilmiştir.

Tablo 6.20 Thyroid veri kümesinde parça değeri parametresi için BK-YBTS sonuçları

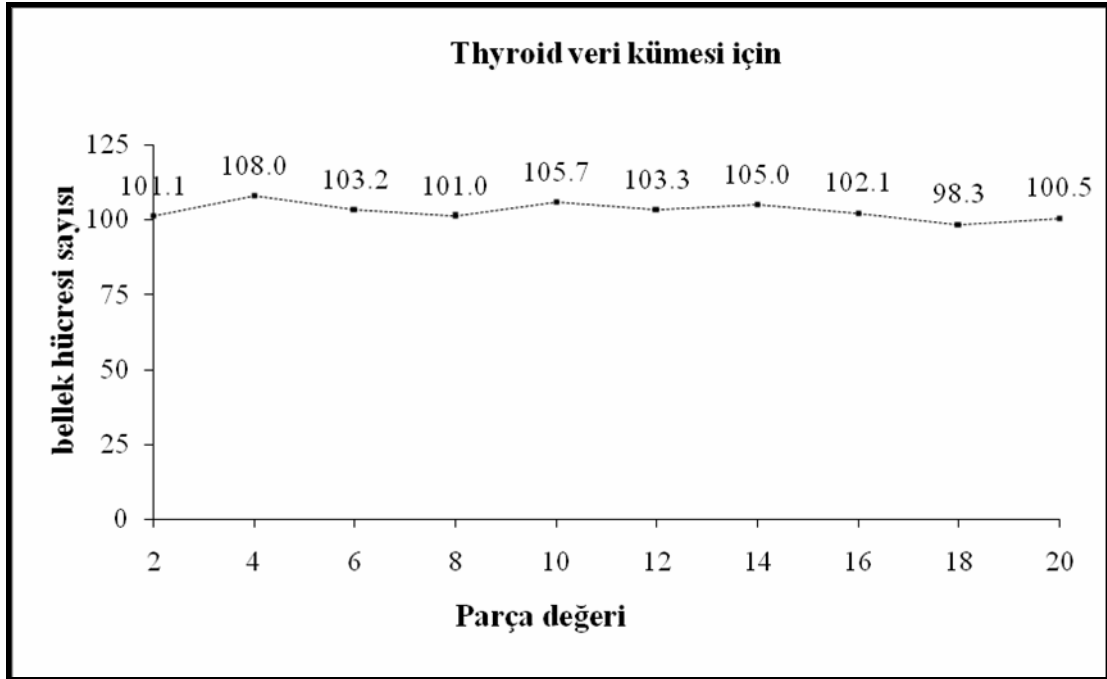
Parça Değeri	Süre (sn)	İterasyon Sayısı	Bellek Hücresi Sayısı	Sınıflandırma Doğruluğu (%)
2	62.1	383.2	101.1	94.4
4	61.8	383.8	108.0	95.2
6	61.8	383.8	103.2	95.4
8	61.8	383.6	101.0	94.8
10	62.5	383.6	105.7	94.8
12	62.2	383.2	103.3	94.6
14	62.1	383.6	105.0	94.8
16	62.1	383.6	102.1	94.9
18	62.0	383.6	98.3	94.6
20	62.2	383.2	100.5	94.5



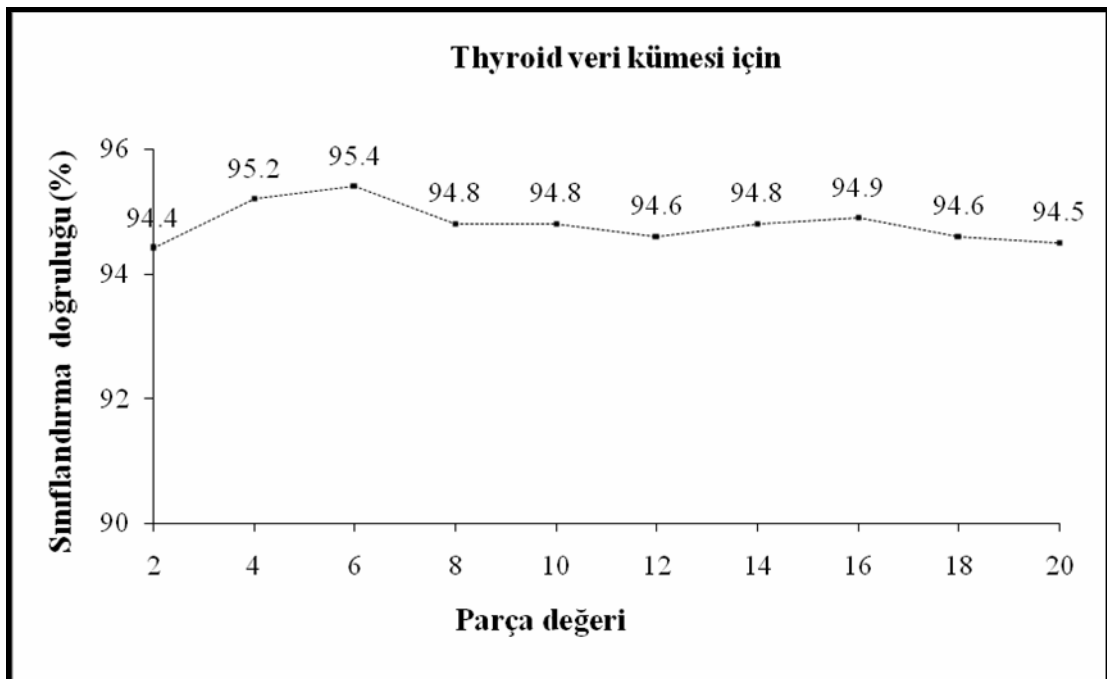
Şekil 6.47 Thyroid veri kümesinde parça değeri parametresinin değişimine göre BK-YBTS'nin çalışma sürelerinin grafiksel gösterimi



Şekil 6.48 Thyroid veri kümesinde parça değeri parametresinin değişimine göre BK-YBTS'nin iterasyon sayılarının grafiksel gösterimi



Şekil 6.49 Thyroid veri kümesinde parça değeri parametresinin değişimine göre BK-YBTS'nin bellek hücreleri sayılarının grafiksel gösterimi



Şekil 6.50 Thyroid veri kümesinde parça değeri parametresinin değişimine göre BK-YBTS'nin sınıflandırma doğruluklarının grafiksel gösterimi

Tablo 6.20'den de anlaşılabileceği gibi *parça değeri* parametresi 6 olduğu durumda, *BK-YBTS* en yüksek sınıflandırma doğruluğu olan % 95.4'e ulaşmıştır.

Parça değeri parametresinin değişimi *BK-YBTS'nin* çalışma süresini, iterasyon sayısını ve bellek hücreleri sayısını çok fazla etkilememiştir.

Thyroid veri kümesinde tüm parametreler için genel olarak *BK-YBTS*, *YBTS'den* daha yüksek sınıflandırma doğruluğuna ulaşmıştır. Bu durum *BK-YBTS* ile beraber önerilen değişikliklerin, *YBTS'ye* göre sınıflandırma performansını arttırdığını göstermektedir.

Tablo 6.21'de *Thyroid* veri kümesi için deneme yanılma yoluyla elde edilen en iyi sınıflandırma doğruluğunu veren parametre değerleri verilmiştir. Tablo 6.22'de ise *Thyroid* veri kümesi için *BK-YBTS'nin* ve literatürdeki diğer sınıflandırıcıların sınıflandırma doğrulukları belirtilmiştir. Tablo 6.22'de *Thyroid* veri kümesi için farklı yazarların farklı sınıflandırma metotları için elde ettikleri sınıflandırma doğrulukları verildiğinden dolayı sınıflandırma doğrulukları en yüksekten düşüğe doğru listelenmemiştir.

Thyroid veri kümesi için *parça değeri* 4 ve 6 olduğu durumda geliştirilen sistem, en yüksek sınıflandırma doğruluğu olan %95.9 değerine ulaşmıştır. Tablo 6.22'den de anlaşılacağı üzere *BK-YBTS'nin* sınıflandırma doğruluğu *Thyroid* veri kümesi için *YBTS'den* daha yüksek elde edilmiştir. Bu sınıflandırma doğruluk değeri, *thyroid* veri kümesi için literatürdeki sınıflandırıcılar arasında en yüksek sınıflandırma doğruluklarından birisidir. Bu sonuçta göstermektedir ki geliştirilen sistem tıbbi teşhis problemlerinde karar vermede uzmanlara güvenli bir şekilde yardımcı olabilecektir.

Tablo 6.21 Thyroid veri kümesi için optimum BK-YBTS parametre değerleri

Parametre Adı	Değeri
<i>Çekirdek hücre sayısı</i>	1
<i>Kaynak sayısı</i>	120
<i>Uyarma eşiği</i>	0.9
<i>Mutasyon oranı</i>	0.1
<i>Duyarlılık eşik ölçüsü (ATS)</i>	0.2
<i>Çoğunluk oylaması için kullanılan k değeri</i>	4
<i>Parça değeri</i>	4 ve 6

Tablo 6.22 Thyroid veri kümesi için BK-YBTS ve literatürdeki diğer sınıflandırıcılar tarafından elde edilen sınıflandırma doğrulukları

Yazar (Yıl)	Metot	Eğitim-test oranı	Doğruluk (%)
Yip ve Webb (1994)	FAFA+C4.5 (Pruned)	10 x CV	94.38
	FAFA+C4.5 (Rules)		94.38
	Einstein		91.91
	FAFA+Einstein		93.34
Cheong ve Yoon (1999)	kNN	70%-30%	96.90
	EACH		95.90
	RPA		96.10
Özyılmaz ve Yıldırım (2002)	MLP with bp	10 x CV	86.33
	MLP with fbp		89.80
	RBF		79.08
	CSFNN		91.14
Albayrak ve Amasyalı (2003)	Hard C-Means	100%-0%	78.10
	Fuzzy C-Means		83.70
Myles ve Brown (2004)	SMC (PLS-QDA)	5 x CV	97.20
	OAC (PLS-QDA)		93.80
	PWC (PLS-QDA)		97.20
	DPM (PLS-QDA)		98.20
Pechenizkiy (2005)	3NN-Par	60%-40%	94.20
	FEDIC-Plain		96.10
	Bayes		94.80
Hassan ve ark. (2006)	HMM	100%-0%	87.91
	SOM		88.84
Kim ve Ghahramani (2006)	LDA	10 x CV	93.44
	SVM (soft)		94.44
	GPC-EP (s, soft)		96.75
	GPC-EP (m, soft)		97.23
Sun (2007)	C4.5 Base	10 x CV	91.57
	C4.5 AdaBoost		91.12
	HPWR Base (Od=3)		91.66
	HPWR AdaBoost (Od=3)		88.17
Tez çalışması	YBTS	10 x CV	94.82
Tez çalışması	BK-YBTS	10 x CV	95.90

6.2.3. Damar tıkanıklığı veri kümesi uygulamaları

Damar tıkanıklığı (Atherosclerosis) yıllar boyunca yayılan bir atar damar hastalığıdır ve hücre dışındaki lipidler, lifli doku, düz kas hücreleri ve kalsiyum toplanması ile ortaya çıkmaktadır. Yağlı tortular atar damarların iç duvarları üzerinde plak formundadır ve atar damar içerisindeki kan akışını önemli derecede azaltacak kadar genişleyebilmektedir. Atar damar daraldıkça daha az kan akmaktadır. Atar damar daha az elastik hale gelip kalp damarları hastalıklarının temel sebebi olan *doku sertleşmesi (sclerosis)* meydana gelmektedir (Hirai ve ark. 1989).

Damarla ilgili hastalıkların teşhisinde birçok test söz konusudur. Bunlar arasında kan testleri, elektrokardiyogram, stres testi, anjiyografi, ultrason ve bilgisayarlı tomografi sayılabilir. Anjiyografi atar damarlarda herhangi bir engel var mı ve varsa bu engelin boyutunu görebilmek için atar damarların içine bakmak için kullanılmaktadır (Stefanadis ve ark. 1990; Dart ve ark. 1991). Bu teknik damarla ilgili hastalıkların varlığını ve ciddiyetini tahmin etmek için en doğru yoldur. Diğer taraftan bu teknikte atar damarların içerisine direkt olarak boya enjekte edilmektedir. Bu nedenle rahatsızlık verici bir tekniktir.

Anjiyografi rahatsız edici ve göreceli olarak yüksek maliyetli olduğu için ultrasonik Doppler sonografi tavsiye edilmektedir. Doppler görüntüleme tekniklerindeki son zamanlardaki gelişmeler sayesinde atar damar sisteminin farklı bölgelerinde akış karakteristiklerini değerlendirmek mümkün olmaktadır. Bu bölgelere örnek olarak aort, koroner, karotis ve çevresel atar damarlar verilebilir. Doppler sinyallerinin analizi, Doppler dalga şeklindeki akış ölçümlerini uygulayan fizyoloji ve bu dalga şeklindeki önemli değişiklikleri tespit eden patoloji hakkında bilgi sağlamaktadır (Evans 2000).

Doppler sonogramları bir zaman serisinin gücünü açıklamaktadır ve böylece Doppler sinyalinin spektral analizi atar damar içerisindeki hız dağılımı hakkında bilgi sağlamaktadır (Muller ve ark. 2001; Evans ve ark. 1989; Vaitkus ve ark. 1988).

Damar tıkanıklığı verisi Erciyes Üniversitesi Tıp Fakültesinden elde edilmiştir. Bu veri kümesinde 114 adet veri söz konusudur. 114 verinin 60'ı hasta, 54'ü sağlıklı kişilerden elde edilmiştir. Veriler Radyoloji bölümünde Toshiba

PowerVision 6000 Doppler Ultrason ünitesi kullanılarak elde edilmiştir. 8 MHz lineer ultrason sondası sol karotis atar damara darbeli ultrason sinyali iletmek için kullanılmıştır. Karotis atar damarlarda Doppler tepkilerini elde etmek için ultrason cihazının ses çıkış portu kullanılmıştır. Ultrason cihazının çıkışındaki ses sinyalleri 44100 Hz'de örneklenmiştir ve bir giriş/çıkış kartı aracılığı ile bilgisayara gönderilmiştir (Kara 1995). Elde edilen veri bir *ADC (Analog to Digital Converter)* kullanılarak sayısal hale getirilmiş ve bir giriş/çıkış kartı ile bilgisayara aktarılmıştır. Doppler verisine güç spektrum tahmini için Welch metodu uygulanmıştır (Dirgenali 2003; Latifoğlu ve ark. 2006; Kodaz ve ark. 2007).

Zaman bazında karotis atar damardan yansıyan Doppler sinyalleri hastalığın varlığı hakkında açık bir bilgi verememektedir. Bu sebepten bu sinyaller sağlıklı kişi ile damar tıkanıklığı olan kişi arasındaki farkları ortaya çıkarmak için frekans bazında analiz edilmektedir. Bu analiz sonucu bir periyot için sonogram 61 örnekle ifade edilmektedir. Bu 61 adet değer *BK-YBTS* için tek bir veriyi ifade etmektedir.

Damar tıkanıklığı veri kümesi için *YBTS*'den farklı olarak *BK-YBTS* ile beraber önerilen değişikliklerin etkisini araştırmak amacıyla *çekirdek hücre sayısı*, *kaynak sayısı*, *uyarma eşiği*, *mutasyon oranı*, *duyarlılık eşik ölçüsü* ve *k* parametreleri için Tablo 6.23'te belirtilen parametre değerleri temel alınarak *YBTS* ve *BK-YBTS* karşılaştırılmıştır.

Tablo 6.23 Damar tıkanıklığı veri kümesi için *YBTS* ve *BK-YBTS*'nin başlangıç parametre değerleri

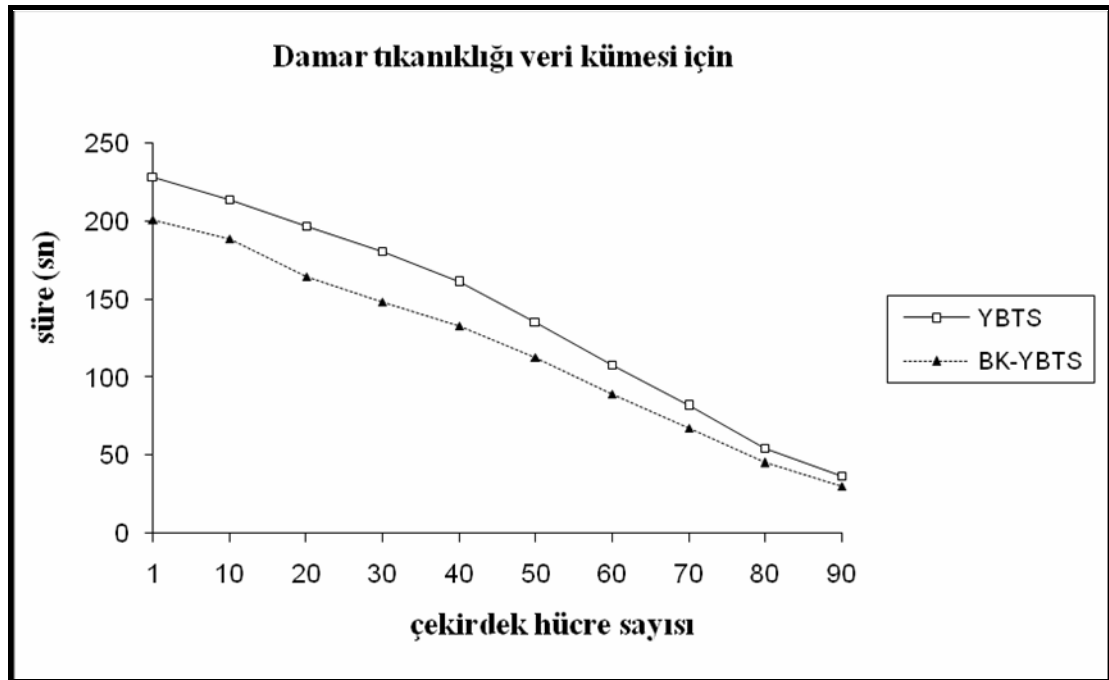
Parametre Adı	Parametre Değeri
<i>Çekirdek hücre sayısı</i>	1
<i>Kaynak sayısı</i>	200
<i>Uyarma eşiği</i>	0.9
<i>Mutasyon oranı</i>	0.1
<i>Duyarlılık eşik ölçüsü</i>	0.2
<i>Çoğunluk oylaması için kullanılan k değeri</i>	1
<i>Parça değeri (BK-YBTS için)</i>	2

Çekirdek hücre sayısı parametresinin aldığı değişik değerler için *YBTS* ve *BK-YBTS* karşılaştırma sonuçları Tablo 6.24'te verilmiştir. *Çekirdek hücre sayısı* parametresinin aldığı değişik değerler için iki sistemin çalışma süreleri açısından karşılaştırılmasının grafiksel olarak gösterimi Şekil 6.51'de, iterasyon sayıları açısından karşılaştırılmasının grafiksel olarak gösterimi Şekil 6.52'de, bellek

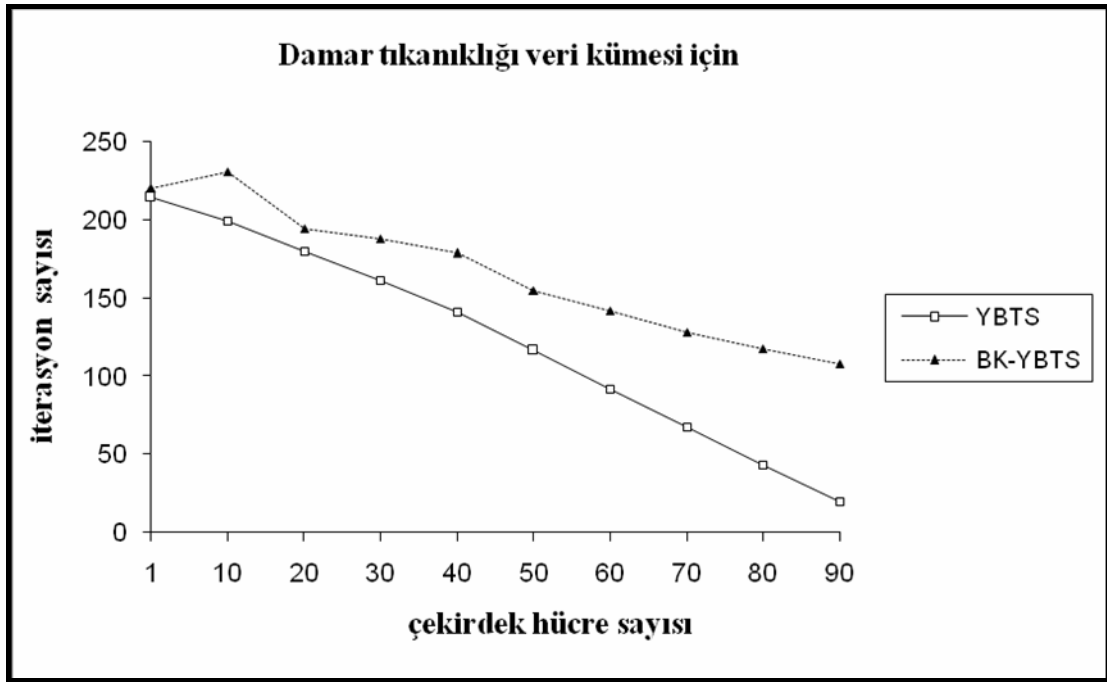
hücreleri sayıları açısından karşılaştırılmasının grafiksel olarak gösterimi Şekil 6.53'te ve sınıflandırma doğrulukları açısından karşılaştırılmasının grafiksel olarak gösterimi Şekil 6.54'te verilmiştir.

Tablo 6.24 Damar tıkanıklığı veri kümesinde çekirdek hücre sayısı parametresi için YBTS ve BK-YBTS karşılaştırma sonuçları

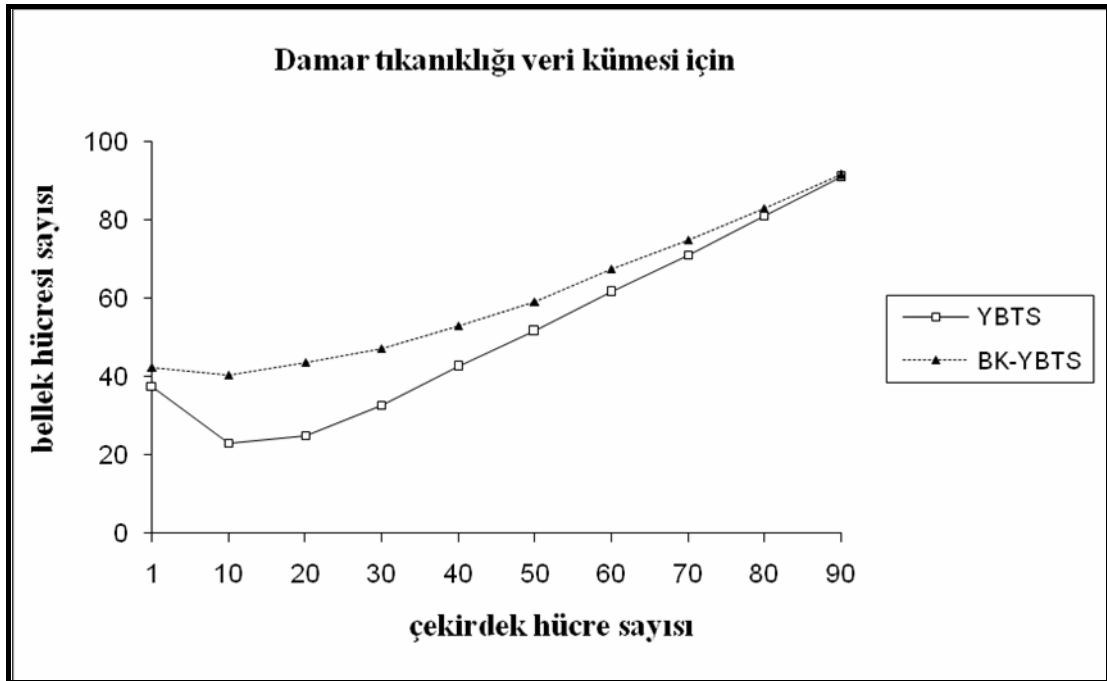
Çekirdek Hücre Sayısı	Süre (sn)		İterasyon Sayısı		Bellek Hücresi Sayısı		Sınıflandırma Doğruluğu (%)	
	YBTS	BK-YBTS	YBTS	BK-YBTS	YBTS	BK-YBTS	YBTS	BK-YBTS
1	228.4	200.7	214.3	220.5	37.5	42.4	99.1	99.1
10	213.8	188.8	199.3	230.5	23.0	40.4	99.1	99.1
20	197.0	164.4	180.0	194.6	25.1	43.8	99.1	99.1
30	180.9	148.4	161.5	188.2	32.8	47.2	99.1	99.1
40	160.9	133.1	140.9	178.7	42.6	53.1	99.1	99.1
50	135.4	112.5	116.4	154.8	51.6	59.2	99.1	99.1
60	107.7	89.5	91.6	141.8	61.6	67.6	99.1	99.1
70	81.6	67.5	67.6	127.8	71.1	75.1	99.1	99.1
80	54.5	45.1	43.4	117.8	81.1	83.2	99.1	99.1
90	36.4	30.3	19.8	107.8	91.0	91.6	99.1	99.1



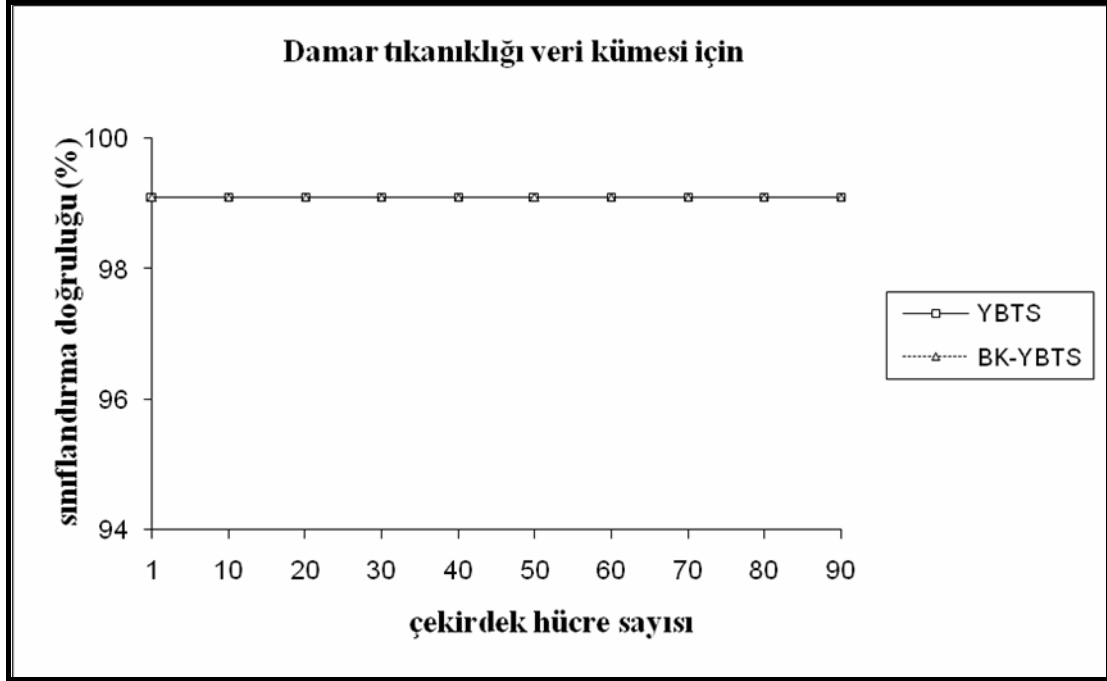
Şekil 6.51 Damar tıkanıklığı veri kümesinde çekirdek hücre sayısı parametresinin değişimine göre YBTS ve BK-YBTS'nin çalışma sürelerinin grafiksel gösterimi



Şekil 6.52 Damar tıkanıklığı veri kümesinde çekirdek hücre sayısı parametresinin değişimine göre YBTS ve BK-YBTS'nin iterasyon sayılarının grafiksel gösterimi



Şekil 6.53 Damar tıkanıklığı veri kümesinde çekirdek hücre sayısı parametresinin değişimine göre YBTS ve BK-YBTS'nin bellek hücreleri sayılarının grafiksel gösterimi



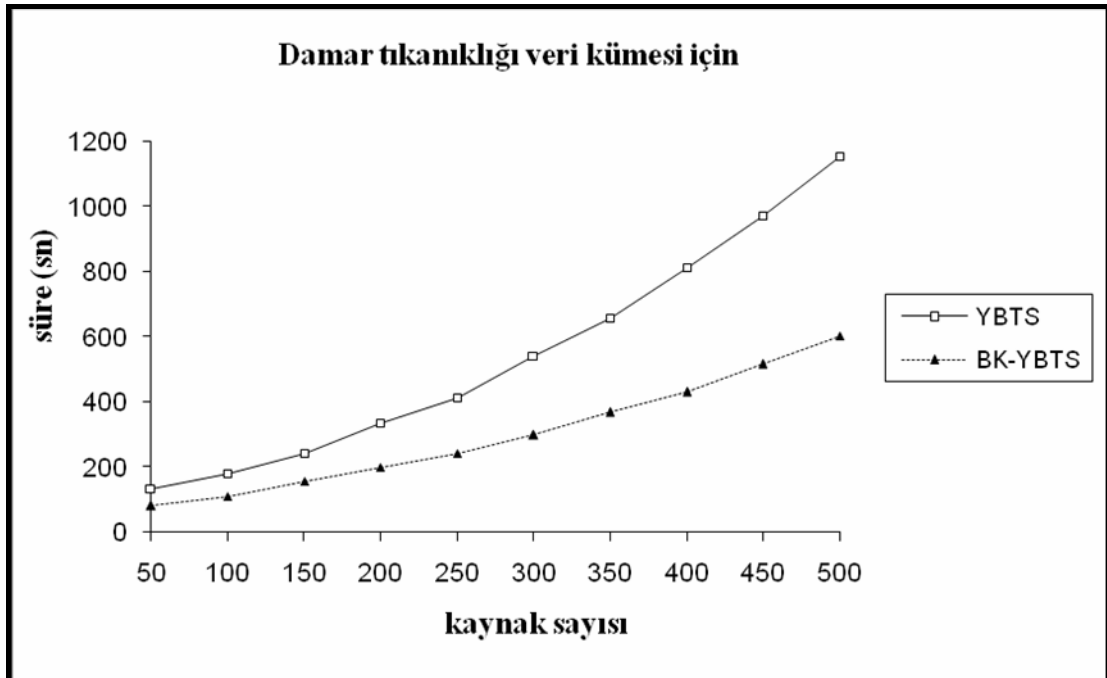
Şekil 6.54 Damar tıkanıklığı veri kümesinde çekirdek hücre sayısı parametresinin değişimine göre YBTS ve BK-YBTS'nin sınıflandırma doğruluklarının grafiksel gösterimi

Çekirdek hücre sayısı arttığı zaman, eğitimin başlangıcında antijenler bellek hücresi ve antikör olarak belirlendiği için doğal olarak *YBTS* ve *BK-YBTS'nin* çalışma zamanı ve iterasyon sayısı azalmış, sistemdeki bellek hücresi sayısı ise artmıştır. *BK-YBTS* ve *YBTS* iterasyon sayısı olarak birbirlerine yakın çalışmakla beraber antikörlere kaynak tahsisi rutininde *YBTS* daha çok zaman harcadığı için *YBTS'nin* çalışma süresi, *BK-YBTS'den* fazla ölçülmüştür. Bellek hücreleri sayısı *BK-YBTS* için *YBTS'den* daha fazla tespit edilmiştir. *Çekirdek hücre sayısı* parametresinin bütün değerleri için *YBTS* ve *BK-YBTS*, %99.1 sınıflandırma doğruluğuna ulaşmıştır.

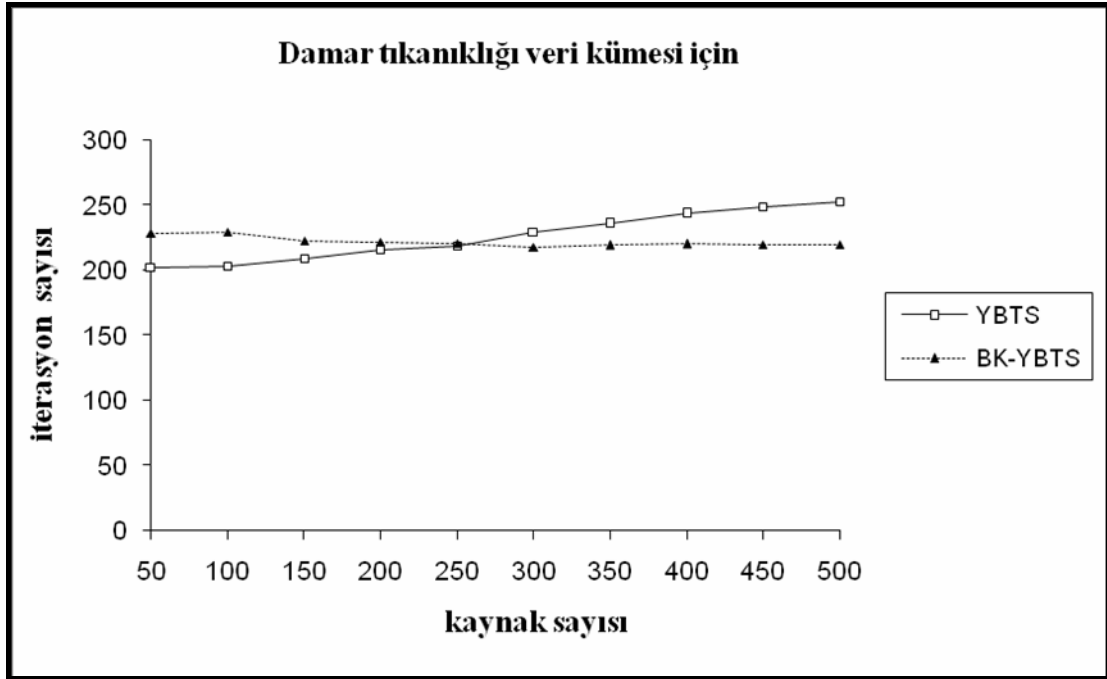
Çekirdek hücre sayısı parametresi için iki sistemin sınıflandırma performansı analiz edildikten sonra *kaynak sayısı* parametresinin değişik değerleri için karşılaştırma sonuçları Tablo 6.25'te verilmiştir. *Kaynak sayısı* parametresinin değişik değerleri için iki sistemin çalışma süreleri, iterasyon sayıları, bellek hücreleri sayıları ve sınıflandırma doğrulukları açısından karşılaştırılmasının grafiksel olarak gösterimleri sırasıyla Şekil 6.55 ile Şekil 6.58 arasında verilmiştir.

Tablo 6.25 Damar tıkanıklığı veri kümesinde kaynak sayısı parametresi için YBTS ve BK-YBTS karşılaştırma sonuçları

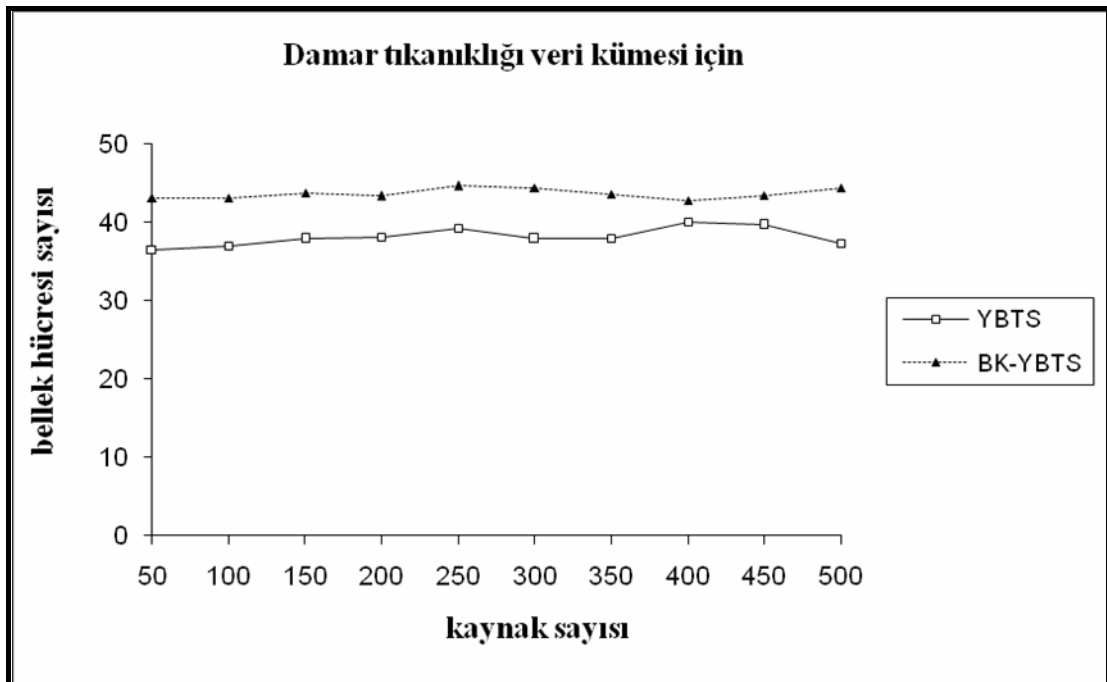
Kaynak Sayısı	Süre (sn)		İterasyon Sayısı		Bellek Hücresi Sayısı		Sınıflandırma Doğruluğu (%)	
	YBTS	BK-YBTS	YBTS	BK-YBTS	YBTS	BK-YBTS	YBTS	BK-YBTS
50	132.5	81.4	202.4	228.2	36.5	43.2	98.2	99.1
100	179.0	110.0	203.2	229.2	37.0	43.2	99.1	99.1
150	241.9	156.4	208.9	222.6	37.9	43.8	99.1	99.1
200	335.7	199.5	215.3	221.5	38.1	43.4	99.1	99.1
250	414.0	241.9	218.7	220.4	39.3	44.7	99.1	99.1
300	540.7	298.0	229.1	217.4	37.9	44.5	99.1	99.1
350	658.3	368.7	235.6	219.1	38.0	43.6	99.1	99.1
400	812.7	430.2	243.4	220.1	40.1	42.8	99.1	99.1
450	972.2	515.8	249.0	220.0	39.7	43.5	98.2	99.1
500	1155.1	604.1	253.0	219.4	37.4	44.4	99.1	99.1



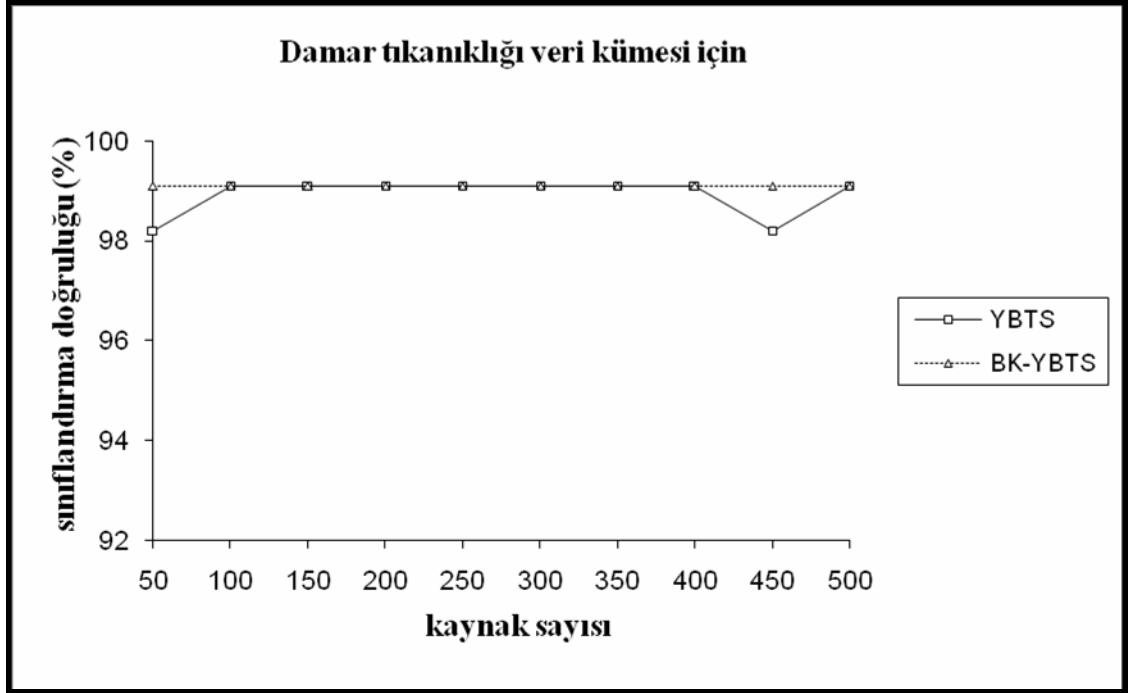
Şekil 6.55 Damar tıkanıklığı veri kümesinde kaynak sayısı parametresinin değişimine göre YBTS ve BK-YBTS'nin çalışma sürelerinin grafiksel gösterimi



Şekil 6.56 Damar tıkanıklığı veri kümesinde kaynak sayısı parametresinin değişimine göre YBTS ve BK-YBTS'nin iterasyon sayılarının grafiksel gösterimi



Şekil 6.57 Damar tıkanıklığı veri kümesinde kaynak sayısı parametresinin değişimine göre YBTS ve BK-YBTS'nin bellek hücreleri sayılarının grafiksel gösterimi



Şekil 6.58 Damar tıkanıklığı veri kümesinde kaynak sayısı parametresinin değişimine göre YBTS ve BK-YBTS'nin sınıflandırma doğruluklarının grafiksel gösterimi

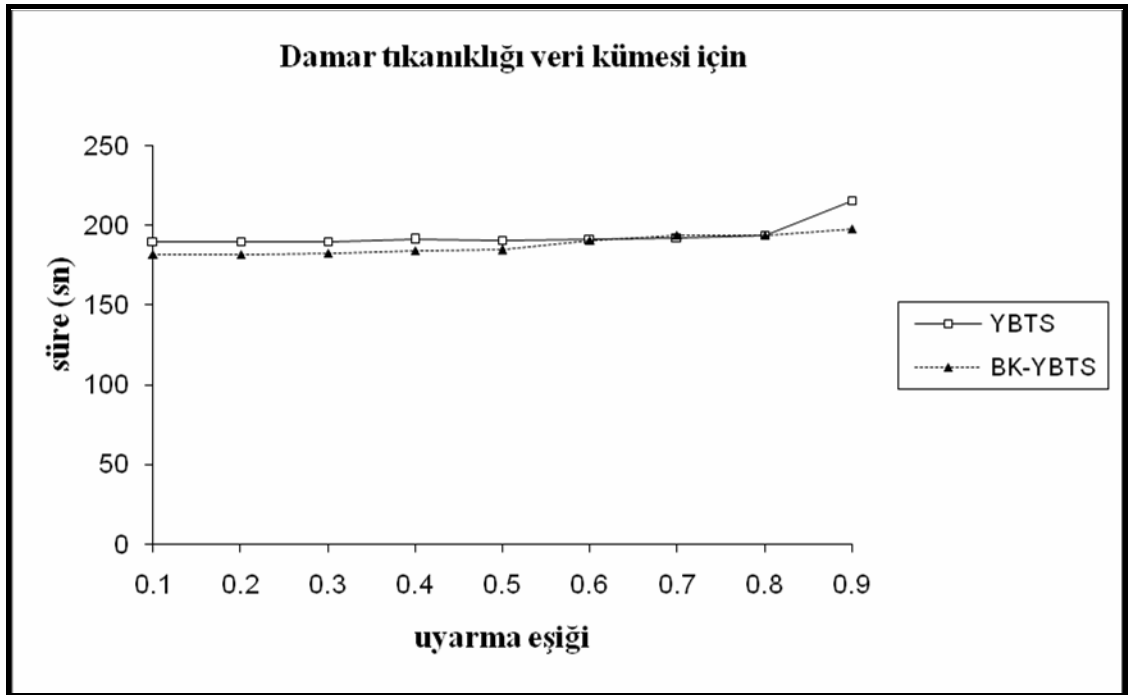
Kaynak sayısı parametresi değeri arttığı zaman, *YBTS* ve *BK-YBTS'nin* çalışma zamanı, iterasyon sayısı ve bellek hücresi sayısı artmıştır. *BK-YBTS'nin* iterasyon sayısı ve çalışma süresi *YBTS'nin* iterasyon sayısı ve çalışma süresinden daha az tespit edilirken, bellek hücreleri sayısı açısından *BK-YBTS'de YBTS'den* daha fazla bellek hücresi oluşmuştur. *Kaynak sayısı* parametresinin iki değeri hariç *YBTS* ve *BK-YBTS* %99.1 sınıflandırma doğruluğuna ulaşmıştır. *Kaynak sayısı* parametresi değeri 50 ve 450 olduğunda *YBTS*, *BK-YBTS'den* daha düşük sınıflandırma doğruluğuna ulaşmıştır.

Kaynak sayısı parametresi için iki sistemin sınıflandırma performansı analiz edildikten sonra *uyarma eşiği* parametresinin değişik değerleri için karşılaştırma sonuçları Tablo 6.26'da verilmiştir.

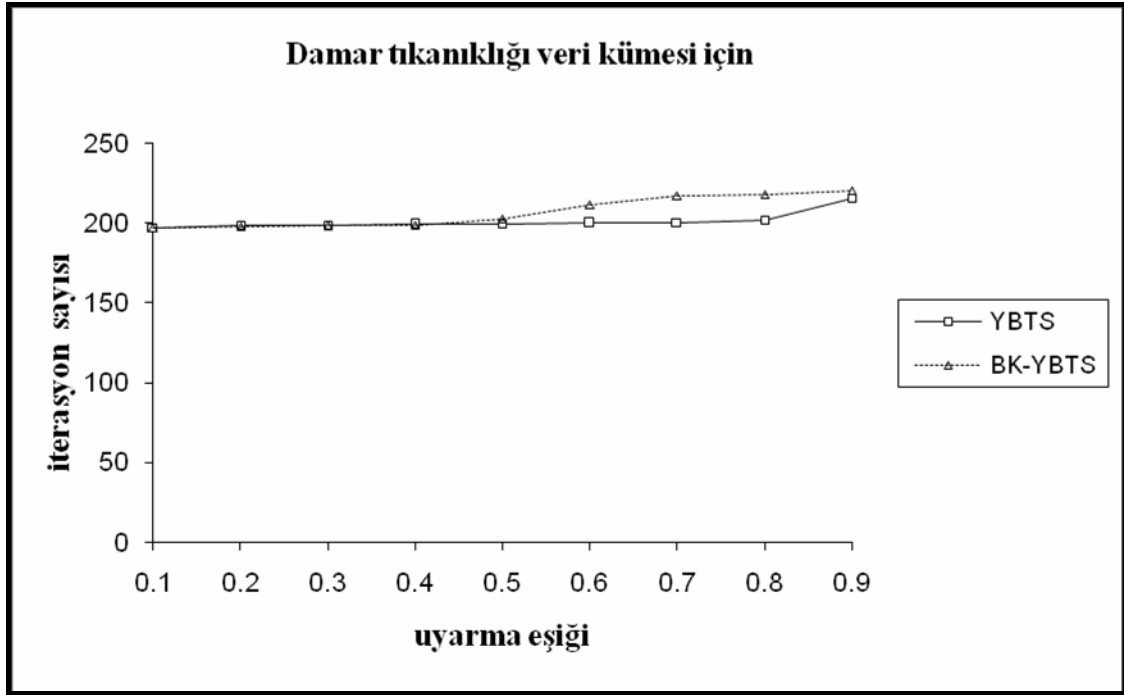
Uyarma eşiği parametresinin değişik değerleri için iki sistemin çalışma süreleri, iterasyon sayıları, bellek hücreleri sayıları ve sınıflandırma doğrulukları açısından karşılaştırılmasının grafiksel olarak gösterimleri sırasıyla Şekil 6.59 ile Şekil 6.62 arasında verilmiştir.

Tablo 6.26 Damar tıkanıklığı veri kümesinde uyarma eşiği parametresi için YBTS ve BK-YBTS karşılaştırma sonuçları

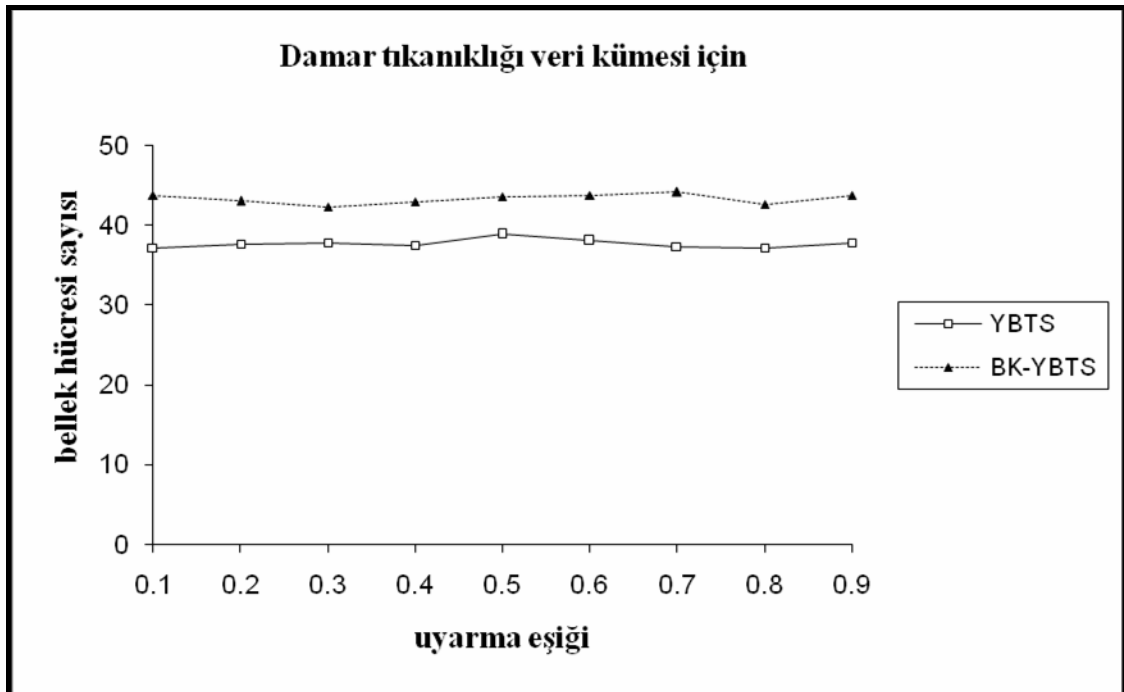
Uyarma Eşği	Süre (sn)		İterasyon Sayısı		Bellek Hücresi Sayısı		Sınıflandırma Doğruluğu (%)	
	YBTS	BK-YBTS	YBTS	BK-YBTS	YBTS	BK-YBTS	YBTS	BK-YBTS
0.1	189.4	181.5	196.8	196.9	37.1	43.7	96.4	99.1
0.2	189.4	181.3	198.2	197.6	37.6	43.0	97.3	99.1
0.3	189.8	182.2	198.2	198.0	37.8	42.3	99.1	99.1
0.4	190.8	184.2	198.9	198.3	37.4	42.9	99.1	99.1
0.5	190.6	184.4	199.1	202.3	38.8	43.6	99.1	99.1
0.6	191.2	190.2	199.7	211.2	38.0	43.7	99.1	99.1
0.7	191.9	193.4	199.9	217.2	37.3	44.1	99.1	98.2
0.8	193.7	193.7	201.6	218.0	37.1	42.6	99.1	99.1
0.9	215.1	197.6	215.1	220.0	37.7	43.7	99.1	99.1



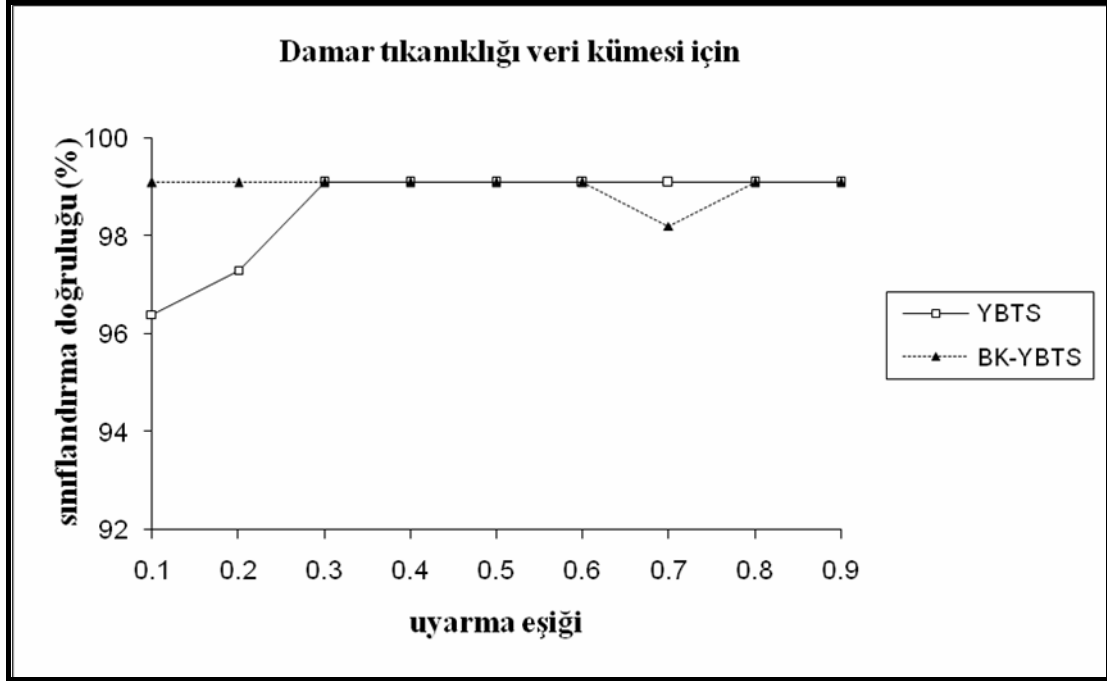
Şekil 6.59 Damar tıkanıklığı veri kümesinde uyarma eşği parametresinin değişimine göre YBTS ve BK-YBTS'nin çalışma sürelerinin grafiksel gösterimi



Şekil 6.60 Damar tıkanıklığı veri kümesinde uyarma eşiği parametresinin değişimine göre YBTS ve BK-YBTS'nin iterasyon sayılarının grafiksel gösterimi



Şekil 6.61 Damar tıkanıklığı veri kümesinde uyarma eşiği parametresinin değişimine göre YBTS ve BK-YBTS'nin bellek hücreleri sayılarının grafiksel gösterimi



Şekil 6.62 Damar tıkanıklığı veri kümesinde uyarma eşiği parametresinin değişimine göre YBTS ve BK-YBTS'nin sınıflandırma doğruluklarının grafiksel gösterimi

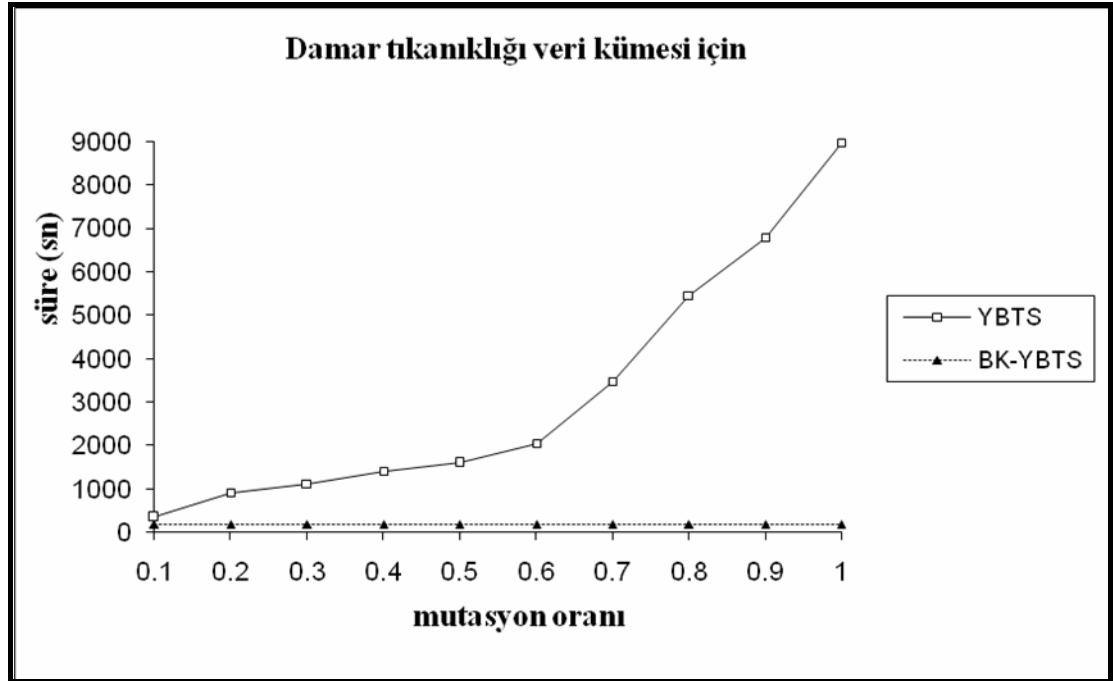
Uyarma eşiği parametresinin üç değeri hariç *YBTS* ve *BK-YBTS* aynı sınıflandırma doğruluğuna ulaşmıştır. *Uyarma eşiği* parametresi değeri 0.7 olduğunda *YBTS*, *BK-YBTS*'den daha yüksek sınıflandırma doğruluğuna ulaşmıştır. *Uyarma eşiği* parametresi değeri 0.1 ve 0.2 olduğunda *YBTS*, *BK-YBTS*'den daha düşük bir sınıflandırma doğruluğuna ulaşmıştır. *Uyarma eşiği* parametresinin değerinin değişimi, her iki sistemde de çalışma zamanını, iterasyon sayısını ve bellek hücresi sayısını çok fazla etkilememiştir. *BK-YBTS*'nin çalışma süresi *YBTS*'nin çalışma süresinden daha az ölçülmüştür. Bellek hücreleri sayısı açısından *BK-YBTS*'de *YBTS*'den daha çok bellek hücresi oluşmuştur.

Uyarma eşiği parametresi için iki sistemin sınıflandırma performansı analiz edildikten sonra *mutasyon oranı* parametresinin değişik değerleri için karşılaştırma sonuçları Tablo 6.27'de verilmiştir.

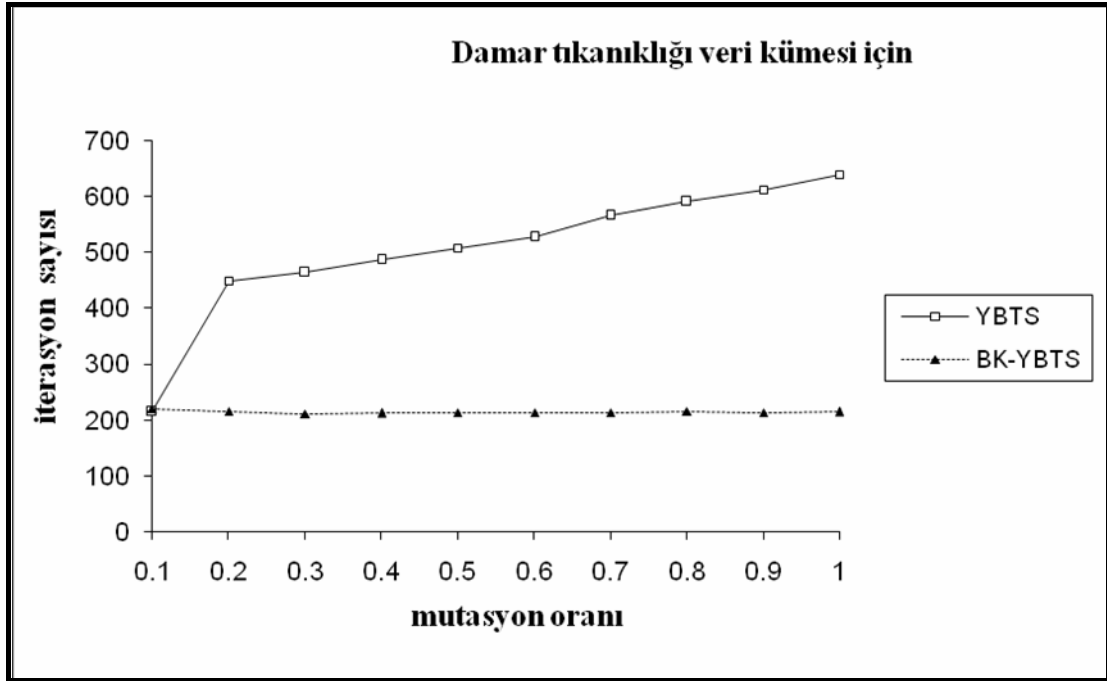
Mutasyon oranı parametresinin değişik değerleri için iki sistemin çalışma süreleri, iterasyon sayıları, bellek hücreleri sayıları ve sınıflandırma doğrulukları açısından karşılaştırılmasının grafiksel olarak gösterimleri sırasıyla Şekil 6.63 ile Şekil 6.66 arasında verilmiştir.

Tablo 6.27 Damar tıkanıklığı veri kümesinde mutasyon oranı parametresi için YBTS ve BK-YBTS karşılaştırma sonuçları

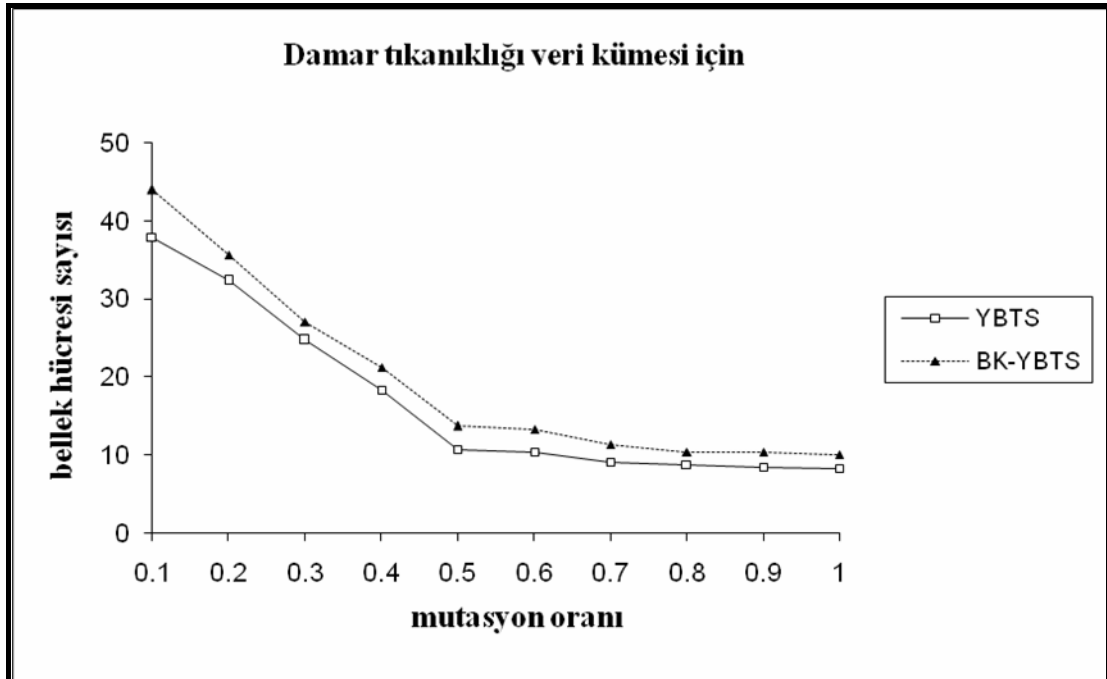
Mutasyon Oranı	Süre (sn)		İterasyon Sayısı		Bellek Hücresi Sayısı		Sınıflandırma Doğruluğu (%)	
	YBTS	BK-YBTS	YBTS	BK-YBTS	YBTS	BK-YBTS	YBTS	BK-YBTS
0.1	340.9	185.9	214.3	220.0	37.8	43.9	99.1	99.1
0.2	915.5	183.3	448.2	216.2	32.3	35.6	99.1	99.1
0.3	1120.6	179.6	463.1	211.3	24.7	27.0	98.2	98.2
0.4	1407.6	180.2	485.7	212.0	18.3	21.2	98.2	97.3
0.5	1593.5	181.1	507.3	213.9	10.7	13.7	97.3	96.4
0.6	2041.2	181.4	526.9	213.1	10.4	13.3	97.3	96.4
0.7	3459.7	182.0	565.3	213.1	9.1	11.4	95.5	95.5
0.8	5448.9	183.7	589.8	215.4	8.8	10.4	96.4	97.3
0.9	6783.2	182.8	611.1	212.7	8.5	10.4	95.5	95.5
1.0	8975.3	183.6	638.9	214.2	8.3	10.1	95.5	96.4



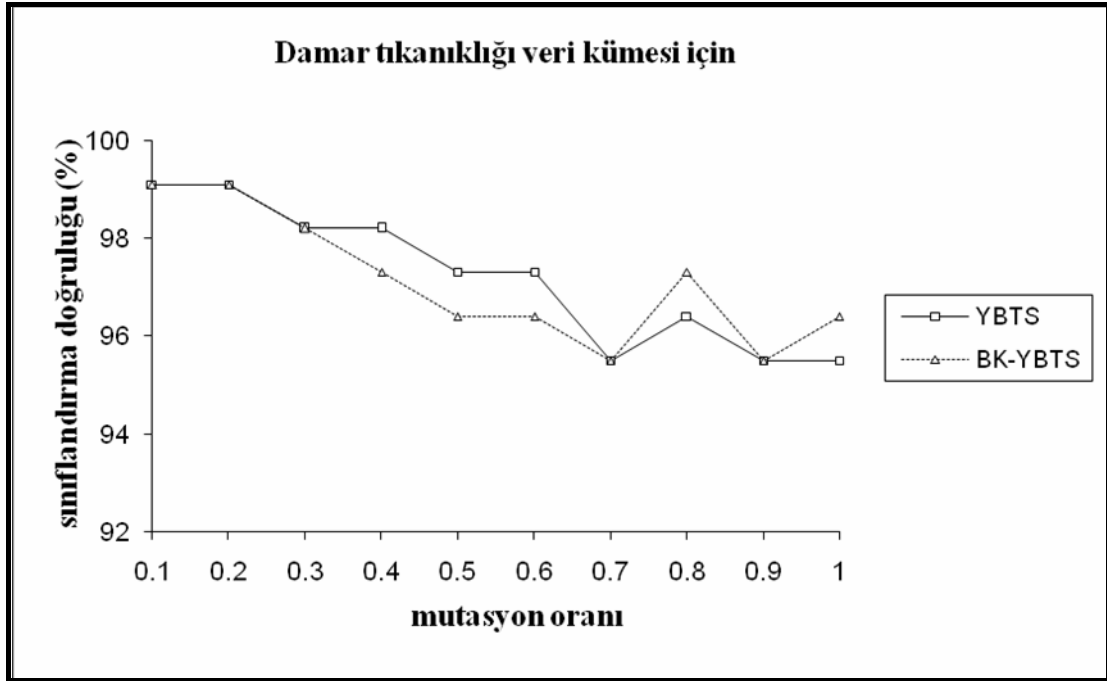
Şekil 6.63 Damar tıkanıklığı veri kümesinde mutasyon oranı parametresinin değişimine göre YBTS ve BK-YBTS'nin çalışma sürelerinin grafiksel gösterimi



Şekil 6.64 Damar tıkanıklığı veri kümesinde mutasyon oranı parametresinin değişimine göre YBTS ve BK-YBTS'nin iterasyon sayılarının grafiksel gösterimi



Şekil 6.65 Damar tıkanıklığı veri kümesinde mutasyon oranı parametresinin değişimine göre YBTS ve BK-YBTS'nin bellek hücreleri sayılarının grafiksel gösterimi



Şekil 6.66 Damar tıkanıklığı veri kümesinde mutasyon oranı parametresinin değişimine göre YBTS ve BK-YBTS'nin sınıflandırma doğruluklarının grafiksel gösterimi

Mutasyon oranı parametresinin 0.8 ve 1.0 değerlerinde BK-YBTS, YBTS'den daha yüksek sınıflandırma doğruluğuna ulaşmıştır. *Mutasyon oranı* değeri 0.4, 0.5 ve 0.6 olduğunda YBTS daha yüksek sınıflandırma doğruluğuna ulaşmıştır. *Mutasyon oranı* arttıkça BK-YBTS'de çalışma zamanında ve iterasyon sayısında önemli bir değişiklik gözlemlenmezken YBTS'de çalışma zamanında ve iterasyon sayısında ciddi artışlar gözlemlenmiştir. Bu durumun sebebi YBTS'de özellik mutasyonunun rasgele değer atanarak yapılmasıdır. BK-YBTS'de ise duyarlılık tabanlı mutasyon mekanizması kullanılmıştır. Bölüm 5.2.2'de YBTS'nin mutasyon mekanizması, Bölüm 5.4.3'te ise BK-YBTS'nin mutasyon mekanizması ayrıntılı olarak açıklanmıştır.

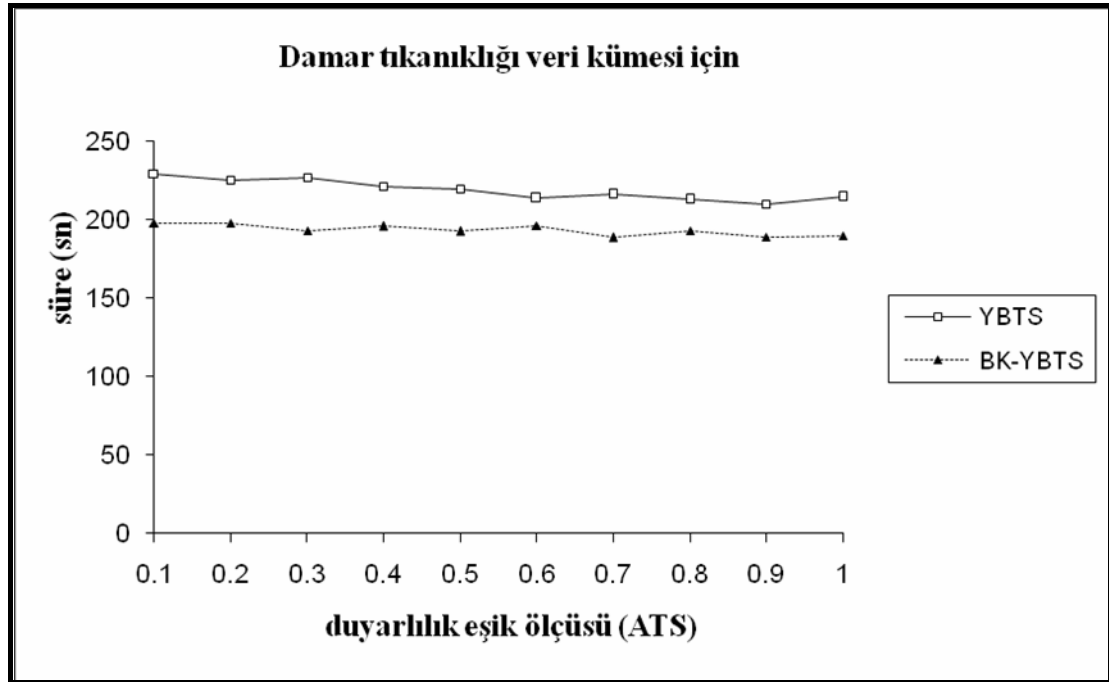
Mutasyon oranı parametresi için iki sistemin sınıflandırma performansı analiz edildikten sonra *duyarlılık eşik ölçüsü* parametresinin değişik değerleri için karşılaştırma sonuçları Tablo 6.28'de verilmiştir.

Duyarlılık eşik ölçüsü parametresinin değişik değerleri için iki sistemin çalışma süreleri, iterasyon sayıları, bellek hücreleri sayıları ve sınıflandırma

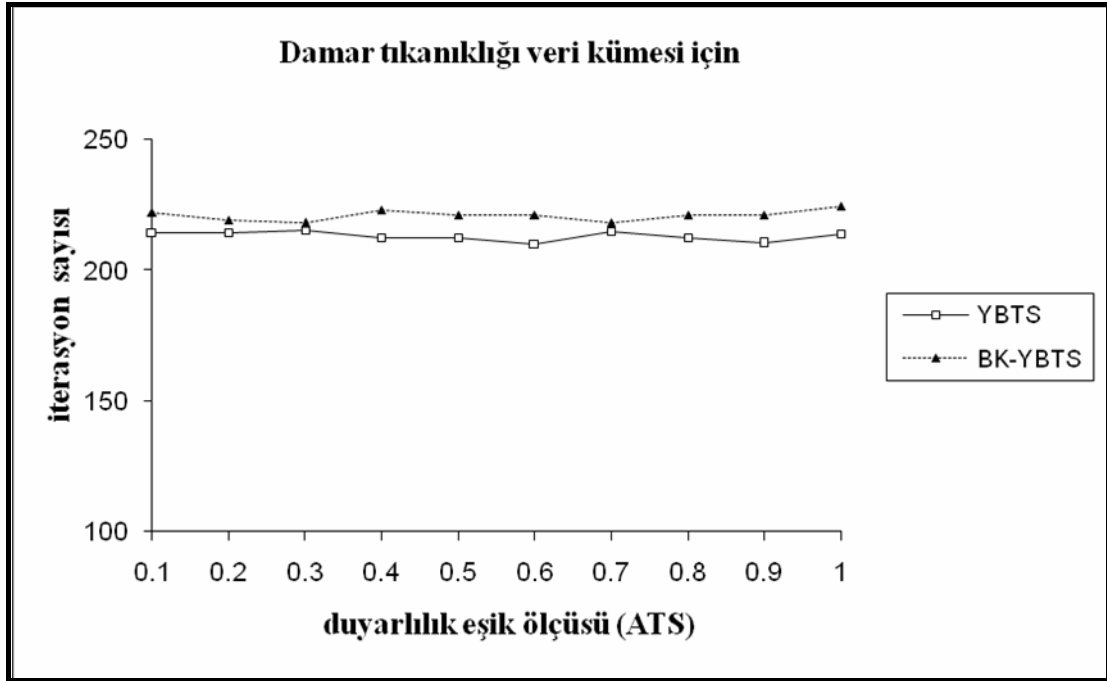
doğrulukları açısından karşılaştırılmasının grafiksel olarak gösterimleri sırasıyla Şekil 6.67 ile Şekil 6.70 arasında verilmiştir.

Tablo 6.28 Damar tıkanıklığı veri kümesinde duyarlılık eşik ölçüsü parametresi için YBTS ve BK-YBTS karşılaştırma sonuçları

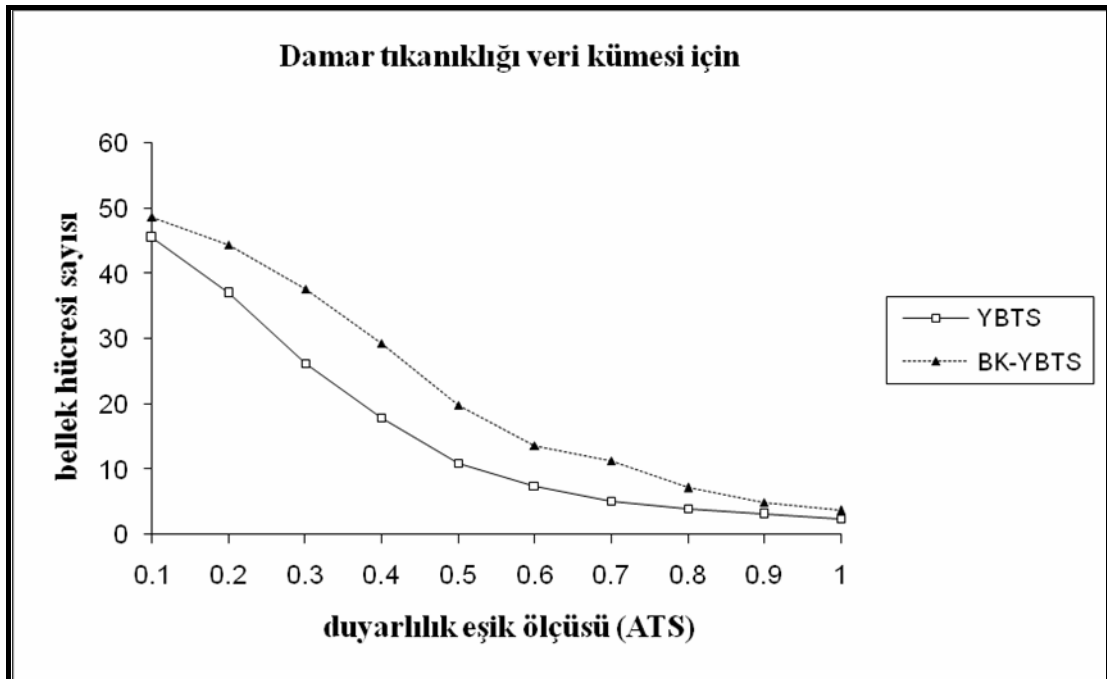
Duyarlılık Eşik Ölçüsü (ATS)	Süre (sn)		İterasyon Sayısı		Bellek Hücresi Sayısı		Sınıflandırma Doğruluğu (%)	
	YBTS	BK-YBTS	YBTS	BK-YBTS	YBTS	BK-YBTS	YBTS	BK-YBTS
0.1	229.7	197.8	214.4	222.1	45.5	48.7	99.1	99.1
0.2	225.6	198.3	214.6	219.3	37.0	44.4	99.1	99.1
0.3	227.0	193.5	215.4	218.1	26.2	37.6	99.1	99.1
0.4	221.2	196.0	212.6	223.1	17.9	29.3	99.1	99.1
0.5	219.6	192.7	212.7	221.1	10.9	19.8	99.1	99.1
0.6	213.8	196.4	210.0	221.4	7.5	13.7	99.1	99.1
0.7	216.2	188.7	215.0	218.3	5.1	11.3	99.1	99.1
0.8	212.9	193.0	212.6	221.4	4.0	7.2	98.2	99.1
0.9	210.3	189.1	210.4	221.3	3.1	5.0	97.3	99.1
1.0	214.6	190.2	214.1	224.5	2.5	3.7	97.3	98.2



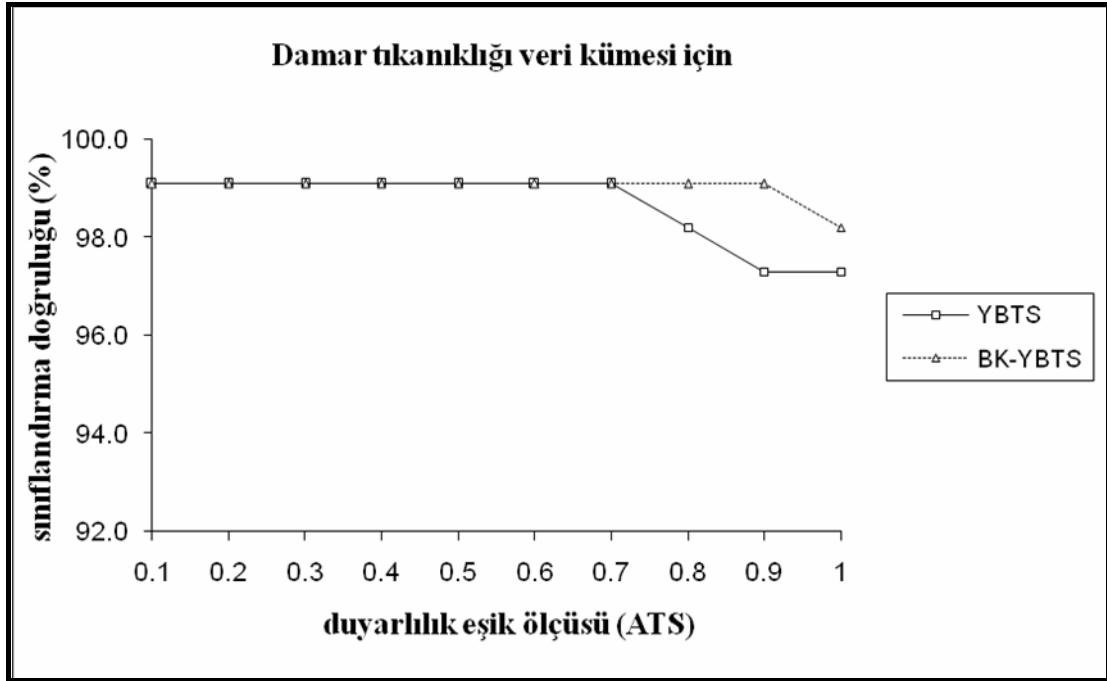
Şekil 6.67 Damar tıkanıklığı veri kümesinde duyarlılık eşik ölçüsü parametresinin değişimine göre YBTS ve BK-YBTS'nin çalışma sürelerinin grafiksel gösterimi



Şekil 6.68 Damar tıkanıklığı veri kümesinde duyarlılık eşik ölçüsü parametresinin değişimine göre YBTS ve BK-YBTS'nin iterasyon sayılarının grafiksel gösterimi



Şekil 6.69 Damar tıkanıklığı veri kümesinde duyarlılık eşik ölçüsü parametresinin değişimine göre YBTS ve BK-YBTS'nin bellek hücreleri sayılarının grafiksel gösterimi



Şekil 6.70 Damar tıkanıklığı veri kümesinde duyarlılık eşik ölçüsü parametresinin değişimine göre YBTS ve BK-YBTS'nin sınıflandırma doğruluklarının grafiksel gösterimi

Duyarlılık eşik ölçüsü parametresi için üç değer hariç, *BK-YBTS* ve *YBTS* aynı sınıflandırma doğruluğuna ulaşmıştır. *Duyarlılık eşik ölçüsü* değeri 0.8, 0.9 ve 1 olduğunda *YBTS*, *BK-YBTS*'den daha düşük sınıflandırma doğruluğuna ulaşmıştır. *Duyarlılık eşik ölçüsü* değeri arttıkça her iki sistemde de çalışma zamanında ve iterasyon sayısında çok az bir oranda azalma gözlemlenmiştir. Tam tersine *duyarlılık eşik ölçüsü* parametresi değeri arttığı zaman her iki sistemde de bellek hücresi sayısında önemli düşüşler gözlemlenmiştir. *Duyarlılık eşik ölçüsü* parametresi bellek hücresi seçiminde aktif rol oynadığı için bu parametrenin değişimi, bellek hücresi sayısını oldukça etkilemiştir. *YBTS*'de duyarlılık eşik ölçüsü kullanımı Bölüm 5.2.4'te, *BK-YBTS* için duyarlılık eşik ölçüsü kullanımı Bölüm 5.4.5'te açıklanmıştır.

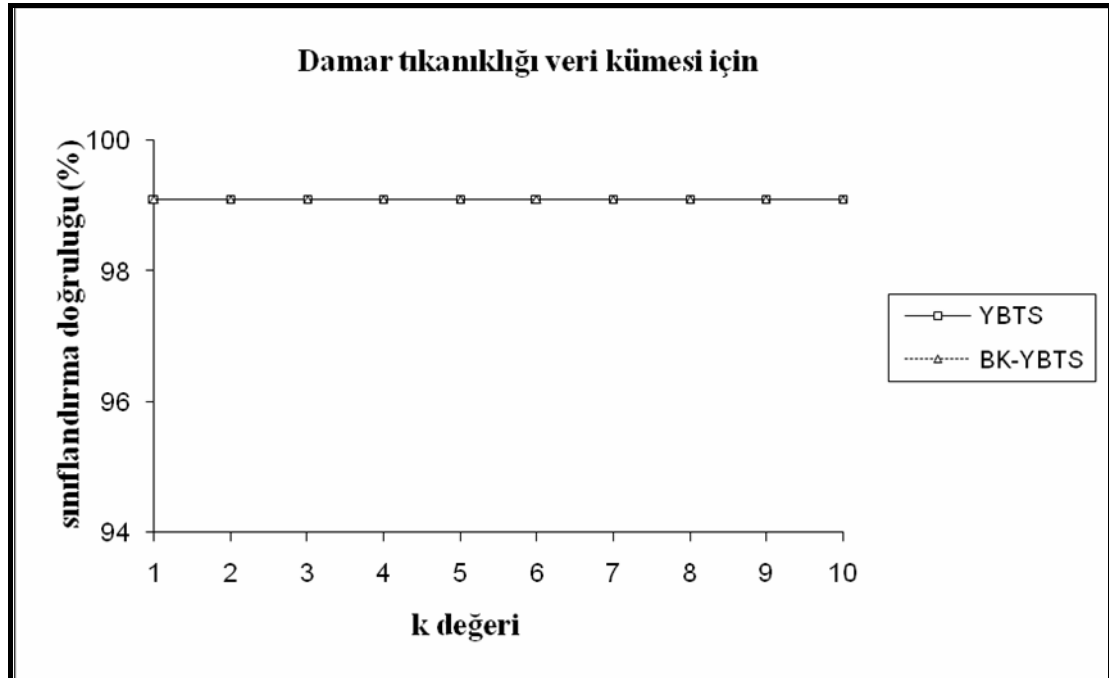
Duyarlılık eşik ölçüsü parametresi için iki sistemin sınıflandırma performansı analiz edildikten sonra *k değeri* parametresinin değişik değerleri için karşılaştırma sonuçları Tablo 6.29'da verilmiştir.

k değeri parametresinin değişik değerleri için iki sistemin sınıflandırma doğrulukları açısından karşılaştırılmasının grafiksel olarak gösterimi Şekil 6.71'de verilmiştir. *YBTS* ve *BK-YBTS* eğitimlerini tamamladıktan sonra oluşan bellek

hücrelerini temel alarak k en yakın komşu yöntemi kullanmaktadırlar. Farklı k değerleri için iki sistem ayrı ayrı çalıştırılmak yerine bir kez çalıştırılmış ve oluşan bellek hücreleri farklı k değerleri ile test verisini sınıflandırmıştır.

Tablo 6.29 Damar tıkanıklığı veri kümesinde k parametresi için YBTS ve BK-YBTS karşılaştırma sonuçları

k değeri	Süre (sn)		İterasyon Sayısı		Bellek Hücresi Sayısı		Sınıflandırma Doğruluğu (%)	
	YBTS	BK-YBTS	YBTS	BK-YBTS	YBTS	BK-YBTS	YBTS	BK-YBTS
1	256.4	194.3	218.0	221.5	36.0	43.2	99.1	99.1
2							99.1	99.1
3							99.1	99.1
4							99.1	99.1
5							99.1	99.1
6							99.1	99.1
7							99.1	99.1
8							99.1	99.1
9							99.1	99.1
10							99.1	99.1



Şekil 6.71 Damar tıkanıklığı veri kümesinde k değeri parametresinin değişimine göre YBTS ve BK-YBTS'nin sınıflandırma doğruluklarının grafiksel gösterimi

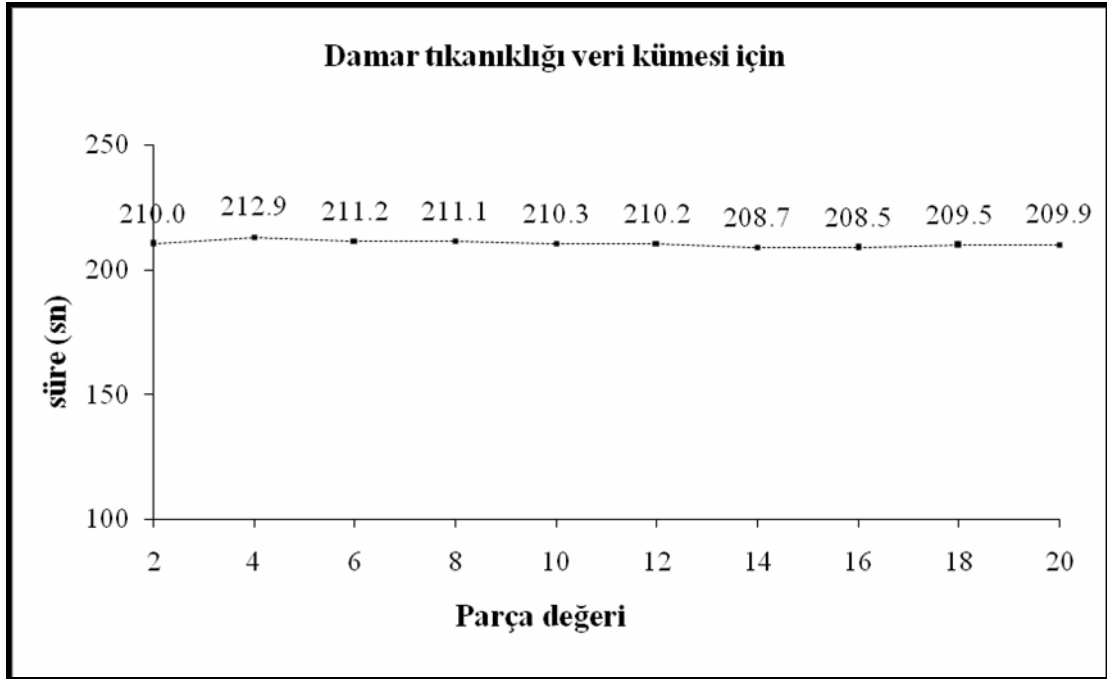
k değeri parametresinin aldığı bütün değerler için *BK-YBTS* ve *YBTS* %99.1 sınıflandırma doğruluğuna ulaşmıştır. Bununla beraber *BK-YBTS*'de ortalama olarak 43.2 adet bellek hücresi oluşurken, *YBTS*'de ortalama olarak 36.0 adet bellek hücresi oluşmuştur. *BK-YBTS*, *YBTS*'den daha kısa sürede çalışmasını tamamlamıştır.

Buraya kadar damar tıkanıklığı veri kümesinde *YBTS* ve *BK-YBTS* için ortak parametreler olan çekirdek hücre sayısı, kaynak sayısı, uyarma eşiği, mutasyon oranı, duyarlılık eşik ölçüsü ve *k değeri* parametrelerinin, sınıflandırma performansına olan etkileri araştırıldıktan sonra sadece *BK-YBTS*'ye özgü olan *parça değeri* parametresinin değişik değerleri için çalışma süreleri, iterasyon sayıları, bellek hücresi sayıları ve sınıflandırma doğrulukları Tablo 6.30'da verilmiştir.

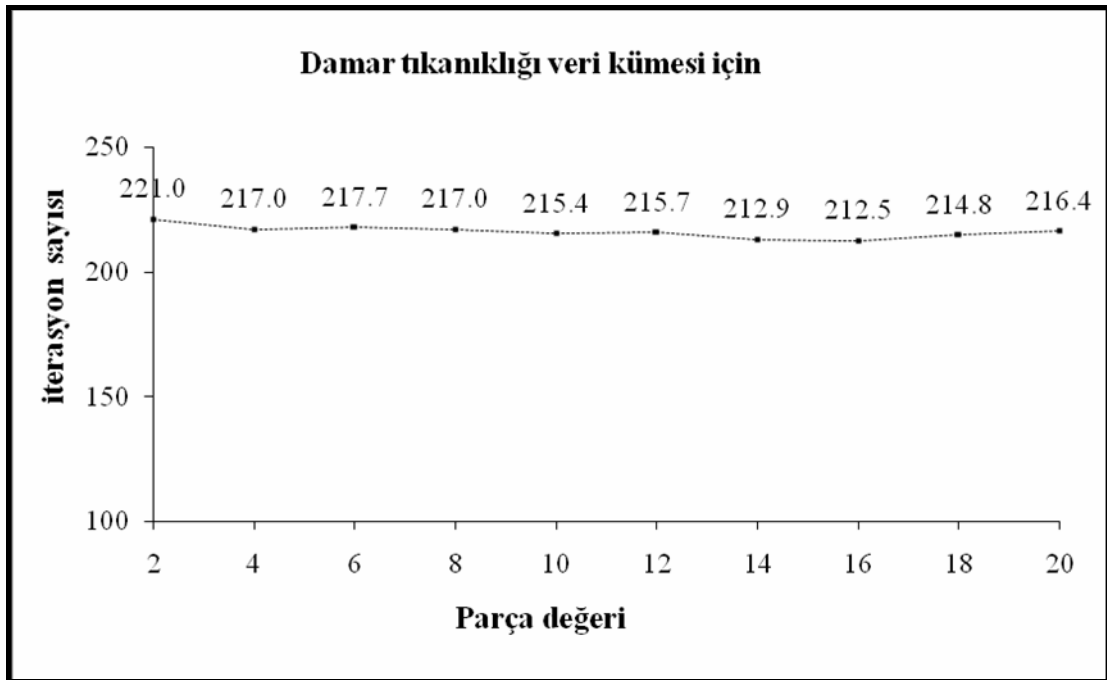
Tablo 6.30 Damar tıkanıklığı veri kümesinde parça değeri parametresi için *BK-YBTS* sonuçları

Parça Değeri	Süre (sn)	İterasyon Sayısı	Bellek Hücresi Sayısı	Sınıflandırma Doğruluğu (%)
2	210.0	221.0	43.9	99.1
4	212.9	217.0	41.2	99.1
6	211.2	217.7	41.8	98.2
8	211.1	217.0	40.7	97.3
10	210.3	215.4	41.1	99.1
12	210.2	215.7	40.9	98.2
14	208.7	212.9	41.5	99.1
16	208.5	212.5	41.3	98.2
18	209.5	214.8	39.2	98.2
20	209.9	216.4	39.2	98.2

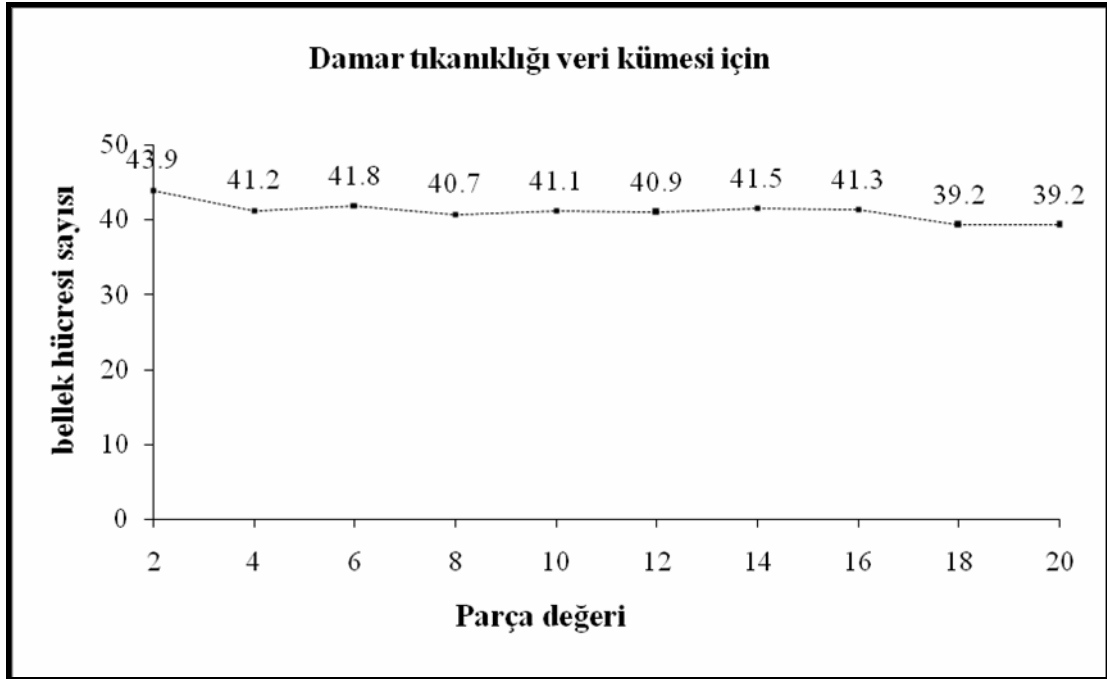
Parça değeri parametresinin aldığı değişik değerler için *BK-YBTS*'nin çalışma süreleri açısından grafiksel olarak gösterimi Şekil 6.72'de, iterasyon sayıları açısından grafiksel olarak gösterimi Şekil 6.73'de, bellek hücreleri sayıları açısından grafiksel olarak gösterimi Şekil 6.74'te ve sınıflandırma doğrulukları açısından grafiksel olarak gösterimi Şekil 6.75'te verilmiştir.



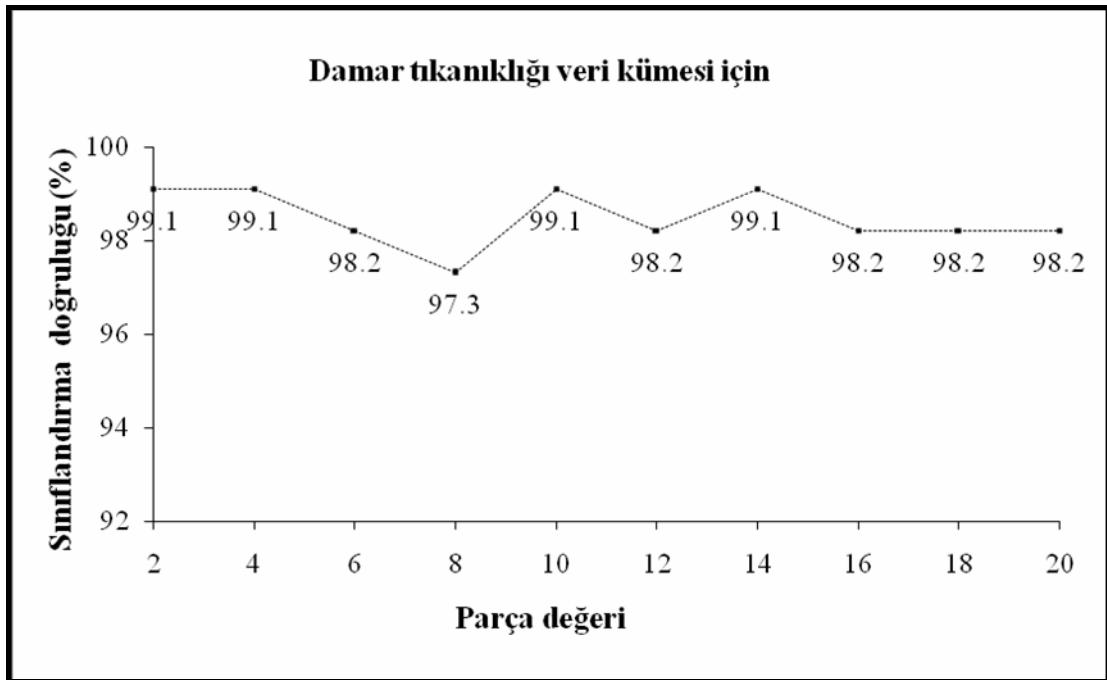
Şekil 6.72 Damar tıkanıklığı veri kümesinde parça değeri parametresinin değişimine göre BK-YBTS'nin çalışma sürelerinin grafiksel gösterimi



Şekil 6.73 Damar tıkanıklığı veri kümesinde parça değeri parametresinin değişimine göre BK-YBTS'nin iterasyon sayılarının grafiksel gösterimi



Şekil 6.74 Damar tıkanıklığı veri kümesinde parça değeri parametresinin değişimine göre BK-YBTS'nin bellek hücreleri sayılarının grafiksel gösterimi



Şekil 6.75 Damar tıkanıklığı veri kümesinde parça değeri parametresinin değişimine göre BK-YBTS'nin sınıflandırma doğruluklarının grafiksel gösterimi

Tablo 6.30'dan da anlaşılabileceği gibi *parça değeri* parametresi 2, 4, 10 ve 14 olduğu durumda BK-YBTS en yüksek sınıflandırma doğruluğu olan % 99.1'e

ulaşmıştır. *Parça değeri* parametresinin değişimi *BK-YBTS'nin* çalışma süresini, iterasyon sayısını ve bellek hücreleri sayısını çok fazla etkilememiştir.

Damar tıkanıklığı veri kümesinde tüm parametreler için genel olarak *BK-YBTS*, *YBTS* benzer sınıflandırma doğruluğuna ulaşmıştır. 114 adet veriden sadece 1 adet veri sınıflandırmada hataya sebep olmaktadır. Bu durum Tablo 6.31'de belirtilmiştir. Veri kümesindeki altıncı parça test kümesi olduğu zaman, 11 adet veriden oluşan bu test kümesinde 1 adet yanlış sınıflandırma söz konusu olmaktadır.

Tablo 6.31 Damar tıkanıklığı veri kümesi için BK-YBTS'nin 10 kat çapraz onaylama yöntemi ile ayrıntılı sınıflandırma doğrulukları

Kat	Eğitim kümesi			Test kümesi			Sınıflandırma Doğruluğu (%)
	Hasta	Normal	Toplam	Hasta	Normal	Toplam	
1	53	50	103	7	4	11	100
2	52	51	103	8	3	11	100
3	56	47	103	4	7	11	100
4	55	48	103	5	6	11	100
5	55	48	103	5	6	11	100
6	55	48	103	5	6	11	99.90
7	50	53	103	10	1	11	100
8	56	47	103	4	7	11	100
9	54	49	103	6	5	11	100
10	54	46	100	6	8	14	100
Ortalama:							99.09

Damar tıkanıklığı veri kümesi üzerinde *YBTS* ve *BK-YBTS* %100'e yakın sınıflandırma doğruluklarına ulaşmıştır. Çalışma süresi açısından *BK-YBTS*, *YBTS'den* daha kısa sürede çalışmıştır. *BK-YBTS* bir çok parametre değeri için %99.1 sınıflandırma doğruluğuna ulaştığı için, en yüksek sınıflandırma doğruluğunu veren parametre değerlerinin verilmesine gerek duyulmamıştır. Tablo 6.32'de *damar tıkanıklığı* veri kümesi için *YBTS*, *BK-YBTS* ve literatürdeki diğer sınıflandırıcıların sınıflandırma doğrulukları belirtilmiştir. Literatürdeki diğer sınıflandırıcılar 10 kat çapraz onaylama yöntemi kullandığı için geliştirilen sistemde de bu eğitim test ayırımı kullanılmıştır. Tablo 6.32'den de görüldüğü gibi en yüksek sınıflandırma

doğruluğuna 61 giriş, 10 gizli katman ve 2 çıkışlı *İleri Beslemeli Geriye Yayılma* algoritması ve *YBTS* ulaşmıştır. Geliştirilen sistemin sınıflandırma doğruluğu bu değere çok yakın tespit edilmiştir. Bu sonuçta göstermektedir ki geliştirilen sistem tıbbi teşhis problemlerinde uzmanlara güvenli bir şekilde yardımcı olabilecektir.

Tablo 6.32 Damar tıkanıklığı veri kümesi için BK-YBTS ve literatürdeki diğer sınıflandırıcılar tarafından elde edilen sınıflandırma doğrulukları

Metot	Sınıflandırma Doğruluğu (%)
<i>MLP + BP, 10 hidden</i>	99.3
<i>YBTS</i>	99.3
<i>BK-YBTS</i>	99.1
<i>C4.5</i>	91.5

6.2.4. Bakteri veri kümesi uygulamaları

Erciyes Üniversitesi Eczacılık Fakültesinde *Cyranose 320* elektronik burun kullanılarak elde edilen bakteri verileri üzerinde geliştirilen sistemin sınıflandırma performansı araştırılmıştır. *Cyranose 320* taşınabilir bir elektronik koku algılayıcı cihazıdır (Dutta ve ark. 2002). Cihaz, 32 ayrı ince film karbon polimer bileşimi sensör içeren bir sensör dizisinden oluşmaktadır. Sensörlerdeki direnç değişimleri sayısal bir örüntü olarak elde edilmektedir. Gerçek zamanlı sensör direnç değerleri *PCNOSE* yazılımı kullanılarak bilgisayar üzerinde bir metin dosyasına aktarılmıştır.

Bakteri denemelerinde 8 farklı bakteri türü araştırılmıştır. Tablo 6.33'te gösterildiği gibi bu bakteri türlerinin dördü *ATCC* (*American Type Culture Collection*) sınıfına aittir (<http://www.lgcpromochem-atcc.com>). Kalan bakteri türleri ise hastalardan klinik ortamında elde edilmiş ve geleneksel metotlarla tanımlanmıştır. Testler laboratuvar şartlarında gerçekleştirilmiştir. Her gün farklı bir bakteri türü test edilmiştir. Bütün bakteri türleri 37 derecede kanlı agar kullanılarak standart petri kaplarında üretilmiştir. Böylece her bakteri türünün üretilmesi aynı şartlarda olmuştur. Petriler etüvde bir gece bekletildikten sonra, etüvden teker teker çıkartılmıştır. Her petri standart bir plastik kilitli poşet içerisinde koku dağılımının

olması için 15 dakika bekletilmiştir. 15 dakika geçtikten sonra *Cyranose 320* veri elde etmek için kullanılmıştır.

Cyranose 320 veri toplamak için kullanılmadan önce sensör haznesinin havası 5 dakika için temizlenmiştir. Daha sonra, bakteriler *Cyranose 320*'nin probu plastik kilitli poşetin içerisine geçirilerek petrilerden veri elde etme işlemine başlanmıştır. Her poşetten 20 örnek alınmıştır. Tablo 6.33'te her bir bakteri türü için kullanılan petri kapları ve örneklerin sayısı verilmektedir. Bu çalışmada 1420 adet örnek elde edilmiştir.

Her bir petri kabından okunan bütün örnekler farklı veri dosyalarında kaydedilmiştir. Bir veri dosyası örnekleme sürecinde 32 sensörden elde edilen direnç değerlerinden oluşmaktadır. Bütün okunan örnekler $(R_{max} - R_o) / R_o$ sensör tepki modeli kullanılarak normalize edilmiştir. Burada R_{max} ; örnek koku sensör tarafından algılandığı zaman sensörün maksimum direnç değeridir. R_o ise, sensör haznesinin kokuya maruz kalmadan önce ortam havası ile temizlenmesi esnasındaki sensörün son direnç değeridir.

Tablo 6.33 Bakteri veri kümesi ayrıntıları

Bakteri Adı	ATCC Sınıfı	Petri Kaplarının Sayısı	Örnek Sayısı
<i>Staphylococcus aureus</i>	ATCC 25923	9	180
<i>Escherichia coli</i>	ATCC 25922	10	200
<i>Pseudomonas aeruginosa</i>	ATCC 27853	8	160
<i>Klebsiella pneumoniae</i>	-	10	200
<i>Acinetobacter baumannii</i>	-	10	200
<i>Candida albicans</i>	ATCC 90028	8	160
<i>Salmonella typhi</i>	-	9	180
<i>Shigella spp.</i>	-	7	140
<i>Toplam</i>		71	1420

Bakteri veri kümesi için YBTS'den farklı olarak BK-YBTS ile beraber önerilen değişikliklerin etkisini araştırmak amacıyla çekirdek hücre sayısı, kaynak sayısı, uyarma eşiği, mutasyon oranı, duyarlılık eşik ölçüsü ve k parametreleri için Tablo 6.34'te belirtilen parametre değerleri temel alınarak YBTS ve BK-YBTS karşılaştırılmıştır.

Tablo 6.34 Bakteri veri kümesi için YBTS ve BK-YBTS'nin başlangıç parametre değerleri

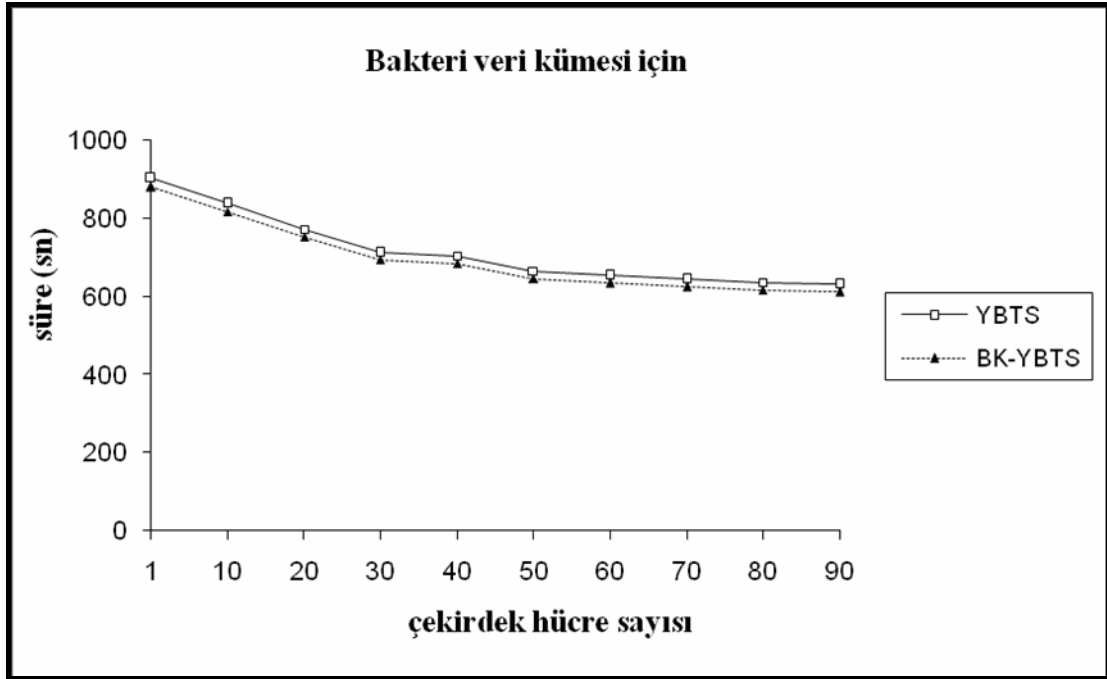
Parametre Adı	Parametre Değeri
Çekirdek hücre sayısı	1
Kaynak sayısı	200
Uyarma eşiği	0.9
Mutasyon oranı	0.1
Duyarlılık eşik ölçüsü	0.2
Çoğunluk oylaması için kullanılan k değeri	1
Parça değeri (BK-YBTS için)	2

Çekirdek hücre sayısı parametresinin aldığı değişik değerler için YBTS ve BK-YBTS karşılaştırma sonuçları Tablo 6.35'te verilmiştir.

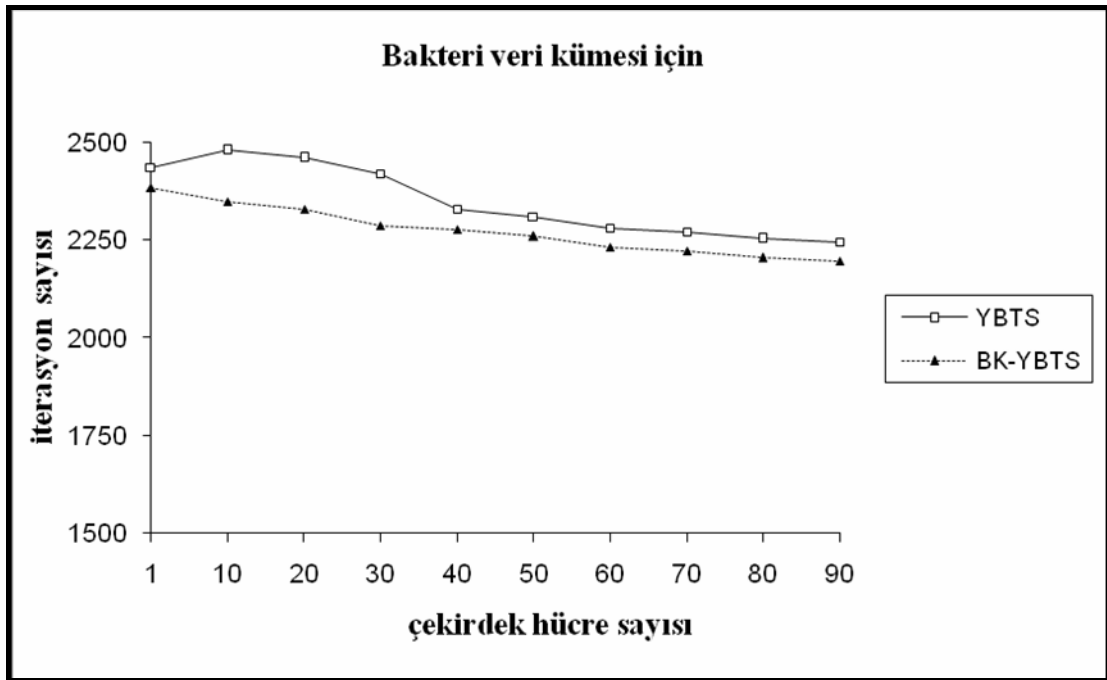
Tablo 6.35 Bakteri veri kümesinde çekirdek hücre sayısı parametresi için YBTS ve BK-YBTS karşılaştırma sonuçları

Çekirdek Hücre Sayısı	Süre (sn)		İterasyon Sayısı		Bellek Hücresi Sayısı		Sınıflandırma Doğruluğu (%)	
	YBTS	BK-YBTS	YBTS	BK-YBTS	YBTS	BK-YBTS	YBTS	BK-YBTS
1	902.5	879.4	2436.6	2383.8	132.2	492.4	84.4	88.5
10	837.7	816.3	2480.0	2348.4	133.1	494.4	85.5	89.2
20	770.4	750.7	2460.0	2329.8	136.4	499.2	84.6	88.2
30	711.1	692.9	2419.8	2286.6	141.2	504.4	84.1	87.4
40	702.1	682.6	2328.9	2278.4	143.1	507.0	83.4	86.6
50	665.3	645.4	2310.3	2260.2	146.4	511.2	83.7	86.5
60	653.2	633.3	2281.5	2232.0	148.7	513.0	83.2	85.9
70	643.4	623.8	2270.9	2221.6	151.1	516.2	82.6	83.2
80	635.0	615.7	2253.5	2204.6	157.2	521.4	82.8	83.4
90	630.3	611.1	2246.3	2197.6	165.4	528.6	82.1	82.6

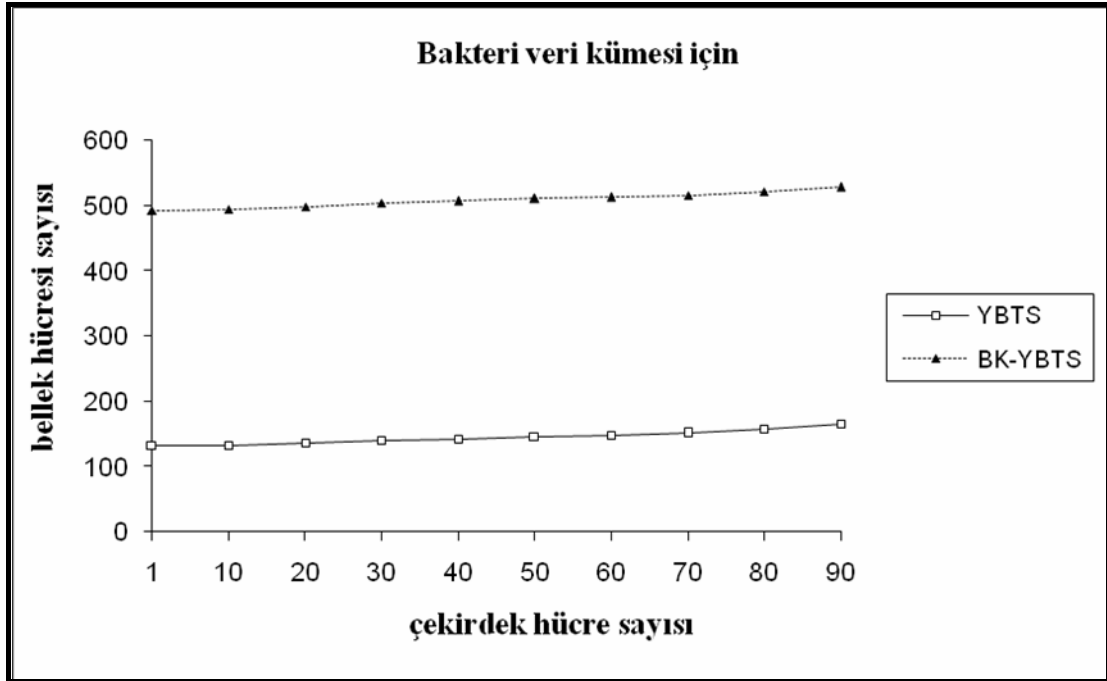
Çekirdek hücre sayısı parametresinin aldığı değişik değerler için iki sistemin çalışma süreleri açısından karşılaştırılmasının grafiksel olarak gösterimi Şekil 6.76'da, iterasyon sayıları açısından karşılaştırılmasının grafiksel olarak gösterimi Şekil 6.77'de, bellek hücreleri sayıları açısından karşılaştırılmasının grafiksel olarak gösterimi Şekil 6.78'de ve sınıflandırma doğrulukları açısından karşılaştırılmasının grafiksel olarak gösterimi Şekil 6.79'da verilmiştir.



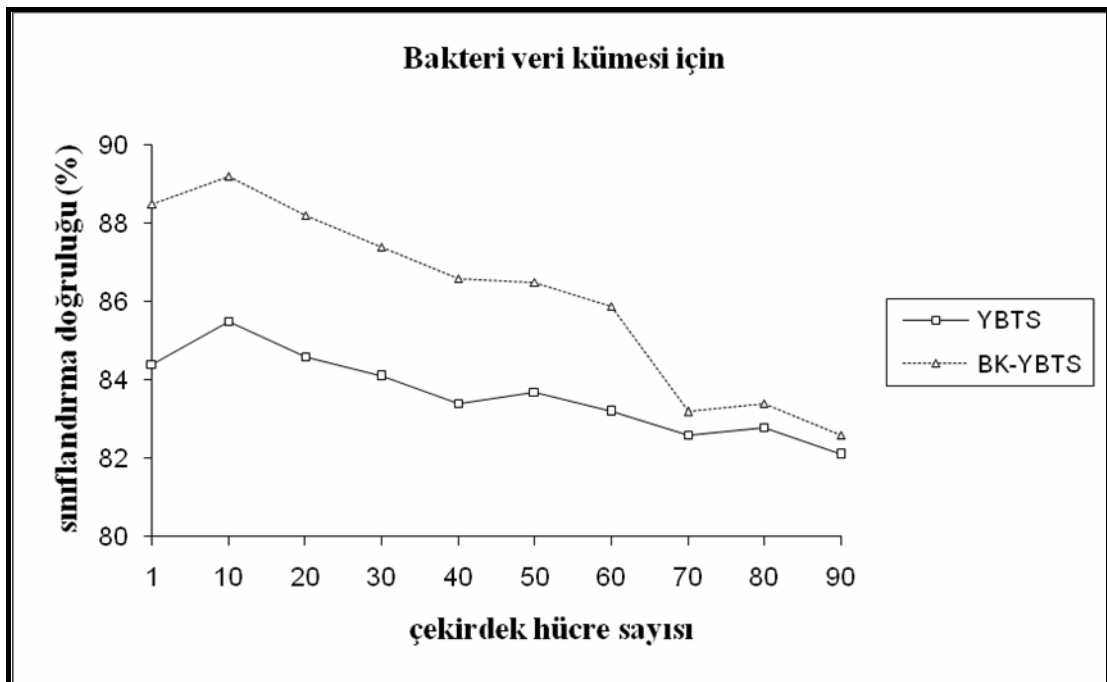
Şekil 6.76 Bakteri veri kümesinde çekirdek hücre sayısı parametresinin değişimine göre YBTS ve BK-YBTS'nin çalışma sürelerinin grafiksel gösterimi



Şekil 6.77 Bakteri veri kümesinde çekirdek hücre sayısı parametresinin değişimine göre YBTS ve BK-YBTS'nin iterasyon sayılarının grafiksel gösterimi



Şekil 6.78 Bakteri veri kümesinde çekirdek hücre sayısı parametresinin değişimine göre YBTS ve BK-YBTS'nin bellek hücreleri sayılarının grafiksel gösterimi



Şekil 6.79 Bakteri veri kümesinde çekirdek hücre sayısı parametresinin değişimine göre YBTS ve BK-YBTS'nin sınıflandırma doğruluklarının grafiksel gösterimi

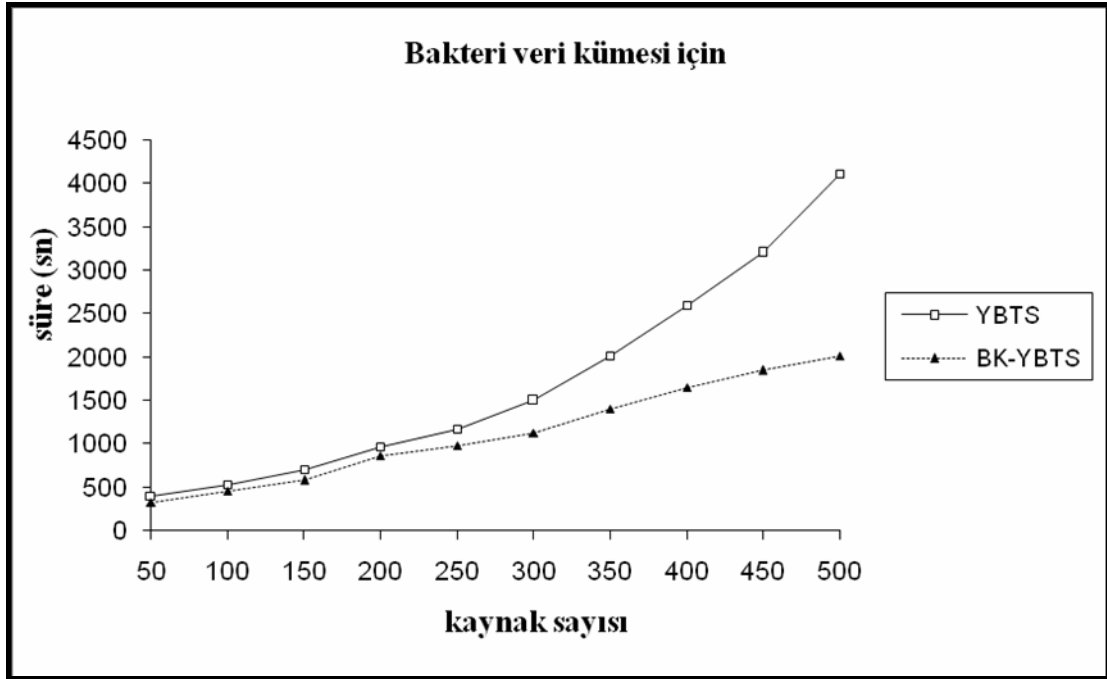
Çekirdek hücre sayısı arttığı zaman, eğitimin başlangıcında antijenler bellek hücresi ve antikör olarak belirlendiği için doğal olarak *YBTS* ve *BK-YBTS*'nin çalışma zamanı ve iterasyon sayısı azalmış, sistemdeki bellek hücresi sayısı ise artmıştır. Bellek hücreleri sayısı *BK-YBTS* için *YBTS*'den daha fazla tespit edilmiştir. *Çekirdek hücre sayısı* parametresinin tüm değerlerinde *BK-YBTS*, *YBTS*'den daha yüksek sınıflandırma doğruluğuna ulaşmıştır.

Çekirdek hücre sayısı parametresi için iki sistemin sınıflandırma performansı analiz edildikten sonra *kaynak sayısı* parametresinin değişik değerleri için karşılaştırma sonuçları Tablo 6.36'da verilmiştir.

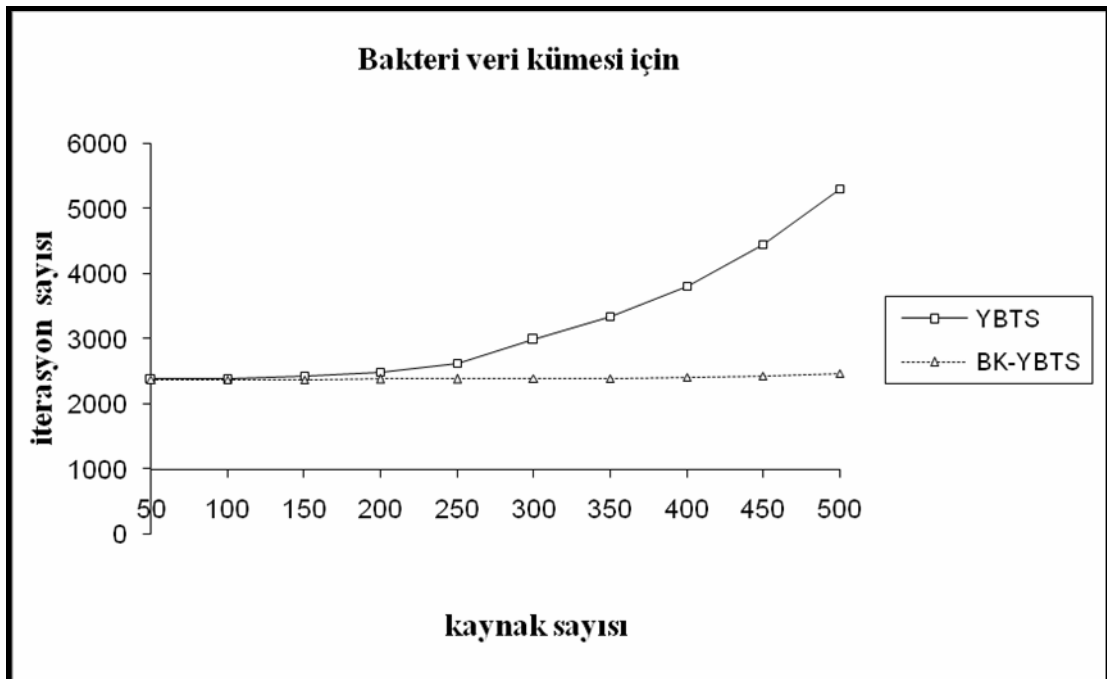
Tablo 6.36 Bakteri veri kümesinde kaynak sayısı parametresi için YBTS ve BK-YBTS karşılaştırma sonuçları

Kaynak Sayısı	Süre (sn)		İterasyon Sayısı		Bellek Hücresi Sayısı		Sınıflandırma Doğruluğu (%)	
	YBTS	BK-YBTS	YBTS	BK-YBTS	YBTS	BK-YBTS	YBTS	BK-YBTS
50	394.5	320.5	2389.9	2371.3	108.6	391.3	78.2	82.1
100	525.4	452.9	2401.7	2375.1	128.9	439.6	82.5	84.1
150	701.3	579.8	2428.2	2379.9	134.3	479.0	85.3	87.3
200	970.2	868.6	2492.8	2383.8	142.1	492.2	84.4	88.5
250	1170.8	980.9	2633.8	2387.2	153.6	517.1	84.6	88.5
300	1496.7	1123.2	2986.0	2390.1	155.8	519.3	85.8	84.9
350	2010.0	1400.5	3345.6	2399.1	162.4	550.5	87.3	86.1
400	2600.5	1650.0	3812.1	2412.5	166.2	563.2	86.5	87.2
450	3201.4	1845.5	4453.2	2432.5	167.1	566.6	83.2	85.2
500	4115.4	2010.4	5312.5	2478.1	171.2	572.1	84.1	85.2

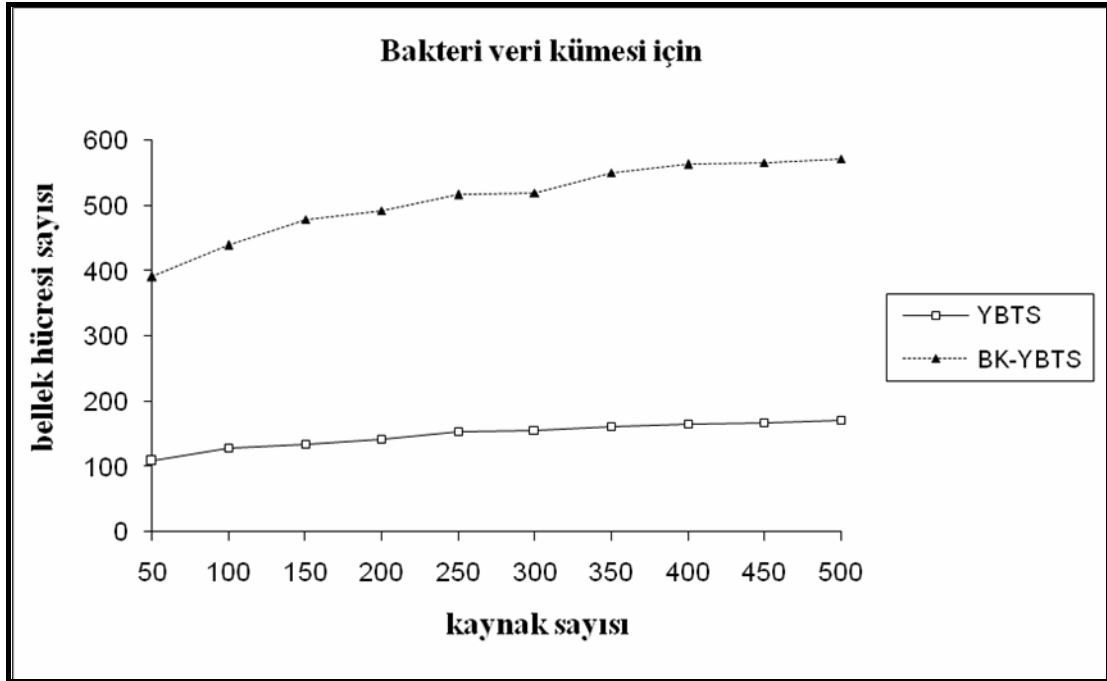
Kaynak sayısı parametresinin değişik değerleri için iki sistemin çalışma sürelerinin grafiksel olarak gösterimi Şekil 6.80'de, iterasyon sayılarının grafiksel olarak gösterimi Şekil 6.81'de, bellek hücreleri sayılarının grafiksel olarak gösterimi Şekil 6.81'de ve sınıflandırma doğruluklarının grafiksel olarak gösterimi Şekil 6.83'te verilmiştir.



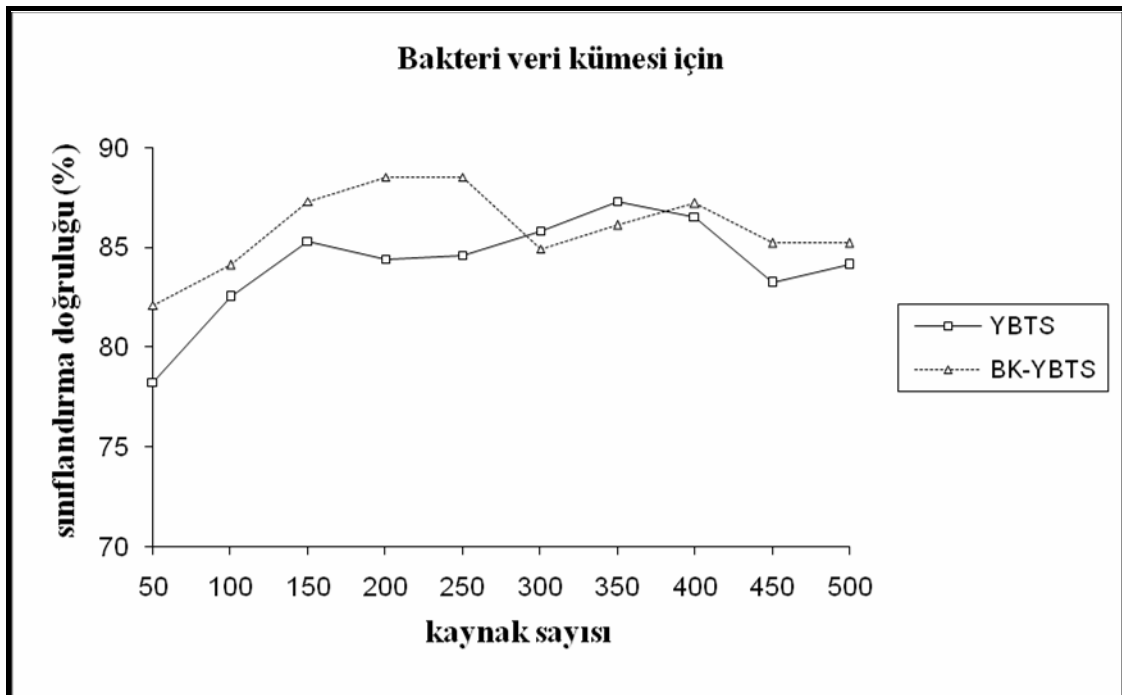
Şekil 6.80 Bakteri veri kümesinde kaynak sayısı parametresinin değişimine göre YBTS ve BK-YBTS'nin çalışma sürelerinin grafiksel gösterimi



Şekil 6.81 Bakteri veri kümesinde kaynak sayısı parametresinin değişimine göre YBTS ve BK-YBTS'nin iterasyon sayılarının grafiksel gösterimi



Şekil 6.82 Bakteri veri kümesinde kaynak sayısı parametresinin değişimine göre YBTS ve BK-YBTS'nin bellek hücreleri sayılarının grafiksel gösterimi



Şekil 6.83 Bakteri veri kümesinde kaynak sayısı parametresinin değişimine göre YBTS ve BK-YBTS'nin sınıflandırma doğruluklarının grafiksel gösterimi

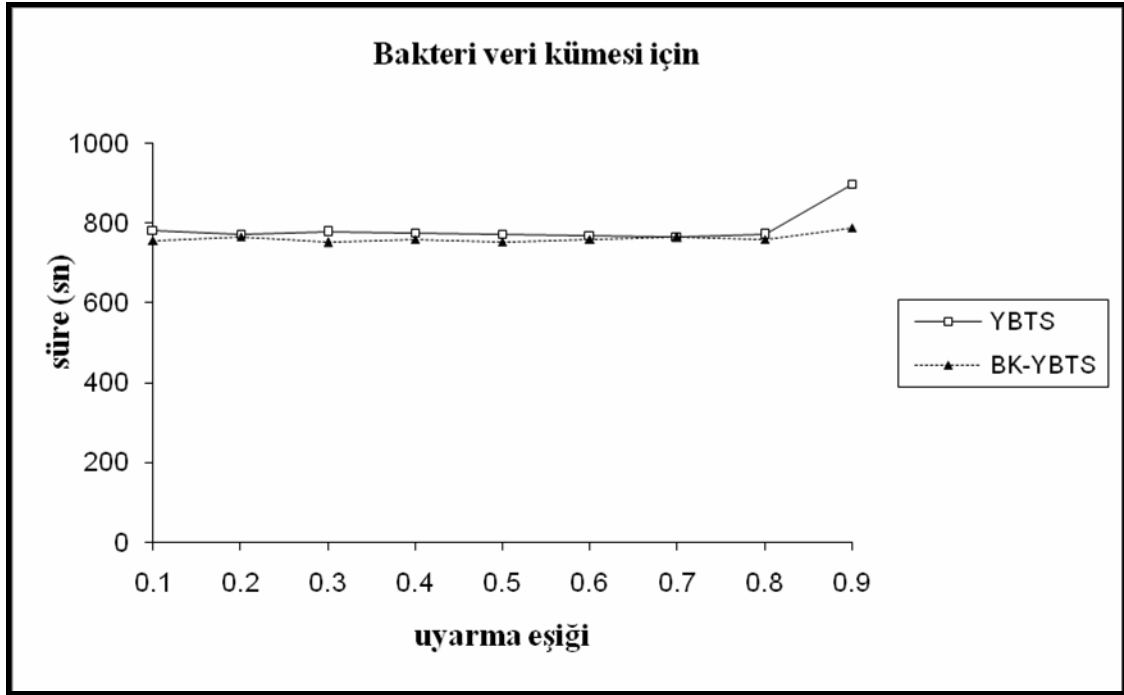
Kaynak sayısı parametresi değeri arttığı zaman, *YBTS* ve *BK-YBTS*'nin çalışma zamanı, iterasyon sayısı ve bellek hücresi sayısı artmıştır. *BK-YBTS*'nin iterasyon sayısı ve çalışma süresi, *YBTS*'nin iterasyon sayısı ve çalışma süresinden daha az tespit edilirken, bellek hücreleri sayısı açısından *BK-YBTS*'de *YBTS*'den daha fazla bellek hücresi oluşmuştur. *Kaynak sayısı* parametresinin iki değeri hariç *BK-YBTS*, *YBTS*'den daha yüksek sınıflandırma doğruluğuna ulaşmıştır. Sadece *kaynak sayısı* parametresi 300 ve 350 değerlerinde *BK-YBTS*, *YBTS*'den daha düşük sınıflandırma doğruluklarına ulaşmıştır.

Kaynak sayısı parametresi için iki sistemin sınıflandırma performansı analiz edildikten sonra *uyarma eşiği* parametresinin değişik değerleri için karşılaştırma sonuçları Tablo 6.37'de verilmiştir.

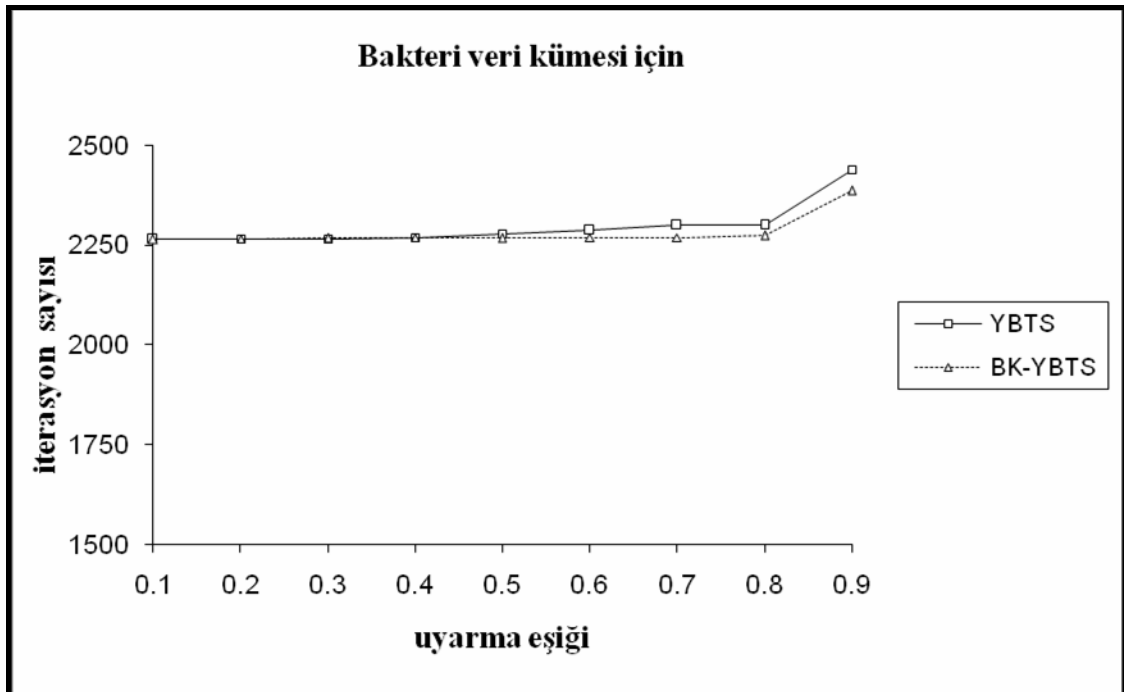
Tablo 6.37 Bakteri veri kümesinde uyarma eşiği parametresi için YBTS ve BK-YBTS karşılaştırma sonuçları

Uyarma Eşiği	Süre (sn)		İterasyon Sayısı		Bellek Hücresi Sayısı		Sınıflandırma Doğruluğu (%)	
	YBTS	BK-YBTS	YBTS	BK-YBTS	YBTS	BK-YBTS	YBTS	BK-YBTS
0.1	780.6	755.3	2263.0	2263.0	135.4	504.8	67.3	70.3
0.2	772.0	764.8	2263.8	2264.0	128.0	510.6	71.8	72.4
0.3	776.1	750.4	2265.8	2266.2	131.0	497.2	76.9	77.2
0.4	773.7	758.5	2267.0	2268.2	132.4	511.6	77.8	77.4
0.5	771.1	752.2	2272.2	2266.6	134.6	498.6	74.0	77.1
0.6	766.9	759.0	2285.8	2268.6	128.0	506.8	77.8	78.5
0.7	764.2	763.3	2298.8	2268.6	140.0	507.0	82.6	83.7
0.8	769.8	758.4	2298.5	2273.0	132.8	502.4	82.9	85.2
0.9	897.5	786.5	2438.2	2385.2	132.2	497.0	84.4	88.5

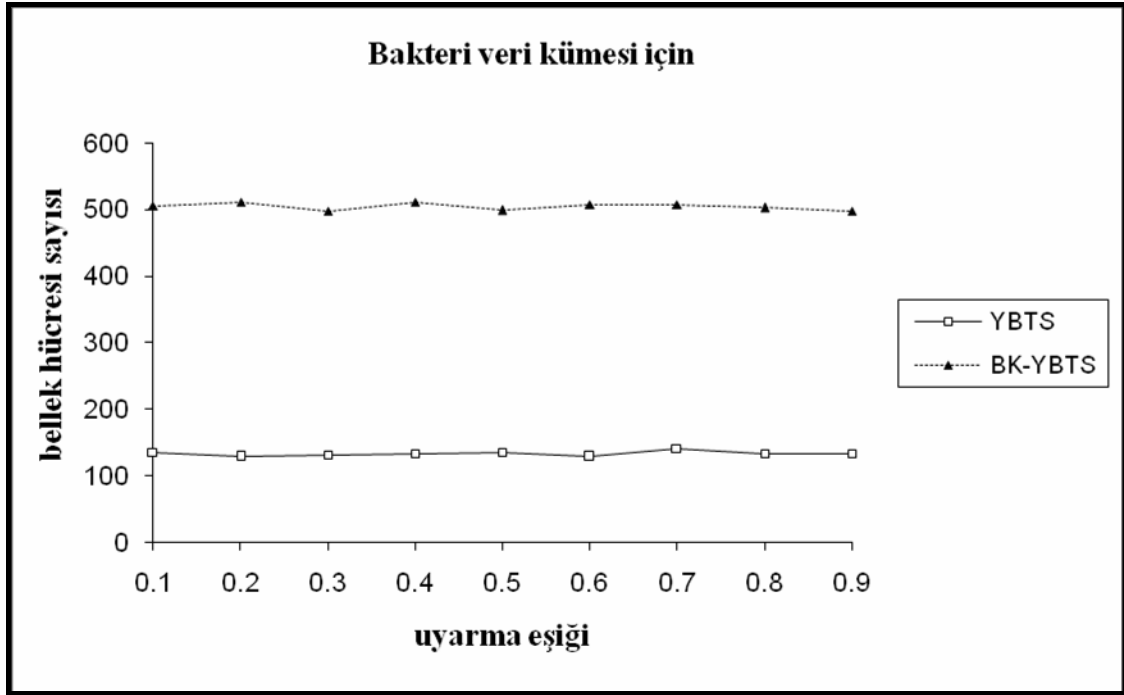
Uyarma eşiği parametresinin değişik değerleri için iki sistemin çalışma süreleri, iterasyon sayıları, bellek hücreleri sayıları ve sınıflandırma doğrulukları açısından karşılaştırılmasının grafiksel olarak gösterimleri sırasıyla Şekil 6.84 ile Şekil 6.87 arasında verilmiştir.



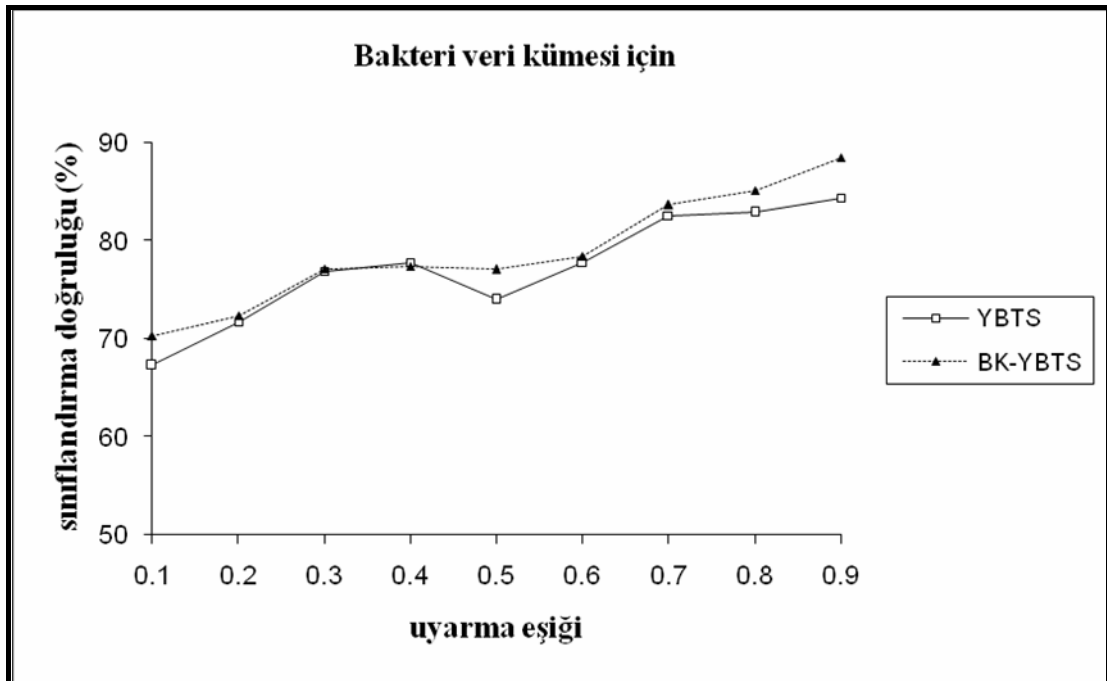
Şekil 6.84 Bakteri veri kümesinde uyarma eşiği parametresinin değişimine göre YBTS ve BK-YBTS'nin çalışma sürelerinin grafiksel gösterimi



Şekil 6.85 Bakteri veri kümesinde uyarma eşiği parametresinin değişimine göre YBTS ve BK-YBTS'nin iterasyon sayılarının grafiksel gösterimi



Şekil 6.86 Bakteri veri kümesinde uyarma eşiği parametresinin değişimine göre YBTS ve BK-YBTS'nin bellek hücreleri sayılarının grafiksel gösterimi



Şekil 6.87 Bakteri veri kümesinde uyarma eşiği parametresinin değişimine göre YBTS ve BK-YBTS'nin sınıflandırma doğruluklarının grafiksel gösterimi

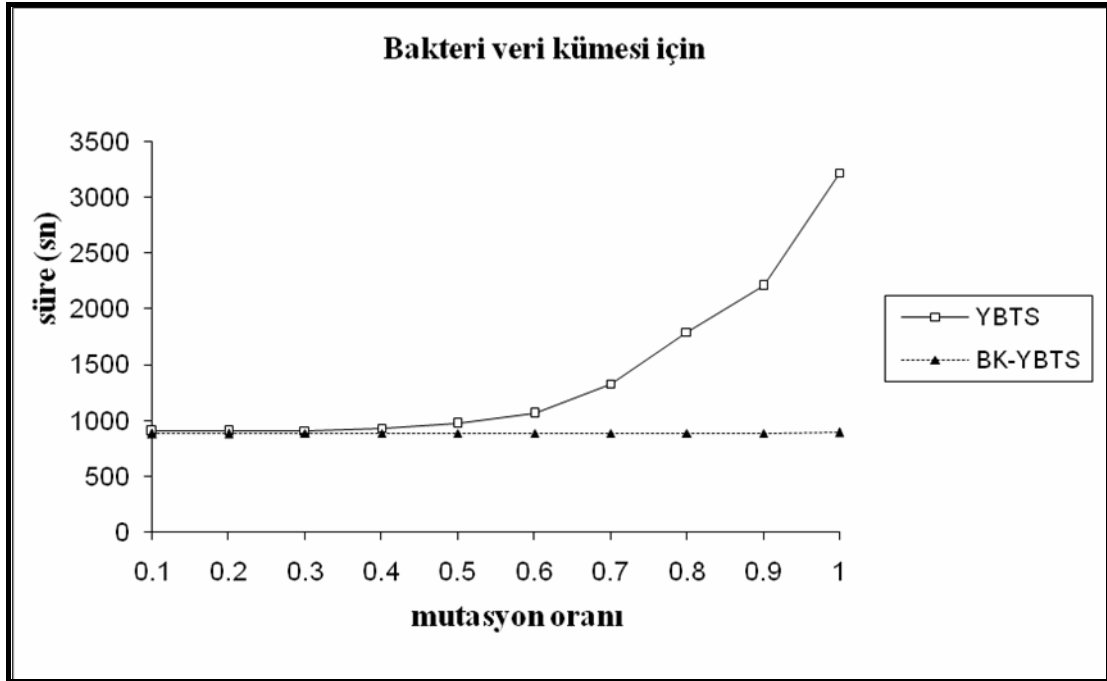
Uyarma eşiği parametresinin bir değeri hariç *BK-YBTS*, *YBTS'den* daha yüksek sınıflandırma doğruluğuna ulaşmıştır. *Uyarma eşiği* parametresi 0.4 değerinde *BK-YBTS*, *YBTS'den* daha düşük sınıflandırma doğruluğuna ulaşmıştır. *Uyarma eşiği* parametresinin değerinin değişimi, her iki sistemde de çalışma zamanını, iterasyon sayısını ve bellek hücresi sayısını çok fazla etkilememiştir. *BK-YBTS'nin* çalışma süresi *YBTS'nin* çalışma süresinden daha az ölçülmüştür. Bellek hücreleri sayısı açısından *BK-YBTS'de* *YBTS'den* daha çok bellek hücresi oluşmuştur.

Uyarma eşiği parametresi için iki sistemin sınıflandırma performansı analiz edildikten sonra *mutasyon oranı* parametresinin değişik değerleri için karşılaştırma sonuçları Tablo 6.38'de verilmiştir.

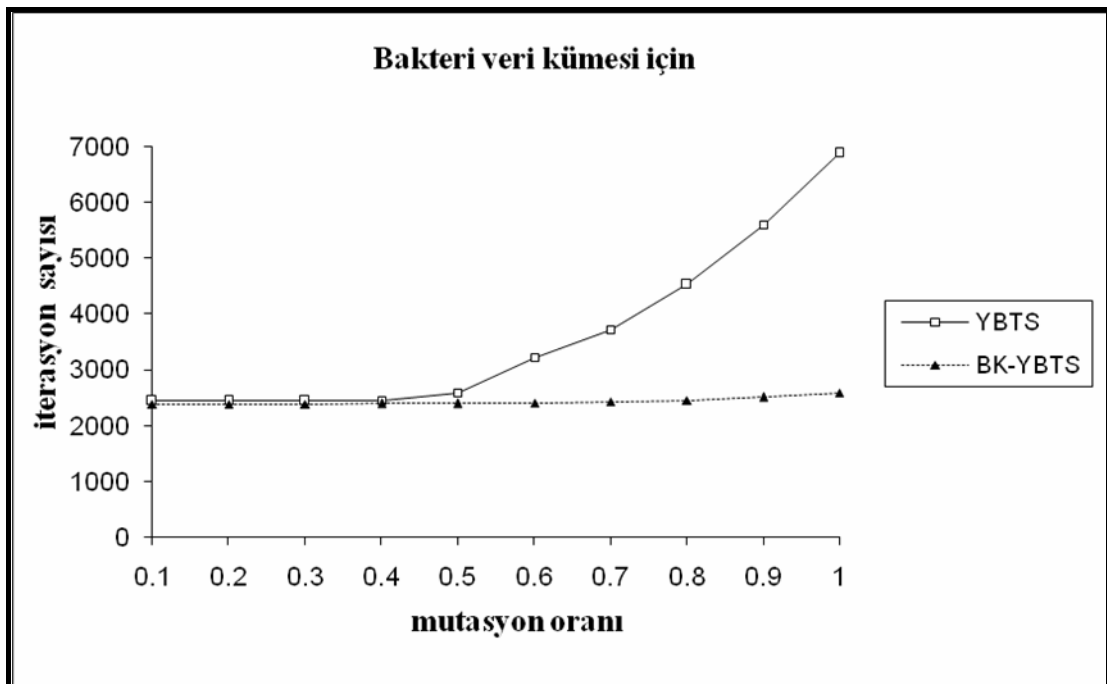
Tablo 6.38 Bakteri veri kümesinde mutasyon oranı parametresi için YBTS ve BK-YBTS karşılaştırma sonuçları

Mutasyon Oranı	Süre (sn)		İterasyon Sayısı		Bellek Hücresi Sayısı		Sınıflandırma Doğruluğu (%)	
	YBTS	BK-YBTS	YBTS	BK-YBTS	YBTS	BK-YBTS	YBTS	BK-YBTS
0.1	902.5	879.4	2436.6	2383.8	132.2	492.4	84.4	88.5
0.2	903.4	879.6	2436.8	2383.8	115.6	401.6	84.7	89.1
0.3	910.5	881.4	2436.8	2383.9	92.3	310.7	83.9	91.1
0.4	931.1	881.8	2441.2	2390.1	72.4	248.6	83.9	89.6
0.5	971.1	882.4	2589.3	2394.1	46.8	180.5	81.2	84.7
0.6	1062.0	883.1	3219.1	2403.3	45.5	175.2	79.5	84.7
0.7	1325.5	883.4	3711.6	2432.3	40.7	153.4	79.8	85.0
0.8	1790.8	886.1	4520.3	2450.4	39.4	142.3	77.2	81.3
0.9	2202.4	889.1	5597.2	2504.6	39.2	142.2	73.8	78.3
1.0	3217.3	891.2	6875.4	2588.5	39.1	139.5	71.2	76.5

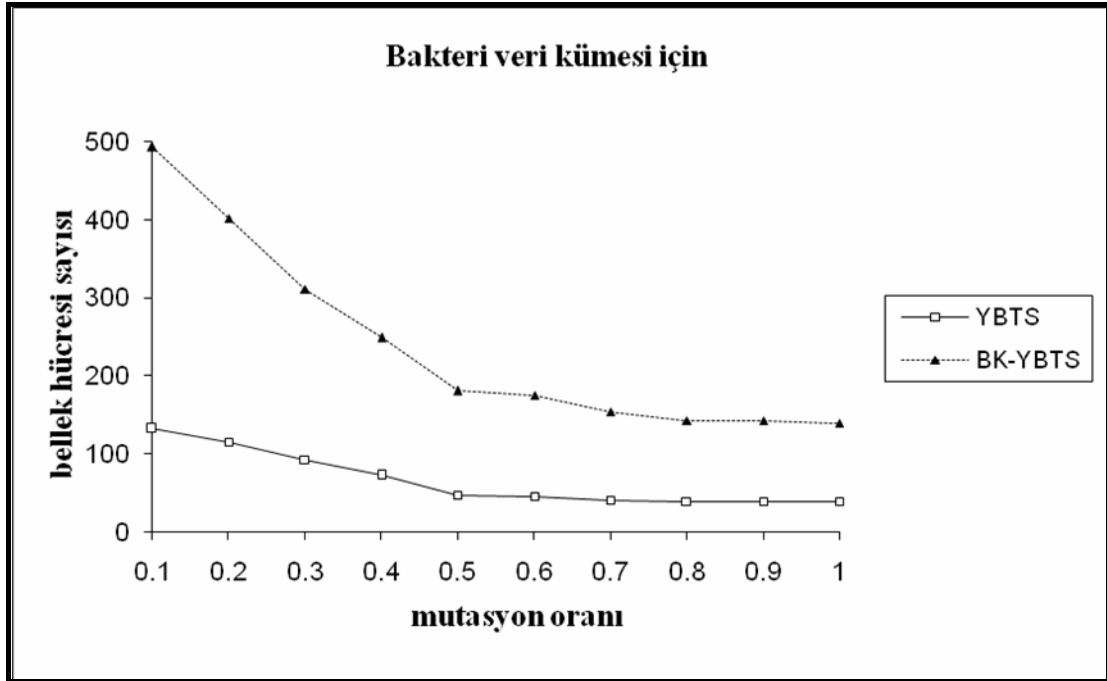
Mutasyon oranı parametresinin değişik değerleri için iki sistemin çalışma süreleri, iterasyon sayıları, bellek hücreleri sayıları ve sınıflandırma doğrulukları açısından karşılaştırılmasının grafiksel olarak gösterimleri sırasıyla Şekil 6.88 ile Şekil 6.91 arasında verilmiştir.



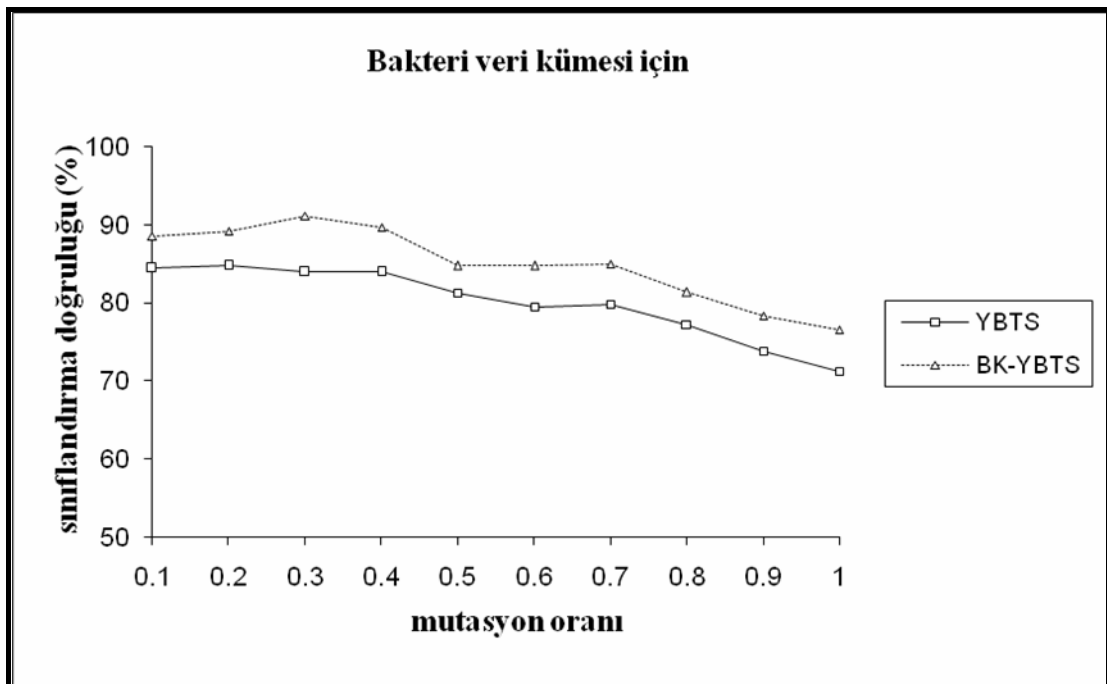
Şekil 6.88 Bakteri veri kümesinde mutasyon oranı parametresinin değişimine göre YBTS ve BK-YBTS'nin çalışma sürelerinin grafiksel gösterimi



Şekil 6.89 Bakteri veri kümesinde mutasyon oranı parametresinin değişimine göre YBTS ve BK-YBTS'nin iterasyon sayılarının grafiksel gösterimi



Şekil 6.90 Bakteri veri kümesinde mutasyon oranı parametresinin değişimine göre YBTS ve BK-YBTS'nin bellek hücreleri sayılarının grafiksel gösterimi



Şekil 6.91 Bakteri veri kümesinde mutasyon oranı parametresinin değişimine göre YBTS ve BK-YBTS'nin sınıflandırma doğruluklarının grafiksel gösterimi

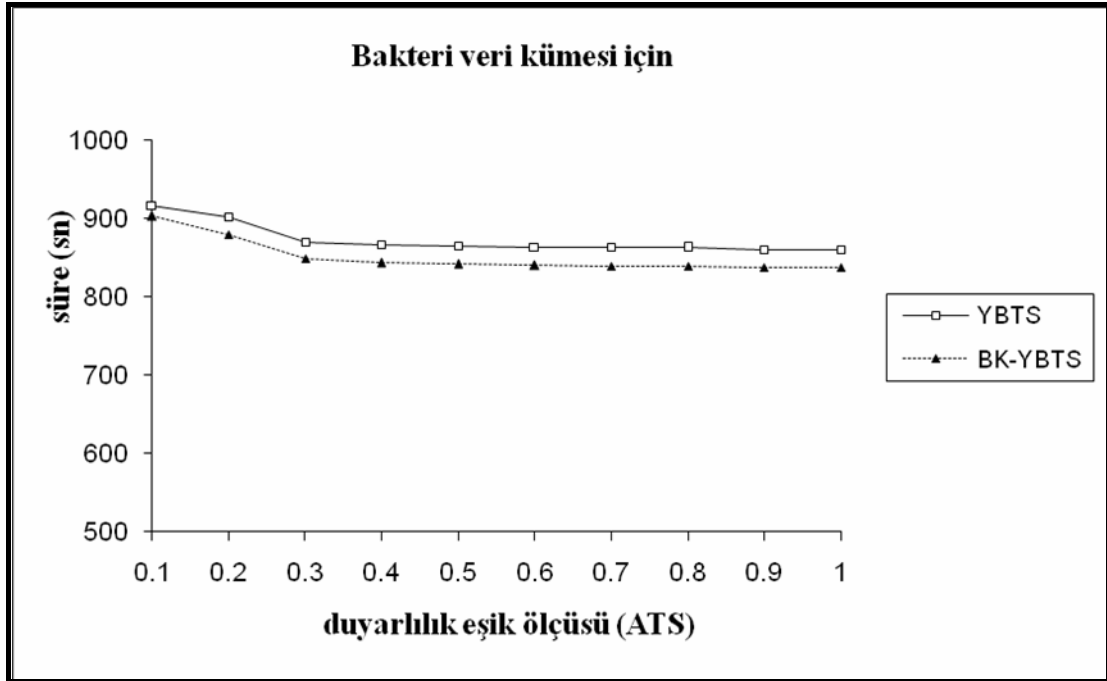
Mutasyon oranı parametresinin bütün değerleri için *BK-YBTS*, *YBTS*'den daha yüksek sınıflandırma doğruluğuna ulaşmıştır. *Mutasyon oranı* arttıkça *BK-YBTS*'de çalışma zamanında ve iterasyon sayısında önemli bir değişiklik gözlemlenmezken *YBTS*'de çalışma zamanında ve iterasyon sayısında ciddi artışlar gözlemlenmiştir. Bu durumun sebebi *YBTS*'de özellik mutasyonunun rasgele değer atanarak yapılmasıdır. *BK-YBTS*'de ise duyarlılık tabanlı mutasyon mekanizması kullanılmıştır. Bölüm 5.2.2'de *YBTS*'nin mutasyon mekanizması, Bölüm 5.4.3'te ise *BK-YBTS*'nin mutasyon mekanizması ayrıntılı olarak açıklanmıştır.

Mutasyon oranı parametresi için iki sistemin sınıflandırma performansı analiz edildikten sonra *duyarlılık eşik ölçüsü* parametresinin değişik değerleri için karşılaştırma sonuçları Tablo 6.39'da verilmiştir.

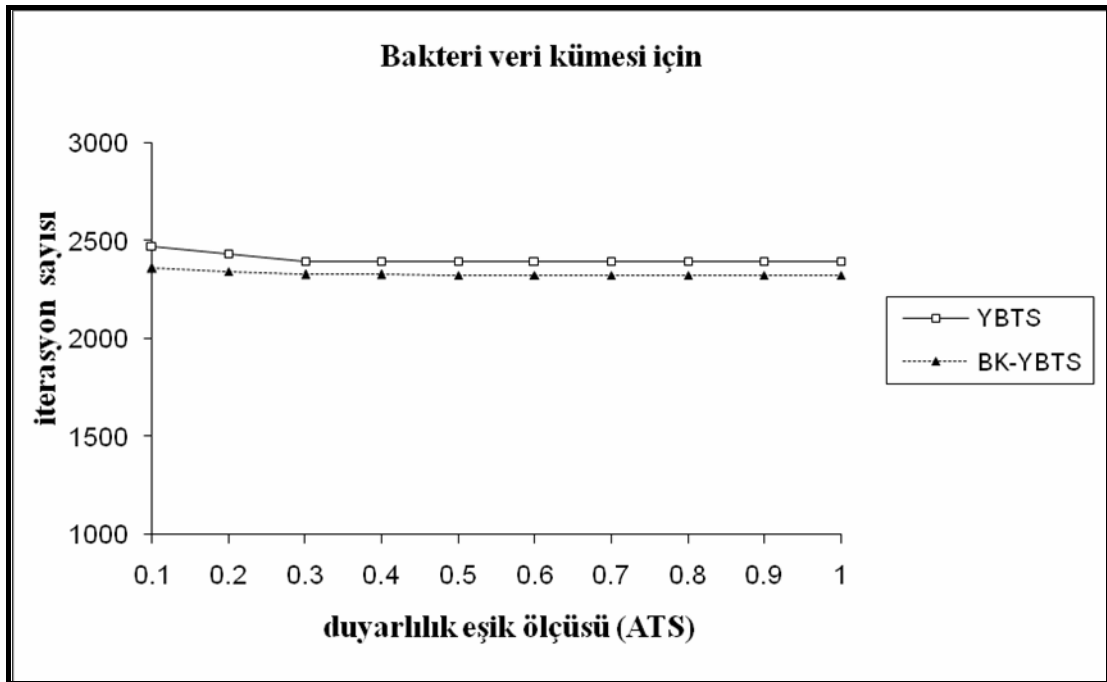
Tablo 6.39 Bakteri veri kümesinde duyarlılık eşik ölçüsü parametresi için YBTS ve BK-YBTS karşılaştırma sonuçları

Duyarlılık Eşik Ölçüsü (ATS)	Süre (sn)		İterasyon Sayısı		Bellek Hücresi Sayısı		Sınıflandırma Doğruluğu (%)	
	YBTS	BK-YBTS	YBTS	BK-YBTS	YBTS	BK-YBTS	YBTS	BK-YBTS
0.1	917.2	903.5	2475.4	2362.2	252.4	650.2	85.2	91.4
0.2	902.5	879.4	2436.6	2343.8	132.2	492.4	84.4	88.5
0.3	870.8	848.9	2398.6	2329.0	81.8	348.3	83.7	88.5
0.4	867.1	843.5	2398.3	2328.8	64.1	275.5	83.7	86.2
0.5	865.4	841.9	2398.2	2328.6	53.9	231.6	82.3	83.6
0.6	864.3	840.4	2398.0	2328.2	47.1	202.2	82.2	82.8
0.7	863.4	839.5	2397.8	2327.8	39.1	165.9	81.1	81.8
0.8	863.2	839.3	2397.7	2327.5	33.9	134.5	80.6	81.1
0.9	860.8	837.9	2397.2	2327.4	31.4	116.5	78.6	79.2
1.0	860.4	837.6	2397.2	2327.3	27.8	110.6	75.2	77.3

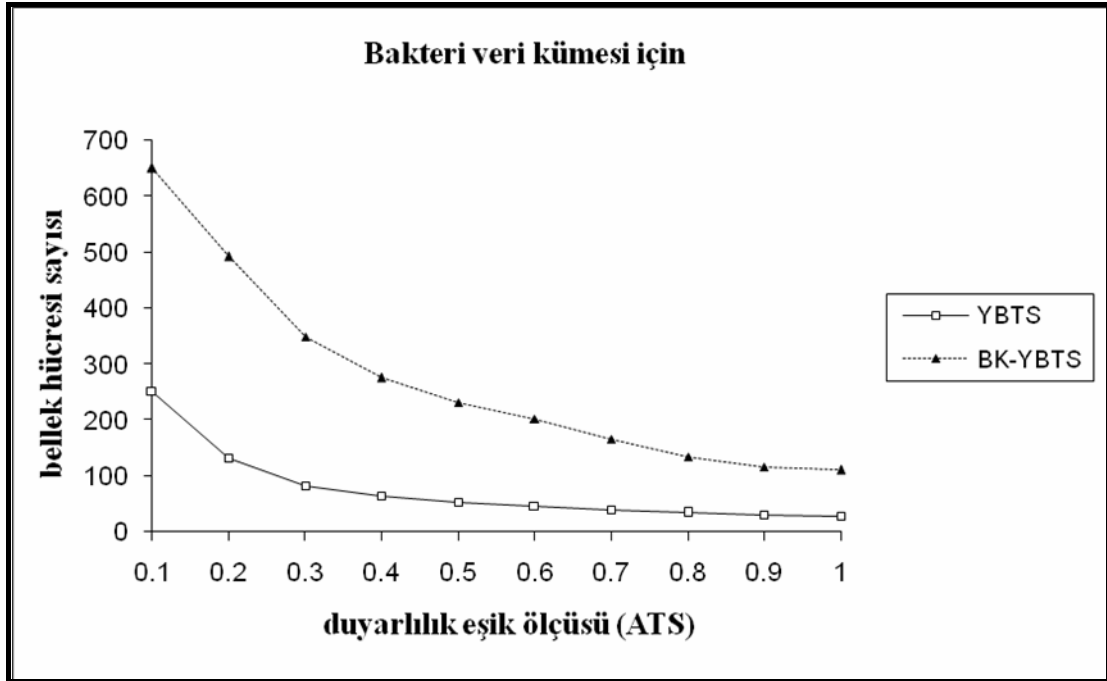
Duyarlılık eşik ölçüsü parametresinin değişik değerleri için iki sistemin çalışma süreleri, iterasyon sayıları, bellek hücreleri sayıları ve sınıflandırma doğrulukları açısından karşılaştırılmasının grafiksel olarak gösterimleri sırasıyla Şekil 6.92 ile Şekil 6.95 arasında verilmiştir.



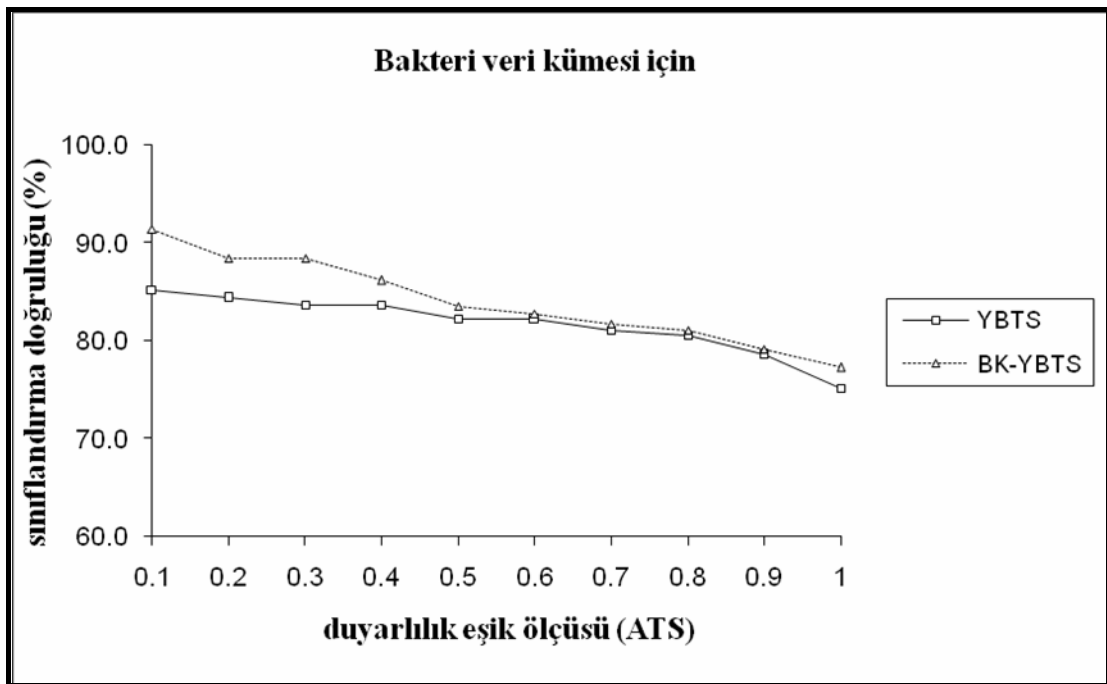
Şekil 6.92 Bakteri veri kümesinde duyarlılık eşik ölçüsü parametresinin değişimine göre YBTS ve BK-YBTS'nin çalışma sürelerinin grafiksel gösterimi



Şekil 6.93 Bakteri veri kümesinde duyarlılık eşik ölçüsü parametresinin değişimine göre YBTS ve BK-YBTS'nin iterasyon sayılarının grafiksel gösterimi



Şekil 6.94 Bakteri veri kümesinde duyarlılık eşik ölçüsü parametresinin değişimine göre YBTS ve BK-YBTS'nin bellek hücreleri sayılarının grafiksel gösterimi



Şekil 6.95 Bakteri veri kümesinde duyarlılık eşik ölçüsü parametresinin değişimine göre YBTS ve BK-YBTS'nin sınıflandırma doğruluklarının grafiksel gösterimi

Duyarlılık eşik ölçüsü parametresinin bütün değerleri için *BK-YBTS*, *YBTS*'den daha yüksek sınıflandırma doğruluğuna ulaşmıştır. *Duyarlılık eşik ölçüsü* değeri arttıkça her iki sistemde de çalışma zamanında ve iterasyon sayısında çok az bir oranda azalma gözlemlenmiştir. Tam tersine *duyarlılık eşik ölçüsü* parametresi değeri arttığı zaman her iki sistemde de bellek hücresi sayısında önemli düşüşler gözlemlenmiştir. *Duyarlılık eşik ölçüsü* parametresi bellek hücresi seçiminde aktif rol oynadığı için bu parametrenin değişimi, bellek hücresi sayısını oldukça etkilemiştir. *YBTS* için duyarlılık eşik ölçüsü kullanımı Bölüm 5.2.4'te, *BK-YBTS* için duyarlılık eşik ölçüsü kullanımı Bölüm 5.4.5'te açıklanmıştır.

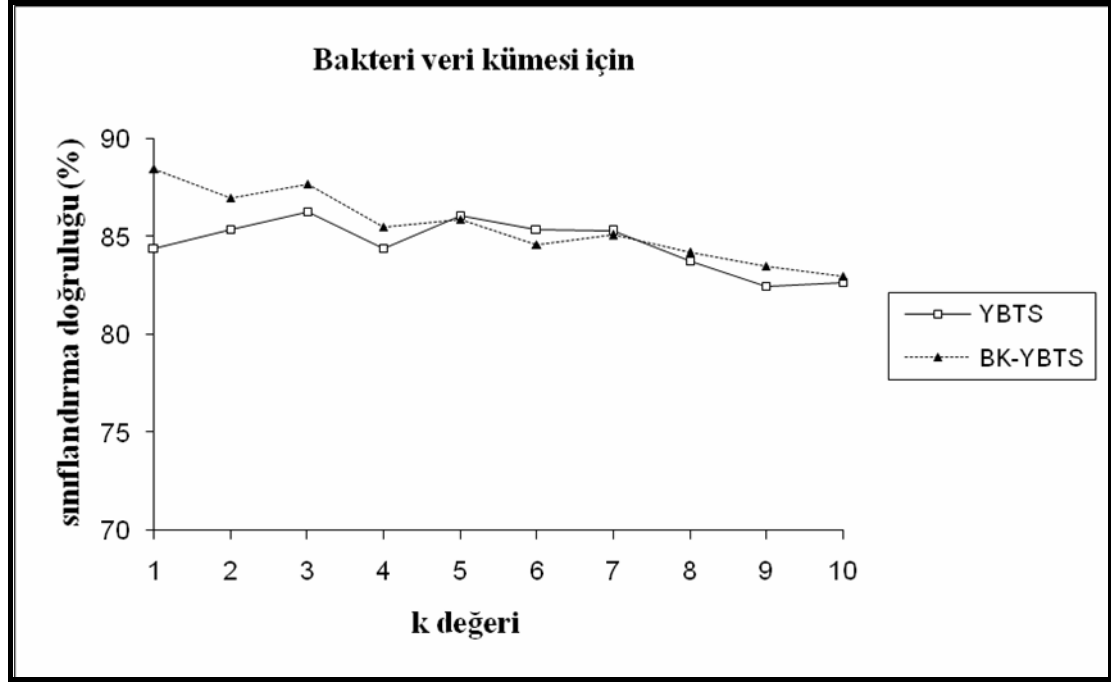
Duyarlılık eşik ölçüsü parametresi için iki sistemin sınıflandırma performansı analiz edildikten sonra *k değeri* parametresinin değişik değerleri için karşılaştırma sonuçları Tablo 6.40'ta verilmiştir.

Tablo 6.40 Bakteri veri kümesinde *k* parametresi için *YBTS* ve *BK-YBTS* karşılaştırma sonuçları

k değeri	Süre (sn)		İterasyon Sayısı		Bellek Hücresi Sayısı		Sınıflandırma Doğruluğu (%)	
	YBTS	BK-YBTS	YBTS	BK-YBTS	YBTS	BK-YBTS	YBTS	BK-YBTS
1	906.8	880.4	2436.9	2385.0	132.7	497.0	84.4	88.5
2							85.4	87.0
3							86.3	87.7
4							84.4	85.5
5							86.1	85.9
6							85.4	84.6
7							85.3	85.1
8							83.8	84.2
9							82.5	83.5
10							82.7	83.0

Test veri kümesini bellek hücreleri üzerinde çoğunluk oylaması yaparak sınıflandırma da kullanılan *k değeri* parametresinin değişik değerleri için iki sistemin sınıflandırma doğrulukları açısından karşılaştırılmasının grafiksel olarak gösterimi Şekil 6.96'da verilmiştir. *YBTS* ve *BK-YBTS* eğitimlerini tamamladıktan sonra oluşan bellek hücrelerini temel alarak *k en yakın komşu* yöntemi kullanılmaktadırlar. Farklı *k*

değerleri için iki sistem ayrı ayrı çalıştırılmak yerine bir kez çalıştırılmış ve oluşan bellek hücreleri farklı k değerleri ile test verisini sınıflandırmıştır. Bundan dolayı farklı k değerleri için iki sistemin çalışma süreleri, iterasyon sayıları ve bellek hücreleri sayıları ortak alınmıştır.



Şekil 6.96 Bakteri veri kümesinde k değeri parametresinin değişimine göre YBTS ve BK-YBTS'nin sınıflandırma doğruluklarının grafiksel gösterimi

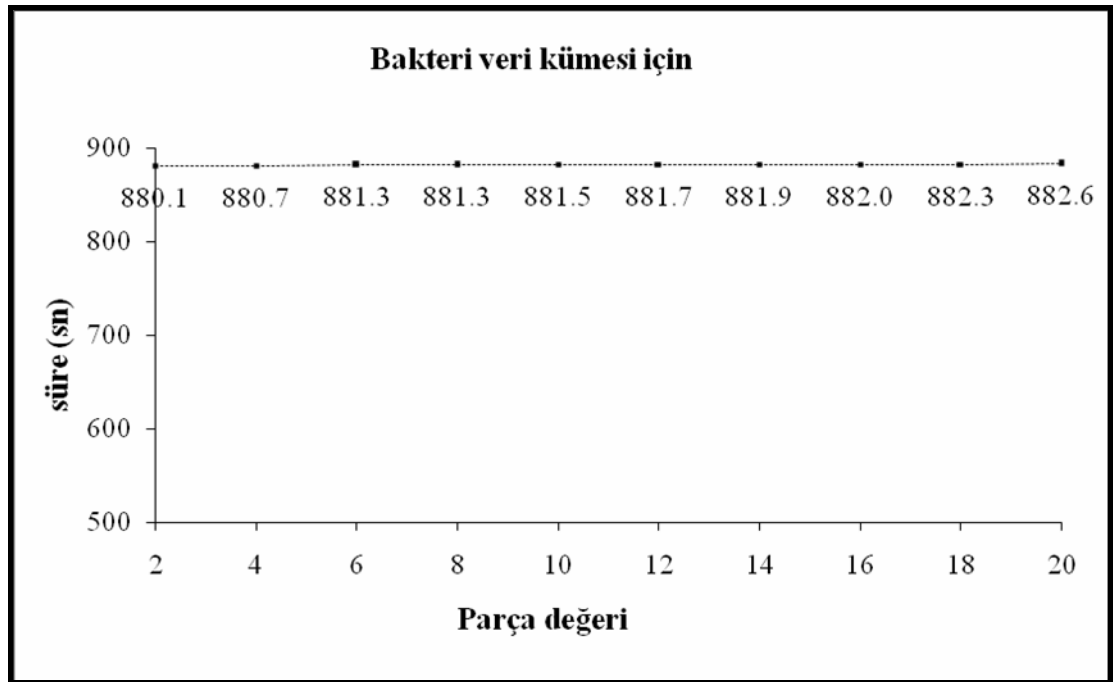
k değeri parametresinin aldığı bir değer hariç *BK-YBTS*, *YBTS*'den daha yüksek sınıflandırma doğruluğuna ulaşmıştır. k parametresi 7 değerini aldığı anda *YBTS*, *BK-YBTS*'den daha yüksek sınıflandırma doğruluğuna ulaşmıştır. Bununla beraber *BK-YBTS*'de ortalama olarak 497.0 adet bellek hücresi oluşurken, *YBTS*'de ortalama olarak 132.7 adet bellek hücresi oluşmuştur.

Buraya kadar *bakteri* veri kümesinde *YBTS* ve *BK-YBTS* için ortak parametreler olan çekirdek hücre sayısı, kaynak sayısı, uyarma eşiği, mutasyon oranı, duyarlılık eşik ölçüsü ve k değeri parametrelerinin, sınıflandırma performansına olan etkileri araştırılmıştır. Bu aşamadan sonra sadece *BK-YBTS*'ye özgü olan *parça değeri* parametresinin değişik değerleri için çalışma süreleri, iterasyon sayıları, bellek hücresi sayıları ve sınıflandırma doğrulukları hesaplanarak Tablo 6.41'de verilmiştir.

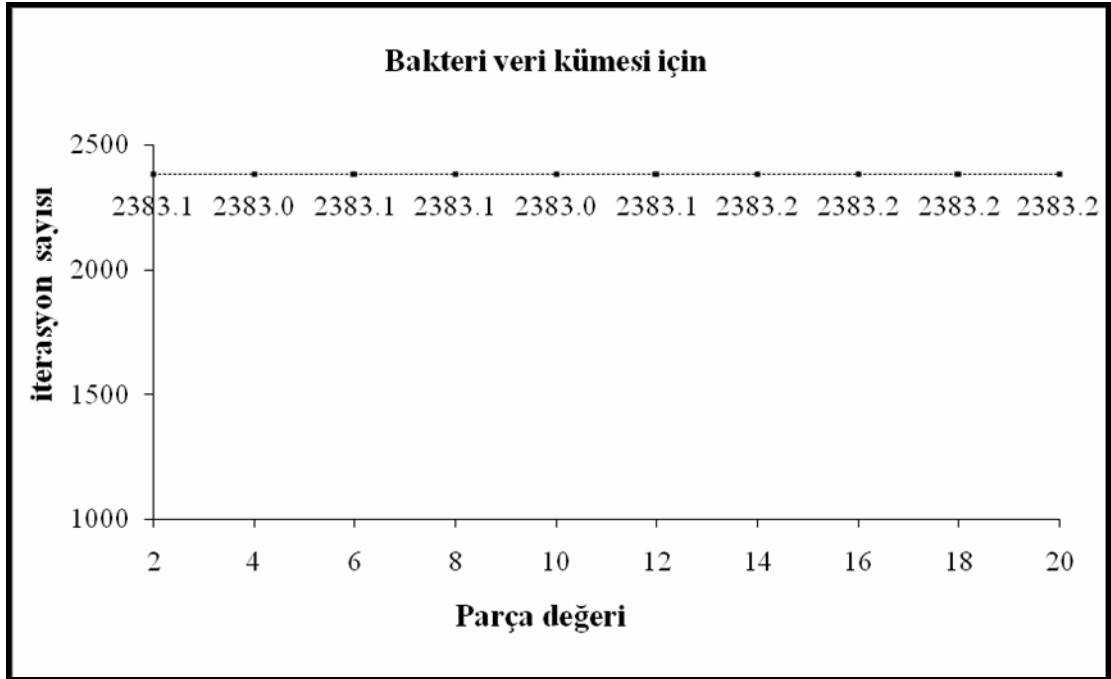
Tablo 6.41 Bakteri veri kümesinde parça değeri parametresi için BK-YBTS sonuçları

Parça Değeri	Süre (sn)	İterasyon Sayısı	Bellek Hücresi Sayısı	Sınıflandırma Doğruluğu (%)
2	880.1	2383.1	492.7	88.5
4	880.7	2383.0	492.4	91.9
6	881.3	2383.1	492.9	91.9
8	881.3	2383.1	492.3	91.5
10	881.5	2383.0	492.9	92.1
12	881.7	2383.1	492.4	91.2
14	881.9	2383.2	492.1	91.9
16	882.0	2383.2	492.1	91.2
18	882.3	2383.2	492.8	90.7
20	882.6	2383.2	491.8	90.5

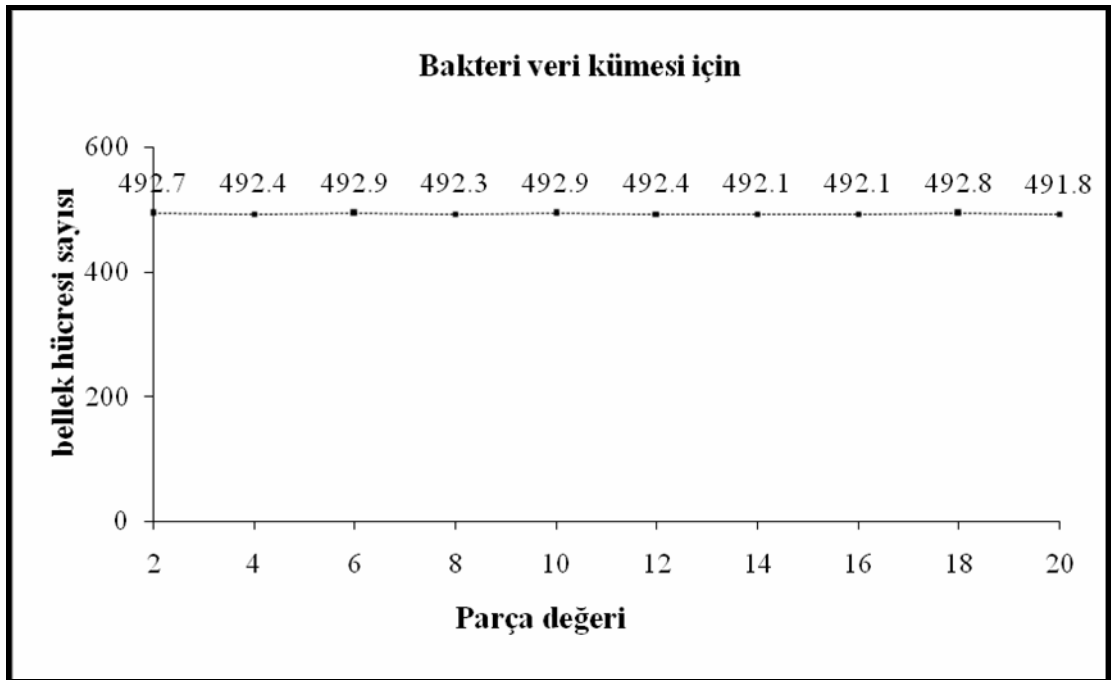
Parça değeri parametresinin aldığı değişik değerler için *BK-YBTS'nin* çalışma süreleri açısından grafiksel olarak gösterimi Şekil 6.97'de, iterasyon sayıları açısından grafiksel olarak gösterimi Şekil 6.98'de, bellek hücreleri sayıları açısından grafiksel olarak gösterimi Şekil 6.99'da ve sınıflandırma doğrulukları açısından grafiksel olarak gösterimi Şekil 6.100'de verilmiştir.



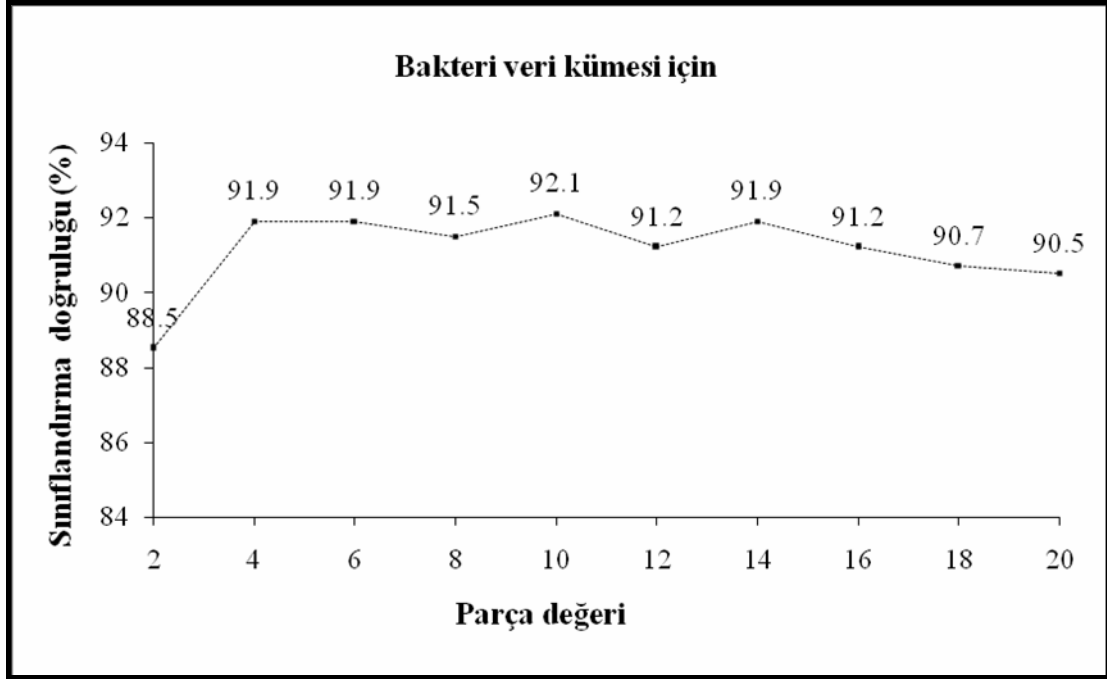
Şekil 6.97 Bakteri veri kümesinde parça değeri parametresinin değişimine göre BK-YBTS'nin çalışma sürelerinin grafiksel gösterimi



Şekil 6.98 Bakteri veri kümesinde parça değeri parametresinin değişimine göre BK-YBTS'nin iterasyon sayılarının grafiksel gösterimi



Şekil 6.99 Bakteri veri kümesinde parça değeri parametresinin değişimine göre BK-YBTS'nin bellek hücreleri sayılarının grafiksel gösterimi



Şekil 6.100 Bakteri veri kümesinde parça değeri parametresinin değişimine göre BK-YBTS'nin sınıflandırma doğruluklarının grafiksel gösterimi

Tablo 6.41'den de anlaşılabileceği gibi *parça değeri* parametresi 10 olduğu durumda, BK-YBTS en yüksek sınıflandırma doğruluğu olan %92.1'e ulaşmıştır. *Parça değeri* parametresinin değişimi BK-YBTS'nin çalışma süresini, iterasyon sayısını ve bellek hücreleri sayısını çok fazla etkilememiştir.

Bakteri veri kümesinde tüm parametreler için genel olarak BK-YBTS, YBTS'den daha yüksek sınıflandırma doğruluğuna ulaşmıştır. Bu durum BK-YBTS ile beraber önerilen değişikliklerin, YBTS'ye göre sınıflandırma performansını arttırdığını göstermektedir. Geliştirilen sistem deneme-yanılma tekniği kullanılarak Tablo 6.42'de belirtilen parametrelerle çalıştırılmıştır. Tablo 6.43'te Bakteri veri kümesi için YBTS, BK-YBTS ve literatürdeki diğer sınıflandırıcıların sınıflandırma doğrulukları belirtilmiştir. Literatürdeki diğer sınıflandırıcılar 5 kat çapraz onaylama yöntemi kullandığı için geliştirilen sistemde de bu eğitim test ayırımı kullanılmıştır.

Tablo 6.42'de belirtilen parametre değerleri ile BK-YBTS, bakteri veri kümesi için %92.4 sınıflandırma doğruluğuna ulaşmıştır. Bu değer YBTS tarafından elde edilen sınıflandırma doğruluğundan daha yüksek bir değerdir. Ayrıca bu veri kümesini sınıflandıran diğer sınıflandırıcılarla karşılaştırılınca en yüksek sınıflandırma doğruluklardan birisine ulaşılmıştır.

Tablo 6.42 Bakteri veri kümesi için BK-YBTS parametre değerleri

Parametre Adı	Değer
<i>Çekirdek hücrelerin sayısı</i>	1
<i>Kaynak sayısı</i>	250
<i>Uyarma eşiği</i>	0.8
<i>Mutasyon oranı</i>	0.1
<i>Duyarlılık eşik ölçüsü (ATS)</i>	0.05
<i>Çoğunluk oylaması için kullanılan k değeri</i>	1
<i>Parça değeri</i>	10

Tablo 6.43 Bakteri veri kümesi için BK-YBTS ve diğer sınıflandırıcılar tarafından elde edilen sınıflandırma doğrulukları

Metot	Doğruluk (%)
<i>Multilayer Perceptron</i>	98.2
<i>J48</i>	93.8
BK-YBTS	92.4
<i>YBTS</i>	89.7
<i>RBF Network</i>	82.6
<i>BayesNet</i>	73.0
<i>Naïve Bayes</i>	68.7
<i>OneR</i>	60.0

6.3. Bütün Veri Kümeleri İçin Elde Edilen Sınıflandırma Sonuçları

Standart makine öğrenmesi veri kümelerinden olan *Iris* ve *Thyroid* veri kümeleri üzerinde geliştirilen sistemin sınıflandırma doğruluğu *YBTS'den* daha yüksek elde edilmiştir. *BK-YBTS'nin* sınıflandırma doğrulukları diğer sınıflandırıcıların doğrulukları ile karşılaştırılınca bir sınıflandırıcı olarak makul sonuçlar elde edilmiştir. Geliştirilen sistemin performansı ayrıca Erciyes Üniversitesi Tıp Fakültesinden alınan *Damar tıkanıklığı (Atherosclerosis)* veri kümesi ve Erciyes Üniversitesi Eczacılık Fakültesinden alınan *Bakteri* veri kümesi üzerinde değerlendirilmiştir. *Bakteri* veri kümesinde de BK-YBTS, YBTS'den daha yüksek sınıflandırma doğruluğuna ulaşmıştır. Sadece *Damar tıkanıklığı* veri kümesinde BK-YBTS, YBTS'den çok az miktarda daha az sınıflandırma doğruluğuna ulaşmıştır. *BK-YBTS'nin* dört adet veri kümesi üzerinde en yüksek sınıflandırma doğrulukları ve bu

sonuçları elde ederken kullanılan parametre değerleri topluca Tablo 6.44'te sunulmuştur.

Tablo 6.44 BK-YBTS'nin bütün veri kümeleri için sınıflandırma doğrulukları ve parametre değerleri

Veri Kümesi	Çekirdek	Kaynak	Uyarma Eşiği	Mutasyon	ATS	k değeri	Parça değeri	Doğruluk (%)
<i>Iris</i>	1	200	0.9	0.2	0.3	3	6	97,3
<i>Thyroid</i>	1	120	0.9	0.1	0.2	4	4 ve 6	95.9
<i>Damar tıkanıklığı</i>	1	200	0.9	0.1	0.2	1	2	99.1
<i>Bakteri</i>	1	250	0.8	0.1	0.05	1	10	92.4

Tablo 6.45'te *YBTS* ve *BK-YBTS'nin* çalışmada kullanılan dört adet veri kümesi için elde ettikleri sınıflandırma doğrulukları verilmiştir. Elde edilen sonuçlar, diğer *YBS* algoritmalarında da geliştirilen sistemle beraber önerilen yaklaşımların kullanılmasının faydalı olacağı konusunda araştırmacılara cesaret vermektedir.

Tablo 6.45 YBTS ve BK-YBTS'nin bütün veri kümeleri için sınıflandırma doğrulukları

Veri kümesi	YBTS sınıflandırma doğruluğu (%)	BK-YBTS sınıflandırma doğruluğu (%)
<i>Iris</i>	96.7	97.3
<i>Thyroid</i>	94.8	95.9
<i>Damar tıkanıklığı</i>	99.3	99.1
<i>Bakteri</i>	89.7	92.4

7. SONUÇLAR VE ÖNERİLER

Bu tez çalışmasında doğal ve yapay bağışıklık sisteminde sergilenen mekanizmalardan esinlenilen yapay bağışıklık sınıflandırıcısı olan *Yapay Bağışıklık Tanıma Sistemi (YBTS)* temel alınarak danışmanlı bir yapay bağışıklık sınıflandırma sistemi olan *Bilgi Kazancı Tabanlı Yapay Bağışıklık Tanıma Sistemi (BK-YBTS)* geliştirilmiştir. *BK-YBTS*'de kaynak rekabeti, mutasyon, klonsal seçim, duyarlılık olgunlaşması ve bellek hücreleri oluşumu gibi bağışıklık sistemi mekanizmaları kullanılmıştır. *BK-YBTS*'nin sınıflandırma doğruluğu makine öğrenmesi için standart veri kümeleri ve gerçek dünya veri kümeleri üzerinde değerlendirilmiştir. Geliştirilen sistemle beraber *YBS* ve diğer Yapay Zekâ teknikleri için yararlı olabilecek çeşitli yaklaşımlar kullanılmıştır. Bu yaklaşımlar aşağıda sıralanmıştır.

- 1- Şekil-uzay gösterimini kullanan *YBS* algoritmaları için uzaklık kriteri kullanımı değiştirilmiştir. Bu algoritmalarda verinin bütün özellikleri uzaklık hesabında aynı etkiye sahiptir. Veri kümesi üzerinde etkisi çok az olan özelliklerle veri kümesi üzerinde çok önemli özellikleri eşit ağırlıkta sayılmaktadır. Özellikle verilerin özellik sayıları büyük olduğunda bu durum sistemin sınıflandırma performansını olumsuz yönde etkilemektedir. Literatürde “*curse of dimensionality (boyutluluk problemi)*” olarak adlandırılan bu durumun üstesinden gelebilmek için veri kümesinde daha önemli olan özelliklere uzaklık hesabında daha yüksek ağırlık verilmiştir. Bu amaçla geliştirilmiş olan *BK-YBTS* bu problemin etkisini minimuma indirmek için veri özelliklerinin bilgi kazancını temel almıştır. Özelliklerin önem derecesi bilgi kazancı hesabıyla belirlenmiştir. Önerilen bilgi kazancı tabanlı özellik ağırlıklandırma yaklaşımında eğer bir özelliğin bilgi kazancı diğer özelliklerin bilgi kazancından yüksekse bu özelliğe diğer özelliklerden daha çok ağırlık verilmiştir. Elde edilen sonuçlarla, bilgi kazancı tabanlı uzaklık kriteri kullanımının daha verimli olduğu tespit edilmiştir.

- 2- Bir diğer yaklaşım olarak, performans artımına katkıda bulunmak amacıyla sistemdeki antikora mutasyon işlemi sırasında antikor veya bellek hücresinin kalitesini arttırmak için rasgele bir değer atamak yerine mutasyona uğrayacak antikoru veya bellek hücresinin duyarlılığını temel alan bir mutasyon mekanizması kullanılmıştır. Yani duyarlılığı yüksek olan antikolar düşük oranda, duyarlılığı düşük olan antikolar yüksek oranda mutasyona uğramıştır. Bu yaklaşımın, yüksek duyarlılıklı hücrelerin hemen çevresindeki arama alanının araştırılmasına izin verdiği tespit edilmiştir. Diğer taraftan daha düşük duyarlılıklı hücreler için daha yüksek mutasyon oranı, daha az verimli alanlardan uzaklaşmayı kolaylaştırmak için arama uzayında daha büyük atlamalara izin verdiği tespit edilmiştir.
- 3- *YBTS'de* kaynaklar antijenle aynı sınıftaki antikolar için en fazla kullanıcı tarafından belirlenmiş olan toplam maksimum kaynağın yarısı olacak şekilde tahsis edilmektedir ve kalan yarısı diğer sınıflara ait antikolar arasında eşit olarak paylaşılmaktadır. *BK-YBTS'de* bu mekanizma veri kümesindeki sınıfların oranı hesaba katılarak değiştirilmiştir. Bu metotla birlikte sistem kaynakları, eğitim kümesinde sınıf dağılımı ile orantılı olarak farklı sınıf antikoları arasında bölünmüştür. Böylelikle geliştirilen sistemin *YBTS'den* daha kısa sürede çalışmasını tamamladığı tespit edilmiştir.
- 4- *BK-YBTS'de* bellek hücresi seçimine karar veren ölçüt eğitim hücrelerinin tamamı arasındaki ortalama duyarlılık değeri yerine aynı sınıfa sahip eğitim hücrelerinin ortalama duyarlılık değerini temel alacak şekilde değiştirilmiştir. Eğitim verilerinin tamamını kullanan geniş bir metottan ziyade aynı sınıfa sahip eğitim örneklerinin ortalama duyarlılığı temel alınarak duyarlılık eşiğine karar verilmiştir. Böylelikle geliştirilen sistemin sınıflandırma performansının arttığı tespit edilmiştir.
- 5- Test verilerinin hangi sınıfta olduğu bellek hücreleri kümesi üzerine *k en yakın komşu* yönteminin uygulanması ile belirlenmektedir. Bu aşamada çoğunluk oylaması yapılırken eşitlik olması durumunda, test verisinin sınıfı olarak daha çok verisi bulunan sınıf atanmıştır. Özellikle sınıf sayılarının eşit

olmadığı veri kümelerinde, bu yaklaşımın sınıflandırma performansını arttırdığı tespit edilmiştir.

BK-YBTS doğal bağışıklık sistemlerinde ve yapay bağışıklık sistemlerinde önceki çalışmaların gözlemlerini temel alan bir danışmanlı öğrenme metodu olarak karşımıza çıkmaktadır. Burada bağışıklık sisteminin çeşitli anahtar bileşenleri kullanılmıştır. Bu bileşenlerden biri veri temsiliinde kullanılan *B* hücrelerinin soyutlanmasıdır. Antijen ve *B* hücresinin antikorları arasındaki reaksiyonlar *bilgi kazancı tabanlı uzaklık* kullanılarak uyarılmıştır. Burada sunulmuş olan çalışmanın temeli bellek hücrelerinin geliştirilmesidir. Bellek hücreleri sınıflandırma için başlıca mekanizmayı sağlamaktadır. Bellek hücrelerinin gelişimi, *YBS* çalışmalarının çoğunda incelenmiş olan klonsal seçim ve duyarlılık olgunlaşması mekanizmaları ile sağlanmıştır.

BK-YBTS'nin performansı standart makine öğrenmesi veri kümelerinden olan *Iris*, ve *Thyroid* veri kümeleri ile gerçek dünya veri kümelerinden olan *Damar tıkanıklığı (Atherosclerosis)* veri kümesi ve *Bakteri* veri kümesi üzerinde değerlendirilmiştir. *YBTS'den* farklı olarak *BK-YBTS* ile beraber önerilen değişikliklerin etkisini araştırmak amacıyla her iki sistemin ortak parametreleri olan *çekirdek hücre sayısı*, *kaynak sayısı*, *uyarma eşiği*, *mutasyon oranı*, *duyarlılık eşiği ölçüsü* ve *k* parametrelerinin değişik değerleri için *YBTS* ve *BK-YBTS'nin* çalışma zamanı, iterasyon sayısı, bellek hücresi sayısı ve sınıflandırma doğrulukları incelenmiştir. Ayrıca *BK-YBTS'ye* özgü *parça değeri* parametresinin değişik değerleri için *BK-YBTS'nin* çalışma zamanı, iterasyon sayısı, bellek hücresi sayısı ve sınıflandırma doğruluğu incelenmiştir. Bu parametrelerin etkisi aşağıda belirtilmiştir.

- *Çekirdek hücre sayısı* arttığı zaman bellek hücreleri ve antikorlar eğitim antijenine benzer yapıda oluşmuştur. Aynı zamanda *çekirdek hücre sayısı* arttığı zaman; sistemdeki bellek hücresi sayısında artış, çalışma süresi ve iterasyon sayısında ise düşüş tespit edilmiştir.
- *Kaynak sayısı* parametresinin, sistemde bulunacak antikor sayısını direkt olarak etkilediği tespit edilmiştir. *Kaynak sayısı* arttıkça sistemde bulunan bellek

hücreleri ve antikorların çeşitliliği artarken, aynı zamanda her iki sistemde de çalışma süresi ve iterasyon sayısının arttığı tespit edilmiştir.

- *Uyarma eşiği* parametresi, belirli bir eğitim antijeninin eğitimi durdurma kriteri olarak kullanılmıştır. *Uyarma eşiği* parametresi, antikor popülasyonunun ne kadar uyarılacağını belirlemiştir. *Uyarma eşiği* değeri arttıkça antikor popülasyonunun uyarılma seviyesinin arttığı tespit edilmiştir.

- *Mutasyon oranı* parametresinin, antikor popülasyonunun çeşitliliğini etkilediği tespit edilmiştir. *Mutasyon oranı* büyük olduğu durumlarda sistemin sınıflandırma doğruluğu ve çalışma zamanı negatif yönde etkilenmiştir. Özellikle *YBTS'de* mutasyon işleminin rasgele yapılması ile hedef arama uzayından uzaklaşma olduğu için sınıflandırma doğruluğu negatif yönde etkilenmiştir. *BK-YBTS'de* duyarlılık tabanlı mutasyon mekanizması kullanıldığı için *YBTS'deki* kadar sınıflandırma doğruluğuna negatif yönde etki gözlemlenmemiştir. Mutasyon oranı arttıkça eğitimi durdurma ölçütü olan uyarma eşiği değerinin, antikorlar arasındaki ortalama uyarma değerinden küçük olma ihtimali azaldığı için çalışma süresinde artış tespit edilmiştir.

- *Duyarlılık eşik ölçüsünün*, bellek hücresi seçimini direkt olarak etkilediği tespit edilmiştir. *YBTS'de ortalama duyarlılık eşiği*, eğitim antijenleri arasındaki ortalama duyarlılığı; *BK-YBTS'de sınıf tabanlı duyarlılık eşiği*, aynı sınıfa sahip eğitim antijenlerinin ortalama duyarlılığını ifade etmektedir. *Duyarlılık eşik ölçüsü* arttıkça sistemde kalan bellek hücresi sayısında azalma tespit edilmiştir. Ayrıca *duyarlılık eşik ölçüsü*, küçük değerler aldığı zaman sınıflandırma doğruluğunda büyük değişiklikler olmadığı; *duyarlılık eşik ölçüsü*, büyük değerler aldığı zaman ise sınıflandırma doğruluğunda azalmalar tespit edilmiştir.

- Eğitim aşaması tamamlandıktan sonra geliştirilen bellek hücreleri sınıflandırmada kullanılmak için uygun duruma gelmektedir. Sistemin son aşamasında *k en yakın komşu* yöntemi uygulanmıştır. Test verilerinin hangi sınıfta olduğu bellek hücreleri kümesi üzerine *k en yakın komşu* yönteminin uygulanması ile belirlenmiştir. Bir verinin sistem tarafından sınıflandırılmasına test verisini en fazla uyaran *k* adet bellek hücrelerinin çıkışlarının bir *çoğunluk oylaması* kullanılarak karar verilmiştir. *YBTS'de* çoğunluk oylaması yapılırken eşitlik olması durumunda test verisinin sınıfı olarak, en küçük indisli sınıf değeri atanmaktadır. *BK-YBTS'de*

çoğunluk oylaması yapılırken eşitlik olması durumunda test verisinin sınıfı olarak, tüm veri kümesinde daha çok verisi bulunan sınıf değeri atanmaktadır. k değeri parametresinin değişik değerleri için çoğunlukla *BK-YBTS*, *YBTS*'den daha yüksek sınıflandırma doğruluğuna ulaştığı tespit edilmiştir.

- *BK-YBTS*'ye özgü bir parametre olan *parça değeri*, veri özelliklerinin bilgi kazancını hesaplamak için sürekli değerli özelliklerin ayrık hale getirilmesinde kullanılmıştır. Hesaplanan bilgi kazancı değerleri, uzaklık hesabında katsayı olarak kullanılmıştır. *Parça değeri* küçük olduğu durumlarda, geliştirilen sistemin daha yüksek sınıflandırma doğruluklarına ulaştığı tespit edilmiştir.

Denemeler sonucunda *Iris* veri kümesinde *BK-YBTS*, %97.3 sınıflandırma doğruluğuna ulaşırken *YBTS*, %96.7 sınıflandırma doğruluğuna ulaşmıştır. *Thyroid* veri kümesinde *BK-YBTS*, %95.9 sınıflandırma doğruluğuna ulaşırken *YBTS*, %94.8 sınıflandırma doğruluğuna ulaşmıştır. *Damar tıkanıklığı* veri kümesinde *BK-YBTS*, %99.1 sınıflandırma doğruluğuna ulaşırken *YBTS*, %99.3 sınıflandırma doğruluğuna ulaşmıştır. Son olarak *Bakteri* veri kümesinde *BK-YBTS* %92.4 sınıflandırma doğruluklarına ulaşırken *YBTS*, %89.7 sınıflandırma doğruluğuna ulaşmıştır. Elde edilen sınıflandırma doğrulukları, literatürdeki sınıflandırıcıların sınıflandırma doğrulukları ile karşılaştırılınca kabul edilebilir sınıflandırma doğruluklarına ulaşıldığı tespit edilmiştir. *BK-YBTS* ve *YBTS*'nin karmaşıklığı, $O(n^2)$ olarak hesaplanmıştır.

Bu sonuçlarla beraber, bağışıklık sisteminde gerçekleşen olayları modelleyerek geliştirilen bu sistemle önerilen yaklaşımların diğer *YBS* algoritmalarına da adapte edileceği gösterilmiştir. Aynı zamanda hazırlanan bu çalışma, ülkemizde bu alanda yapılan ilk akademik çalışmalardan biri olarak, gelecekte bu konuda çalışacak olan araştırmacılar için kaynak bir çalışma olmuştur.

Gerçekleştirilen uygulamalar sonucunda, problemin en iyi çözümünü sağlayan parametrelerinin belirlenmesinde güçlüklerle karşılaşmıştır. Deneme yanılma yöntemiyle en uygun parametre değerleri belirlenmiştir. Ayrıca karşılaşılan bir başka güçlük ise biyolojik olarak bağışıklık sisteminde gerçekleşen bazı mekanizmaların tam olarak anlaşılamaması olmuştur.

Gelecekte bu alanda çalışacak araştırmacılar için öneriler şu şekilde sıralanabilir. Geliştirilen sistemin uygulanacağı problem konusunda uzman kişiler

aracılığıyla, veri kümesi daha verimli bir şekilde ayrık hale dönüştürülebilir ve dolayısıyla daha anlamlı bilgi kazancı değerleri elde edilebilir. Diğer yapay zekâ teknikleri kullanılarak mevcut sistemden daha verimli olacak karma sistemler geliştirilebilir. Ayrıca şu an için biyolojik olarak tam olarak anlaşılamayan bağışıklık sisteminde gerçekleşen bazı mekanizmaların, gelecekte anlaşılmasıyla beraber daha verimli YBS algoritmalarının geliştirilmesi mümkün olabilecektir.

8. KAYNAKLAR

- Ada G.L., Nossal G., 1987, "The Clonal Selection Theory", Scientific American, Vol. 257(2): 50-57.
- Aho A., Hopcroft J.E., Ullman J. D., 1975, "The Design and Analysis of Algorithms", Addison-Wesley, Reading, Massachusetts, USA.
- Albayrak S., Amasyalı F., 2003, "Fuzzy C-Means Clustering on Medical Diagnostic Systems", International XII. Turkish Symposium on Artificial Intelligence and Neural Networks-TAINN 2003.
- Allen P.M., 1987, "Antigen Processing at the Molecular Level", Immunol. Today, Vol. 8: 270-273.
- Allen D., Cumano A., Dildrop R., Kocks C., Rajewsky K., Rajewsky N., Roes J., Sablitzky F., Siekevitz M., 1987, "Timing, Genetic Requirements and Functional Consequences of Somatic Hypermutation During B-cell Development", Immunological Reviews, Vol. 96: 5-22.
- Balthrop J., Forrest S., Glickman M.R., 2002, "Revisting LISYS: Parameters and Normal Behavior", In The Proceedings of The Special Sessions on Artificial Immune Systems in The 2002 Congress on Evolutionary Computation, IEEE World Congress on Computational Intelligence, Honolulu, Hawaii.
- Bellman R., 1961, "Adaptive Control Processes: A Guided Tour", Princeton University Press, Princeton, New Jersey, USA.
- Bersini H., Varela F., 1990, "Hints for Adaptive Problem Solving Gleaned from Immune Network", In Parallel Problem Solving from Nature, H.P. Schwefel and H. M'hlenbein (Eds.), Publisher -Springer-Verlag, 343 - 354.
- Beyer K., Goldstein J., Ramakrishnan R., Shaft U., 1999, "When is "Nearest Neighbor" Meaningful?", Lectures Notes in Computer Science LNCS, 1540: 217-235.
- Blahut R.E., 1987, "Principles and Practice of Information Theory", Addison-Wesley, Reading, Massachusetts, USA.
- Burnet M., 1959, "The Clonal Selection Theory of Acquired Immunity", University Press, Cambridge.
- Carter J.H., 2000, "The Immune System As a Model for Pattern Recognition and Classification", Journal of the American Medical Informatics Association, Vol. 7(1): 28-41.
- Cendrowska J., 1990, "Knowledge Acquisition for Expert Systems: Inducing Modular Rules from Examples", PhD Thesis, The Open University.
- Cheong T.S., Yoon C.H., 1999, "A Memory Based Classifier Using The Recursive Partition Averaging", IEEE Tencon, 1038-1041.

- Chun J.S., Cho D.H., Kim M.K., 1997, "A study on Comparison Between Immune Algorithm and The Other Algorithms. In: Motor Design", Int. Conf. on Intelligent System Application to Power Systems ISAP'97, Seoul, 588-592.
- Coomans D., Broeckeaert I., Jonckheer M., Massart D.L., 1983, "Comparison of Multivariate Discrimination Techniques for Clinical Data Application to The Thyroid Functional State", *Methods Information in Medicine*, Vol. 22: 93-101.
- Coutinho A., 1989, "Beyond Clonal Selection and Network", *Immun. Rev.*, Vol. 110: 63-87.
- Dart A.M., Lacombe F., Yeoh J.K., Cameron J.D., Jennings G.L., Laufer E., Esmore D.S., 1991, "Aortic Distensibility in Patients with Isolated Hyper-Cholesterolaemia, Coronary Artery Disease, or Cardiac Transplant", *Lancet*, 270 -273.
- Dasgupta D., Forrest S., 1995, "Tool Breakage Detection in Milling Operations using a Negative-Selection Algorithm", Technical Report CS95-5, Department of Computer Science, University of New Mexico.
- Dasgupta D., KrishnaKumar K., Wong D., Berry M., 2004, "Negative Selection Algorithm for Aircraft Fault Detection", Published in The Proceedings of The 3rd International Conference on Artificial Immune Systems, Italy, September 13-16.
- D'haeseleer P., Forrest S., Helman P., 1996, "An Immunological Approach to Change Detection: Algorithms, Analysis, and Implications", Published in The Proceedings of The 1996 IEEE Symposium on Computer Security and Privacy, IEEE Computer Society Press, Los Alamitos, CA, 110-119.
- 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.*, Vol. 155: 295-333.
- De Castro L.N., Von Zuben F.J., 1999, "Artificial Immune Systems: Part I- Basic Theory and Applications", Technical Report, School of Computing and Electrical Engineering, State University of Campinas, Brazil, No. TR DCA 01/99, December 1999.
- De Castro L.N., Von Zuben F.J., 2000a, "The Clonal Selection Algorithm with Engineering Applications", In The Proceedings of Genetic and Evolutionary Computation Conference (GECCO), Workshop on Artificial Immune Systems and Their Applications, Las Vegas, USA, 36-37.
- De Castro L.N., Von Zuben F.J., 2000b, "Artificial Immune Systems: Part II- A Survey of Applications", Technical Report, Department of Computer Engineering and Industrial Automations, State University of Campinas, Brazil, No. DCA-RT 02/00, February 2000.
- De Castro L.N., Von Zuben F.J., 2001, "Ainet: An Artificial Immune Network for Data Analysis". (Full version, pre-print). Book Chapter in *Data Mining: A Heuristic Approach*, Hussein A. Abbass, Ruhul A. Sarker, and Charles S. Newton (Eds.), Idea Group Publishing, USA.
- De Castro L.N., Timmis J., 2002a, "Artificial Immune Systems: A New Computational Intelligence Approach", Springer-Verlag.

- De Castro L.N., Von Zuben F.J., 2002b, "Learning and Optimization Using The Clonal Selection Principle", In The Special Issue on Artificial Immune Systems of The journal IEEE Transactions on Evolutionary Computation, Vol. 6 (3): 239-251.
- Dirgenali F., 2003, "Karotid atar damar Doppler sinyallerinden damar sertliğinin erken dönemde teşhisi", Erciyes Üniversitesi, Fen Bilimleri Enstitüsü, Yüksek Lisans Tezi.
- Dutta R., Hines E.L., Gardner J.W., Boilot P., 2002, "Bacteria Classification Using Cyranose 320 Electronic Nose", BioMedical Engineering Online, 1-4.
- Evans D., 2000, "Doppler Signal Analysis", Ultrasound in Med. & Biol., Vol. 26(1): 13-15.
- Evans D.H., McDicken W.N., Skidmore R., Woodcock J.P., 1989, "Doppler Ultrasound: Physics", Instrumentation and Clinical Applications, Chichester: Wiley, New York, USA.
- Farmer J.D., Packard N.H., Perelson A.S., 1986, "The Immune System, Adaptation, and Machine Learning", Physica 22D, 187-204.
- Fayyad U.M., Irani K.B., 1993, "Multi-interval Discretization of Continuous Valued Attributes for Classification Learning", In Proceedings of The 13th International Joint Conference on Artificial Intelligence, Los Altos, CA, Morgan Kaufmann, 1022:1027.
- Forrest S., Perelson A.S., Allen L., Cherukuri R., 1994, "Self-Nonself 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.
- Forrest S., Hofmeyr S.A., 1999, "John Holland's Invisible Hand: An Artificial Immune System", presented at the Festschrift held in honor of John Holland, University of Michigan, USA.
- Greensmith J., 2003, "New Frontiers For An Artificial Immune System", MSc Thesis, Multidisciplinary Informatics, University of Leeds, UK.
- Haykin S., 1999, "Neural Networks: A Comprehensive Foundation", Prentice Hall, 2nd Ed., Upper Saddle River, NJ, USA.
- Hamaker J., Boggess L., 2004, "Non-Euclidean Distance Measures in AIRS, an Artificial Immune Classification System", Proceedings of the 2004 Congress on Evolutionary Computation, 1067-1073.
- Hassan R., Nath B., Kirley M., 2006, "A Data Clustering Algorithm Based On Single Hidden Markov Model", Proceedings of the International Multiconference on Computer Science and Information Technology, 57-66.
- Hirai T., Sasayama S., Kawasaki T., Yagi S., 1989, "Stiffness of Systemic Arteries in Patients with Myocardial Infarction, A Noninvasive Method to Predict Severity of Coronary Atherosclerosis", Circulation, 78 -86.
- Hoffmann, G.W., 1986, "A Neural Network Model Based on The Analogy with the Immune System", J. Theor. Biol., Vol. 122: 33-67.

- Hofmeyr S.A., Forrest S., 1999, "Immunity by Design: An Artificial Immune System", In The Proceedings of The Genetic and Evolutionary Computation Conference (GECCO), Orlando, Florida, USA, 1289-1296.
- Hofmeyr S., 2000, "An Interpretative Introduction to The Immune System", In Design Principles for The Immune System and Other Distributed Autonomous Systems. I. Cohen and L. Segel(eds.). Oxford University Press.
- Hofmeyr S., Forrest S., 2000, "Architecture for An Artificial Immune System", Evolutionary Computation Journal, Vol. 8(4): 443-473.
- Hodgkin P.D., 1998, "Role of Cross-Reactivity in The Development of Antibody Responses", The Immunologist, 6/6: 223-226.
- Holte R.C., 1993, "Very Simple Classification Rules Perform Well on Most Commonly Used Datasets", Machine Learning, Vol. 11(1): 63-90.
- Hunt J.E., Cooke D.E., 1996, "Learning Using An Artificial Immune System", Journal of Network and Computer Applications, Vol. 19: 189-212.
- Hunt J., Timmis J., Cooke D., Neal M., King C., 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.
- <http://archive.ics.uci.edu/ml/> son erişim 20 Aralık 2007.
- <http://www.endocrineweb.com/thyroid.html>, son erişim 7 Kasım 2007.
- <http://www.lgcpromochem-atcc.com>, son erişim 7 Kasım 2007.
- Igawa K., Ohashi H., 2007, "Discrimination-Based Artificial Immune System: Modeling the Learning Mechanism of Self and Non-Self Discrimination for Classification", Journal of Computer Science 3 (4): 204-211.
- Janeway C.A., 1992, "The Immune System Evolved to Discriminate Infectious Nonself from Noninfectious Self", Imm. Today, 13(1): 11-16.
- Janeway C.A., Travers P., Walport M., Capra J.D., 1999, "Immunobiology: The Immune System in Health and Disease", 4th Edition, Garland Publishing.
- Jerne N.K., 1973, "The Immune System", Scientific American, 229(1): 52-60.
- Jerne N.K., 1974, "Towards A Network Theory of The Immune System", Annals of Immunology. (Inst. Pasteur) Vol. 125C: 373-389.
- Ji Z., Dasgupta D., 2006, "Applicability Issues of The Real-Valued Negative Selection Algorithms", In The Proceedings of International Conference on Genetic and Evolutionary Computation (GECCO), Seattle, Washington, USA, 111-118.
- Jiao L.C., Wang L., 2000, "A Novel Genetic Algorithm Based on Immunity", IEEE Trans. Systems, Man and Cybernetics. 30(5): 552-561.
- Kamber H.J., 2000, "Data Mining Concepts and Techniques", Morgan Kaufmann Publishers, 1st Ed., San Francisco, USA.

- Kara S., 1995, "A Study of Mitral and Tricuspid Valve Blood Flows by Autoregressive Spectral Analysis Method and Doppler Unit", PhD Thesis, Institute of Science of Erciyes University.
- Kim J., Bentley P., 1999, "Negative Selection and Niching by An Artificial Immune System for Network Intrusion Detection", In The Proceedings of International Conference on Genetic and Evolutionary Computation Conference (GECCO), Morgan-Kaufmann, Orlando, Florida, USA, 149-158.
- Kim H.C., Ghahramani Z., 2006, "Bayesian Gaussian Process Classification with The EM-EP Algorithm", IEEE Transactions On Pattern Analysis and Machine Intelligence, Vol. 28 (12): 1948-1959.
- Knight T., Timmis J., 2001, "Assessing The Performance of The Resource Limited Artificial Immune System AINE", Computing Laboratory, University of Kent at Canterbury, Technical Report 3-01.
- Knuth D.E., 1968, "The Art of Computer Programming", Vol. 1: Fundemantal Algoritms, Addison-Wesley, Reading, Massachusetts, USA.
- Kodaz H., Kara S., Latifoğlu F., Güneş S., 2007, "A New Hybrid Classifier System: Information Gain-Based Artificial Immune Recognition System", Experimental Techniques, Vol. 31 (6): 36-43.
- Latifoğlu F., Kodaz H., Kara S., Güneş S., 2006, "Medical Application of Artificial Immune Recognition System (AIRS): Diagnosis of Atherosclerosis from Carotid Artery Doppler Signals", Computers in Biology and Medicine, Vol. 37: 1092-1099.
- Marwah G., Boggess L., 2002, "Artificial Immune Systems for Classification: Some Issues", 1st International Conference on Artificial Immune Systems (ICARIS-2002), University of Kent at Canterbury, UK, September 9-11, 2002.
- Mason D., 1998, "Antigen Cross-Reactivity: Essential in the Function of TCRs", The Immunologist, Vol. 6/6: 220-222.
- Matzinger P., 1994, "Tolerance, Danger, and The Extended Family", Annual Review of Immunology, Vol. 12: 991-1045.
- Meng L.J., Putten P., Wang H., 2005, "A Comprehensive Benchmark of The Artificial Immune Recognition System (AIRS) ", In proceedings Advanced Data Mining and Applications, First International Conference, ADMA 2005, Wuhan, China, July 22-24, 575-582.
- Michalski R.S., Kodratoff Y., 1990, "Research in Machine Learning: Recent Progress, Classification of Methods, and Future Directions", Machine Learning Vol.3, Morgan Kaufmann, San Mateo, CA, 3-30.
- Mitchell T.M., 1997, "Machine Learning", The McGraw-Hill Comp. Press, New York, USA.
- Mori K., Tsukiyama M., Fukuda T., 1993, "Immune Algorithm with Searching Diversity and its Application to Resource Allocation Problem", The Trans. of The Institute of Electrical Engineers of Japan, Vol.113-C (10): 872-878.

- Muller M., Ciccotti P., Reiche W., Hagen T., 2001, "Comparison of Color Flow Doppler Scanning, Power Doppler Scanning, and Frequency Shift for Assessment of Carotid Artery Stenosis", *J. Vasc. Surg.*, Vol. 34(6): 1090–1095.
- Myles A.J., Brown S.D., 2004, "Decision Pathway Modeling", *Journal of Chemometrics*, Vol. 18: 286-293.
- Nasraoui O., Gonzalez F., Dasgupta D., 2002, "The Fuzzy Artificial Immune System: Motivations, Basic Concepts, and Application to Clustering and Web Profiling", Published at IEEE International Conference on Fuzzy Systems. Published in The Proceedings of The IEEE World Congress on Computational Intelligence, Hawaii, May 12-17.
- Nossal G.J.V., 1993, "The Molecular and Cellular Basis of Affinity Maturation in The Antibody Response", *Cell*, Vol. 68: 1-2.
- Oprea M., Forrest S., 1999, "How The Immune System Generates Diversity: Pathogen Space Coverage with Random and Evolved Antibody Libraries", In The Proceedings of The Genetic and Evolutionary Computation Conference (GECCO), Orlando, Florida, USA, 1651-1656.
- Özyılmaz L., Yıldırım T., 2002, "Diagnosis of Thyroid Disease Using Artificial Neural Network Methods", Proceedings of The 9th International Conference on Neural Information Processing (ICONIP'02), Orchid Country Club, Singapore, Vol. 4: 2033-2036.
- Pazzani M., 1995, "An Iterative Improvement Approach for The Discretization of Numeric Attributes in Bayesian Classifiers", Proceedings of The First International Conference on Knowledge Discovery and Data Mining (KDD-95), Montreal, Canada, 228-233.
- Pechenizkiy M., Tsymbal A., Puuronen S., Patterson D.W., 2007, "Feature Extraction for Dynamic Integration of Classifiers", *Fundamenta Informaticae*, Vol. 77(3): 243-275.
- Perelson A., Hightower R., Forrest S., 1996, "Evolution (And Learning) of V-region Genes", *Research in Immunology*, Vol. 147: 202-208.
- Perelson A.S., Mirmirani M., Oster G.F., 1978, "Optimal Strategies in Immunology II. B Memory Cell Production", *J. Math. Biol.*, Vol. 5: 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.*, Vol. 81: 645-670.
- Perelson A.S., 1989, "Immune Network Theory", *Immunological Review*, Vol. 110: 5-36.
- Perelson A.S., Weisbuch G., 1997, "Immunology for Physicists", *Rev. of Modern Physics*, 69(4): 1219-1267.
- Polat K., Şahan S., Kodaz H., Güneş S., 2005a, "A New Classification Method for Breast Cancer Diagnosis: Feature Selection Artificial Immune Recognition System (FS-AIRS)", *Lecture Notes in Computer Science LNCS*, 3611: 830 – 838.

- Polat K., Şahan S., Kodaz H., Güneş S., 2005b, "Outdoor Image Classification Using Artificial Immune Recognition System (AIRS) with Performance Evaluation by Fuzzy Resource Allocation Mechanism", Lecture Notes in Computer Science LNCS, 3691: 81 - 87.
- Polat K., Şahan S., Kodaz H., Güneş S., 2007, "Breast Cancer and Liver Disorders Classification Using Artificial Immune Recognition System (AIRS) with Performance Evaluation by Fuzzy Resource Allocation Mechanism", Expert Systems with Applications, Vol. 32(1): 172-183.
- Quinlan J.R., 1986, "Induction of Decision Trees", Machine Learning, 1(1): 81-106.
- Rensberger B., 1996, "In Self-Defense", Life Itself, Oxford University Press, 212-228.
- Sathyanath S., Sahin F., 2001, "An AIS Approach to A Color Image Classification Problem in A Real Time Industrial Application", In The Proceedings of The IEEE International Conference on Systems, Man and Cybernetics, Arizona, USA, Vol. 4: 2285-2290.
- Secker A., Freitas A.A., Timmis J., 2003, "A Danger Theory Inspired Approach to Web Mining", Published in the proceedings of Second International Conference on Artificial Immune Systems (ICARIS), September 1-3, 2003, Napier University, Edinburgh, UK.
- 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, 321-343.
- Smith D.J., Forrest S., Hightower R.R., Perelson S.A., 1997, "Deriving Shape Space Parameters from Immunological Data", J. Theor. Biol., Vol. 189: 141-150.
- Sprent J., 1994, "T and B Memory Cells", Cell, Vol. 76: 315-322.
- Stefanadis C., Stratos C., Boudoulas H., Kourouklis C., Toutouzas P., 1990, "Distensibility of The Ascending Aorta, Comparison of Invasive and Non-invasive Techniques in Healthy Men and in Men with Coronary Artery Disease", Eur Heart J., 990 -996.
- Stibor T., Timmis J., Eckert C., 2005, "A Comparative Study of Real-Valued Negative Selection to Statistical Anomaly Detection Techniques", Lecture Notes in Computer Science LNCS, 3627: 262-275.
- Sun Y., 2007, "Cost-Sensitive Boosting for Classification of Imbalanced Data", PhD. Thesis, Waterloo University, Canada.
- Sutton R.S., Barto A.G., 1998, "Reinforcement Learning An Introduction", MIT Press, Cambridge, Massachusetts, USA.
- Şahan S., 2004, "Yapay Bağışıklıkta Yeni Algoritmalar", Yüksek Lisans Tezi, Selçuk Üniversitesi, Fen Bilimleri Enstitüsü.
- Şahan S., Kodaz H., Güneş S., Polat K., 2004, "A New Classifier Based on Attribute Weighted Artificial Immune System (AWAIS)", Lecture Notes in Computer Science LNCS, 3280: 11-20.

- Şahan S., Polat K., Kodaz H., Güneş S., 2005, "The Medical Applications of Attribute Weighted Artificial Immune System (AWAIS): Diagnosis of Heart and Diabetes Diseases", *Lectures Notes in Computer Science LNCS*, 3627: 456-468.
- Şahan S., Polat K., Kodaz H., Güneş S., 2007, "A New Hybrid Method Based on Fuzzy-Artificial Immune System and k-nn Algorithm for Breast Cancer Diagnosis", *Computers in Biology and Medicine*, Vol. 37(3): 415-423.
- Tarlinton D., 1998, "Germinal Centers: Form and Function", *Current Opinion in Imm.*, Vol. 10: 245-251.
- Taylor R.J., 1984, "Predation", Chapman and Hall, New York.
- Timmis J., 2000, "On Parameter Adjustment of The Immune Inspired Machine Learning Algorithm AINE", Technical Report 12-00, Computing Laboratory, University of Kent at Canterbury, Canterbury, Kent, CT2 7NF.
- Timmis J., Neal M., Hunt J., 1999, "Data Analysis Using Artificial Immune Systems, Cluster Analysis and Kohonen Networks: Some Comparisons", In *The Proceedings of IEEE International Conference on Systems, Man, and Cybernetics*, Tokyo, Japan, Vol. 3: 922-927.
- Timmis J., Neal M., 2000, "Investigating The Evolution and Stability of A Resource Limited Artificial Immune System", In *The Proceedings of International Conference on Genetic and Evolutionary Computation Conference (GECCO), Special Workshop on Artificial Immune Systems*, Las Vegas, Nevada, USA, 40-41.
- Timmis J., Neal M., Hunt J., 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", PhD. Thesis, University of Wales, Aberystwyth.
- Xu L., Chow M., Timmis J., Taylor L.S., 2007, "Power Distribution Outage Cause Identification With Imbalanced Data Using Artificial Immune Recognition System (AIRS) Algorithm", *IEEE Transactions on Power Systems*, Vol. 22(1): 198-204.
- UCI Repository of Machine Learning Databases. <http://archive.ics.uci.edu/ml/> son erişim 20 Aralık 2007.
- Vaitkus P.J., Cobbold R.S., Johnston K.W., 1988, "A Comparative Study and Assessment of Doppler Ultrasound Spectral Estimation Techniques Part II: Methods and Results", *Ultrasound Med. Biol.*, Vol. 14(8): 673-688.
- Watanabe Y., Ishiguro A., Uchikawa Y., 1998, "Decentralized Behavior Arbitration Mechanism for Autonomous Mobile Robot Using Immune Network", In *Artificial Immune Systems and Their Applications*, Edited by D. Dasgupta, Berlin: Springer, 187-209.
- Watkins A., 2000, "An Immunological Approach to Intrusion Detection", In *Proceedings of The 12th Annual Canadian Information Technology Security Symposium*, Ottawa, Canada, 447-454.

- Watkins A., 2005, "Exploiting Immunological Metaphors in The Development of Serial, Parallel, and Distributed Learning Algorithms", PhD Thesis, Kent University.
- Watkins A., 2001, "AIRS: A Resource Limited Artificial Immune Classifier", M.S. Thesis, Mississippi State University.
- Watkins A., Timmis J., 2002, "Artificial Immune Recognition System (AIRS): Revisions and Refinements", In 1st International Conference on Artificial Immune Systems (ICARIS), University of Kent at Canterbury, UK, September 9th-11th, 2002.
- Watkins A., Boggess L.C., 2002, "A Resource Limited Artificial Immune Classifier", In The Proceedings of The Special Sessions on Artificial Immune Systems in Congress on Evolutionary Computation, IEEE World Congress on Computational Intelligence, Honolulu, Hawaii, May 2002.
- Wierzchon S.T., Kuzelewska U., 2002, "Stable Clusters Formation in An Artificial Immune System", First International Conference on Artificial Immune Systems (ICARIS), University of Kent at Canterbury, UK, September 9-11, 68-75.
- Yip S.P., Webb G.I., 1994, "Empirical Function Attribute Construction in Classification Learning", Joint Conference on Artificial Intelligence (AI'94), Singapore, 29-36.
- Zhang G., Berardi L.V., 1998, "An Investigation of Neural Networks in Thyroid Function Diagnosis", Health Care Management Science, 29-37.